PERGAMON

Available online at www.sciencedirect.com
SCIENCE DIRECT®

Vision Research 43 (2003) 2539–2558

Vision Research

www.elsevier.com/locate/visres

Do humans optimally integrate stereo and texture information for judgments of surface slant?

David C. Knill *, Jeffrey A. Saunders

Center for Visual Sciences, University of Rochester, 274 Meliora Hall, Rochester, NY 14627, USA

Received 2 December 2002; received in revised form 22 April 2003

Abstract

An optimal linear system for integrating visual cues to 3D surface geometry weights cues in inverse proportion to their uncertainty. The problem of integrating texture and stereo information for judgments of planar surface slant provides a strong test of optimality in human perception. Since the accuracy of slant from texture judgments changes by an order of magnitude from low to high slants, optimality predicts corresponding changes in cue weights as a function of surface slant. Furthermore, since humans show significant individual differences in their abilities to use both texture and stereo information for judgments of 3D surface geometry, the problem admits the stronger test that individual differences in subjects' thresholds for discriminating slant from the individual cues should predict individual differences in cue weights. We tested both predictions by measuring slant discrimination thresholds and stereo/texture cue weights as a function of surface slant for multiple subjects. The results bear out both predictions of optimality, with the exception of an apparent slight under-weighting of texture information. This may be accounted for by factors specific to the stimuli used to isolate stereo information in the experiments. Taken together, the results are consistent with the hypothesis that humans optimally combine the two cues to surface slant, with cue weights proportional to the subjective reliability of the cues.

1. Introduction

Vision provides a number of independent cues to the three-dimensional layout of objects and scenes—stereo, motion, texture, shading, etc. While individual cues by themselves provide uncertain information about a scene, under normal conditions multiple cues are available to an observer. By efficiently integrating information from all available cues, the brain can derive more accurate and robust estimates of three-dimensional geometry (i.e. positions, orientations, and shapes in three-dimensional space). One complication that makes cue integration a hard problem is that the reliability of the information provided by different cues can change in a-priori unpredictable ways as a viewer moves or as surfaces change position and orientation in a scene. In order to most accurately interpret multiple cues, the visual system should combine the information provided by the cues in a way that accounts for these changes in their relative reliability.

Fig. 1 illustrates the effect of cue uncertainty on the optimal interpretation of a pair of visual cues to depth. The information provided by each cue is characterized by the likelihood function derived from the image information for that cue. The spread, or variance, of the likelihood function is a measure of the uncertainty of the data. Assuming that the image data associated with each cue are conditionally independent (e.g. the noise on one set of measurements is independent of the noise on the other), the joint likelihood function for the two cues together is simply the product of the individual likelihood functions. The result is a likelihood function whose peak is biased toward the more reliable of the two cues. When likelihood functions are Gaussian, the peak of the joint likelihood function is a weighted average of the peaks of individual likelihood functions, with weights inversely proportional to the variances of the likelihood functions. Thus, an optimal integration system will arrive at an interpretation that is, on average, a weighted sum of the interpretations from each cue individually, with more weight given to the more reliable cue.

In this paper, we test whether the human visual system integrates stereo and texture information to estimate surface slant in a statistically optimal way. In

* Corresponding author.
E-mail address: knill@cvs.rochester.edu (D.C. Knill).

0042-6989/$ - see front matter © 2003 Elsevier Ltd. All rights reserved.
doi:10.1016/S0042-6989(03)00458-9

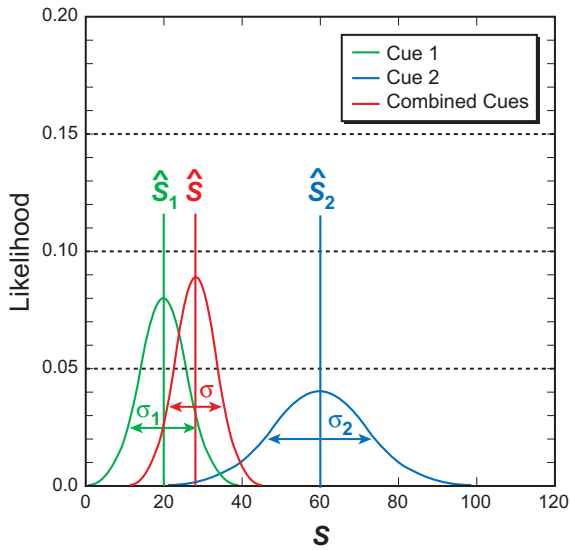

Fig. 1. The information provided by a cue about a scene S is given by its likelihood function, $p(I|S)$, where I is the image data associated with the cue (e.g. disparities for stereo or the flow field for structure-from-motion). The likelihood function for a combination of cues is, under some independence assumptions, simply the product of the likelihood functions for each cue. The peak of the joint likelihood function for the two cues, $\hat{S}$ is biased toward the peak of the narrower likelihood function. The variance of the joint likelihood function, σ^2, is smaller than the variances of either of the individual likelihood functions, σ_1^2 or σ_2^2. This reflects the reduction in uncertainty that is gained by combining multiple sources of information.

particular, we test the hypothesis that human observers are "subjectively" ideal observers for this perceptual task. A subjectively ideal observer is one that weights cues in inverse proportion to their subjective uncertainty—the uncertainty with which the observer can make inferences from individual cues. Several things make the problem of integrating stereo and texture information for slant perception a particularly interesting problem for testing optimal integration.

First, we can reasonably expect the relative uncertainties of texture and stereo information about slant to vary as a function of the slant itself. The uncertainty in the information provided by texture is known to decrease by an order of magnitude as slant increases from 0° to 70° (Knil, 1998a, 1998b). How the uncertainty of stereo information behaves as a function of slant is somewhat less clear; however, Banks, et al. computed theoretical reliability curves for slant from stereo based on an assumption of fixed noise levels on horizontal disparity, vertical disparity and horizontal vergence and found that the predicted reliability varied little over a wide range of slants (Banks, Hooge, & Backus, 2001). While this result may not hold exactly for large field of view stimuli, in which disparity noise can be expected to vary as a function of relative depth away from fixation, it strongly suggests that the relative uncertainties of texture and stereo cues to slant will vary significantly as a function of slant—the very parameter being estimated. This differs from the more commonly studied situation in which cue uncertainty varies with changes in an unrelated scene dimension (e.g. stereo information improves at closer viewing distance, motion parallax information improves with increased head motion) or is made to vary by adding visually apparent noise in one or the other cue (Ernst & Banks, 2002). Unlike in these situations, in which ancillary cues exist to help determine cue uncertainty (Landy, Maloney, Johnston, & Young, 1995), changes in cue uncertainty that result from changes in slant cannot be estimated independently of the slant itself.

Second, large individual differences exist in subjects' abilities to use stereo information for judging depth; thus, we are likely to find large differences in what would be each individual's optimal cue combination rule for texture and stereo. We can use these individual differences to test whether the relative weighting of texture and stereo for each subject is consistent with their subjective uncertainties for the two cues—a strong prediction of the subjective ideal observer hypothesis.

Finally, previous quantitative tests of optimal cue integration have studied how the brain integrates information from different sensory modalities—auditory and visual (Gharamani, Wolpert, & Jordan, 1997), proprioceptive and visual (van Beers, Sittig, & Denier van der Gon, 1999), or visual and haptic (Ernst & Banks, 2002)—rather than within-modality integration. Within-modality integration may have different properties than cross-modal integration. For example, cross-modal integration may involve selective allocation of attentional resources, whereas attention cannot be easily deployed selectively between different, spatially coincident visual cues when both are available (except by artificial means such as closing one eye to eliminate the stereo cue).

Our research followed an experimental strategy similar to that taken by Ernst and Banks in their study of visual–haptic cue integration (Ernst & Banks, 2002). We first measured individual subjects' slant discrimination thresholds for stimuli containing only one or another of the studied cues. These provided measures of the subjective uncertainty in each cue. Applying optimal estimation theory, we used these thresholds to predict the pattern of weights that each subject should give to stereo and texture cues as a function of surface slant. Using a cue perturbation paradigm, we measured the actual weights that characterize subjects' combination rules for integrating stereo and texture cues to slant and compared these to the weights predicted by the discrimination thresholds.

1.1. Optimal cue integration

Several sources provide good tutorial introductions to optimal linear cue integration; in particular, showing

how the weights in a linear model relate to the underlying uncertainty in a set of cues (see, for example, Blake, Bulthoff, & Sheinberg (1993) or Landy et al. (1995)). Here, we introduce the concept of optimal cue integration beginning from a somewhat more general perspective. The concept of an ideal observer from statistical estimation theory is central to understanding the theoretical underpinnings of cue integration. An ideal observer is an estimator that combines information from multiple cues so as to minimize a pre-defined error function on the estimated parameters. We use the standard definition of an ideal observer as one that minimizes the mean squared error of its estimates (for unbiased observers, this is necessarily a minimum variance estimator). The ideal observer bases its estimates on a *posterior* conditional probability density function, $p(\vec{S}|\vec{I})$, on the parameter being estimated, $\vec{S}$, given a set of image data, $\vec{I}$. Assuming a flat prior probability density function on $\vec{S}$, [1] the posterior density function is proportional to the likelihood function, $p(\vec{I}|\vec{S})$.

As illustrated in Fig. 1, the joint likelihood for a pair of cues, $\vec{I}_1$ and $\vec{I}_2$ is simply the product of the likelihood functions for each individual cue,

$$p(\vec{I}_1, \vec{I}_2|\vec{S}) = p(\vec{I}_1|\vec{S})p(\vec{I}_2|\vec{S}). \tag{1}$$

When the likelihood functions for the two cues are Gaussian, the joint likelihood function is Gaussian as well. The mean of the joint likelihood function, $\widehat{S}$, is a weighted sum of the means of the individual likelihood functions, $\widehat{S}_1$ and $\widehat{S}_2$,

$$\widehat{S} = w_1\widehat{S}_1 + w_2\widehat{S}_2, \tag{2}$$

where the weights, w_i, are in inverse proportion to the variances of the individual cue likelihood functions (Rao, 1973),

$$w_i = \frac{1/\sigma_i^2}{1/\sigma_1^2 + 1/\sigma_2^2}. \tag{3}$$

The variance of the joint likelihood function, σ^2 is given by (Rao, 1973)

$$\sigma^2 = \frac{1}{1/\sigma_1^2 + 1/\sigma_2^2}. \tag{4}$$

These relationships lead naturally to an implementation of an ideal integrator as one that computes a weighted average of the outputs of independent estimators for each of the individual cues available in an image (see Fig. 2). Rather than take such a mechanistic point of view to cue integration, we consider the system for estimating surface slant to be a black box with inputs coming from stereo and texture and an output giving some representation of surface orientation. We are agnostic as to the algorithm that the system uses to integrate the cues but would like to test whether the system is optimal (we take up the issue of mechanism in the general discussion section). The optimality hypothesis, in this context, predicts certain consistency relationships between the statistics of the slant estimates generated under different cue conditions. The two specific predictions are, first, that the variance in slant estimates derived from images containing both cues is related to the variance in slant estimates derived from images containing only one or another of the cues by Eq. (4), and, second, that the average estimated slant for images in which the slants suggested by stereo and texture conflict will be a weighted average of the slants suggested by the individual cues (assuming an unbiased estimator), with the weights related to the variance of slant estimates derived from individual cues according to Eq. (7).

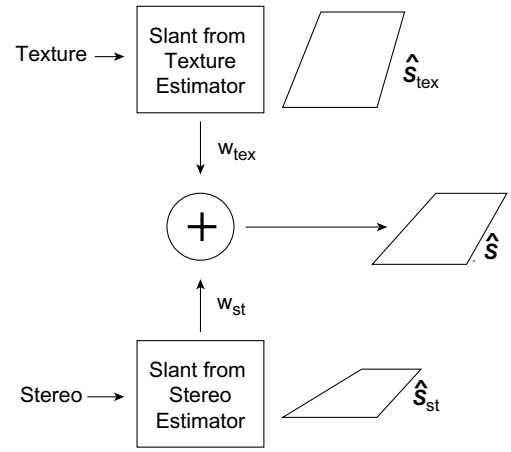

Fig. 2. The classic model of linear cue integration assumes independent modules for estimating a scene parameter like surface slant from each cue. The estimates derived from each cue are presumed to be weighted and summed to arrive at a final estimate. This point of view leads to questions of the form, "how does the visual system determine the weights to give to each cue?" As described later in the general discussion section, such an explicit embodiment of cue weights in the system need not exist for a system to be optimal.

1.2. Previous work on optimal cue integration

Numerous studies have shown that subjects give different weights to cues under different stimulus conditions. For example, recent psychophysical studies have shown that the human visual system gives a progressively lower weight to stereo information as vergence distance increases (Johnston, Cumming, & Parker, 1993). This seems rational, as the reliability of stereo information about relative depth along a surface decreases with increasing distance away from the observer (Banks et al., 2001). The same is true for motion—when the number of frames of a motion sequence is reduced to two, the weight that subjects give to motion cues for 3D shape is reduced (Johnston, Cumming, & Landy, 1994).

[1] The effect of a non-flat prior is minimal when the image data is much more constraining than one's prior knowledge of scenes.

These results are qualitatively consistent with the predictions of optimal integration.

Another approach to modulating cue reliability has been to add noise to the visual features underlying a cue, either naturally (e.g. by increasing the randomness in surface textures prior to projection (Knill, 1998c; Young, Landy, & Maloney, 1993)), or less naturally (e.g. adding motion jitter to texture elements in a motion display (Young et al., 1993)). As predicted, increasing the noisiness of a cue reduces the weight that subjects appear to give to the cue when combined with other, uncorrupted cues.

Results like these are qualitatively consistent with optimal integration of purely visual cues, but have not quantitatively tested for optimality. One exception in the vision domain was an experiment by Jacobs (Jacobs, 1999), in which subjects' variances in shape settings for motion-only and texture-only stimuli were used to predict their biases in shape settings for combined cue stimuli. Jacobs showed that subjects' shape settings for multiple cue stimuli could be accurately predicted by a linear integration model with weights set using Eq. (4), combined with a free parameter for the variance and mean of the subjective prior. This data provides indirect evidence for optimal integration, but Jacobs did not actually measure cue weights, nor did he find the best fitting set of weights to compare with the variance measures. Whether or not subjects used a quantitatively optimal integration strategy in the experiment is left unclear.

In the domain of cross-modal integration, a number of studies have directly addressed the predictions of optimal cue integration. Gharamani et al. (1997) studied the optimality of visual—auditory integration for target localization. He found that, while localization was dominated by vision, subjects appeared to give a small weight to auditory cues (inconsistent with complete visual capture). Unfortunately, differences in visual and auditory cue reliability across conditions were not large enough to provide a strong test of optimality. More recently, Ernst and Banks (2002) tested for optimal visual–haptic cue integration in object size judgments by adding different levels of external visual noise to virtually displayed three-dimensional blocks. This allowed them to artificially vary the reliability of visual cues to object size over a large enough range to quantitatively test the predictions of an optimal integrator. Ernst and Banks found that visual and haptic size discrimination thresholds accurately predicted the weights that subjects gave to visual and haptic cues for size judgments when simultaneously viewing and grasping objects.

The present study tested the predictions of an optimal integrator for intra-modal (i.e. visual) cues to depth, when the relative reliability of the cues changes naturally as a function of the surface geometry being estimated and when one might expect large individual differences that allow a strong test of the hypothesis that humans are subjectively optimal observers.

1.3. Specific psychophysical predictions

In order to operationalize the predictions of anoptimal integration model, we used slant discrimination performance as an empirical measure of cue uncertainty. We measured subjects' slant difference thresholds for discriminating the slants of surfaces depicted by stimuli containing only texture or stereo information individually or a combination of both cues. Assuming small amounts of decision noise and a weak prior on the expected slant, discrimination thresholds can be directly related to standard deviation parameters in the linear Gaussian model, so that we can express Eq. (4) in terms of the experimentally measured thresholds,

$$\frac{1}{T_{\text{st–tex}}(S)^2} \approx \frac{1}{T_{\text{st}}(S)^2} + \frac{1}{T_{\text{tex}}(S)^2}, \tag{5}$$

where $T_{\text{st–tex}}(S)$ is a subjects' threshold for discriminating surface slant from stimuli containing both stereo and texture cues, expressed as a function of the base slant, S, around which the threshold is measured. $T_{\text{st}}(S)$ is the threshold obtained using stimuli containing only stereo cues and $T_{\text{tex}}(S)$ is the threshold obtained using stimuli containing only texture cues.

Individual cue thresholds also predict the relationship between the average perceived slant of cue conflict stimuli and the slants suggested by each cue individually. For an optimal integrator, the weights accorded individual cues in a linear model are given by Eq. (3), which can be expressed in terms of thresholds as

$$w_{\text{st}}(S) \approx k\frac{1}{T_{\text{st}}(S)^2}, \tag{6}$$

$$w_{\text{tex}}(S) \approx k\frac{1}{T_{\text{tex}}(S)^2}, \tag{7}$$

or

$$\frac{w_{\text{st}}(S)}{w_{\text{tex}}(S)} \approx \frac{T_{\text{tex}}(S)^2}{T_{\text{st}}(S)^2}. \tag{8}$$

The weights, like the thresholds, can change as a function of slant, S.

We set out to test these predictions by measuring discrimination thresholds and cue weights for a number of subjects at a range of surface slants. In particular, we tested (a) whether or not slant discrimination thresholds for single cue stimuli measured at different surface slants accurately predict discrimination thresholds for combined cue stimuli, (b) whether or not the single cue thresholds predict differences in cue weights as a function of surface slant, and (c) whether or not individual

differences in the same slant discrimination thresholds predict individual differences in cue weights.

2. Overview of experimental logic

We ran seven naive subjects in two experiments each to test for subjective optimality. The first experiment measured subjects' slant difference thresholds for discriminating surface slant from stimuli containing only texture cues, only stereo cues or both. We measured thresholds for test slants ranging from 0° to 70° away from the fronto-parallel. We used this data to test the perceptual uncertainty predictions of an optimal integrator model as embodied in Eq. (5).

We then ran the same subjects in a standard cue perturbation experiment to measure the weights in a linear model relating the perceived slants as suggested by stereo and texture cues individually to the perceived slant of combined cue stimuli. In this experiment, test stimuli were generated with small conflicts between the stereo and texture cues. Subjects made slant discrimination judgments comparing the cue conflict stimuli to stimuli with consistent cues. Using this data, we estimated the weights in a linear model characterizing the perceived slant of a stimulus as a weighted sum of the slants suggested by the texture and stereo cues. This allowed us to test the prediction embodied in Eq. (8) relating discrimination thresholds to cue weights.

The biggest problem we faced was to generate stimuli that isolated stereo cues (for the stereo-only stimulus condition). Texture-only stimuli were easy to generate—subjects viewed projections of randomly tiled textures with one eye patched. Combined stereo-texture stimuli were similarly generated by having subjects view the same stimuli, projected in stereo, using both eyes. To generate stereo-only stimuli, we used large arrays of very small, randomly positioned dots rendered on a receding planar surface (see Fig. 3). Technically, these stimuli contained texture density cues to a surface's orientation; however, we reasoned that since humans appear not to effectively use texture density to judge surface slant (Buckley, Frisby, & Blake, 1996; Knill, 1998c) and since the rendered dots were so small as to make the size and foreshortening cues nearly undetectable, these stimuli had no *subjectively* useful texture information. An alternative approach would have been to use textures that were constrained to have a uniform density in the front-parallel plane. Such stimuli, however, would not have eliminated the texture density cue, but rather have provided a constant, conflicting cue that surfaces were fronto-parallel. In most experimental conditions, this would have corresponded to a large, unnatural cue conflict, raising the possibility that subjects might resort to unknown non-linear cue integration strategies in the discrimination task. For this reason, we chose to use textures that were uniform in the plane of each test surface, creating cue-consistent conditions. A control experiment showed that subjects were so much poorer at discriminating slant from monocular views of the random-dot textures than they were from binocular views that the density cue could have only had a minimal effect on measured discrimination thresholds in the binocular viewing condition, confirming our intuition.

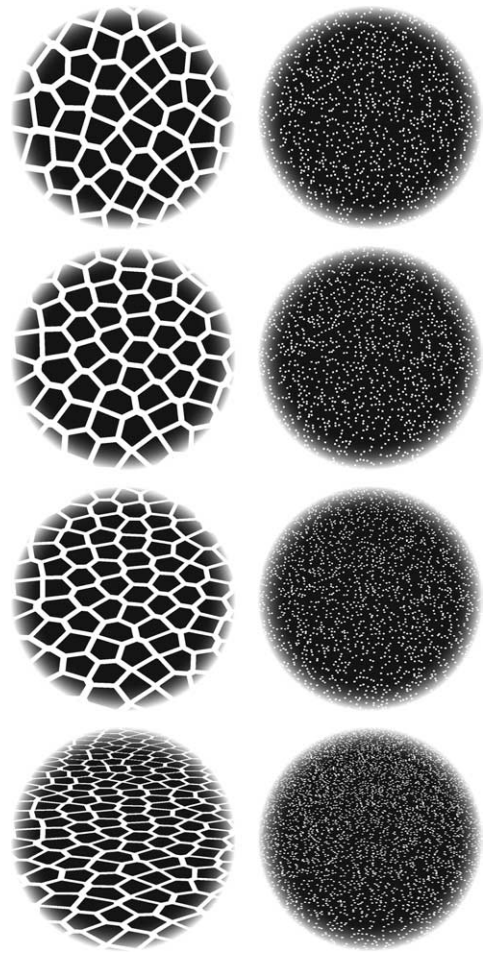

Fig. 3. Example stimuli used in the experiment. Stimuli are projected at 0°, 30°, 50°, and 70° from top to bottom. Note that the random-dot stimuli appear to have little if any slant. The blurry borders reflect the visually blurred boundaries of the occluders, as seen by subjects.

3. Experiment 1: Slant discrimination

3.1. Methods

3.1.1. Stimuli

Stimuli simulated perspective views of planar, textured surfaces that were slanted relative to the frontal image plane. Surface slant varied, but tilt direction was always vertical (i.e. the gradient of surface depth relative

to the viewer was vertical in the cyclopean projection). The slant of the virtual surfaces was conveyed by some combination of texture and/or stereo information (see Fig. 3). Three cue conditions were tested in the experiment:

- Stereo and texture—Stimuli were stereoscopically rendered views of a surface covered with a texture composed of Voronoi polygons. The textures were generated by computing the Voronoi tiling for a set of randomly positioned points in the plane, and then shrinking each polygon by 20% around its center of mass. To increase the regularity of texel spacing, a stochastic diffusion algorithm was applied to random initial positions before constructing the Voronoi tiling (see Knill, 1998b; Rosenholtz & Malik, 1997).
- Texture-only—Stimuli in the texture-only condition were identical to the stereo and texture stimuli, except that only one eye's view was presented, with the other eye patched, so that no stereo information was available.
- Stereo-only—Stimuli were stereoscopic views of a surface densely covered with small randomly positioned planar dots. The random-dot texture was chosen to minimize texture information and isolate stereo information (see the control experiment below).

Nineteen Voronoi and nineteen random-dot textures were generated in advance of the experiment. Each trial used a randomly chosen pair of two different textures from these pre-generated sets. Prior to mapping a texture onto a slanted surface, the texture was randomly oriented in the plane, effectively increasing the number of test textures. This also counterbalanced the effects of any global compression that may have been present by chance in the limited set of sample textures (which could have created biased slant judgments). Both Voronoi and dot textures were constructed as wrap-around textures—for stimuli with high surface slants the textures were repeated as necessary to fill the field of view. The periodicity in the textures is not readily apparent, as can be seen in Fig. 3.

Voronoi textures consisted of 400 elements. These were scaled prior to mapping them onto a test surface so that the textures would have a density of 0.25 texels/cm^2 and an average polygon diameter of 2.1 cm as measured on the surface. For a texel at the fixation point, this diameter corresponds to approximately a 2° visual angle. For the dot textures, samples consisted of 1600 elements, scaled to have a density of 6.0 texels/cm^2 and dot diameters of 0.11 cm (0.11° visual angle at the fixation point, on average). In the stereo conditions, subjects could theoretically discriminate surface slant based only on the difference in depth at the top (or bottom) of a pair of stimuli. Similarly, in the texture-only condition, subjects could make judgments based on the difference in texture density at the top (or bottom) of a pair of stimuli. In order to minimize the effectiveness of these cues, we randomized the depths of the surfaces displayed within a trial by ±4 cm around a mean distance of 60 cm at the point of fixation (at the center of the stimulus). This randomized the texture density in the image, since the density was held constant on the surface.

Displays included a small spherical fixation target (rendered without shading) in the center of the display at the depth of the test surface in a stimulus. The fixation point was scaled to have a diameter of 0.2° of visual angle. The fixation point appeared prior to stimulus presentation to allow subjects to establish fixation. Because we randomized the absolute depth of surfaces within a trial, the fixation target was made visible during the delay between pairs of stimuli in a trial, positioned at the depth of the succeeding surface. That is, after the first stimulus surface disappeared, the fixation mark moved in depth to the depth of the second stimulus surface. This facilitated proper fixation prior to the presentation of each test stimulus. The fixation mark remained on during the stimulus presentation.

3.1.2. Apparatus

Visual displays were presented in stereo from a computer monitor viewed through a mirror (Fig. 4), using CrystalEyes shutter glasses to present different stereo views to the left and right eyes. Circular apertures were positioned in front of each eye, at a distance of 6–8 cm, to limit the field of view for each eye to a 15° region around the fixation point. By placing the occluders near the eyes, we also eliminated spurious frame effects of

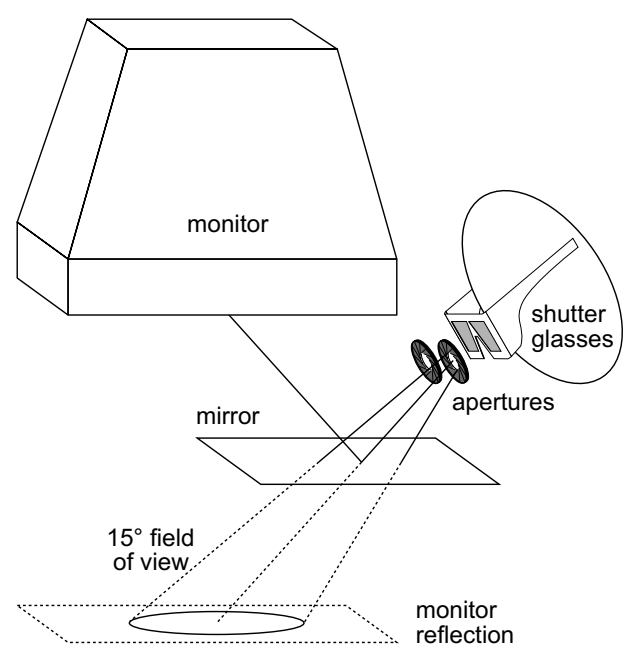

Fig. 4. Schematic of the apparatus used in the experiment.

viewing surfaces through an artificial occluder at the same depth as the surface.

In stereo mode, the monitor had a refresh rate of 120 or 60 Hz for each eye's view, and a pixel resolution of 1024×768. The stimuli and feedback were all drawn in red to take advantage of the comparatively faster red phosphor of the monitor and prevent inter-ocular cross-talk. The virtual surface of the monitor reflected through the mirror was slanted relative to the viewer, and any depth cues that cannot be simulated using stereo shutter glasses, such as accommodative gradients, would be consistent with the slant of the reflected monitor surface. In the experiment, the angle between the monitor surface normal and the viewer's line of sight was approximately 40° (varying slightly between subjects), which was near the middle of the range of test slants used for stimuli.

At the start of each experimental session, we used an optical alignment procedure to calibrate the virtual environment. The backing of the half-silvered mirror was temporarily removed, so that subjects could simultaneously see both the reflection of the monitor and a small optical marker, which was tracked in 3D by an Optotrak 3020 system. A sequence of visual locations were cued by dots on the monitor, and subjects aligned the marker with the cued locations. Cues were presented monocularly, and matches were performed in separate sequences for left and right eyes. Thirteen positions on the monitor were cued, and each position was matched twice at different depth planes. The combined responses for both eyes were used to estimate the plane of the virtual monitor surface (the reflected image of the monitor behind the mirror) and the left and right eye positions in 3D space. These parameters allowed us to render geometrically correct images of left and right eye views of a stimulus surface for each individual subject. It also automatically accounted for any drift in the 3D orientation of the mirror between experimental sessions. After the calibration procedure, a rough test was performed, in which subjects moved the marker while it was visible through the half-silvered mirror and checked that a rendered dot moved with the marker appropriately. Calibration was deemed acceptable if deviations were less than approximately 1–2 mm. Otherwise, the calibration procedure was repeated.

3.1.3. Procedure

Subjects performed a two-alternative forced-choice slant discrimination task. On each trial, subjects were presented with a successive pair of surfaces, and judged whether the first or second surface was more slanted. Slant was defined to be the signed angle between the surface normal and the line of sight to a cyclopean eye mid-way between a subjects' left and right eyes. For positive slants, the tops of stimulus surfaces appeared to recede in depth; for negative slants, the bottoms appeared to recede in depth.

Subjects were presented with some examples to demonstrate the task, and in the first experimental session performed a short block of practice trials with feedback to ensure that they understood the procedure. On any given trial, one of the pair of surfaces was the test stimulus, set to one of four test slants (0°, 30°, 50°, 70°) and the other was a probe stimulus. The order of test and probe stimuli was randomized within blocks. The probe stimuli had slants that varied around the test slants, chosen using an adaptive staircase procedure. Prior to each trial, all the previous responses from trials in the same condition were used to compute maximum likelihood estimates of the point of subjective equality between the first and second stimuli, PSE, and the threshold, T. The new probe value was randomly chosen from within a small range around either the estimated 25% point ($\mathrm{PSE} - T$) or the estimated 75% point ($\mathrm{PSE} + T$). A-priori estimates of the mean and variance of PSE and thresholds were combined with the data, which served to constrain the choice of initial probes when few or no previous trials are available. These a-priori values were set manually between experimental sessions based on offline fits of the data.

Trials began with a 250 ms presentation of the fixation point alone, followed by a pair of slanted surfaces, each displayed for 1000 ms. Between pairs of surfaces, there was a 500 ms delay with a blank screen and new fixation point, presented at the depth of the second stimulus. After both surfaces were presented, the display remained blank until the subject made a response, which initiated the next trial. Except for the initial practice trials in the first session, no feedback was given.

Trials were self-paced, and subjects were encouraged to take breaks as necessary. Subjects performed three blocks of trials in each 1-h experimental session, corresponding to the three cue conditions: texture-only, stereo-only or stereo-and-texture. In the texture-only condition, the unused eye was covered with an eye patch. The order of conditions was randomized across sessions, and the randomized order was varied across subjects. Each block consisted of 256 trials, corresponding to 64 trials for each of the four test slant conditions. The experiment consisted of 6 sessions, scheduled on separate days over the course of 2–3 weeks. The data from the first session of each subject was discarded from the final analysis, to prevent any initial learning effects from biasing the results. Pooling across the remaining sessions yielded a total of 320 trials per subject for each of the 12 (3×4) combinations of cue condition and test slant.

3.1.4. Subjects

Seven undergraduates at the University of Rochester served as subjects. All subjects were naive to the goals of

the experiment and to vision research in general. All had normal or corrected-to-normal vision and no known problems with stereo vision. Performance on the stereo-only conditions (combined with the control experiment showing the weakness of the monocular cues in those stimuli) showed that all subjects could make reasonable use of stereo for depth judgments.

3.1.5. Data analysis

For each test slant, the raw data was organized into arrays specifying the number of trials on which subjects reported the second stimulus to be more slanted than the first stimulus, as a function of the slant difference between the two stimuli. In pilot experiments, we found that some naive subjects have a significant guessing rate (e.g. because of attentional lapses). This was reflected in psychometric functions that leveled off at points below 1.0 and above 0.0. In order to correct for guessing, we fitted a modified cumulative Gaussian psychometric function to each subject's data in which the probability of selecting a comparison stimulus was assumed to be a mixture of an underlying Gaussian discrimination process and a random guessing process. Writing subjects' decision as

$$D = \begin{cases} 1; & \text{Comparison stimulus judged more slanted,} \\ 0; & \text{Test stimulus judged more slanted.} \end{cases} \tag{9}$$

The psychometric model was

$$p(D = 1|\Delta S) = (1 - p)\Gamma(\Delta S; \mu, \sigma) + pq, \tag{10}$$

$$p(D = 0|\Delta S) = 1 - p(D = 1|\Delta S), \tag{11}$$

where ΔS is the difference in slant between the first and second stimulus, μ is the mean of the cumulative Gaussian, σ is the standard deviation of the cumulative Gaussian, p is the probability that a subject guessed on any given trial and q is the probability that a subject guessed the comparison stimulus, given that he or she guessed at all. The mean parameter, μ, is a measure of the point of subjective equality between first and second stimuli. It accommodates effects like perceptual drift in the remembered slant of the first stimulus. A corrected 75% threshold can be computed from the standard deviation parameter σ. The corrected threshold reflects the 75% threshold difference in slant between test and comparison stimuli that a subject would have in a 2-AFC choice without guessing and without a temporal order bias in slant judgments (reflected by the μ parameter).

Guessing parameters for each subject were assumed to be constant across conditions within an experiment. Parameters for the psychometric model (thresholds, σ, biases, μ and guessing parameters, p and q) were computed from maximum likelihood fits to the raw data. The likelihood of a subject making a decision, D_{ij}, on trial i, for test slant j can be expressed as

$$\mathscr{L}_{i,j} = 1 - D_{i,j} + (2D_{i,j} - 1)[(1 - p)\Gamma(\Delta S_{i,j}; \mu_j, \sigma_j) + pq], \tag{12}$$

where $\Delta S_{i,j}$ is the difference in slant between two stimuli on trial i of the jth test slant condition, and μ_j and σ_j are, respectively, the bias and threshold parameters for the jth test slant condition. The likelihood function for the entire set of a given subject's data is then given by

$$\mathscr{L} = \Pi_{j=1}^{4} \Pi_{i=1}^{N} \mathscr{L}_{i,j}, \tag{13}$$

where N is the number of trials in each condition. The standard error of our parameter estimates can be derived from the covariance matrix of the likelihood function, $\mathscr{L}$, for the psychometric model parameters (the standard error for each parameter estimate is the square root of the corresponding diagonal element of the covariance matrix). We used the standard approximation of the covariance function as the inverse of the Hessian of the log-likelihood function, computed at the maximum of the likelihood function (Rao, 1973) (asymptotically correct for an infinite number of data points).

3.2. Results

Fig. 5 shows sample plots of the best fitting 75% thresholds (corrected for guessing) for three subjects. Note that with the exception of one data point for subject 3, the threshold for combined cue stimuli was lower than or equal to the thresholds measured for the individual cue stimuli. An optimal integrator would show thresholds that varied lawfully as a function of the thresholds for the single cue stimuli. Eq. (5) expresses this lawful relationship,

$$\frac{1}{\widehat{T}_{\text{st-tex}}(S)^2} \approx \frac{1}{T_{\text{st}}(S)^2} + \frac{1}{T_{\text{tex}}(S)^2}, \tag{14}$$

where $\widehat{T}_{\text{st-tex}}(S)$ is the threshold for discriminating surface slant from stimuli containing both stereo and texture cues predicted by an optimal integrator of the two cues. Fig. 6 shows average thresholds for each cue condition as a function of surface slant along with the average of the combined cue thresholds that would be predicted by an optimal integrator for each observer. The measured combined cue thresholds do not differ significantly from those predicted from the individual cue thresholds by an optimal model.

The guessing rate for subjects was on average 0.16 ± 0.13, indicating a high variance in attentional focus. The average value of the q parameter (the probability of selecting the second stimulus, given that a subject was guessing) was 0.41 ± 0.25, with the high

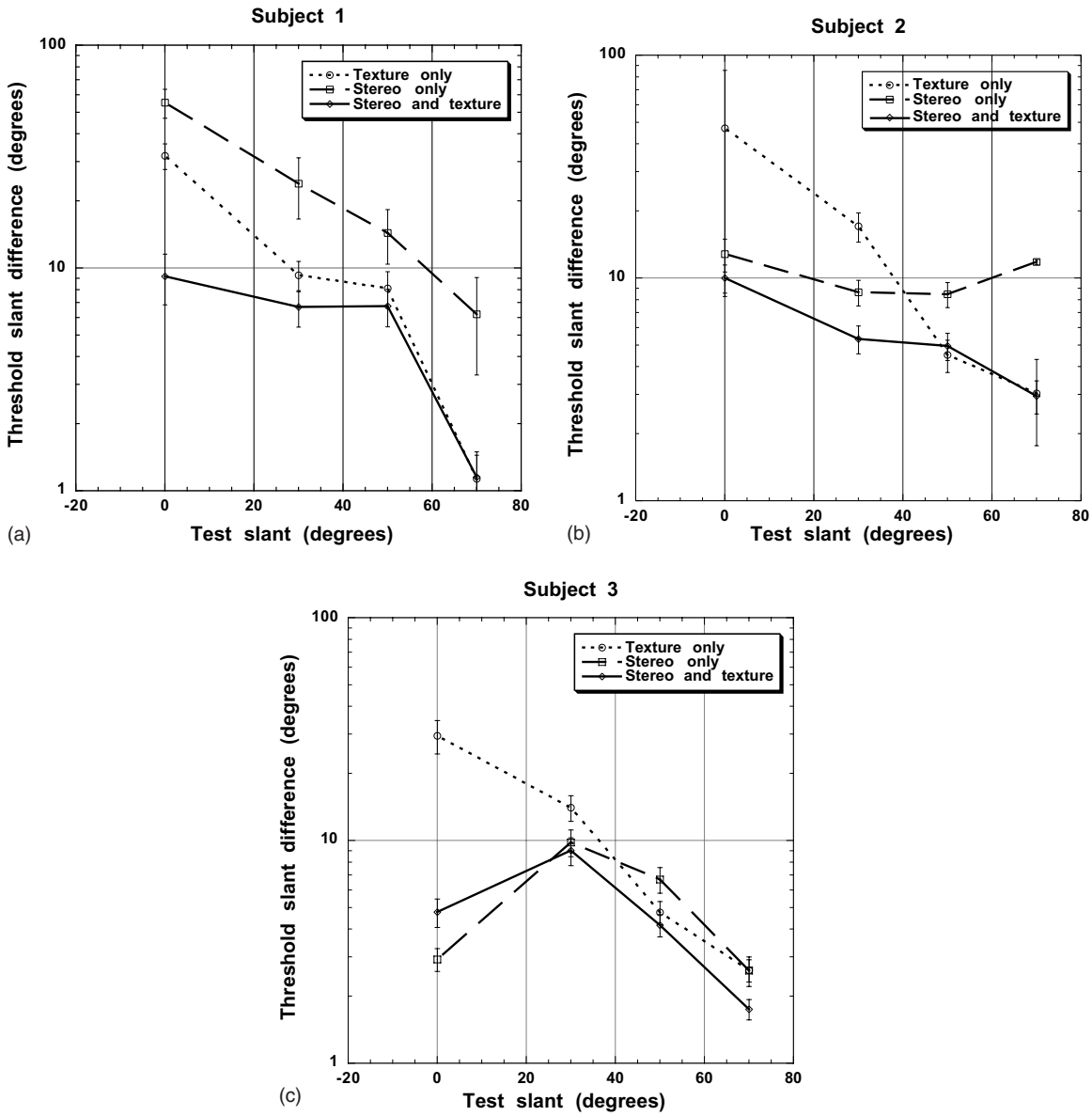

Fig. 5. Slant discrimination thresholds for three subjects. With the exception of subject 3 in the 0° slant condition, thresholds for combined cue stimuli (solid line) are below the thresholds for single cue stimuli or are equal to the lowest of the single cue thresholds. Error bars were computed from the likelihood functions derived from the data for the psychometric model parameter fits—they correspond to the standard error of the threshold estimates.

standard deviation again reflecting a large variance between subjects in guessing strategy.

3.3. Discussion

The first effect that jumps out from the threshold data is that, while both texture and stereo cues become more reliable indicators of slant as surface slant is increased, they do so at markedly different rates. At low slants, near the fronto-parallel, stereo is significantly more reliable than texture, but at test slants of 50° and 70°, subjects, on average, are better able to make slant judgments from texture information than from stereo information. This trend is consistent across all subjects tested here, though subjects differ somewhat in their average ability to use the two cues. Given individual differences in human stereo-acuity, these individual differences are not surprising. The decrease in slant-from-texture thresholds as a function of slant is consistent with earlier results using similar stimuli (Knill, 1998b) and with the theoretical analysis showing large differences in the theoretical reliability of texture information between surfaces at large slants and surfaces at low slants.

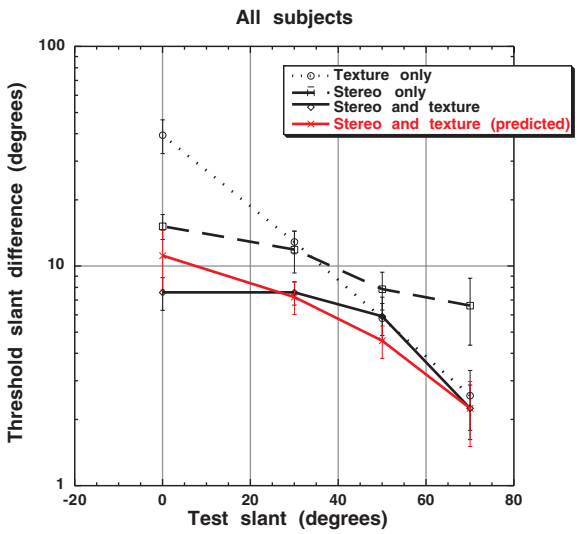

Fig. 6. Average slant discrimination thresholds for all three cue conditions. The average of the combined cue thresholds predicted from the single cue thresholds is shown in red. Error bars are the standard error of the mean computed by averaging subjects' individual thresholds.

As noted in the introduction, determining such theoretical predictions for stereo disparity information is more difficult. It requires assumptions about the underlying measures that contribute to slant-from-disparity judgments (e.g. absolute disparity vs. disparity gradients) and the levels of internal noise corrupting those measurements. Assuming constant levels of noise on horizontal disparity, vertical disparity and vergence angle, Banks, et al. measured predicted reliability curves (the inverse of threshold curves) for slant-from-disparity as a function of slant and distance from the viewer (Banks et al., 2001). They found very small effects of slant on their reliability measures, less than those found here. From their results, we would have expected flatter threshold functions for slant-from-stereo; however, a more complete noise model (for example, which accounts for changes in noise levels as a function of absolute disparity) could well change the theoretical predictions. What the current results suggest, regardless of the source of uncertainty in slant-from-disparity judgments, is that humans should give progressively more weight to texture as the slant of a surface increases. Many of the subjects tested here would ideally give more weight to texture information than stereo information at high slants.

The results are broadly consistent with the hypothesis that subjects, on average, optimally integrated stereo and texture cues to surface slant. The one slant condition that shows some deviation from the prediction is the 0° slant condition. For six out of seven subjects, combined cue thresholds for the 0° slant condition were significantly lower than predicted by the single cue thresholds under an optimal integration model. Subject 3 in Fig. 5 was the only one of the seven subjects not to show some super-additivity. Informal subject reports suggested a potential reason for the apparent super-additivity. The sign of slant for monocular, textured stimuli at low slants often appeared ambiguous to subjects—while appearing slanted away from the fronto-parallel, the surfaces were bistable; they appeared to be receding either at the top or the bottom of the surface. Previous studies of slant perception from texture (Knill, 1998a, 1998b) suggest why this bi-modality might occur. These studies have shown that subjects strongly rely on a local foreshortening cue in texture patterns—using the local deviation of textures from isotropy to estimate slant. Since the local foreshortening of a texture is the same for local slants of opposite sign (a circle projects to the same ellipse from slants of 45° and −45°), this cue by itself does not disambiguate the direction (sign) of slant. Other gradient-based cues such as scaling are needed to disambiguate the direction of slant. If these cues are unreliable, as they are at low slants, the likelihood function for slant from texture would not be Gaussian as assumed in the linear integration model (and in the psychometric model), but rather would be bimodal with peaks at positive and negative values of slant. Li and Zaidi, for example, have described examples in which scaling information in a stimulus is not enough to disambiguate the sign of surface slant (Li & Zaidi, 2002). This uncertainty would greatly exaggerate the uncertainty in the absolute magnitude of slant from texture for small slants. We, therefore, expect that the threshold measures derived for the monocular texture stimuli are exaggerated, leading to an underestimate of the predicted combined cue thresholds at 0°. [2] Since the stereo cue effectively disambiguates the sign of slant in the combined-cue stimuli, The combined cue likelihood function is unimodal and the added uncertainty caused by the "phantom" mode in the texture likelihood function disappears (see Knill (2003) for a longer discussion of this phenomenon).

A more central concern for interpreting the threshold data is that stimuli in what we have referred to as the stereo-only condition contained texture information about surface slant. Looking at the stimuli in Fig. 3 suggests that this information was not perceptually salient. To insure that this was indeed the case, we ran a control experiment with two naive subjects to measure their ability to make slant judgments from monocular views of these stimuli.

[2] The proportional error on threshold estimates for the 0° texture-only condition was significantly higher than for the other slants. It was typically between 10% and 20% for non-zero slants, but all standard errors on threshold estimates for the 0° slant condition were greater than 30%.

4. Control experiment

We repeated the discrimination experiment using two cue conditions—binocular views of the random-dot textures (equivalent to the stereo-only stimuli in experiment 1) and monocular views of the same random-dot textures. Since we were interested in measuring the degree to which texture cues influenced slant judgments in the random-dot stimuli in experiment 1, we interleaved the two types of stimuli within experimental blocks. Monocular conditions were generated by displaying only the left eye's view of the dot stimuli, with the right eye's view set to a black screen. In all other respects, the methods were the same as in experiment 1. Two naive undergraduates served as subjects in the experiment.

Fig. 7 shows the results of fitting thresholds to the monocular and binocular conditions of the control experiment. While both subjects could perform the task under binocular viewing, in most conditions, they were effectively at chance under monocular viewing. We've plotted the thresholds as 90° for conditions in which thresholds were unfittable simply as a point of comparison with the thresholds from binocular viewing. In fact, in those conditions, the fitted thresholds were effectively infinite. We were able to fit thresholds to subject 2's data in the 30° and 50° conditions, but these thresholds were more than 4 times the thresholds found under binocular viewing, indicating that even were the subject to have used texture information in the binocular viewing condition, it would have contributed only minimally to their performance.

5. Experiment 2: Measuring cue weights

The average threshold data provides some power for testing the optimality hypothesis; however, the uncertainty in threshold estimates is large relative to the small improvements in thresholds predicted for most conditions. This makes it impossible to use this data to test whether the hypothesis of subjective optimality predicts individual differences in thresholds. The predicted relationship between single cue thresholds and cue weights provides a more promising approach to test optimality. The clearest prediction of the threshold data is that subjects should weight texture information more heavily as the slant of a surface increases. The large individual differences in relative thresholds across the single cue conditions also support the stronger test of whether or not individual variations in cue uncertainty predict individual differences in cue weighting. The second experiment was designed to measure the effective weights that subjects gave to stereo and texture cues when making slant judgments. For each test slant used in experiment 1, we created eight cue conflict test stimuli, with one cue (either texture or stereo) simulated so as to suggest the test slant and the other cue simulated so as to suggest a slant that differed from the test slant by $\pm\Delta$ or $\pm 2\Delta$, where Δ was chosen separately for each test slant to be a weakly discriminable slant difference (based on the discrimination thresholds). Subjects performed the same discrimination task used in experiment 1, with probe stimuli containing consistent stereo and texture cues to slant. We fit a psychometric model to the data

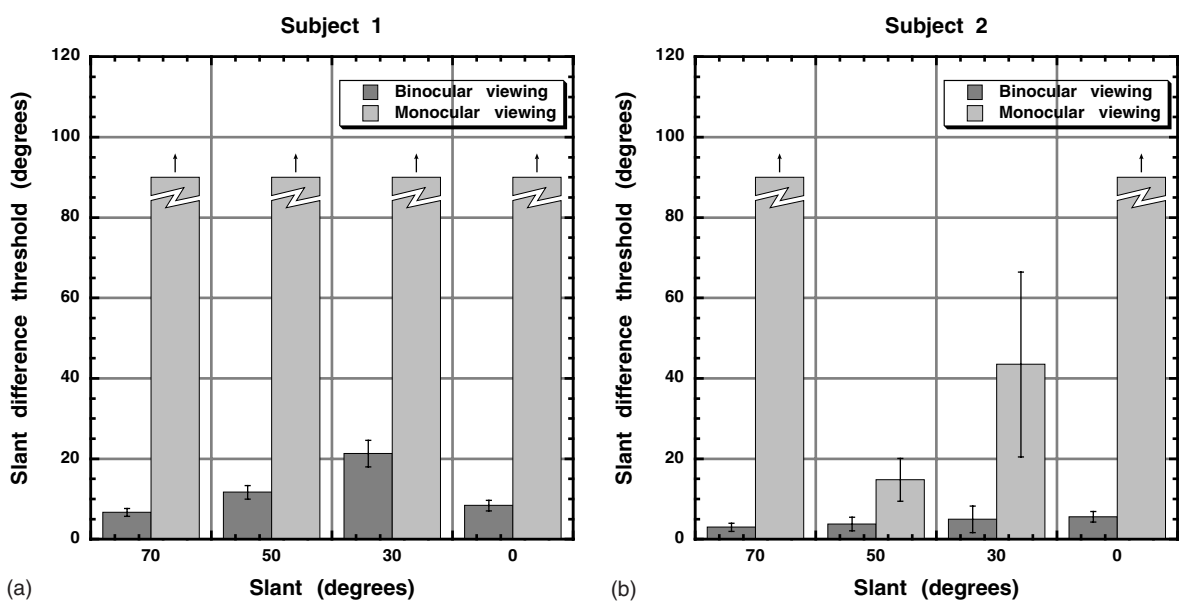

Fig. 7. Slant discrimination thresholds for two subjects in the control experiment. The broken bars with arrows denote conditions in which thresholds were unfittable—subjects performed essentially at chance in these conditions. Error bars reflect the standard error in estimates of subjects' thresholds, estimated using the same method used in experiment 1.

that assumed that for each test slant, subjects based judgments on a weighted sum of the slants suggested by texture and stereo cues.

5.1. Methods

5.1.1. Stimuli

Consistent cue stimuli were identical to stimuli from the texture and stereo condition in experiment 1: binocular images simulating left and right eye's perspective views of a planar surface covered with an isotropic Voronoi texture, slanted away from the viewer in the vertical direction. For these stimuli, the slant specified by stereo and texture was always the same. Test stimuli were generated so that the stereo cue suggested one slant (S_{st}) and the texture cue suggested a different one (S_{tex}). This was done by rendering a distorted planar, Voronoi texture at the stereo slant, S_{st}. The texture was distorted before mapping onto the surface so that when projected from the stereo slant to a point midway between a subjects' two eyes (the cyclopean view), the texture suggested the texture slant, S_{tex}. We determined the texture distortion in two stages. First, we projected positions of texture vertices for a cyclopean view of a surface with slant S_{tex}. We then back-projected these points from the cyclopean eye's projection onto a surface with slant S_{st} to generate the new, distorted texture vertices.

5.1.2. Procedure

The task and procedure were the same as in experiment 1. As before, subjects made forced-choice discriminations between the slants of successive pairs of surfaces. From the perspective of the subject, the only difference was that there were no longer different cue conditions—all stimuli were viewed binocularly and contained planar, Voronoi textures, as in the combined cue condition in experiment 1. In the test stimuli of experiment 2, the slants specified by texture and stereo were independently varied, so that the two slant cues had small conflicts between them. On any given test trial, one of the two slant cues specified the test slant, chosen from the set {0°, 30°, 50°, 70°}, and the other cue specified a slant that differed from the test slant by $\{-2\Delta, -\Delta, \Delta, 2\Delta\}$. The value for Δ varied across subjects and base slants. We chose it to be 1/2 the magnitude of the discrimination threshold measured from the combined cue stimuli in experiment 1. The threshold measure we used to set Δ, however, was derived without taking into account attentional lapses, as we did for final estimates of thresholds (as reported here for experiment 1); thus, the values we chose varied somewhat from what was intended (see the caption for Table 1 for an extended discussion of this point).

Table 1 shows the values of Δ used to create cue conflict stimuli for all seven subjects and all four test slants. Also shown in the table are d' values for each value of Δ, computed for each subject from the texture-only and stereo-only texture thresholds measured in experiment 1. The d' values reflect the discriminability of the stereo and texture cues within a stimulus. Note that with a few exceptions, the d' values are near the planned-for level of 1/2.

For the initial sessions of the first two subjects, a staircase was used to choose probe slants, as in experiment 1. We noticed that the staircase was not very effective: because there were few trials per condition, the probe choices were dominated by a priori settings. For the remaining sessions of the first two subjects, and for all sessions of the other subjects, we switched to a method of constant stimuli, with probe slants set manually to span a range around a point of subjective equality expected from equal weighting of the cues. Subjects performed the experiment across six 1-hour experimental sessions, scheduled on separate days. Each session consisted of three blocks of 256 trials, and the 32

Table 1
The values of Δ used to create cue conflict stimuli in the experiment

Subject	Δ				d'			
	70°	50°	30°	0°	70°	50°	30°	0°
S1	1.5°	6.0°	7.0°	12.0°	0.3368	0.5144	0.3853	0.2658
S2	3.5°	7.5°	8.0°	5.5°	0.2839	0.7246	0.4797	0.1862
S3	2.0°	3.0°	4.0°	7.0°	0.2320	0.4423	0.2957	0.0072
S4	2.0°	4.0°	4.0°	4.0°	1.6022	0.8897	0.6938	0.0753
S5	1.3°	3.3°	5.3°	8.0°	0.3121	0.5319	0.5623	0.1968
S6	2.0°	5.0°	2.0°	1.0°	0.7651	0.8613	0.1651	0.0477
S7	1.0°	2.0°	4.0°	5.0°	0.4687	0.5973	0.4318	0.03

Values were chosen based on an initial approximate estimate of subjects' slant discrimination thresholds for combined stereo-texture stimuli. The proper measure for determining the size of an appropriate cue conflict is the d' computed for discriminating the slant suggested by the texture pattern from the slant suggested by the stereo disparity pattern. We derived these measures from the single cue slant discrimination thresholds measured in experiment 1. The values fluctuate around an average value 0.44, in part because we used the combined stereo-texture cue thresholds as a heuristic measure to set the conflicts and in part because the initial psychometric fit used to set the conflicts had not been optimized (e.g. by accounting for attentional lapses).

conditions (4 test slants × 8 conflicts) were randomly inter-mixed within each block. Across sessions, this yielded a total of 144 trials per condition for each subject.

5.1.3. Subjects

The seven subjects from experiment 1 participated in this experiment.

5.1.4. Data analysis

The data analysis was similar to the first experiment with one important difference. The psychometric decision model was modified to replace the slant difference term, ΔS, with a weighted average of the slant difference suggested by each cue, $w_{\text{tex}}\Delta S_{\text{tex}} + (1 - w_{\text{tex}})\Delta S_{\text{st}}$. The resulting psychometric decision model is

$$p(D = 1|\Delta S_{tex}, \Delta S_{\text{st}}) = (1 - p)\Gamma(w_{\text{tex}}\Delta S_{\text{tex}} + (1 - w_{\text{tex}})\Delta S_{\text{st}}; \mu, \sigma) + pq, \quad (15)$$

where ΔS_{tex} is the difference in slant suggested by texture between the first and second stimulus and ΔS_{st} is the difference in slant suggested by stereo information. w_{tex} is the weight given by the observer to the texture cue, constrained to lie between 0 and 1. Implicit in the equation is the assumption that the weights given to stereo and texture cues sum to 1. By including the weights in the full psychometric function fit, we gain more statistical power than would be obtained by first finding points of subjective equality for each cue combination condition and then using linear regression to estimate the weights.

Since the likelihood function over the weight parameter was highly non-Gaussian, due to the boundaries at 0 and 1, we used bootstrapping (Davison, 1997) to fit error bars to the weight estimates. We repeated the psychometric fits 1000 times, each time resampling (with replacement) the individual trial data. The standard deviation of the repeated estimates of the texture weight parameter provided a measure of the standard error of our estimate.

5.2. Results

Fig. 8 shows subjects' texture weights as a function of surface slant (stereo weights would be given by $w_{\text{s}} = 1 - w_{\text{tex}}$). As predicted by the threshold data, all subjects show a strong trend to weight texture information more heavily as surface slant increases. Using Eq. (8) and assuming that the cue weights sum to one, we computed the texture weights predicted by subjects' discrimination thresholds. Fig. 9 plots the texture weights predicted by three subjects' slant discrimination thresholds along with the weights measured in experiment 2 (the same subjects shown in Fig. 5). Fig. 10 shows averages across the seven subjects of both the measured and predicted weights. Subjects' texture weights increase as a function of surface slant as predicted by single cue thresholds ($F(3,6) = 6.6$, $p < 0.05$). On average, subjects appear to underweight texture by a small amount, as compared to the weights predicted by discrimination thresholds; however, this difference did not reach significance (average difference = 0.12, $F(1,6) = 3.2$, $p > 0.05$).

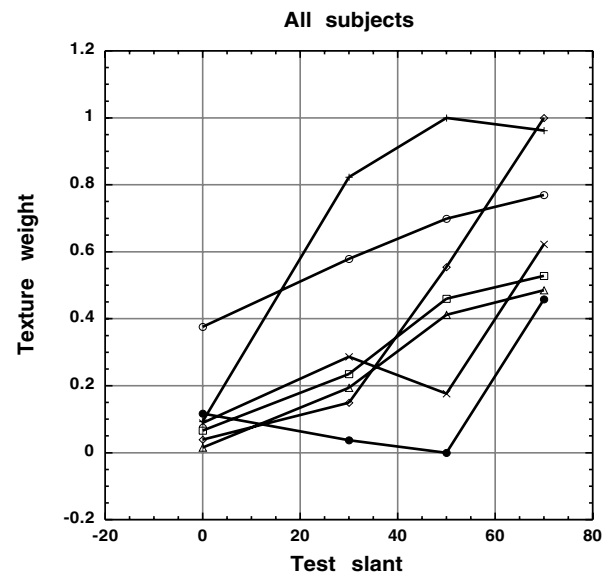

Fig. 8. Texture cue weights, w_{tex} (stereo weights are $1 - w_{\text{tex}}$) as a function of surface slant for all seven subjects.

5.2.1. Individual differences

The data clearly show that changes in discrimination thresholds for slant from texture and slant from stereo as a function of surface slant predict, on average, the weights subjects give to the two cues. That is, on average, subjects appear to weight the two cues optimally. How well do the predictions hold at the individual level? In order to assess this, we measured the correlations between measured and predicted texture weights for each subject. These are shown by the dark grey bars in Fig. 11. Correlations varied from 0.325 to 0.96.

A resampling procedure was used to estimate the standard errors of the correlation coefficient measures. On each iteration of the procedure, a new set of single cue thresholds and texture weights was chosen from the measured error distributions on those parameters. The random threshold samples were then used to compute predicted texture weights (using Eq. (8)), which were then correlated with the random samples of measured weights. The standard deviations of the resulting correlation coefficients provided a measure of the standard error in our estimates of the coefficients. With the exception of subject 1, all correlations were significantly

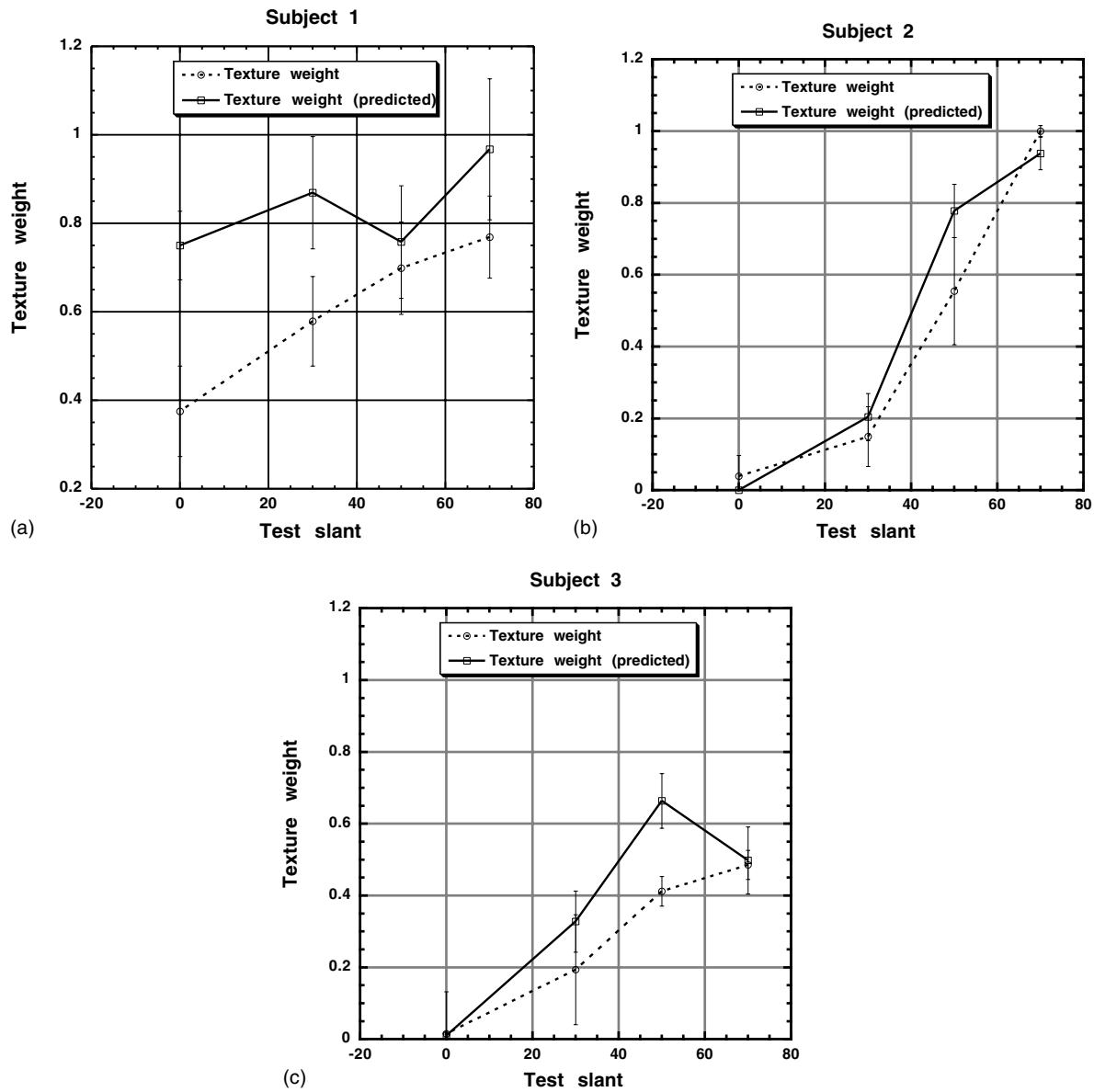

Fig. 9. Plots of both measured and predicted texture cue weights for the same three subjects shown in Fig. 5.

greater than 0 at the $p < 0.05$ level, and most were much more significant than that.

These results would seem to indicate that the optimal model fit some subjects' data (higher correlation coefficients) better than others. The measured correlations, however, depend not only on the fit of the model, but also on the uncertainty in our estimates of thresholds, from which we derived the weights predicted by the optimal model, and in our estimates of subjects' texture cue weights. Larger levels of uncertainty in our estimates of a subject's thresholds and weights (as reflected in their std. errors) will lead to smaller correlation coefficients. We therefore measured the correlations that we would have expected to measure if subjects were in fact optimal, given the uncertainty in our estimates of thresholds and weights.

To do this, we used a resampling technique in which we associated with each subject an ideal observer whose cue weights were related to its *true* discrimination thresholds by Eq. (8), but for whom the experimentally measured thresholds and weights were corrupted by the noise equivalent to the standard error of the experimentally measured values. We do not, however, know subjects' true thresholds, but rather can only compute a likelihood functions for these thresholds, given the experimental data. We therefore used a bootstrap procedure to repeatedly sample possible values for the true thresholds from the computed likelihood functions. For each sample of a possible set of thresholds, we computed the correspondingly optimal texture weights. This provided threshold/weight pairs for the set of ideal integrators that fit the data from experiment 1. For each of

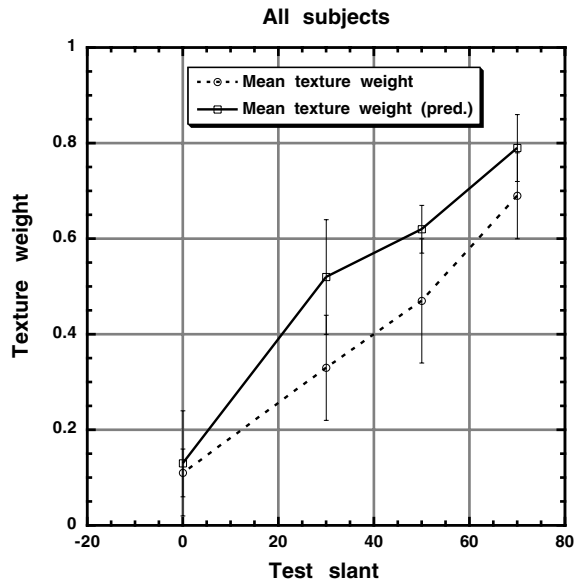

Fig. 10. Average measured texture weights as a function of test slant compared with the average weights predicted from the discrimination threshold data.

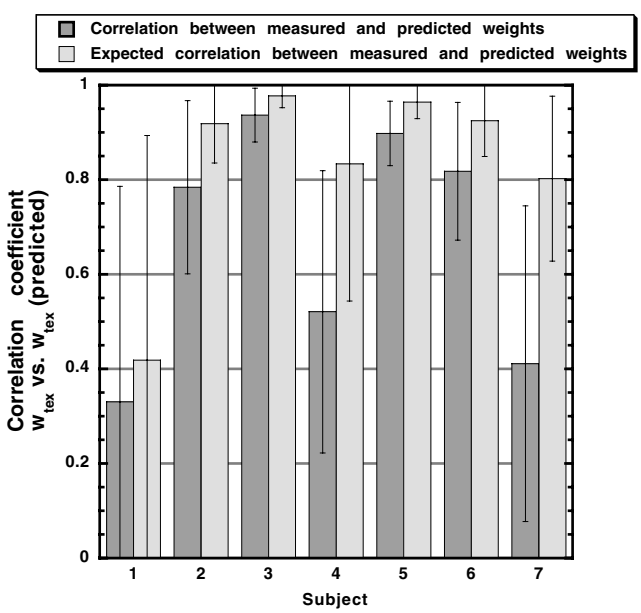

Fig. 11. The dark grey bars show the correlation between the measured and predicted texture weights across test slants for each subject. The light grey bars show the correlation that would be obtained assuming that subjects' texture cue weights were optimally related to their true slant discrimination thresholds, taking into account the noisiness of the measurements (see text for details).

these possible "true" values for the thresholds and weights, we generated simulated samples of the thresholds and weights that we might have measured in our experiment (again using the likelihood function derived from the data in experiment 1). This, finally, provided an estimate of the threshold and weight pairs that we would have measured were we to have run the experiment many times over on any of the optimal integrators whose thresholds fit the data for a given subject in experiment 1. For each of the simulated experiments, we measured the correlation between the weights measured in the experiment and the weights computed by applying Eq. (8) to the thresholds measure in that experiment. This corresponds to a sample of the correlation that we might have measured had we run the experiment over again on an ideal integrator constrained to have thresholds fitting the data measured for a given subject. We repeated this resampling process 10,000 times to compute the average correlation coefficient that we would have expected to measure from an ideal integrator given the noisiness in our own experimental data.

The light grey bars in Fig. 11 show the correlations between measured and predicted texture weights that we would expect to have obtained form an ideal integrator constrained by the uncertainty in threshold measurements for each subject. The error bars show the std. deviation in the correlations computed across the simulated experiments, and reflect the amount of variation we might expect in the correlations we would measure for each subject were we to repeat the experiment multiple times. To a large extent, variations in the correlations measured for each subject follow those that would be predicted by the uncertainty in subjects' threshold data. We can therefore infer that the optimal integration model predicts relative changes in texture weights across slant about as well as the uncertainty in our experimental data would allow.

The previous analysis shows that for each subject, relative changes in measured texture weights are well-predicted by an optimal integrator model. We can push the question of optimality even further by asking whether the differences in the weights that individual subjects give to texture are well predicted by individual differences in their thresholds within any given test slant condition. Fig. 12 shows scatter plots of subjects' measured texture weights vs. the weights predicted from their single cue threshold data, with each slant highlighted in a different color. The green diamonds, for example, show the measured texture weight at 30° for all seven subjects, plotted as a function of the weight predicted by their discrimination thresholds. Looking separately at each color, shows that, for each test slant, individual differences in texture weights do appear to covary with individual differences in thresholds.

To quantify this effect, we measured, for each test slant, the correlation between measured and predicted texture weights across the seven subjects. Fig. 13 shows the measured correlations as dark grey bars. All four correlation coefficients were significantly greater than zero at the $p < 0.05$ level. Using a procedure exactly analagous to that used in the previous analysis, we computed the correlations that would be predicted were subjects to have been truly optimal, given the uncertainty in our threshold and weight measures. These

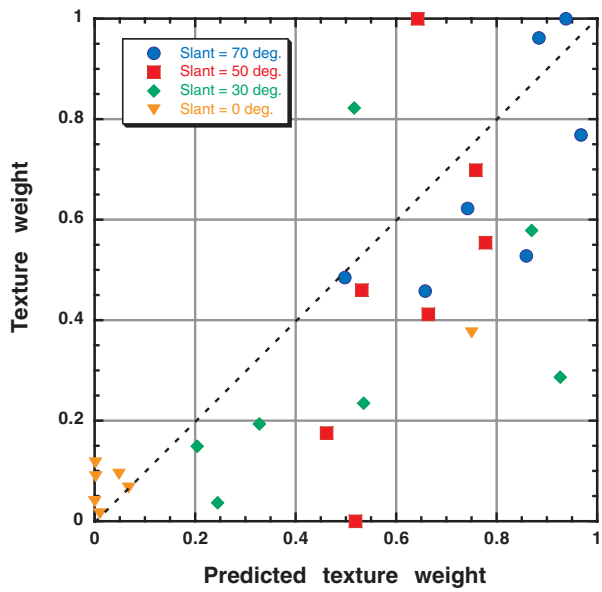

Fig. 12. Scatter plot of measured texture weights as a function of predicted texture weights. The dashed curve shows the predicted linear relationship (slope = 1) between the two. Weights for each test slanted are highlighted in a different color to show that for each test slant, the slant discrimination thresholds (from which predicted weights were derived) predict individual differences in subjects' weights.

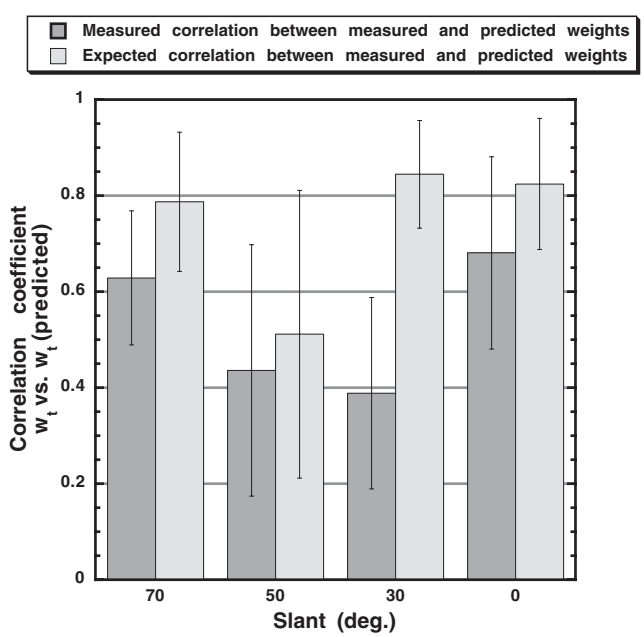

Fig. 13. The dark grey bars show the correlation between the measured and predicted texture weights across subjects for each test slant. The light grey bars show the correlation that would be obtained assuming that subjects' texture cue weights were optimally related to their true slant discrimination thresholds, taking into account the noisiness of the measurements (see text for details).

values are shown as light grey bars. On average, the correlations between measured and predicted weights are somewhat lower than those that would have been predicted by the optimal model, but only marginally so (except at 30°).

6. General discussion

The weight given by subjects to texture information increased dramatically with increasing surface slant. This increase was largely predicted by slant discrimination thresholds at each slant, which show that the subjective uncertainty in slant from texture becomes less than the uncertainty in slant from stereo at high slants (slants greater than 30°, on average). Moreover, individual differences in subjects' cue weights are well correlated with individual differences in their slant discrimination thresholds. The results are thus generally consistent with the hypothesis that humans integrate texture and stereo cues to surface slant in a subjectively optimal way. The one possible deviation from optimality in the data is that subjects tended to give slightly less weight to texture than would be predicted by the discrimination data. Before discussing the implications of these results, however, we need to critically evaluate some of the assumptions of our analysis in light of the data.

6.1. Modeling assumptions

6.1.1. The Gaussian discrimination model

The psychometric model we used to model subjects' judgments effectively assumed that perceived slant from both texture and stereo are corrupted by Gaussian noise that has constant variance within the range of slants used to create stimuli around each test value. Subjects' thresholds, however, are not constant as a function of slant, indicating that the uncertainty in perceived slant for any given stimulus is skewed around that slant. This is particularly true for the texture cue, for which discrimination thresholds shrink by more than an order of magnitude from 0° to 70°. Thus, for texture-only stimuli, the underlying noise model should have increasing variance with slant. Unfortunately, the amount of data collected in the experiments did not support reliable estimates of a skew parameter in the psychometric model (as was done, for example, in Knill, 1998b). The threshold measures, therefore, reflect an average uncertainty around the test slant.

One implication of this is that the optimal model for combining texture and stereo cues is not linear. Rather, the linear weights are a first-order fit to the non-linear combination rule around each test slant. For cue conflict stimuli in which the stereo information is fixed to suggest one slant, we should, in theory, be able to measure smaller weights for the texture cue when the texture cue suggests a smaller slant than when it suggests a larger slant. Again, the data did not support accurate measures

of this type of asymmetry in the weights. Our measurements should be treated as first-order effects near a given test stimulus. Some of the difference between predicted and measured weights may be due to the nonlinearity of the truly optimal model. More focused tests would be needed to test this possibility.

6.1.2. The relationship between thresholds and cue uncertainty

A second assumption of our analysis was that slant discrimination thresholds accurately reflect subjects' perceptual uncertainty about slant. In particular, the predictions derived from the threshold data were based on the assumption that thresholds are proportional to the standard deviation of internal slant estimates. In reality, discrimination thresholds will reflect other sources of uncertainty such as high-level decision noise. We have modeled some of this explicitly by including parameters in our psychometric model for attentional lapses and guessing, but other forms of high-level noise probably corrupt subjects' judgments. The common way to model such high-level effects is to assume that the decision process effectively adds an independent noise source to perceptual estimates. Assuming that decision noise corrupts the integrated estimate of slant derived from all available cues in the image, the presence of such noise changes the predicted relationship between thresholds and cue weights.

Thresholds should be modeled as being proportional to the total noise in the system, given by

$$T_i(S) = k\sqrt{\sigma_i^2(S) + \sigma_N^2(S)}, \tag{16}$$

where $T_i(S)$ is the discrimination threshold for a test slant, S, under cue condition i (e.g. stereo-only, texture-only or stereo-and-texture), $\sigma_i(S)$ is the standard deviation of the internal estimates of slant under this cue condition and $\sigma_N(S)$ is the standard deviation of an additive noise source that models the effects of high-level decision uncertainty.

The predictions that we have shown for cue weights were derived by effectively assuming that $\sigma_N(S)$ was negligible and could be set to 0. To understand the effects of additive decision noise on the predicted relationship between thresholds and cue weights in our simplified model, we simulated an ideal observer with several variants of decision noise (constant variance and variance proportional to the variance of the slant estimate derived from the stimulus). Even high levels of decision noise had only a small effect on the relationship between the ideal observer's true weights and those predicted from the incorrect assumption that thresholds are not affected by decision noise. Fig. 14 shows an example in which the decision noise variance $\sigma_N^2(S)$ was assumed to be equal to the variance of the internal estimate of slant derived from a combined cue stimulus. Thus, while subjects in the experiment undoubtedly were effected by some amount of high-level decision noise, this noise was unlikely to have significantly impacted the measured relationship between thresholds and cue weights.

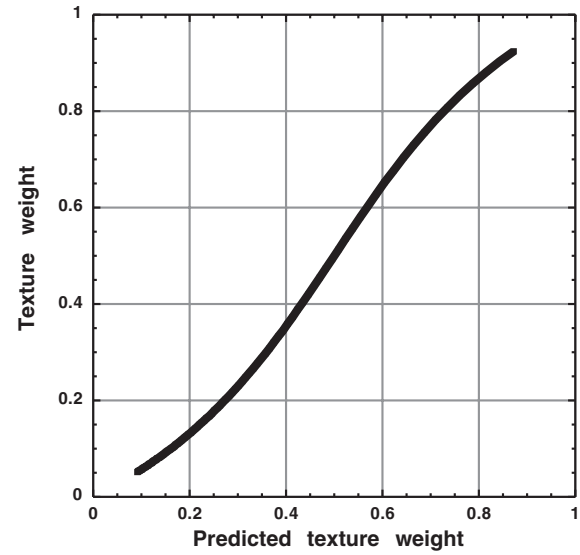

Fig. 14. We simulated an ideal observer whose texture weights were given by Eq. (3) and whose uncertainty in slant estimates from texture and stereo varied over a large range across test slants. Discrimination thresholds at each test slant were assumed to be determined by the slant uncertainty in a given cue condition plus an additive noise factor reflecting high-level noise. For this simulation, the standard deviation of the high-level noise was set to equal the standard deviation of slant estimates derived from the combined cue stimuli at that slant. According to this model, discrimination thresholds are given by $T_{tex}(S) = k\sqrt{\sigma_{tex}^2(S) + \sigma_N^2(S)}$ and $T_{st}(S) = k\sqrt{\sigma_{st}^2(S) + \sigma_N^2(S)}$, with the additive noise variance set to $\sigma_N^2(S) = (1/\sigma_{tex}^2(S) + 1/\sigma_{st}^2(S))^{(-1)}$ (the variance of the slant estimates derived from optimal integration of texture and stereo cues). The graph plots the texture weights of an ideal observer (with $w_{tex} = \sigma_{st}^2(S)/\sigma_{st}^2(S) + \sigma_{tex}^2(S)$) as a function of the weights predicted from the approximation, $w_{tex} = T_{st}^2(S)/T_{st}^2(S) + T_{tex}^2(S)$. Even though the decision noise level was high, the curve does not deviate very strongly from a linear slope of 1.

6.1.3. Generalizing from random-dot stereoscopic stimuli

A serious concern for our interpretation of the threshold data is the degree to which the thresholds measured for the stimuli containing stereoscopic views of random-dot textures accurately reflected the stereo uncertainty in the stimuli used to estimate cue weights—stereoscopic views of randomly tiled textures. The control experiment effectively dealt with the issue of the texture information contained in the random-dot stimuli. To the extent that it was used, it would not significantly impact our predictions. A more serious concern is that the stereo information in the randomly tiled texture stimuli may have been qualitatively better than is available in the random-dot stimuli. Were this true, our estimates of stereo cue uncertainty in the stimuli used to

estimate cue weights would be higher than the true values. This could explain why subjects appear to give less weight to texture (hence, more weight to stereo) than would be predicted by our threshold data. The fact that subjects do not perform measurably better in the combined cue stimuli than predicted by the single cue threshold data argues against this interpretation; however, it remains a possibility, since the super-additive effect of having improved stereo information in the combined-cue stimuli could have been counteracted by the sub-additive effects of any putative high-level decision noise.

6.1.4. Learning

The analysis presented here relies on subjects' cue weights remaining stable over the time course of both experiments. Jacobs and colleagues have performed a number of experiments showing that subjects can effectively modify the weights that they give to visual cues over a short-time scale, when given feedback, either haptic (Atkins, Fiser, & Jacobs, 2001) or auditory (Jacobs & Fine, 1999), that is consistent with one of the cues in a set of cue conflict stimuli. That such learning could occur here sees unlikely, as subjects receive no feedback in either part of the experiment. It remains possible, however, that experiencing a large number of single cue stimuli in the first experiment could lead to a change in cue weights over time. Similarly, experience of the cue conflict stimuli could potentially lead to changes in weights that would violate the stationarity assumptions of our analysis. Since no feedback was given in either of the experiments and, at least in experiment 2, all cue conflict stimuli were inter-mixed in experimental sessions, it is unclear how such learning would occur or what changes such learning would lead to. One possibility is that subjects simply become better at using either texture or stereo information over the time course of experiment 1—a form of passive perceptual learning. Since thresholds were estimated assuming stationarity over time, it is possible that the threshold estimates are a biased reflection of the uncertainty that applies to subjects' interpretation of slant in experiment 2. We have looked at threshold estimates derived form the first half of experiment 1 as compared to the second half and found no consistent pattern across subjects; however, the reliability of the data make fine learning effects impossible to pull out of this analysis. Beyond this type of effect no rational principles exist to suggest a particular pattern for weight changes, thus, we expect that our stationarity assumptions are, at least to a first approximation, reasonable.

6.2. Underlying mechanisms

We took pains in the introduction to remain agnostic about the mechanisms underlying cue integration. In part, this was because psychophysical measurements of cue weights do not, in themselves, tell us much about mechanism. More importantly, we believe that interpreting the linear model as a direct reflection of computational structures built into visual processing is somewhat implausible. The problem considered here, in which the uncertainty of a pair of cues varies with the scene parameter being estimated, highlights this—the notion of a system explicitly adjusting cue weights based on cue uncertainty seems to require ancillary cues (e.g. vergence angle for depth, measures of the noisiness in image measurements, etc.) for measuring this uncertainty. Such ancillary cues are not available in the context of the current phenomenon—estimating cue uncertainty requires an implicit estimate of slant, as the two covary so strongly. Performing this computation independently of estimating slant would appear to be inefficient at best.

Alternatively, separate modules for slant from texture and slant from stereo could output estimates of uncertainty along with their estimates of slant. These uncertainty estimates could be explicitly used to adjust the weights used to combine the two estimates. Of course, this approach would only support linear integration and would be difficult to reconcile with problems that require non-linear cue interactions (Knill, 2003; Saunders & Knill, 2001; Yuille & Bulthoff, 1996; Yuille & Clark, 1993). Several modern theories of neural population coding provide an alternative approach in which apparent re-weighting of cues results implicitly from combining separate population codes derived from each cue that implicitly code estimator uncertainty. The most straightforward approach would be to use population codes to represent likelihood functions (Zemel, Dayan, & Pouget, 1998). Appropriate combination strategies would then support the "multiplication" of individual cue likelihood functions to arrive at a joint likelihood function for any given scene parameter.

Ernst and Banks described a particularly simple model for this, in which different neural populations code object size as estimated from different cues. The firing rates of cells tuned to different object sizes would directly code the likelihood of that size. Simple multiplication of the firing rates of two such populations would give a new population code in which the joint log-likelihood function would be represented by the firing rates in a "higher-level" population of cells. As they noted, this specific instantiation of a population code for likelihood functions has many limitations; however, it effectively conveys the general form such a computation might take. Recently, Deneve, Latham, and Pouget (2001) have proposed an alternative form of neural cue integration in which a dynamic network with a middle layer of basis function units can be shown to compute maximum likelihood estimates of scene parameters from multiple cues, even when the integration is inherently

non-linear. All of these ideas have in common the property that cue uncertainty is computed and represented in populations of neurons and that computations on these populations implicitly take this uncertainty into account. Certainly, this provides a more parsimonious account of the current data than one in which separate systems exist to estimate cue uncertainty.

6.3. Implications for depth perception in the natural world

The paper has focused primarily on the broad question of whether the visual system optimally integrates multiple visual cues to estimate 3D surface geometry. The results, however, also speak to the basic question of when texture cues will significantly contribute to human perception of three-dimensional spatial layout. While some researchers have found small weights for texture relative to stereo, others have found larger weights. Rather than being contradictory, the results elucidate those stimulus conditions in which texture information is and is not an effective cue to 3D surface geometry. It is clear from these and other results (Frisby, Buckley, & Freeman, 1992, 1996) that texture is a highly salient cue to planar surface orientation when surfaces are slanted significantly away from the fronto-parallel. Other researchers have studied texture and stereo cue integration for surface curvature. These results suggest that texture is a weak cue when surfaces curve in a plane aligned with the line of sight (e.g. when lines of curvature project to straight lines in the image, as with an upright cylinder) (Johnston et al., 1993). When surfaces are oriented so that the surface curves in a direction not aligned with such a plane, the curvature becomes more apparent in the curvilinear distortion of textures and texture becomes a stronger cue (Frisby et al., 1996). We have performed a number of ideal observer simulations which suggest that this is a simple reflection of the informational structure of texture patterns, however, it may also reflect specific mechanisms tuned to apparent flow in projected texture patterns (Knill, 2001; Li & Zaidi, 2001; Zaidi & Li, 2002). In previous work we have also shown that the skew symmetry in projections of planar symmetric figures provides a stronger cue to surface orientation at high slants (Saunders & Knill, 2001). Finally, Tittle and colleagues have shown that texture and shading information dominate for judgments of curvature magnitude, while stereo disparity information dominates for judgments of the local shape index (reflecting the change from elliptical through cylindrical to hyperbolic surfaces) (Tittle, Norman, Perotti, & Phillips, 1997). Taken together, these results indicate that pictorial cues like texture and contour can provide strong cues to surface layout—sometimes stronger than stereo—even at small viewing distances, but that their importance depends on their relative uncertainty for the scene property of interest.

6.4. Conclusions

The subjective uncertainty of both stereo and texture information for surface slant varies as a function of surface slant itself. The effect is strongest for texture, which is unreliable at low slants but very reliable at high slants. For all subjects, the ratio of the texture cue uncertainty to stereo cue uncertainty decreases (texture becomes more reliable) as surface slant increases. This predicts that subjects should effectively give progressively more weight to texture information as surface slant increases when estimating slant. Our data confirms that subjects behave in exactly this way. Subjects' only deviation from optimality is that they give somewhat more weight to stereo on average than the threshold data would predict. While this may reflect some degree of suboptimality in the visual system, it might also reflect a mismatch between the stereo information in the stimuli used to measure stereo uncertainty and the stimuli used to measure cue weights. Subjects also show large individual differences both in the uncertainty with which they can make slant judgments from individual cues and in the relative weights that they give to the cues. Much of the variance in the weight differences, however, is accounted for by the differences in subjective cue uncertainty. Taken together the results of the current experiments are consistent with the hypothesis that the human visual system is a subjectively ideal cue integrator; that is, that its cue integration behavior is determined by the low level uncertainty in its ability to use individual cues as information about slant.

References

Atkins, J. E., Fiser, J., & Jacobs, R. A. (2001). Experience-dependent visual cue integration based on consistencies between visual and haptic percepts. *Vision Research, 41*, 449–461.

Banks, M. S., Hooge, I. T. C., & Backus, B. T. (2001). Perceiving slant about a horizontal axis from stereopsis by Martin S. Banks. *Journal of Vision, 1*(2), 55–79.

Blake, A., Bulthoff, H. H., & Sheinberg, A. (1993). Shape from texture: ideal observers and human psychophysics. *Vision Research, 33*(12), 1723–1737.

Buckley, D., Frisby, J., & Blake, A. (1996). Does the human visual system implement an ideal observer theory of slant from texture? *Vision Research, 36*(8), 1163–1176.

Davison, A. C. (1997). *Bootstrap methods and their application*. Cambridge, England: Cambridge University Press.

Deneve, S., Latham, P. E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience, 4*(8), 826–831.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature, 415*(6870), 429–433.

Frisby, J. P., Buckley, D., & Freeman, J. (1992). Experiments on stereo and texture cue combination in human vision using quasi-natural viewing. In G. A. Orban, & H. H. Nagel (Eds.), *Artificial and biological visual systems*. Berlin: Springer-Verlag.

Frisby, J. P., Buckley, D., & Freeman, J. (1996). Stereo and texture cue integration in the perception of planar and curved large real

surfaces. In: T. Inui, & J. L. McClelland (Eds.), *Attention and performance XVI: information and integration in perception and communication*.

Gharamani, Z., Wolpert, D. M., & Jordan, M. I. (1997). Computational models of sensori-motor integration. In P. G. Morasso, & V. Sanguineti (Eds.), *Self-organization, computational maps, and motor control*. Amsterdam: Elsevier Press.

Jacobs, R. A. (1999). Optimal integration of texture and motion cues to depth. *Vision Research, 39*, 3621–3629.

Jacobs, R. A., & Fine, I. (1999). Experience-dependent integration of texture and motion cues to depth. *Vision Research, 39*, 4062–4075.

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research, 34*(17), 2259–2275.

Johnston, E. B., Cumming, B. G., & Parker, A. J. (1993). Integration of depth modules—stereopsis and texture. *Vision Research, 33*(5–6), 813–826.

Knill, D. C. (1998a). Surface orientation from texture: ideal observers, generic observers and the information content of texture cues. *Vision Research, 38*, 1655–1682.

Knill, D. C. (1998b). Discriminating surface slant from texture: comparing human and ideal observers. *Vision Research, 38*, 1683–1711.

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

Knill, D. C. (2001). Contour into texture: the information content of surface contours and texture flow. *Journal of the Optical Society of America A, 18*(1), 12–36.

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

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: in defense of weak fusion. *Vision Research, 35*(3), 389–412.

Li, A., & Zaidi, Q. (2001). Information limitations in perception of shape from texture. *Vision Research, 41*(12), 1519–1534.

Li, A., & Zaidi, Q. (2002). Isotropic textures convey distance not 3-D shape. *Journal of Vision, 2*(7), 112.

Rao, C. (1973). *Linear statistical inference and its applications*. New York: John Wiley and Sons.

Rosenholtz, R., & Malik, J. (1997). Shape from texture: isotropy or homogeneity (or both)? *Vision Research, 37*(16), 2283–2294.

Saunders, J., & Knill, D. C. (2001). Perception of 3D surface orientation from skew symmetry. *Vision Research, 41*(24), 3163–3185.

Tittle, J. S., Norman, J. F., Perotti, V. J., & Phillips, F. (1997). The perception of scale-dependent and scale-independent surface structure from binocular disparity, texture and shading. *Perception, 26*, 147–166.

van Beers, R. J., Sittig, A. C., & Denier van der Gon, J. J. (1999). Integration of proprioceptive and visual position information: an experimentally supported model. *Journal of Neurophysiology, 81*, 1355–1364.

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*(18), 2685–2696.

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

Yuille, A. L., & Clark, J. J. (1993). Bayesian models, deformable templates and competitive priors. In L. Harris, & M. Jenkin (Eds.), *Spatial vision in humans and robots*. Cambridge, England: Cambridge University Press.

Zaidi, Q., & Li, A. (2002). Limitations on shape information provided by texture cues. *Vision Research, 42*(7), 815–835.

Zemel, R. S., Dayan, P., & Pouget, A. (1998). Probabilistic interpretation of population codes. *Neural Computation, 10*(2), 403–430.