Learning Bayesian priors for depth perception

David C. Knill

Center for Visual Science, University of Rochester,
Rochester, NY, 14627

How the visual system learns the statistical regularities (e.g. symmetry) needed to interpret pictorial cues to depth is one of the outstanding questions in perceptual science. We test the hypothesis that the visual system can adapt its model of the statistics of planar figures for estimating 3D surface orientation. In particular, we test whether subjects, when placed in an environment containing a large proportion of randomly shaped ellipses, learn to give less weight to a prior bias to interpret ellipses as slanted circles when making slant judgments of stereoscopically viewed ellipses. In a first experiment, subjects placed a cylinder onto a stereoscopically viewed, slanted, elliptical surface. In this experiment, subjects received full haptic feedback about the true orientation of the surface at the end of the movement. When test stimuli containing small conflicts between the circle interpretation of as figure and the slant suggested by stereoscopic disparities were intermixed with stereoscopically viewed circles, subjects gave the same weight to the circle interpretation over the course of five daily sessions. When the same test stimuli were intermixed with stereoscopic views of randomly shaped ellipses, however, subjects gave progressively lower weights to the circle interpretation of test stimuli over five daily sessions. In a second experiment, subjects showed the same effect when they made perceptual judgments of slant without receiving feedback, showing that feedback is not required for learning. We describe a Bayesian model for combining multiple visual cues to adapt the priors underlying pictorial depth cues that qualitatively accounts for the observed behavior.

Keywords: Bayes, prior, learning, cue integration, depth perception, slant perception

Introduction

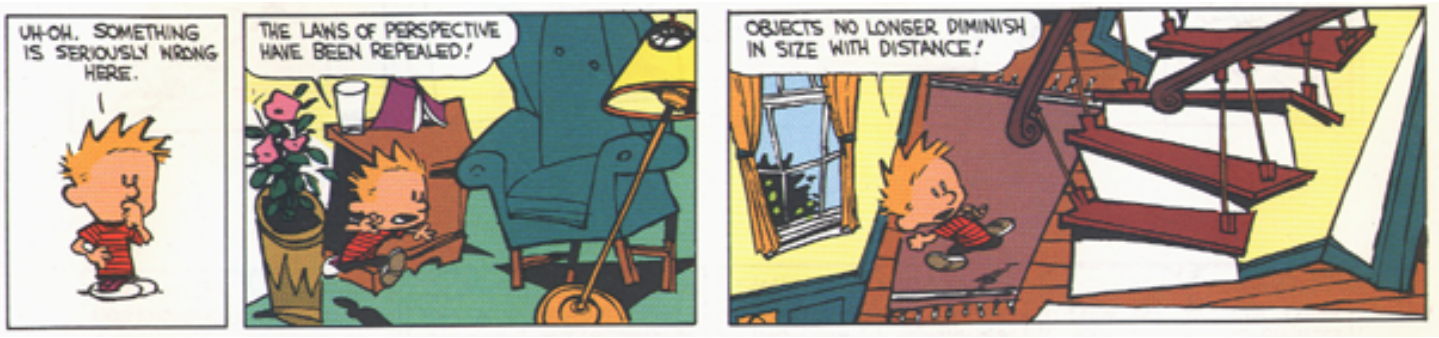


Figure 1. While Calvin interprets his inability to accurately perceive the 3D layout of the room to a failure of perspective, nothing in the drawing actually violates the laws of perspective projection. What the artist has primarily done is to randomize normally highly regular objects in the scene. The inability to apply normally used priors on these objects renders the pictorial cues to depth unreliable.

Statistical regularities in the environment play a vital role in our ability to perceive the three-dimensional layouts of scenes (see figure 1). We would have a hard time navigating the environment if polygons weren't often symmetric, ellipses weren't often circles and edges were never parallel. Pictorial cues to depth, which are usually more than adequate to create strong 3D percepts, depend critically on prior knowledge of these regularities. For example, texture cues rely on prior assumptions that surface textures are homogeneous (Malik and Rosenholtz 1997; Knill 1998) and isotropic (Witkin 1981; Garding 1993), contour cues rely on assumptions about figural symmetry (Kanade 1981; Saunders and Knill 2001)} or how contours follow the curvature of underlying surfaces (Stevens 1981; Knill 2001) and shading relies on assumptions about the light source and about surface reflectance (Ikeuchi and Horn 1981; Ramachandran 1988).

Learning to interpret pictorial cues to depth requires learning the statistical regularities that make particular visual features informative about depth relations in a scene. One way the brain could learn these regularities would be to corre-

late haptic (touch) feedback with visual features. Adams, et. al., for example, have recently shown that subjects use haptic feedback from active exploration of shaded objects in the world to adjust the prior bias on light source direction used to interpret shape-from-shading (Adams, Graf et al. 2004). Thus, subjects appear to be able to adapt biases in their internal model of the statistics of scenes. The Calvin and Hobbes' cartoon in figure 1 illustrates somewhat hyperbolically (and comically) a somewhat different problem that commonly occurs in our world – a problem created by the fact that we operate in a variety of environments, each with different levels of statistical regularity. Since statistical regularities in the environment determine the reliabilities of monocular depth cues, cue reliability can vary from environment to environment as the underlying statistics change. This implies that subjects should appear to weight monocular cues differently in different environments.

We explored the hypothesis that the human visual system can adapt its internal model of the statistical regularities that shape the information provided by a pictorial depth cue; in particular, we tested whether the influence of a pictorial cue on subjects' perceptual estimates of slant decreases as subjects learn the statistics of a highly randomized stimulus ensemble. We hypothesize that even when viewing an object that has the structure that normally makes a pictorial cue informative (e.g. is symmetric, circular or otherwise regular) in a less regular environment, the visual system will rely less on the pictorial cue relative to stereopsis, because objects are less likely overall to have that structure. Thus, for example, the foreshortened shapes of figures in monocular images of squares or circles should contribute less to perception in such an environment than in a more regular one.

We focused our research on the figural cue of foreshortening - a relatively simple but powerful cue to 3D surface orientation. Figure 2 illustrates the foreshortening cue. It is evidenced most strongly in images of elliptical figures. While a given ellipse in an image could have been projected from an entire family of ellipses in a scene, each with a different 3D orientation, human observers are biased to interpret such figures as having been projected from circles in the world. The term "foreshortening cue" typically refers to this way of interpreting elliptical images. Applied more generally, it refers to interpreting arbitrary figures in the image as having projected from figures in the world that are statistically isotropic (Brady and Yuille 1984). This bias has a strong effect on perceived slant even in the presence of conflicting binocular cues (Knill 2005). The bias to interpret ellipses as circles reflects an internalized model of shape statistics for figures in the world. In effect, human observers implement a prior on aspect ratios that is very tightly concentrated around 1. We hypothesize that after exposure to a world that contains a broader distribution of elliptical shapes, human observers will adapt their prior model appropriately. This in turn should lead to an apparent down-weighting of the foreshortening cue relative to the information about surface orientation provided by binocular disparities – even for images of circles themselves.

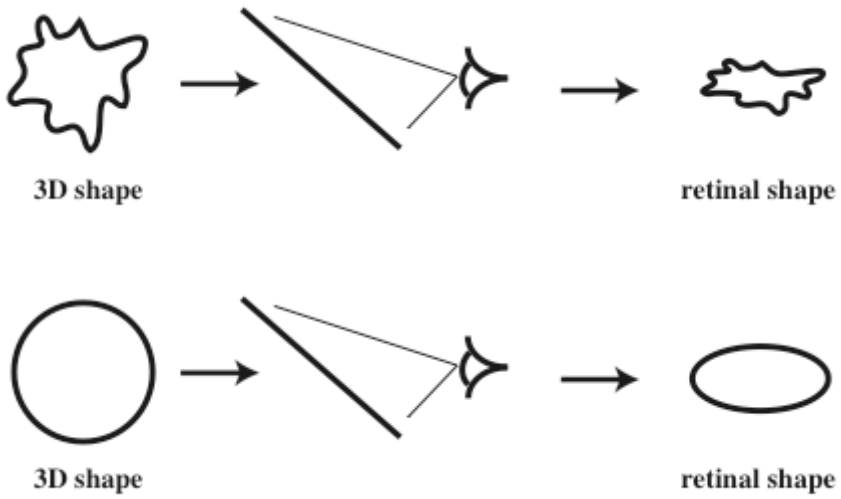


Figure 2. Planar figures, when viewed at a slant, project to compressed figures in the retinal image. A statistical tendency for figures to be isotropic (distributed evenly in all directions) renders the shapes of figures in the retinal image informative about 3D surface orientation. A figure's orentation suggests its tilt and a figure's aspect ratio suggests its slant away from the fronto-parallel.

Background

A number of studies have shown that subjects adapt the weights that they give to different visual cues after receiving haptic feedback that is consistent with one cue being a more reliable indicator of surface shape or orientation than another (Ernst, Banks et al. 1999; Atkins, Fiser et al. 2001). Researchers have generally interpreted these latter results as reflecting a mechanism that changes cue weights based on correlations between cues and objective reality, as indicated by haptic feedback. While such mechanisms may well underlie these effects, new results on robust cue integration suggest another adaptation mechanism that could underlie apparent changes in cue weights. We have recently shown that the weight that subjects give to the foreshortening cue relative to stereopsis is smaller at large cue conflicts than at small conflicts. The observed changes are consistent with a Bayesian model in which the foreshortening cue is interpreted using a mixture of priors on figure shapes in the world – some proportion are assumed to be isotropic (have an aspect ratio of 1) and another proportion are assumed to be drawn from an ensemble of figures with random aspect ratios (Knill 2007). When conflicts between the foreshortening cue (the slant suggested by an isotropic interpretation of a figure) and stereoscopic cues are large, the influence of the isotropic component of the prior model is smaller and that of the random model larger than when conflicts are small. This appears experimentally as a difference in the weights that subjects appear to give to the foreshortening cue.

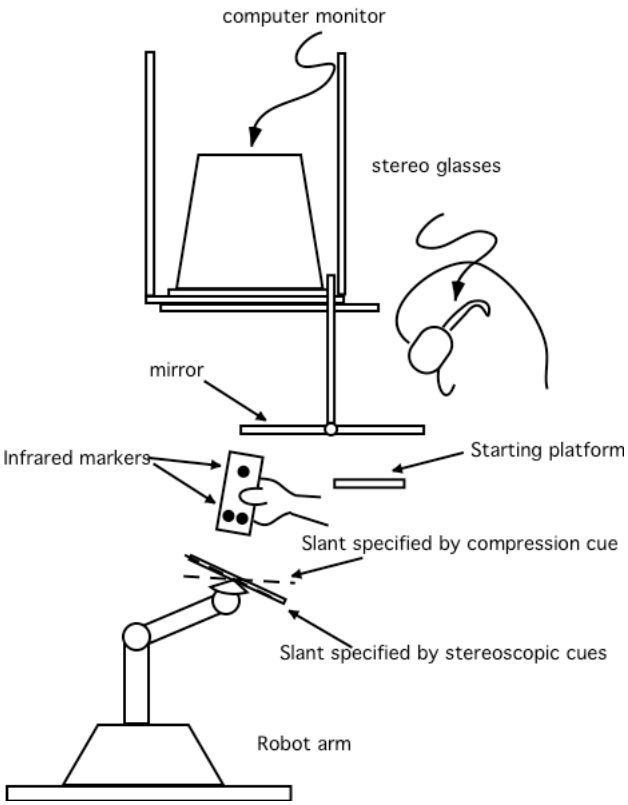
Subjects' robust non-linear cue integration behavior has several implications for studies of learning. First, it indicates that at large cue conflicts, the strong prior assumptions that underlie the informativeness of monocular depth cues may not be much "in play." The brain may effectively use different prior constraints for interpreting one or another depth cue at large conflicts than it does at small conflicts. This complicates the analysis of cue re-weighting studies, which have used stimuli with large cue conflicts for training with haptic feedback. On a more positive note, the Bayesian model for robust cue integration suggests a novel learning mechanism that would lead indirectly to apparent changes in cue weights. An optimal Bayesian estimator that incorporates a mixture of priors for interpreting a cue can, in the presence of large conflicts with other cues, more accurately estimate the properties of objects that determine the informativeness of the cue. For example, when stereoscopically viewing a slanted figure of an ellipse with an aspect ratio very different from 1 (creating a large conflict between foreshortening and stereopsis), the estimator will more accurately estimate the shape of the figure than when viewing an ellipse with an aspect ratio near 1 (small cue conflict). This is because the stereoscopic cues effectively disambiguate the shape of the figure in the large cue conflict stimulus. Estimates of figure shape derived from large cue conflict stimuli can potentially drive an adaptive mechanism for modifying the visual system's internal model of the statistics of figure shape without haptic feedback obtained from interacting with objects. For example, it could operate by adjusting an internal estimate of the relative proportion of circles and random ellipses in an environment. This would in turn lead to apparent changes in cue weights even for small cue conflict stimuli (Knill 2007).

Overview of experiments

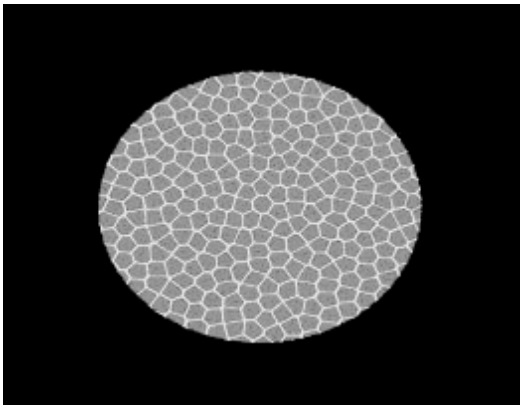
We performed four experiments to test the hypothesis that subjects can adapt their internal models of the prior statistics of figure shape and that these adaptive changes lead to changes in the apparent weights that subjects give to the foreshortening cue relative to stereoscopic cues to slant. The first experiment was initially run with the goal of testing whether haptic feedback about surface slant could drive adaptive changes in cue weights when subjects view stimuli with only small cue conflicts between foreshortening and stereoscopic cues to slant; that is, when one can confidently treat subjects' cue integration as linear. The idea was to mimic the kinds of cue discrepancies that arise from noise alone and to determine if haptic feedback consistent with stereoscopic cues could drive adaptive changes in cue weights in this stimulus regime. Subjects performed a controlled visuomotor task in which they placed a cylinder on a stereoscopically viewed, slanted figure of a textured ellipse (see figure 3). A robot arm placed a real surface co-aligned with the surface suggested by the stereoscopic slant cues in the visual display; thus, subjects received haptic feedback at the end of their movements that was always consistent with stereoscopic cues, but was somewhat decorrelated from foreshortening cues. Under the conditions of experiment 1, a purely correlational adaptation mechanism could lead to adaptive changes in cue weights; however, the alternative mechanism described above, in which figure shape estimates derived from a robust Bayesian estimator adapt subjects' internal model of the statistics of figure shapes, should not show an adaptation effect. This is because the stimuli were all consistent, within the limits of sensory noise, with viewing slanted circles. Subjects ran in the experiment over five days to track adaptive changes in their cue weights over time.

In the context of the current paper, experiment 1 served primarily as a control for the second experiment, in which we modified the statistics of the figure ensemble from which stimuli were drawn for the same experimental task. Since no significant changes in cue weights were found in experiment 1, we could be more certain that any observed changes in cue

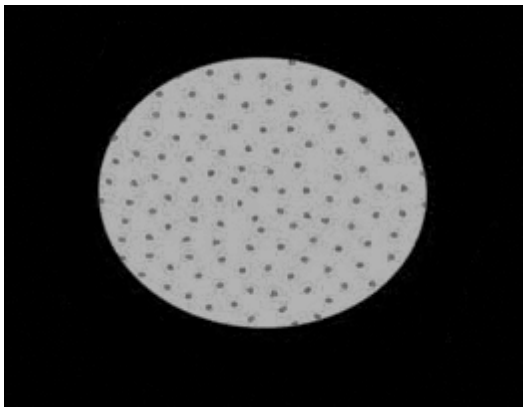
weights found in experiment 2 would not have been due to general adaptation mechanisms, like subjects getting accustomed to the virtual display or the motor task. Experiment 2 was designed to test whether showing subjects stimuli from a more random ensemble of figure shapes would lead to a gradual decrease in the weights that subjects give to foreshortening cues, as predicted by the hypothesis that subjects use robust estimates of figure shape to adapt their internal model of shape statistics. The experiment essentially repeated the conditions of experiment 1, but the test stimuli used to compute cue weights in experiment 2 (with small cue conflicts) were embedded with stimuli containing images of randomly shaped ellipses presented at different stereoscopic slants. In order to enhance the role of figure shape as a slant cue, we modified the textures by shrinking the randomly shaped tiles used in the experiment 1 stimuli to very small sizes, making the textures random fields of randomly shaped dots (see figure 3c). Placed in this environment, subjects showed a significant decrease in the weight that they appeared to give to the foreshortening cue over the course of five days.



(a)



(b)



(c)

Figure 3. (a) The apparatus used for experiments 1 and 2. (b) An example of an elliptical figure as it would appear to a subject in experiment 1. (c) An example stimulus in experiment 2, in which the size of the texture elements has been shrunk to render it a random dot texture – a stimulus that provides stereoscopic disparity slant cues without significant texture cues.

Experiment 2 was essentially a laboratory replication of the situation Calvin find himself in the cartoon strip shown in figure 1, with experiment 1 as a control. In both of these experiments, subjects received haptic feedback about the "true" state of the world, as indicated by stereoscopic disparities. Experiments 3 and 4 were designed to test whether haptic feedback was necessary for the adaptation effects to appear. In these experiments, subjects made perceptual slant judgments for stereoscopically viewed images of slanted, planar figures drawn from one of two sets of ellipses. Subjects received no feedback about their performance. In experiment 3, most of the stimuli were circles; thus, the experimental "world" in which subjects performed the task was highly regular. In experiment 4, stimuli were drawn from an irregular world containing randomly shaped ellipses. Intermixed with these "training" stimuli were test stimuli with small conflicts between the foreshortening cue and stereoscopic cues to slant. These were used to compute cue weights. The two experiments were matched in every way excepting the composition of the training stimuli. Subjects ran in five experimental sessions over

five consecutive days. Subjects showed no significant change in cue weights over time in experiment 3 (the regular world condition), but a significant decrease in weights in experiment 4 (the irregular world condition).

Experiment 1

Methods

Stimuli and procedure

Visual displays were presented in stereo on a computer monitor viewed through a mirror (figure 3) using CrystalEyes shutter glasses to present different stereo views to the left and right eyes. The left and right eye views were accurate perspective renderings of the simulated environment. Displays had a resolution of 1024x768 pixels and a refresh rate of 118Hz (59 Hz for each eye's view). Stimuli were drawn in red to take advantage of the comparatively faster red phosphor and were rendered in the center of the virtual image of the CRT in 3D space. Stimuli consisted of planar, elliptical disks filled with 120 small, randomly positioned, randomly shaped "dots". The disks had a long radius of 6 cm, so that on average, they subtended 12 degrees of visual angle. Subjects viewed the stimuli from a distance of approximately 50 cm. Randomly tiled textures were created from a set of random Voronoi patterns. To create the tile pattern, we first created twenty sets of random lattices of points by sampling from a stochastic reaction-diffusion process. The process effectively perturbed the positions of the points in the lattices away from a rectangular grid. The resulting lattices represented a trade-off between a completely random selection of points in the plan and a regular lattice that would have created linear perspective cues. The points in each random lattice were used to create a Voronoi pattern – a collection of polygons centered on the points in the lattice that tiled the plane. The randomly shaped polygons generated in this way were then shrunk to a width of 0.88 cm (~1 degree of arc), on average, to create a set of randomly shaped tiles.

The experimental apparatus consisted of a start plate off to the right of the display, the cylinder that subjects moved and a target plate mounted on the end of the robot arm placed under the mirror. The start and target plates were connected through a simple circuit to a plate mounted on the bottom of the cylinder. Subjects were instructed to place the cylinder on the start plate after completing each trial. The beginning of a trial was triggered by closing the circuit between the bottom of the cylinder and the start plate. At this point, the robot arm moved the target plate to the chosen orientation and after a period of 1 second, a new target stimulus – a stereoscopic rendering of a texture filled ellipse (as in figure 3b) - was displayed. After another 750 msec., an audible beep was given to signal the subject to move the cylinder and place it flush onto the target surface. Closing the circuit between the bottom of the cylinder and the target plate signaled the end of the trial. 1-1/2 seconds after the go signal, the target stimulus disappeared, signaling to subjects that they could move back to the starting plate. Subjects did not see the cylinder. This process was repeated until the end of a block. Infrared markers were placed on the cylinder and tracked by an Optotrak at 120 Hz throughout subjects' movements. This allowed us to compute the orientation of the cylinder as a function of time.

For each experimental session, subjects performed the object placement task for two classes of stimuli, randomly intermixed. *Test* stimuli were images of ellipses generated so that the foreshortening cues suggested a slant that differed by 5 degrees from the slant suggested by the stereoscopic cues. Both the foreshortening and stereoscopic cues in these stimuli suggested a tilt of 90 degrees; that is, the figures appeared to be rotated around a horizontal axis. Cue conflicts were generated by first projecting a circle filled with an isotropic Voronoi texture at the slant specified for the foreshortening cues into the image plane corresponding to a cyclopean view mid-way between a subjects' two eyes, and then back-projecting the resulting image onto a surface at the slant specified by the stereoscopic cues. Stereoscopic projection of the resulting figure and texture created a stimulus whose monocular foreshortening cues were consistent with one slant and whose disparities were consistent with another slant. Subjects' performance on the test stimuli (how they oriented the cylinder when placing it on the surface) provided the data for computing cue weights in a session. Test stimuli were generated around both 25 and 35 degrees and consisted of the slant pairs, *[(20, 25), (25, 20), (30, 25), (25, 30)]* and *[(30, 35), (35, 30), (40, 35), (35, 40)]*.

Non-test stimuli were presented at slants of *20, 25, 30, 35, 40,* and *45* degrees, also slanted around a horizontal axis. Non-test stimuli were circles filled with isotropic Voronoi textures, so that the foreshortening and stereoscopic cues were consistent. In the first session of the experiment, the orientation of the physical target surface positioned by the robot arm was set to a slant randomly chosen from the range defined by the two cues; thus, it was not more or less consistent with either the foreshortening or the stereoscopic slant cues. Data from the first sessions were used to derive baseline estimates of cue weights. The following four sessions were "training" sessions, in which the physical target surface was always positioned at the slant suggested by the stereoscopic cues. The haptic feedback that subjects received at the end of their

movements in the training sessions was perfectly correlated with the stereoscopic cues in the stimuli, but more weakly correlated with the foreshortening cues provided by the figure's shape and the texture pattern.

Experimental sessions lasted from five to seven days depending on whether or not a subject's sessions extended over a weekend. Many subjects, therefore, had a two-day break during the experiment. The timing of this break varied randomly across subjects. Sessions consisted of four blocks of trials and lasted approximately one hour. Each block contained 80 test stimuli and 60 non-test stimuli.

Data analysis

The slant of the cylinder just prior to making contact with the target surface provided a measure of subjects' visuomotor slant estimates. We refer to this as the contact slant. While subjects' were able to adjust both the tilt and slant of the cylinder as they moved it, we found little variance in the contact tilt of the cylinder on the test trials used to compute cue weights. Subjects' mean contact tilt on test trials was 87.9 degrees, with a std. deviation of only 2.8 degrees and there was no pattern in subjects' tilt settings as a function of cue conflict. These observations justify the use of only contact slant for computing cue weights. Subjects' contact slants on test trials provided the data for computing weights for the foreshortening and stereoscopic cues.. Cue weights were determined by regressing the slant setting on each trial against the slants suggested by the two cues using the equation,

$$\hat{\sigma} = \alpha\left[w_{compression}\sigma_{compression} + (1 - w_{compression})\sigma_{stereo} \right] + \beta \,, \qquad (1)$$

where $\hat{\sigma}$ is a subject's slant setting, $\sigma_{compression}$ is the slant suggested by the foreshortening cue on that trial, σ_{stereo} is the slant suggested by stereoscopic cues on that trial and $w_{compression}$ is the normalized weight that subjects give to the foreshortening cue relative to the stereoscopic cues. α and β are constants that capture multiplicative and constant biases in subjects' slant settings.

Subjects

Subjects were 6 undergraduates at the University of Rochester who were naive to the goals of the experiment. Subjects had normal or corrected to normal vision and had normal stereoscopic vision.

Results

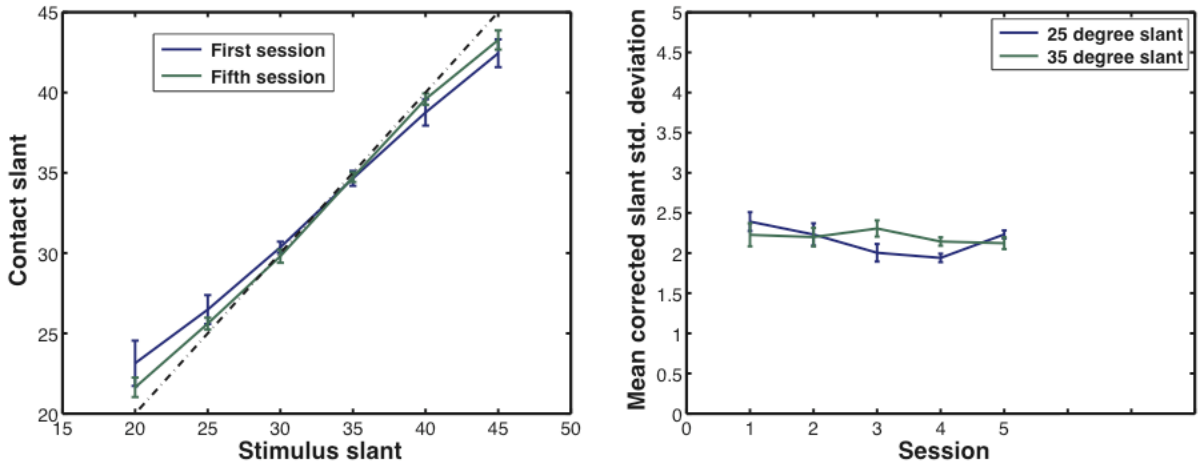
Figure 4a shows subjects' performance on the non-test (no-conflict) stimuli. Subjects' contact slants were close to the true stimulus slants for non-cue conflict stimuli. The slopes of the best-fitting linear function relating contact slant to stimulus slant changed from 0.79 to 0.89 from session 1 to session 5, but these changes were not statistically significant ($T(5) = 1.3$; $p = .25$). Subjects' performance showed little variable error. Figure 4b shows the average std. deviations in contact slants for the cue conflict stimuli used to calculate cue weights, as a function of session number. The results show that subjects performed the task with very high accuracy.

Figure 5 shows the average weights that subjects gave to the foreshortening cues as a function of time (session). Subjects' weights changed very little from the first to last session of the experiment. This is reflected in the results of a three-way ANOVA on foreshortening cue weight as a function of session, test slant and subject. Only test slant (($F(1,49) = 19.78$, $p < .001$) and subject ($F(5,49) = 22.48$, $p < .001$) had main effects on cue weights. The effect of session was not significant ($F(4,49) = 1.71$, $p > .16$). The effect of test slant was as expected from previous studies on similar or equivalent cues – subjects gave more weight to foreshortening cues at high (35 degree) slants than at low (25 degree) slants (Knill and Saunders 2003; Hillis, Watt et al. 2004; Knill 2005).

Discussion

The results show no significant change in cue weights despite the fact that subjects received haptic feedback in the last four sessions that was perfectly correlated with the slant suggested by stereopsis, but was less correlated with the slant suggested by foreshortening cues. Two factors might explain the lack of learning found here as compared to previous studies. First, because subjects in the current experiment received training over the course of five days, re-exposure to the

natural environment between experimental sessions may have suppressed any learning that might have occurred. Second, because stimuli in the experiment only had small cue conflicts (and many had no conflict), the correlational signals needed to drive adaptive changes in cue weights based on haptic feedback would have been small here.



(a) (b)

Figure 4: (a) Mean contact slant as a function of the slant of the non-cue conflict stimuli for the first (pre-training) and last sessions of experiment 1. A small bias toward the upright (approximately 38 degrees of slant in these experiments) appears in the plots. (b) The average standard deviations of subjects' contact slants in the cue conflict trials as a function of experimental session. Standard deviations were calculated for the stimuli used to compute cue weights around 25 and 35 degrees, respectively. Error bars in the figure represent the std. error of the mean of the std. deviations computed across subjects.

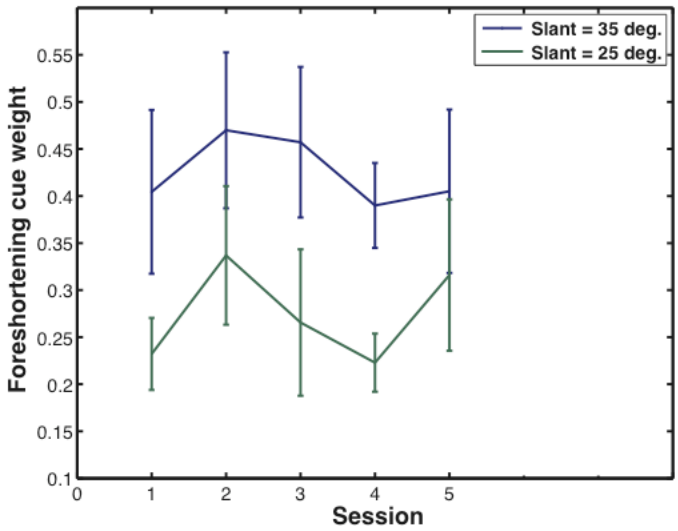


Figure 5: (a) Average foreshortening cue weights as a function of session number for the two test slant conditions. Error bars are the standard errors of the means computed across subjects.

Experiment 2

In experiment 2, subjects performed the same task as experiment 1 with a few differences in the stimuli. In experiment 1, non-test (cue-consistent) stimuli were circles presented at a range of slants. In experiment 2, we used similar stimuli for the first session, but used randomly shaped ellipses for non-test stimuli in the last four sessions. These figures

had aspect ratios ranging from 0.5 to 1 and were oriented at random angles within the plane defined by the stimulus slant. Because these stimuli largely define the statistics of the figure ensemble that subjects view in the experiment, we refer to them as training stimuli. Reframing the stimulus ensemble in experiment 1 in this context, we would say that that experiment contained training stimuli that were all circles, embedded with ellipses with aspect ratios close to 1 (small cue-conflict stimuli). In this stimulus context, subjects showed no change in cue weights over time. Experiment 2 tested whether, when exposed to an environment containing a large number of randomly shaped ellipses, subjects' behavior would show a decrease in the contribution of the foreshortening cue to their movements (a decrease in foreshortening cue weight). Because of the similarities between the two experiments, the methods section details only those aspects of experiment 2 that were different from experiment 1.

Methods

Stimuli

The textures used in the stimuli in experiment 1 were modified to reduce the salience of texture cues in the stimuli. The randomly shaped polygons that made up the textures were shrunk to an average width of 0.22 cm (~15 minutes of arc). This gave the textures the appearance of random arrays of dots (see figure 3c). Because the "dots" were really randomly shaped polygons, the local figural cues provided by the dots were minimized.

Test stimuli with 5 degree conflicts between foreshortening and stereoscopic cues were generated around only one base slant - 35 degrees – because the low weights that subjects gave to foreshortening cues at 25 degrees in experiment 1 would make adaptive changes in cue weights harder to detect in the results. Having a smaller number of test stimuli also allowed us to increase the proportion of stimuli that were random ellipses during training. Test stimuli consisted of the slant pairs (for the foreshortening and stereoscopic cues) *[(30, 35), (35, 30), 40, 35), (35, 35), (35, 40)]*.

As in experiment 1, training stimuli were figures projected stereoscopically at slants ranging from 20 – 45 degrees. In the first session, the figures were all circles. In the following four sessions, training stimuli were randomly shaped ellipses with aspect ratios drawn from a uniform distribution between 0.5 and 1 and with random orientations in the plane. Thus, subjects were exposed to a large proportion of randomly shaped ellipses in experiment 2. 74% of figures in the last four sessions were randomly shaped ellipses projected at slants ranging from 20 – 45 degrees. 26% were circles or near-circles projected at a stereoscopic slants between 30 and 40 degrees (the cue conflict stimuli).

Subjects

Subjects were eight undergraduates at the University of Rochester. All were naïve to the purposes of the experiment and had normal or corrected-to-normal vision.

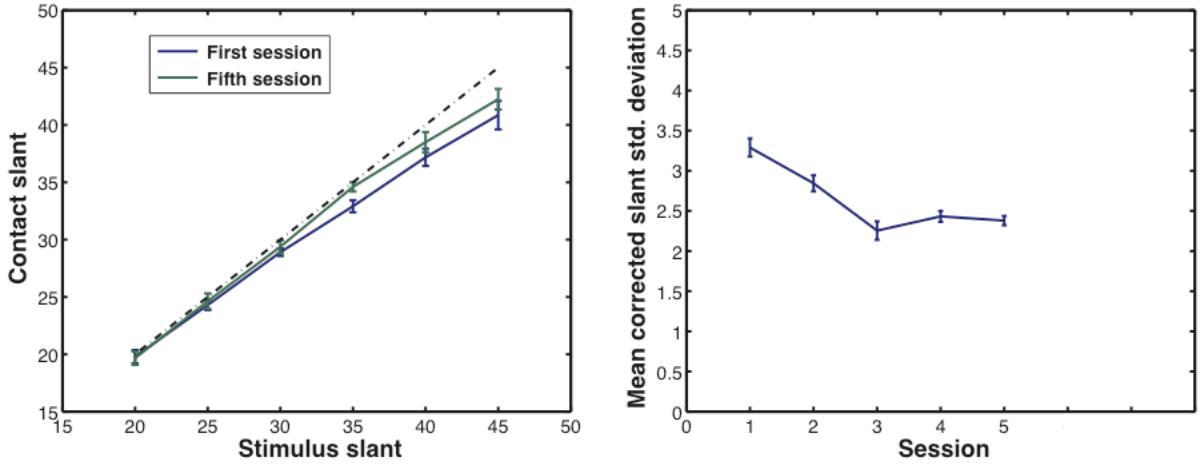
Results

Figure 6a shows subjects' performance on the non-test (no-conflict) stimuli. Subjects' contact slants were close to the true stimulus slants for non-cue conflict stimuli. The slopes of the best-fitting linear function relating contact slant to stimulus slant changed from 0.85 to 0.91 from session 1 to session 5 these changes were not significant ($T(7) = 1.18$; $p = .28$). Figure 6b shows the average std. deviation of subjects' contact slants for the cue conflict stimuli. In experiment 2, subjects showed somewhat more variable error in performance in the early sessions than in experiment 1, but subjects' asymptotic variable error was very similar in the two experiments. The results, like those of experiment 1, show high accuracy for performing the motor task.

Figure 7 shows the average foreshortening cue weights computed from subjects' data in experiment 2 as a function of session. Shown for comparison on the same graph are the average foreshortening cue weights from experiment 1. The weights that subjects' gave to the foreshortening cue decreased sharply as a function of session number in experiment 2. In order to quantify the adaptation effect, we fit a decaying exponential to individual subjects' cue weights using a non-linear least-squares-regression. The adaptation function took the form

$$w(t) = w_0 e^{-kt} \tag{2}$$

where $w(t)$ is a subjects' foreshortening cue weight computed from the data in session t, w_0 is the foreshortening cue weight after the first, baseline session ($t=0$) and k quantifies the rate of adaptation. The solid curves in Figure 7a are exponential functions parameterized by the average of w_0 and k across the subjects in each experiment. Figure 7b shows the average rate of adaptation, k, for each experiment. The rate of adaptation was significantly greater in experiment 2 than in experiment 1 ($T(12) = 4.83$, $p < .001$)..



(a) (b)

Figure 6: (a) Mean contact slant as a function of the slant of the stimulus surfaces for the non-cue conflict stimuli in the first (pre-training) and last sessions of the experiment 2. (b) The average standard deviations of subjects' contact slants for the cue conflict stimuli as a function of experimental session. The error bars in the figure represent the std. error of the mean of the corrected std. deviations computed across subjects.

Discussion

The lack of learning in experiment 1 demonstrates that neither subject's interpretation of the figural cue nor the texture cues changed with simple exposure to the stimuli or the task; for example, as would have resulted from generally adapting to the virtual stereoscopic displays. One potential complication in the interpretation of results is that the information provided by the texture pattern embedded within the figures was made consistent with the foreshortening cue in the test stimuli used to compute cue weights. Thus, the foreshortening cue weights shown in figure 6 represent a combination of figure and texture cues to slant. In experiment 2, in which the shapes of the figures used for training stimuli were randomized, the texture patterns remained consistent with the stereoscopic cues in the training stimuli. Thus, the learning effect shown by the change in weights in experiment 2 should logically have been due to changes in how subjects interpreted the figural cues, not the texture cues. Moreover, any contribution of the texture cues to subjects' performance would have mitigated the learning effect, since these cues were consistent with the stereoscopic cues and with the haptic feedback in the training stimuli.

The fact that the textures used in the stimuli for experiments 1 and 2 were different means that experiment 1 is not a perfect control for the learning effect found in experiment 2. Note, however, that subjects' foreshortening cue weights were almost exactly the same in the first sessions of both experiments. It therefore appears that either the texture information was similarly strong in both experiments or that subjects gave it relatively little weight compared to the foreshortening information provided by the figure. The latter interpretation is consistent with previous results from our lab using the same task. This study showed that subjects gave significantly less weight to the foreshortening information provided by slanted, random figures filled with Voronoi textures of the type used in experiment 1 here (relative to stereopsis) than to the foreshortening information provided by slanted circles filled with the same textures (Knill 2005), arguing that the primary carrier of slant information in the stimuli was the shape of the projected figure.

Subjects' absolute errors showed no significant change with training in both experiments 1 and 2, but their variable error in experiment 2 was higher in the initial sessions than in later sessions. This effect, however, as exaggerated by an

outlier subject whose variable error in his contact slant in the first session of experiment 2 was unusually high (std. deviation > 6.2). No such outlier subjects were apparent in the data from experiment 1.

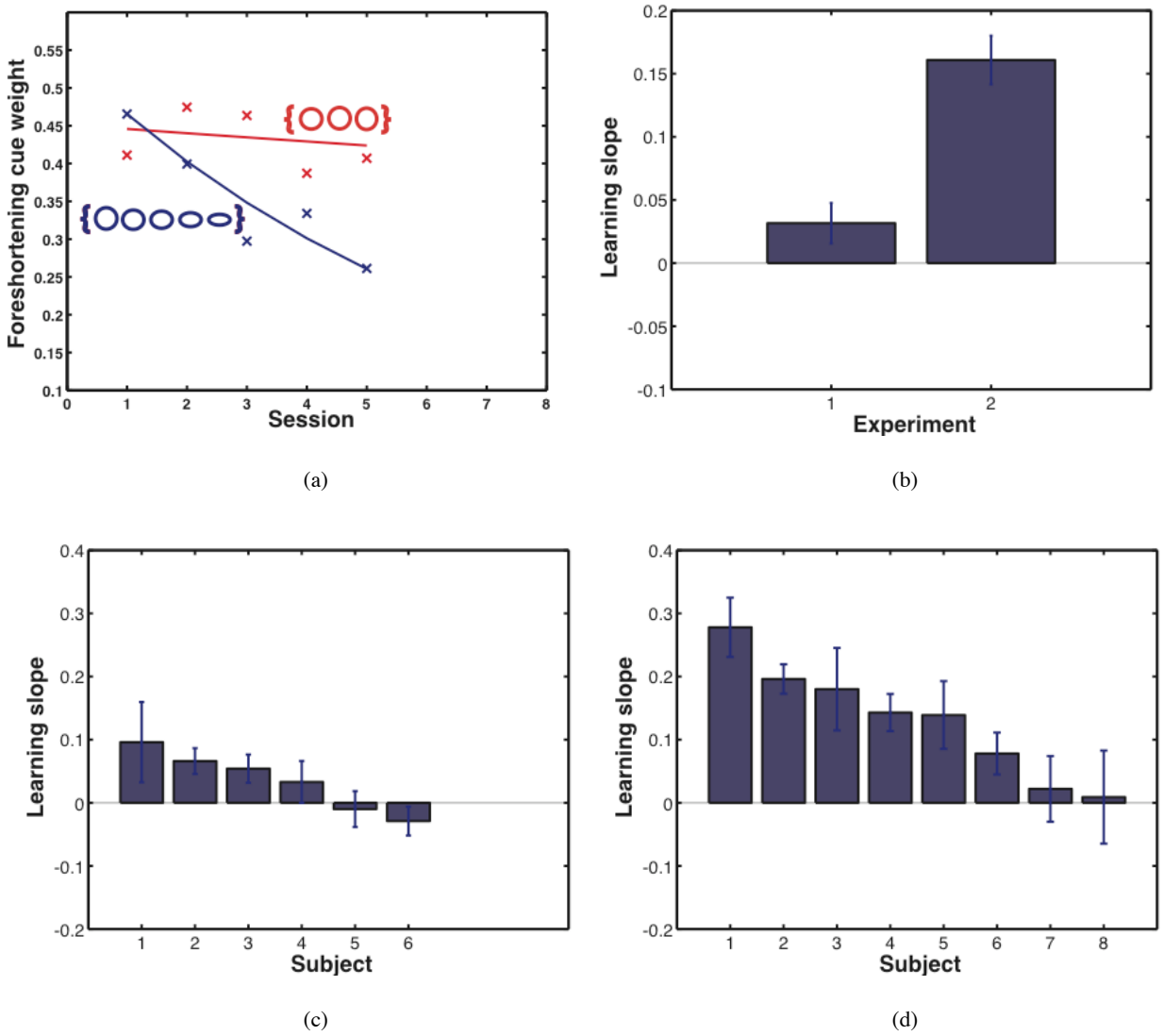


Figure 7. (a) Foreshortening cue weights computed from each session in the two experiments, with the average, best fitting decaying exponential shown as solid curves (red- experiment 1, blue – experiment 2). The ellipses drawn above each curve illustrate the distributions of shapes that subjects were exposed to in the two experiments. (b) The average rate of decay - the time constant in the exponential fit - across subjects for experiments 1 and 2. (c and d) The rate of decay for individual subjects in experiments 1 (c) and 2 (d).

In both experiments, subjects received haptic feedback that was consistent with the stereoscopic cues to orientation but was often inconsistent with the foreshortening cues. One account for the results could be that the visual system uses haptic feedback as a training signal to directly estimate the uncertainty associated with stereoscopic and foreshortening cues, respectively. This would lead to a change in the weights that the visual system gives to different visual cues. Similarly, the visual system could use the relative correlations between the haptic feedback and the two visual cues to adjust the cue weights directly. This idea has previously been posited to account for observed changes in cue weights observed in other experiments that manipulated the covariation of haptic feedback with different visual cues (Ernst, Banks et al. 1999; Atkins, Fiser et al. 2001). According to this account, the random shapes used in experiment 2 created large cue conflicts between the stereoscopic and foreshortening cues. Since the haptic feedback is yoked to the stereoscopic cues, subjects might simply learn that the foreshortening cue is an unreliable indicator of 3D surface orientation. In experiment 2, the average difference between the true orientation of the surface and the orientation suggested by the foreshortening cue in the training sessions was 30.8^{o}, creating a large error signal that could drive this kind of learning. In experiment 1, by contrast, the average error signal was only 2.7^{o}.

The introduction outlined an alternative account – that subjects adapt an internal model of the statistics of figure shapes that underlie the foreshortening cue. Such changes would indirectly lead to corresponding changes in the internal estimate of cue reliability, which could lead to a change in cue weights. In theory, the latter mechanism does not require feedback derived from interacting with the environment – whether that comes in the form of haptics or motor error signals. A robust cue integrator would, when presented with large cue conflicts between stereoscopic cues and the foreshortening cue, recognize that the figure was not a circle, rely mostly on the stereoscopic cues to estimate slant (Knill 2003) and derive an estimate of figure shape that could be used to adjust an internal model of the shape statistics of the environment. Experiments 3 and 4 test the prediction that feedback gained from interactions with objects in the environment are not needed to drive the adaptive changes in cue weights observed in experiment 2.

Experiments 3 and 4

Experiments 3 and 4 replicated the logic of experiments 1 and 2, but used a perceptual rather than motor task to estimate cue weights. Rather than have subjects place an object on a surface, subjects made an explicit perceptual judgment of the orientation of the surface displayed in a stimulus by adjusting a stereoscopically presented line probe to appear perpendicular to the surface. No feedback was given about the true orientation of the surface in the stimulus. We used the matched orientations of the line probe in place of the contact slants of the cylinder to compute foreshortening cue weights. We also regularized the timing of experimental sessions so that they were run on five consecutive days for all subjects. Finally, the textures used in both experiments 3 and 4 were the same random dot textures used in experiment 2.

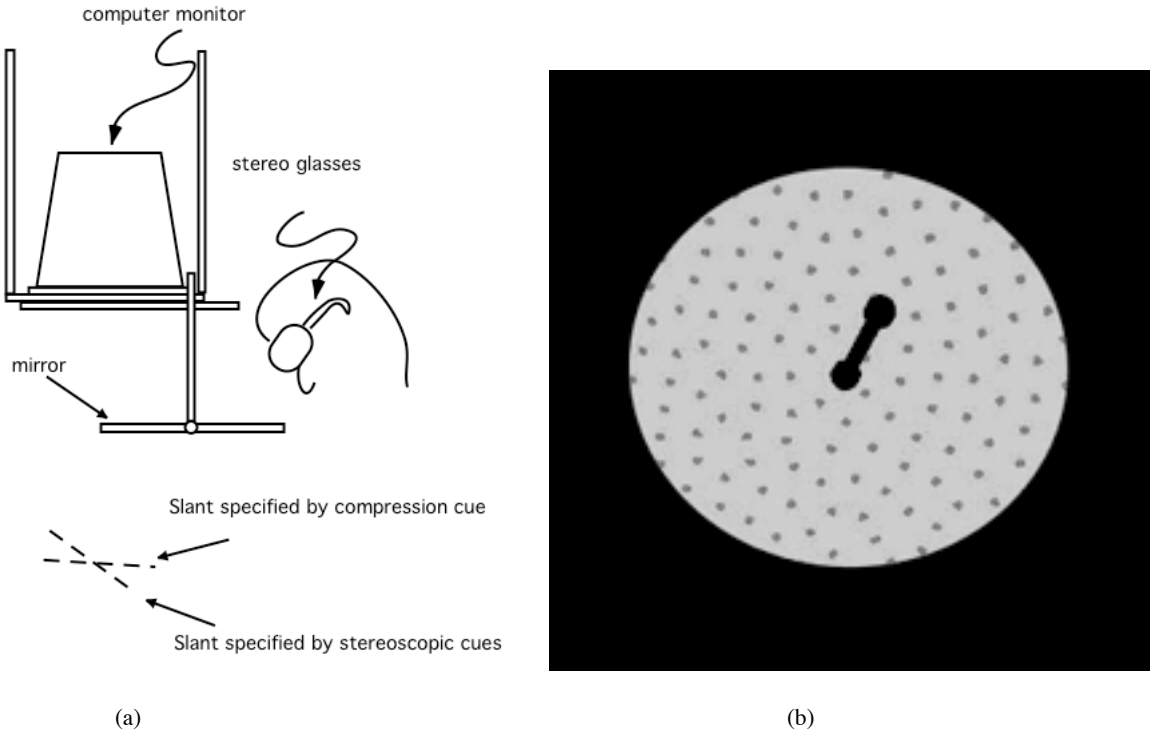


Figure 8. (a) Subjects viewed stimuli stereoscopically through a mirror, so that the stimulus appeared in three dimensions below the mirror. All stimuli in the experiment looked similar to the one in (b) – ellipses filled with a random array of dots. Subjects adjusted the orientation of a 3-dimensional probe placed on the center of the elliptical surface to appear perpendicular to the surface.

Methods

We used the same apparatus used in experiments 1 and 2, without the robot arm, to display stimuli in experiments 3 and 4. Figure 8 shows an example of the stimulus used in experiments 3 and 4. The line probe was presented in stereo. Subjects used the computer mouse to adjust the 3D orientation of the probe. Movement of the mouse was mapped to the sphere to allow full 3D rotations of the line probe. On each trial, the initial orientation of the probe was randomly selected

from an annular region on the view sphere centered on the stereoscopically defined orientation of the stimulus surface. To do this, we randomly selected an initial orientation from the view sphere subject to the constraint that $90^o < \theta < 30^o$, where θ is the angle between the surface normal and the probe. In both experiments, test stimuli contained cue conflicts around 35 degrees and consisted of the slant pairs *[(30, 35), (35, 30), (40, 35), (35, 40), (35, 35)]*. As in experiments 1 and 2, other, "training" stimuli were presented at slants of 15, 20, 25, 30, 35, 40 and 45 degrees. In the first sessions of both experiments 3 and 4, the training stimuli were all circles. In the following four sessions in experiment 3, they remained circles. In experiment 4, however, the training stimuli in the last four sessions were ellipses with aspect ratios drawn from a uniform distribution between 0.5 and 1 and with random orientations in the plane (experiment 2). The textures used in experiments 3 and 4 were the same as those used in experiment 2.

Subjects ran in five sessions each. Each session consisted of four blocks of trials, with each block containing 6 trials for each of the five test stimuli and 12 trials for each of the seven non-test, training stimuli. Data from the first session, in which the stimuli were the same in the two experiments, was used to calculate baseline measures of cue weights. The next four "training" sessions contained different sets of non-test training stimuli in the two experiments as described above. Experimental sessions were run over five consecutive days.

We used subjects' probed settings in place of the contact slants to compute cue weights from the test trials in each session, using equation (1). Different sets of eight naïve subjects were used in the two experiments. As in the first two visuomotor experiments, subjects were able to adjust both the probe's slant and tilt.

Results

Figures 9 a and b show subjects' performance on the non-test (no-conflict) stimuli for experiments 3 and 4. The slopes of the best-fitting linear function relating contact slant to stimulus slant changed from 1.69 to 1.24 from session 1 to session 5 in experiment 3 and from 1.41 to 0.97 in experiment 4. Both changes were significant (Experiment 3 – $T(7) = 3.45$, $p < .01$; Experiment 4 – $T(7) = 5.05$, $p < .002$). Figures 9 c and d shows the average std. deviation of subjects' contact slants for the cue conflict stimuli in the two experiments. Subjects' variable error in the matching task was higher by approximately 75% than their variable error in the motor task. Average tilt estimates for the cue conflict stimuli were 90.63 and 89.87 in the two experiments, with average std. deviations of 3.05 and 3.23 in the two experiments, respectively.

Figure 10 shows foreshortening cue weights calculated from subjects' slant estimates in experiments 3 and 4. The results qualitatively replicate those found in experiments 1 and 2. Subjects clearly show a greater effect of exposure time in experiment 4 (random environment) than in experiment 3 (regular environment) as shown by comparing the average learning rate derived from the exponential fits to the weights (two-tailed T test, $T(14) = 2.76, p < .02$). Figures 10 c and d show the learning rates fit to each subjects' data in the two experiments.

Discussion

The data from experiments 3 and 4 replicated those from experiments 1 and 2. The apparent weight that subjects appeared to have given to the foreshortening cue decreased over time when subjects viewed images of figures drawn from a set that contains a large proportion of randomly shaped ellipses. The results of experiment 3, in which figures were all either circles or ellipses with aspect ratios close to one, shows that the change in weights fund in experiment 4 was not a result simply of experience with the experimental task. Taken together, the results show that the learning phenomenon did not depend on feedback from interactions with the environment. The change in cue weights found in experiment 4 could not have derived from the type of correlation mechanism proposed to explain other adaptive changes in cue weights. A more plausible explanation is that subjects' visual systems adapted their internal prior on figures to match the irregularity of the world. This effect appeared both in subjects' explicit judgments of surface orientation in a purely perceptual task without any feedback to guide learning and in subjects' motor behavior when orienting their hands to place an object on a slanted surface.

Experiment 3 Experiment 4

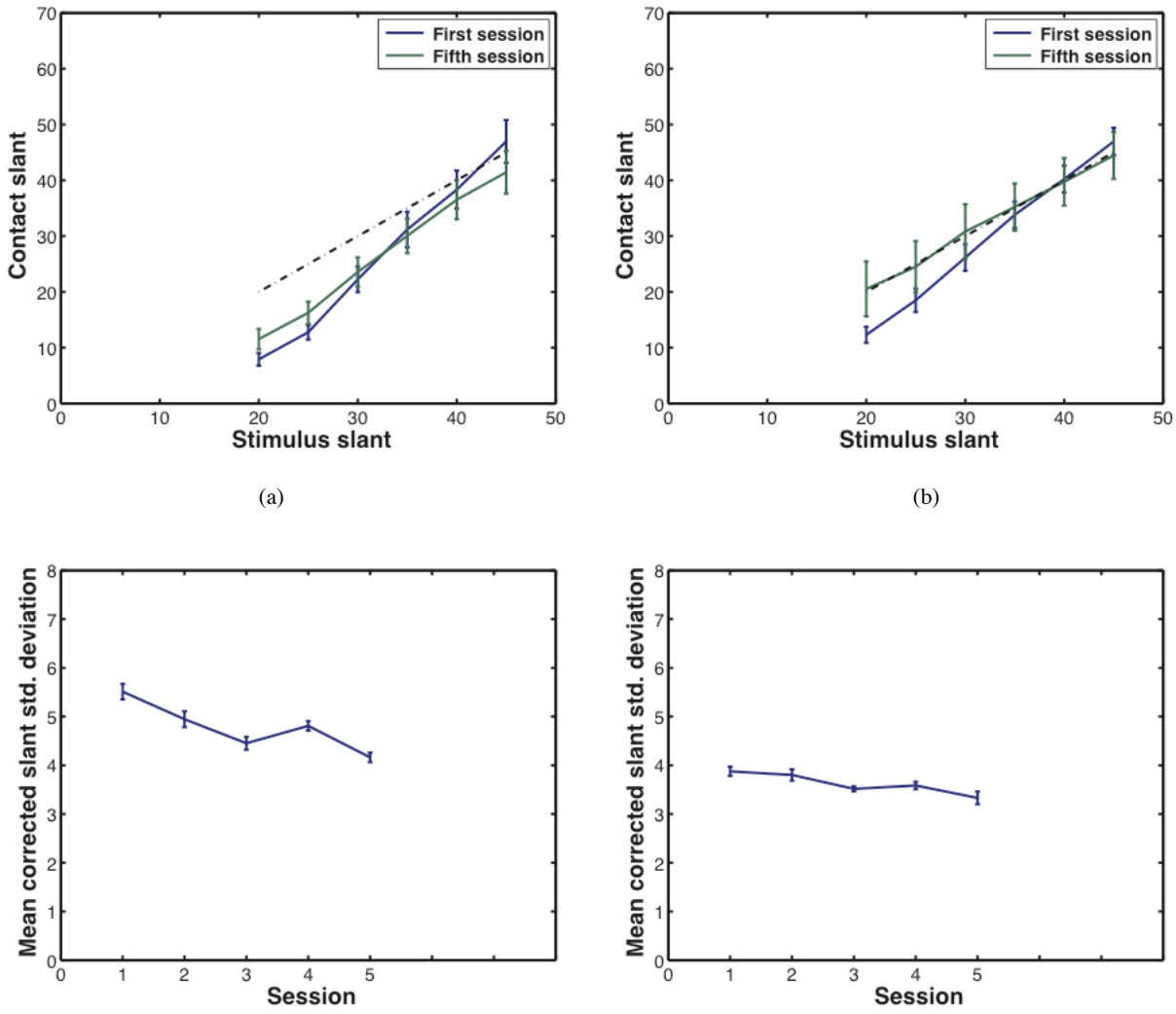


Figure 9: (a and b) Mean slant estimates for training stimuli in experiments 3 and 4, shown for the first and final sessions. (c and d) Mean std. deviations in slant estimates for the cue conflict stimuli used to estimate cue weights in experiments 3 and 4. Error bars are standard errors of the means calculated across subjects.

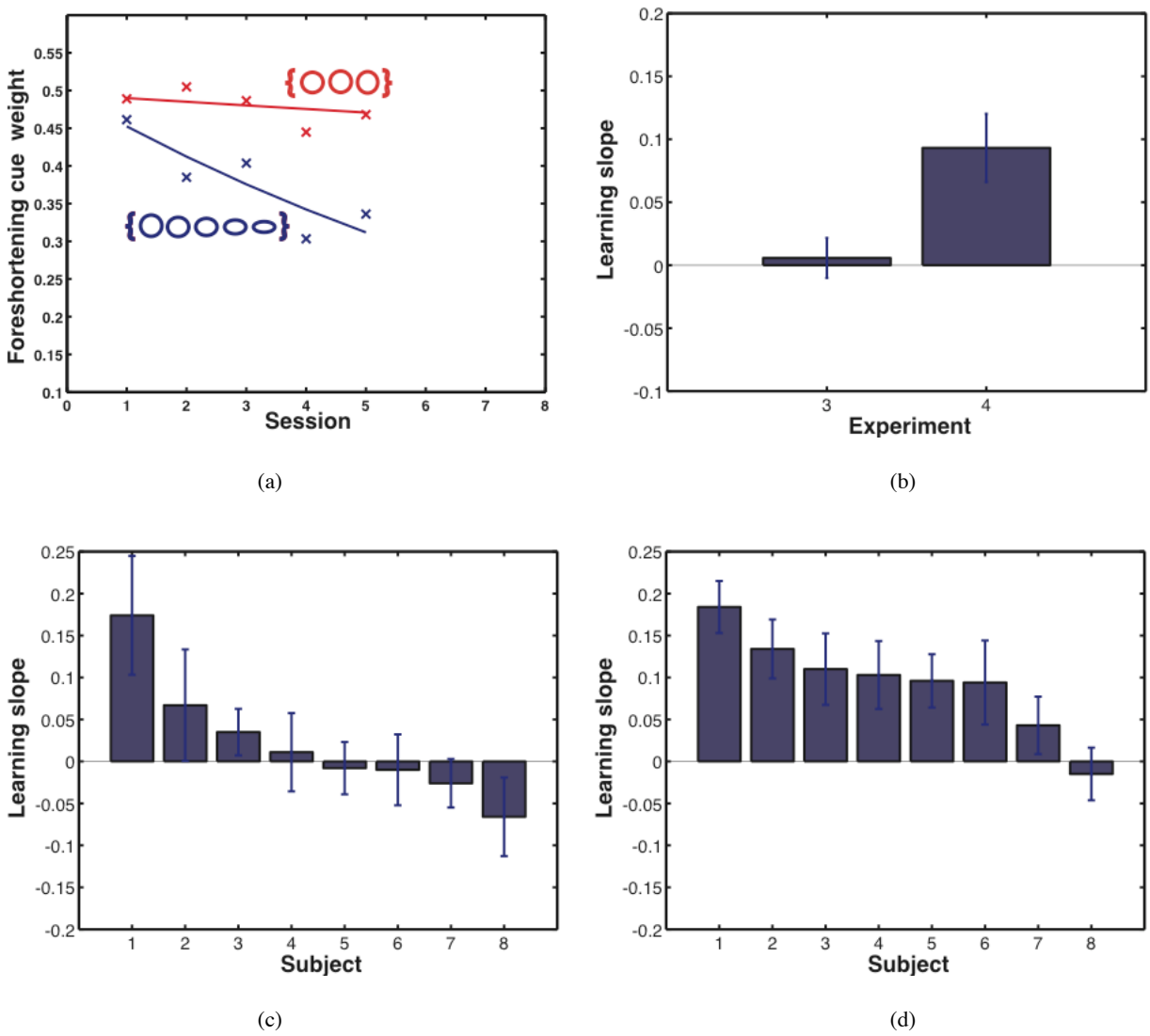


Figure 10: (a) Foreshortening cue weights computed from each session in experiments 3 and 4, with the average, best fitting decaying exponential shown as solid curves (red- experiment 3, blue – experiment 4). (b) The average rate of decay - the time constant in the exponential fit - across subjects for experiments 3 and 4. (c and d) The rate of decay in foreshortening cue weights measured for each subjects in experiments 3 (c) and 4 (d). Error bars in (c) and (d) represent the standard deviations of the maximum likelihood estimates of the slopes of the exponential functions fit to subjects' weights.

A learning model

We hypothesize that during visual exposure to a world that contains a broader distribution of elliptical shapes, human observers adapt their prior model appropriately. This in turn leads to an apparent down-weighting of the foreshortening cue relative to the information about surface orientation provided by binocular disparities – even for images of circles themselves. In order for this to work, the visual system cannot a-priori assume that all elliptical figures are circles in the world. Such an assumption would preclude the system from learning a new distribution of aspect ratios. Rather it would have to allow the possibility that at least some shapes could be elliptical.

Simple intuition suggests that the visual system does this. Not all elliptical figures are interpreted as circles; that is, when other cues like binocular disparities suggest an orientation that is very inconsistent with a circle interpretation, observers perceive a figure to be an elongated ellipse oriented at a slant and tilt close to that suggested by the binocular disparities. Subjects' unsolicited reports from experiments 2 and 4 are consistent with this phenomenon. They consistently

commented, after running in the second session of the experiments that we had added ellipses to the stimulus set, which they had seen as containing entirely circles in the first sessions. The simplest explanation for this behavior is that the visual system infers the 3D orientations and shapes of figures and that it does so using a mixed prior distribution on the aspect ratios of figures in the world. It assumes that a large percentage of ellipses in the world have an aspect ratio equal to one (are circles), but that some proportion are drawn from a significantly broader distribution of aspect ratios. Such a model can account for several experimentally observed effects, including an apparent down-weighting of monocular cues for stimuli containing large conflicts between the monocular cues and binocular cues (Knill 2006) and bimodal switching of the perceived slant of a cue conflict stimulus that is yoked to changes in the perceived shape of a stimulus (van Ee, Adams et al. 2003).

Incorporating this type of estimator into a learning model allows the system to adapt its internal model of the prior distribution of aspect ratios by (1) Estimating slant and aspect ratio using the information provided by the shape of the ellipse in the image and other cues like stereopsis and (2) using the estimated aspect ratios to update the internal model of the distribution of aspect ratios in the world, at least in a particular context. The estimator will be strongly biased to see circles for images of shapes that are close to being circular, but will be accurate for shapes that differ significantly from being circular.

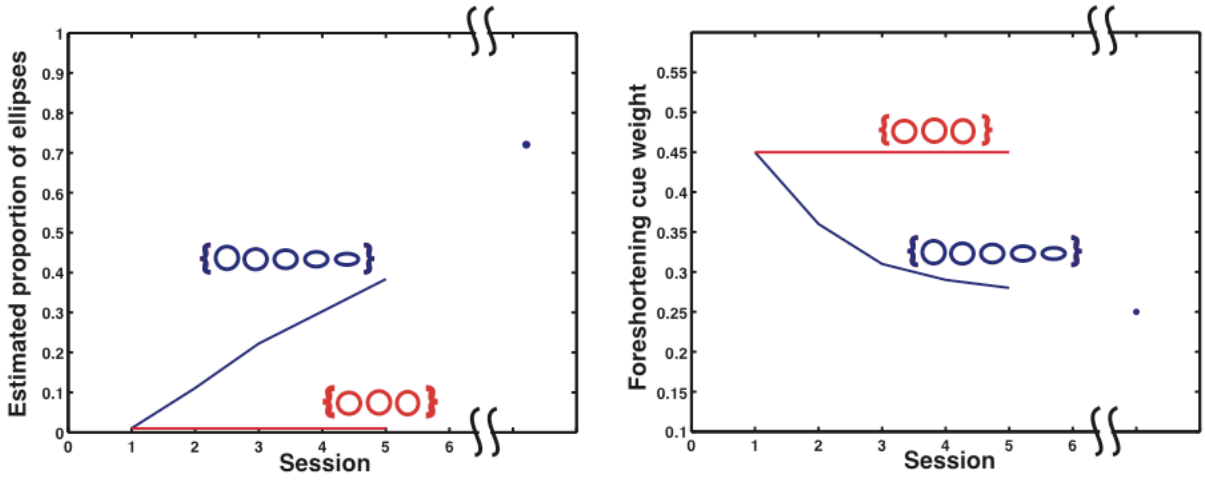
Given that the human visual system seems to have internalized a mixed prior on aspect ratios in the environment, learning can lead to two types of changes in the prior – adapting the mixing proportions in the model; that is, the relative proportion of circles and non-circular ellipses, and adapting the shape of the distribution of non-circular ellipses. In (Knill 2006), we show that increasing the proportion of ellipses assumed by the model decreases the apparent weight that the model gives to the foreshortening cue as measured using stimuli with small cue conflicts. Increasing assumed spread of aspect ratios in the random ellipse model has minimal effects on apparent cue weights. We therefore describe a form of the model that adapts its estimate of the relative mixtures of ellipses and circles in the environment.

We model learning as occurring through a Markovian estimation process akin to a Kalman filter. In order to accommodate learning, the observer must in some way assume that the mixture proportions in the prior model can change from environment to environment or over time. We developed a model in which the observer assumes that the mixing proportions can change over time according to a slow random walk. For simplicity, the random walk is not parameterized by real time, but by the number of occurrences of an elliptical stimulus in a scene. While a better model might be to assume point changes in the prior that occur when changing from one environment to another, the assumption of slow, continuous changes allows the observer to update its internal prior based only on information in the current stimulus. The optimal learner maintains an internal model of the probability distribution of the mixing parameter, conditioned on the previously viewed images of ellipses. It works by iterating through two steps. The first step is an estimation step in which the current internal model of the prior is used to estimate the slant and aspect ratio of a surface from the information provided by the shape of the projected ellipse and whatever other cues are available – in our case, stereopsis. The second step is the learning step in which the model uses the likelihood of different aspect ratios computed in the first step to update the prior on the mixing proportion. When the likelihood function is concentrated around 1, the internal model of the distribution on this parameter shifts toward a higher proportion of circles. When it is concentrated around a different aspect ratio, the prior shifts toward a higher proportion of non-circular ellipses. Learning occurs when multiple cues are available to disambiguate the shape of the figure that appears in an image. Appendix A gives the mathematical details of the model.

Figure 11 shows the performance of the learning model when presented with stereoscopic images of slanted ellipses randomly drawn from two different distributions of aspect ratios. The simulated observer begins by assuming that almost all figures are circles (a high estimate on the proportion of circles). When placed in a world that contains many non-circular ellipses whose aspect ratios are close to 1, the estimator always estimates the shapes of figures to be circular, because of its strong prior belief that most figures are circular. Thus, no learning occurs. This is illustrated by the red curve in Figure 11a When placed in a world in which the aspect ratios of figures vary over a larger range, the estimator sometimes correctly estimates the aspect ratios of figures (the light gray region shows the stimulus conditions that lead to less biased estimates of aspect ratio). This drives up its estimate of the proportion of non-circular ellipses in the image. The estimator and the learner work synergistically – as the internal estimate of the proportion of non-circular ellipses goes up, the estimator's interpretation of a figure's aspect ratio becomes less and less biased toward circular, which, in turn increases the proportion of stimuli that can drive learning. When left to run long enough, the model converges on the correct proportion of non-circular ellipses (the lone point in the right of Figure 11a).

One would intuitively expect that adaptations in the prior distribution of aspect ratios would influence how an observer integrates figural shape information with other cues like stereopsis. We presented the model with the same stimuli used in experiments 3 and 4. We regressed the model's slant estimates against the slants suggested by the foreshortening cue (the circle interpretation) and the stereoscopic cues in each simulated experimental session to estimate the influence of the foreshortening cue on the model's slant estimates, characterized as the normalized weight given to the foreshortening cue slant in the linear model. We repeated the simulation 100 times and computed the average result across all simulated experiments. As shown in Figure 11b, the "weight" given to the foreshortening cue by the model decreases from an initial

value of .45 to an asymptotic value of .25. As suggested earlier, the optimal estimator learns that it is acting in a more random environment and therefore relies less on the pictorial cue of foreshortening, though this happens indirectly through the influence of the learned distribution of aspect ratios.



(a) (b)

Figure 11. Results of simulating the optimal learning model described in the text. (a) Simulations of a model that begins assuming that 99% of ellipses in the world are circular. When presented with images of ellipses with aspect ratios equal to or near 1 (similar to experiments 1 and 3), the model shows no learning (red curve), essentially because it cannot reject the hypothesis that the ellipses are circular. When presented with images of a mixture of 26% circles or near-circles (aspect ratios near 1) and 74% ellipses drawn from a broad range of aspect ratios (0.5 to 1, as in experiments 2 and 4), the model adapts its internal estimate of the proportion of non-circular ellipses in the environment (the blue curve), eventually converging on the correct proportion (the blue dot). (b) The near-circle stimuli can be used to estimate the effective weight that the model gives to the foreshortening cue for ellipses that are close to being circular. This drops monotonically from a beginning value of .45 until it converges on an asymptotic value of .25 (the blue dot on the right of (b)).

General discussion

Learning

The data from experiments 2 and 4 clearly show that human observers progressively lower the "weight" that they give to interpreting an elliptical figure as a circle when viewing a sequence of images that suggest an environment in which a large proportion of ellipses are not circular. In our laboratory version of Calvin's world as depicted in figure 1, subjects behave as if they have adapted their internal prior on figures to match the irregularity of the world. This effect appears both in subjects' explicit judgments of surface orientation in a purely perceptual task without any feedback to guide learning and in subjects' motor behavior when orienting their hands to place an object on a slanted surface.

Central to the learning model presented here is the assumption that humans incorporate mixed priors on object parameters to interpret visual depth cues. Our perceptual phenomenology is consistent with this idea – we appear to have categorical percepts of parameters like figure shape, texture homogeneity, object rigidity (for structure form motion), etc.. Data that show an apparent downweighting of pictorial cues in the presence of large cue conflicts is consistent with this type of model (Knill 2006). When presented with stimuli containing large cue conflicts between a cue like stereopsis and a pictorial cue, a Bayesian model switches the prior model that it uses to interpret the pictorial cue to a less constrained model – effectively downweighting the cue. A more theoretical argument can be gleaned from the accuracy with which subjects perform the object placement task (which was considerably better than the perceptual matching task). Subjects' performance in this task prior to training was consistent with a visuomotor variability of only 2.5 degrees. Given that subjects gave almost equal weights to the foreshortening and stereoscopic cues, this places a lower bound on the variability with which they could have interpreted slant from the foreshortening cue of 3.5 degrees (If subjects integrate cues optimally, the std. deviation of estimates from one cue should be greater than the std. deviation of estimates from both cues

by a factor of 1.414). Experiments that measured how well humans discriminate the aspect ratios of ellipses in the image suggest that the standard deviation of humans' sensory estimates of aspect ratio in the conditions of this experiment are somewhere between 0.02 and 0.04 (Regan and Hamstra 1992). Using the lower of these values (0.02) to parameterize the noise model in an optimal Bayesian estimator of slant from the foreshortening cue (that assumes all figures are circles) results in an estimator standard deviation of approximately 3 degrees. Were subjects' sensory uncertainty in aspect ratio higher, the uncertainty in slant-from-foreshortening would be correspondingly higher. This, then, provides a rough lower limit on slant-from-foreshortening uncertainty under an assumption that one is viewing a circle. The close correspondence between this lower limit and subjects' performance argues that subjects must have effectively been imposing a hard constraint that the figures were circles, since any broadening of the prior would have led to greater estimator variance.

The visuomotor task differed from the perceptual task in that subjects in that task obtained feedback from interacting with the world that could be used for learning, since the orientation of the physical surface on which subjects placed the object was correlated with the orientation suggested by stereo disparities during learning. In this sense, the visuomotor experiments were similar in design to previous experiments that showed downweighting of pictorial cues in response to incongruent haptic feedback (Ernst, Banks et al. 1999; Atkins, Fiser et al. 2001). The current results suggest a possible reinterpretation of the earlier results. In those experiments, haptic feedback provided a sensory cue that could have disambiguated subjects' internal estimates of object parameters on which the pictorial cues depended, possibly causing a change in the internal prior on these parameters. For example, in the Ernst, et. al. study, subjects relied slightly less on texture cues to judge surface slant relative to stereoscopic cues after viewing cue-conflict stimuli in which they received haptic feedback consistent with the stereo cues compared to when they received haptic feedback consistent with the texture cues. The cue conflicts used in the experiment were large enough (30 degrees) for the haptic cues to trigger a reinterpretation of the surface texture from homogeneous (informative) to inhomogeneous (uninformative) when they were made consistent with the stereo-specified slant. This could have triggered a re-learning of the texture prior. Thus, haptic cues in these experiments may have driven learning by serving as a cue that indirectly disambiguated subjects' percepts of surface texture regularity. It also seems quite plausible, however, that both the correlative learning model described earlier and the type of prior learning described here can co-exist.

The learning demonstrated here is perceptual in nature, in the sense that it reflects changes in the mechanisms that interpret sensory input to derive estimates of scene properties. However, it is different in kind from most perceptual learning phenomena. Classic perceptual learning phenomena involve improvements in subjects' abilities to discriminate basic sensory dimensions (e.g., position, orientation) with repeated presentation of particular stimuli (Karni and Sagi 1993; Fahle, Edelman et al. 1995). These changes can often be attributed in part to adapting low-level sensory representations to the first-order statistics of image features. Recent results have further shown that subjects can learn second-order statistical contingincies between image features (likelihoods of co-occurrence) (Fiser and Aslin 2001; Fiser and Aslin 2002) and that infants can learn bimodal distributions of auditory features to support phonemic categorization (Maye, Merker et al. 2002). These results show adaptations of sensory representations to statistics of the sensory input. The current results demonstrate a form of visual learning that reflects dynamic tuning of inferential processes to the statistics of the world.

While we have demonstrated that subjects can learn, without haptic feedback, a new prior on the shapes of figures in the world, we would not suggest that they carry this new prior into visual processing in their normal activities outside the lab. Presumably, after leaving the lab each day, they experience figures that conform to the statistical prior that they brought into the lab on the first day. What is striking about the results is that they do not show a re-adaptation of the prior from session to session. That is, the results show a monotonic decrease in the weight that subjects give to the foreshortening cue despite the large breaks between experimental sessions. This suggests that subjects have learned a context-specific prior, which further suggests that our visual systems are flexible enough to apply different prior models to scenes in different contexts.

Primacy of stereopsis in the model

The learning model presented here assumes in some sense that stereopsis has a special place in the pantheon of depth cues; in particular, that when faced with large conflicts between pictorial cues and stereoscopic cues, subjects will necessarily down-weight the pictorial cues. In the Bayesian model, the primacy of stereopsis derives from the fact that pictorial cues can be interpreted according to any of several prior assumptions about scenes, at least some of which are generic, making the cues significantly less informative about surfaces. The same would not seem to hold true of stereopsis, which relies only on geometric relationships in the viewing geometry. In the context described in this paper – stereoscopic viewing of slanted figures – the primacy of stereopsis appears to hold. Subjects, when faced with large conflicts between stereopsis and foreshortening cues "re-interpret" the foreshortening cues by seeing the figures as slanted, non-circular ellipses. This is certainly what subjects report phenomenally, remarking after the second session of the experiments I which

randomly shaped ellipses were used as training stimuli that we had changed the stimuli to include ellipses as well as circles.

We should point out that stereopsis is not immutable. Wallach, et. al. showed that when subjects view rotating objects through a telestereoscope (effectively increasing interocular distance) subjects adapt to interpret stereoscopic depth in accord with the structure suggested by motion cues (Wallach, Moore et al. 1963). Similarly, when subjects wear a magnifying lens over one eye for several days, they adapt their interpretations of slant and shape from stereo in a way suggesting a the assumed relationship between disparity and relative depth (Epstein and Morgan 1970; Adams, Banks et al. 2001). These effects, however, reflect slow adaptations to constant directional changes in the gain factor needed to infer relative depth from disparities. They may reflect the operation of adaptation processes that exist to adjust for drift in the multiplicative relationship between disparity and relative depth. In the current experiment, no consistent directional change in this relationship existed – sometimes foreshortening cues suggested a slant greater than stereopsis, sometimes they suggested a slant less than stereopsis. Thus, classical adaptation mechanisms meant to maintain calibration between stereopsis in the real world would not lead to the effects seen here. This type of re-calibration is fundamentally different from the adaptive learning described here, in which, subjects learn to adjust their priors on figure shape. The former leads to changes in perceptual bias, the latter to changes in cue reliability.

Differences between motor and perceptual behavior

Subjects behavior in the visuomotor and perceptual experiments was very similar. In particular, the measured cue weights in the first, baseline session in all experiments showed no significant differences. This stands in contrast to a recent report that subjects give more weight to monocular slant cues for perceptual judgments than for the object placement task used here (Knill 2005). The previous results were obtained using largely similar stimuli and tasks. The specific difference between the current results and the previous report is that the visuomotor cue weights measured here are essentially equal to those measured previously, but the monocular cue weights measured perceptually here are lower than those measured in the previous experiments. While it is unclear wht this would be the case, two possibilities present themselves. First, stimuli in the previously reported experiments contained highly salient texture cues, provided by Voronoi textures within the figures that hadn't been shrunk to small dots. Thus, the previously reported differences could result from texture cue processing specifically. Second, the perceptual matching tasks used previously was subtly different from the one used here. Specifically, subjects in the earlier perceptual experiments were only allowed to adjust the slant of the test probe around a horizontal axis, while the subjects in these experiments adjusted both the slant and tilt of the probe. Why this would lead to such a large change in cue weights is unclear, however, it may have affected the timing of subjects' responses and it is known that stereoscopic cues take some time to process after initial presentation of a stimulus. Thus, the earlier results may have reflected an interaction between task type and stimulus exposure time rather than between task type and cue weights, per se.

Appendix A

We derived a two-step Bayesian model of observer performance in the experiments described here. The model observer iterates through an estimation step and a learning step on each trial. In the estimation step, the observer uses its current model of the prior distribution of aspect ratios to estimate the slant of a stimulus from the information provided by the shape of the ellipse in the retinal image and from stereoscopic cues. For purposes of simulation, we took the shape of the ellipse in the image to be the shape of the ellipse that would have been projected to a cyclopean eye placed between the left and right eyes (an average of the shapes in the left and right eyes). In the learning step, the observer uses the information provided about the aspect ratio of the ellipse in the world by the two cues (and the previous prior) to update its internal model of the prior distribution of aspect ratios. We assumed a mixed prior distribution of aspect ratios, A, with the form

$$p(A) = \lambda \frac{1}{A\sqrt{2\pi}\sigma} e^{-(\log A)^2/2\sigma_A^2} + (1-\lambda)\partial(A-1), \qquad \text{(A.1)}$$

where λ is the proportion of ellipses in the world that are non-circular. $\partial(A-1)$ is a dirac function that takes the value 0 for $A \neq 1$. σ_A is the standard deviation of the log aspect ratio for non-circular ellipses. In all of the simulations, we set $\sigma_A = 0.5$, which created a distribution that dropped off sharply at $A = 0.5$ and $A = 2.0$, approximating the uniform dis-

tribution of aspect ratios used in the learning experiments. The mixing parameter, λ, is the only free parameter in the prior distribution. Learning occurs by adjusting an estimate of λ on each trial.

Estimation

The general form for the Bayesian estimator is derived by assuming that the image data on each trial, t, is given by a deterministic function of the scene with some corrupting additive noise

$$I_t = f(X) + \omega_t , \tag{A.2}$$

where *I* represents the image measurements and *X* represents the parameters describing a scene (at least all parameters that influence *I*) and ω is a random variable representing noise in the sensory system. An optimal Bayesian observer bases its estimate of *X* on the posterior distribution

$$p_t\left(X|I_t\right) = k\left(I_t|X\right)p_t(X) \tag{A.3}$$

where *k* is a constant that normalizes the distribution. $p_t(X)$ is the current model of the prior distribution on the scene. This is updated from trial to trial in the learning step described below. Assuming that the prior has a parametric form parameterized by a set of parameters λ, we can approximate the posterior using

$$p_t\left(X|I_t\right) \approx k\left(I_t|X\right)p\left(X;\hat{\lambda}_t\right) \tag{A.4}$$

where $\hat{\lambda}_t$ is the current best estimate of λ.

For the problem of combining two cues– the shape of the ellipse in the retinal image and the stereoscopic cues (treated as one cue) - to estimate slant, *X* represents the slant and aspect ratio the ellipse in the world. We will therefore replace *X* with two variables for slant and aspect ratio – *S* and *A*. Assuming that the prior on slant is fixed (and broad), we can write (A.4) as

$$p_t\left(S,A\left|\alpha,\vec{\delta}\right.\right) \approx kp\left(\alpha|S,A\right)p\left(\vec{\delta}|S\right)p\left(A;\hat{\lambda}_t\right)p(S), \tag{A.5}$$

where α represents the aspect ratio of the ellipse in the retinal image and δ represents a vector of disparity measurements. For simplicity, we are assuming that the surface is rotated around the horizontal axis (the tilt is fixed) and that ellipse is oriented so that one of its major axes is aligned with the horizontal (so we can ignore the effects on projection of the spin of the ellipse in the plane). For the problem of estimating slant, we integrate (A.5) over all possible aspect ratios, giving as our slant estimator

$$p_t\left(S\left|\alpha,\vec{\delta}\right.\right) \approx k\left[\int\left(\alpha|S,A\right)p\left(A;\hat{\lambda}_t\right)dA\right]p\left(\vec{\delta}|S\right)p(S), \tag{A.6}$$

where we have re-arranged terms somewhat to separate the likelihood function for the monocular cue provided by the shape of the ellipse in the retinal image (the term in brackets), which depends on the prior on elliptical aspect ratios, from the likelihood function for stereo. This gives the usual Bayesian form for integrating two cues – the product of likelihood functions for each cue multiplied by the prior on the scene parameter being estimated.

The prior distribution on A, $p\left(A;\hat{\lambda}_t\right)$ is given by (A.1). The likelihood function for the measured shape of the ellipse in the retinal image was derived by assuming that the sensory measurement is corrupted by Gaussian noise. In this case, the projection function is given by

$$\alpha = A\cos S + \omega, \tag{A.7}$$

giving for the likelihood function

$$p\left(\alpha_t \middle| S, A\right) = e^{-(\alpha_t - A\cos S)^2/2\sigma_\alpha^2}, \tag{A.8}$$

where σ_α is the standard deviation of the noise on the sensory measurements of α. We treated the stereoscopic cues as providing an unbiased estimate of slant corrupted by Gaussian noise, so we can write the likelihood function for stereo as

$$p\left(\vec{\delta}_t \middle| S\right) = e^{-(S-\hat{S}_{stereo})^2/2\sigma_{stereo}^2}; \qquad \hat{S}_{stereo} = \arg\max_S\left[p\left(\vec{\delta}_t \middle| S\right)\right] \tag{A.9}$$

in simulations, we sampled values of $\hat{S}_{stereo}$ from a Gaussian distribution with mean equal to the true slant of the stimulus and a standard deviation σ_{stereo}.

In simulations, we replicated the conditions of each experiment, so that on a given trial, the measured aspect ratio of the ellipse in the image was given by (A.7), and the estimate of slant from stereo was drawn from a Gaussian distribution as described above. The prior on aspect ratio was initially parameterized by $\lambda = 0.01$ and was updated using the learning procedure described below. The noise parameters for the simulations were chosen to be consistent with subjects performance prior to learning. Subjects performed the motor task used in experiments 3 and 4 with an average standard deviation of 2.5 degrees in the slant of the cylinder at contact (and no significant bias) and gave approximately equal weights to the foreshortening and stereo cues for surfaces at the test slant of 35 degrees. The model performed equivalently (with $\lambda = 0.01$) when the noise parameter for the aspect ratio measurement was set to $\sigma_\omega = .025$ and the noise parameter for slant estimates from stereo was set to $\sigma_\omega = 3.5^o$. The noise parameter on aspect ratio measurements was remarkably close to measurements of human subjects' thresholds for discriminating aspect ratio (Regan and Hamstra 1992), We used these noise parameters throughout all simulations.

To compute foreshortening cue weights for the model, we ran the model on the same stimulus conditions used in the experiments and used the estimates of slant collected over each simulated session in the regression procedure described for the human subjects. We simulated eight subjects and averaged the resulting weights across "subjects".

Learning

The model learns a "prior" on aspect ratios by adjusting it's internal estimate of the mixing parameter, λ, in equation (A.1) from trial to trial. The model assumes that λ can change slowly over time according to the equation

$$\lambda_{t+1} = \lambda_t + \eta_t \tag{A.10}$$

where η_t is a white noise process with standard deviation σ_η. The slant estimator described above requires an estimate of λ_t. To do this, the model updates an internal model of $p\left(\lambda_t \middle| I_{t-1}, I_{t-2}, \ldots, I_1\right)$ on each trial based on the information in the image on that trial, I_t. For simplicity, we will write this as $p_{t|t-1}\left(\lambda_t\right)$. From equation (A.10), this can be written as

$$p_{t|t-1}(\lambda_t) = p_{t-1|t-1}(\lambda_{t-1}) * N(0,\sigma_\eta), \tag{A.11}$$

where * is the convolution operator and $N(0,\sigma_\eta)$ is a normal distribution with mean 0 and standard deviation σ_η. The distribution $p_{t-1|t-1}(\lambda_{t-1})$ is be updated according to the equation

$$p_{t-1|t-1}(\lambda_{t-1}) = kp(I_{t-1}|\lambda_{t-1})p_{t-1|t-2}(\lambda_{t-1}) \tag{A.12}$$

where $p(I_{t-1}|\lambda_{t-1})$ is the likelihood of seeing the image I_{t-1}, given a prior on aspect ratios parameterized by λ_{t-1}. This is given by

$$p(I_{t-1}|\lambda_{t-1}) = k\iint p(I_t|A,S)p(A|\lambda_{t-1})p(S)dAdS \tag{A.13}$$

(A.11) and (A.13) together, provide iterative update equations for $p_{t|t-1}(\lambda_t)$. We use the mode of $p_{t|t-1}(\lambda_t)$ as the estimate of λ to use for estimating slant on each trial. The only free parameters in the model are the standard deviation of the white noise process assumed for λ, which we set to 0.01, and the initial distribution $p_{1|0}(\lambda_1)$, which we set to be a normal with mean equal to 0.01 (an initial estimate that 99% of ellipses are circular) and standard deviation equal to 0.001. The behavior of the model is largely indifferent to these parameters, though changing the initial standard deviation for the prior on λor the standard deviation of the random walk by large amounts affects the learning rate (e.g. setting the standard deviation of $p_{1|0}(\lambda_1)$to 0 turns off learning).

Bibliography

Adams, W. J., M. J. Banks, et al. (2001). "adaptation to three-dimensional distortions in human vision." Nature Neuroscience **4**: 1063-1064.

Adams, W. J., E. W. Graf, et al. (2004). "Experience can change the 'light-from-above' prior." Nature Neuroscience **7**(10): 1057-1058.

Atkins, J. E., J. Fiser, et al. (2001). "Experience-dependent visual cue integration based on consistencies between visual and haptic percepts." Vision Research **41**(4): 449-461.

Brady, M. and A. Yuille (1984). "An extremum principle for shape from contour." IEEE Transactions on Pattern Analysis and Machine Intelligence **PAMI-6**(3): 288-301.

Epstein, W. and C. L. Morgan (1970). "Adaptation to uniocular image magnification; modification of the disparity-depth relationship." American Journal of Psychology **83**: 322-329.

Ernst, M. O., M. S. Banks, et al. (1999). "Haptic feedback affects slant perception." Investigative Ophthalmology & Visual Science **40**(4): S802-S802.

Fahle, M., S. Edelman, et al. (1995). "Fast Perceptual-Learning in Hyperacuity." Vision Research **35**(21): 3003-3013.

Fiser, J. and R. N. Aslin (2001). "Unsupervised statistical learning of higher-order spatial structures from visual scenes." Psychological Science **12**(6): 499-504.

Fiser, J. and R. N. Aslin (2002). "Statistical learning of new visual feature combinations by infants." Proceedings of National Academy of Science, USA **99**(24): 15822-15826.

Garding, J. (1993). "Shape from texture and contour by weak isotropy." J. of Artificial Intelligence **64**: 243-297.

Hillis, J. M., S. J. Watt, et al. (2004). "Slant from texture and disparity cues: optimal cue combination." Journal of Vision **4**(12): 967-992.
Ikeuchi, K. and B. K. P. Horn (1981). "Numerical Shape from Shading and Occluding Boundaries." Artificial Intelligence **17**(1-3): 141-184.
Kanade, T. (1981). "Recovery of the three-dimensional shape of an object from a single view." Artificial Intelligence **17**: 409 - 460.
Karni, A. and D. Sagi (1993). "The Time-Course of Learning a Visual Skill." Nature **365**(6443): 250-252.
Knill, D. C. (1998). "Discrimination of planar surface slant from texture: human and ideal observers compared." Vision Res **38**(11): 1683-711.
Knill, D. C. (2001). "Contour into texture: information content of surface contours and texture flow." J Opt Soc Am A Opt Image Sci Vis **18**(1): 12-35.
Knill, D. C. (2003). "Mixture models and the probabilistic structure of depth cues." Vision Res **43**(7): 831-54.
Knill, D. C. (2005). "Reaching for visual cues to depth: The brain combines depth cues differently for motor control and perception." Journal of Vision **5**(2): 103-115.
Knill, D. C. (2006). "Humans implement a Bayes' optimal scheme for robust, nonlinear cue integration." Journal of Vision **Submitted**.
Knill, D. C. (2007). "Robust Bayesian cue integration: a model and psychophysical evidence." Journal of Vision **In press**.
Knill, D. C. and J. A. Saunders (2003). "Do humans optimally integrate stereo and texture information for judgments of surface slant?" Vision Res **43**(24): 2539-58.
Malik, J. and R. Rosenholtz (1997). "Computing local surface orientation and shape from texture for curved surfaces." International Journal of Computer Vision **23**(2): 149-168.
Maye, J., J. F. Merker, et al. (2002). "Infant sensitivity to distributional information can affect phonetic discrimination." Cognition **82**: B101-B111.
Ramachandran, V. S. (1988). "Perceiving Shape from Shading." Scientific American **259**(2): 76-83.
Regan, D. and S. J. Hamstra (1992). "Shape discrimination and the judgment of perfect symmetry: dissociation of shape from size." Vision Res **32**(10): 1845-1864.
Saunders, J. A. and D. C. Knill (2001). "Perception of 3D surface orientation from skew symmetry." Vision Res **41**(24): 3163-83.
Stevens, K. A. (1981). "The visual interpretation of surface contours." Artificial Intelligence **17**: 47-73.
van Ee, R., W. J. Adams, et al. (2003). "Bayesian modeling of cue interaction: bistability in stereoscopic slant perception." J Opt Soc Am A Opt Image Sci Vis **20**(7): 1398-406.
Wallach, H., M. E. Moore, et al. (1963). "Modification of stereoscopic depth perception." American Journal of Psychology **76**: 191-204.
Witkin, A. P. (1981). "Recovering Surface Shape and Orientation from Texture." Artificial Intelligence **17**(1-3): 17-45.