

The accuracy and reliability of perceived depth from linear perspective as a function of image size

Jeffrey A. Saunders

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



Benjamin T. Backus

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



We investigated the ability to use linear perspective to perceive depth from monocular images. Specifically, we focused on the information provided by convergence of parallel lines in an image due to perspective projection. Our stimuli were trapezoid-shaped projected contours, which appear as rectangles slanted in depth. If converging edges of a contour are assumed to be parallel edges of a 3D object, then it is possible in principle to recover its 3D orientation and relative dimensions. This 3D interpretation depends on projected size; hence, if an image contour were scaled, accurate use of perspective predicts changes in perceived slant and shape. We tested this prediction and measured the accuracy and precision with which observers can judge depth from perspective alone. Observers viewed monocular images of slanted rectangles and judged whether the rectangles appeared longer versus wider than a square. The projected contours had varying widths (7, 14, or 21 deg) and side angles (7 or 25 deg), and heights were varied by a staircase procedure to compute a point of subjective equality and 75% threshold for each condition. Observers were able to reliably judge aspect ratios from the monocular images: Weber fractions were 6–9% for the largest rectangles, increasing to as high as 17% for small rectangles with high simulated slant. Overall, the contours judged to be squares were taller than the projections of actual squares, consistent with perceptual underestimation of depth. Judgments were modulated by image size in the direction expected from perspective geometry, but the effect of size was only about 20–30% of what was predicted. We simulated the performance of a Bayesian ideal observer that integrated perspective information with an a priori bias toward compression of depth and which was able to qualitatively model the pattern of results.

Keywords: linear perspective, depth perception, slant perception, picture perception, Bayesian model, vision

Introduction

We can perceive 3D structure from photographs and linear perspective pictures in an effortless and stable manner, despite the absence of depth cues like binocular disparities or motion parallax. The effectiveness of pictures is only possible because they reproduce regularities present in real-world views of a structured environment. A static monocular image typically has many regularities that can be used to infer 3D structure, called *pictorial depth cues* (for taxonomies, see Cutting and Vishton, 1995, or Kubovy, 1986). For example, in Figure 1, depth can be inferred from the gradient of size for the square tiles or from the convergence (in the image) of lines that recede in the world. These cues are typically correlated, but each is conditional on different assumptions: that squares lying on a surface have constant size in the world or that lines on the surface are parallel in the world. In this article, we focus on the information provided by the latter cue, which we will refer to as *perspective convergence*.

Psychophysical studies of 3D perception from perspective convergence date back more than 50 years. Early studies demonstrated that observers make systematic slant

judgments from minimal stimuli that contain perspective convergence (Clark, Smith, & Rabe, 1955, 1956; Freeman, 1966a, 1966b; Rosinski, Mulholland, Degelman, & Farber, 1980; Smith, 1967; Stavrianos, 1945) and that convergence can dominate other slant information in cue conflict situations (Attneave & Olson, 1966; Banks & Backus, 1998; Braunstein & Payne, 1969; Gillam, 1968; Smith, 1967). Recent slant-from-texture experiments by Andersen, Braunstein, and Saidpour (1998) and Todd, Thaler, and Dijkstra (2005) included conditions in which convergence was the only available cue and found that these stimuli were effective for supporting slant judgments. Modulations in perspective convergence have also been found to contribute to the perceived 3D shape of curved surfaces (Li & Zaidi, 2000).

Although it is clear that linear perspective can contribute to 3D perception, significant clarification is still needed as to how the visual system uses this information. No attempts have been made to measure the reliability of perceived depth from solely perspective convergence information, except in the special case of surfaces that are close to frontal (Freeman, 1966a; Gillam, 1968). A number of studies have measured the ability to discriminate slant-from-texture, but for these studies, perspective convergence was either absent (Knill, 1998a, 1998b;

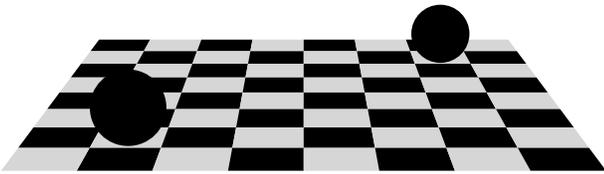


Figure 1. Example of an image in which depth can be perceived from pictorial cues. The distorted checkerboard pattern is seen as a slanted rectangular surface covered with uniform texture. Perceived slant in depth of the surface could be based on different assumptions about the texture, such as that the squares are uniform sized in the world or that the converging lines in depth are parallel edges in the world.

Knill & Saunders, 2003) or confounded with other texture information (Hillis, Watt, Landy, & Banks, 2004; Rosas, Wichmann, & Wagemans, 2004; Saunders & Backus, 2006). There have been studies that measure the accuracy of perceived slant when only convergence information is available (Andersen et al., 1998; Rosinski et al., 1980; Smith, 1967; Todd et al., 2005). However, these data allow only limited inferences about whether perspective convergence is used in a geometrically consistent manner.

The goal of the experiments reported here was to attain a more quantitative psychophysical measure of the perception of depth from perspective convergence. Specifically, we measured the accuracy and reliability of 3D judgments when the only information indicating depth was perspective convergence and tested whether perceived depth changed in a geometrically consistent way in response to size scaling of images. We used a match-to-known-standard paradigm that indirectly probed perceived depth: Subjects were presented with images of slanted rectangles and judged whether the perceived 3D shape was taller or wider than a square. As we illustrate in the next section, this task can be performed using monocular information, and the correct response depends on image size.

A similar aspect ratio judgment task has been used for 2D rectangular shapes viewed frontally, and observers exhibit discrimination thresholds as low as 1% (Regan, Hajdúr, & Hong, 1996; Regan & Hamstra, 1992; Zanker & Quenzer, 1999). Like the 2D task, the 3D task depends on the ability to measure aspect ratio and to make a decision. However, the 3D task also forces the observer to interpret perspective convergence as slant. Thresholds are, in general, lower when observers are instructed to make 2D judgments than when they are instructed to make 3D judgments, indicating that performance in the 3D task is limited by the observer's ability to recover slant from perspective convergence (see Experiment 1).

Size dependence of perspective information

Consider the geometry of lines projected onto a planar image. If a pair of lines is parallel in the world, they

project to lines that converge to a vanishing point in the image. If multiple sets of parallel lines are present, with different orientations but lying on the same planar surface, then their multiple vanishing points uniquely define a horizon and thereby specify 3D orientation of the surface, that is, its *slant* and *tilt* (Stevens, 1983) relative to a line of sight that intersects the plane. Once the slant and tilt are known, it would then be possible to reconstruct the shape of the object up to a scale factor for distance (e.g., by back projecting onto the slanted plane).

It is unnecessary to compute the horizon line explicitly to use the convergence cue. For example, it would suffice to learn a direct mapping from differences in projected line orientations to possible parallel interpretations in 3D. Figure 2 illustrates the relationship between perspective convergence and 3D orientation for an image of a slanted rectangle. In this special case, where one pair of edges remains parallel in the image (“converging” to a point at infinite distance in the horizontal direction), there is a simple relationship between slant and perspective convergence:

$$\tan(a) = \tan(w/2) \times \tan(s), \quad (1)$$

where a is the angle of the converging side edges relative to the tilt direction (i.e., $\tan(a)$ is the slope relative to vertical), s is the egocentric slant of the surface (as measured at the origin), and w is the angular width measured through the center. This formulation follows that of Freeman (1966b; equivalent formulations include Braunstein & Payne, 1969, and Flock, 1962). A similar relation holds for more generic poses but would involve both the slant and tilt of the surface. We describe perspective convergence in terms of projection onto an image plane because it is convenient; the same geometric

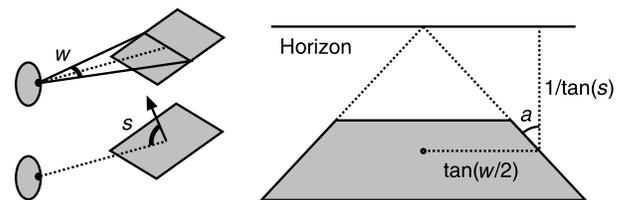


Figure 2. Illustration of the relationship between perspective convergence and surface slant for the projected image of a slanted rectangular surface, aligned with its direction of tilt. We define the angular width w and slant s to be relative to a reference point at the center of the 3D object (left panels) and assume a projection plane with a distance of 1. The sides of the projected contour converge to a point on the horizon (right panel), which is a distance of $\tan(90 \text{ deg} - s)$, or $1/\tan(s)$, away from the central reference point. Note that the reference point can be identified in the image as the intersection between the diagonals of the projected contour. The slope of the sides relative to vertical, $\tan(a)$, is equal to $\tan(w/2) \times \tan(s)$.

relationship could also be described in spherical coordinates.

An interesting aspect of the geometry is that the visual angle subtended by an object factors into its 3D interpretation (i.e., the parameter w in Equation 1). Accurate computation of slant in depth from perspective would therefore require the visual system to know the absolute angular size of a projected image. One way to intuitively understand the size dependence is to think of perspective convergence as identifying the location of the horizon within a projected image. Surface slant is the complement of the visual angle between a reference point and the horizon. To obtain this visual angle from a projected contour, one must also know the visual angle subtended by the contour itself, which we will refer to as its *projected size*. As illustrated in Figure 3, similar shapes, when scaled to different sizes, correspond to different 3D interpretations. Thus, if the visual system does use perspective convergence in a geometrically correct way, one would expect that rescaling a perspective image would change perceived 3D structure in predictable ways. As projected size increases, the perceived aspect ratio of the rectangle (length–width) should decrease, and

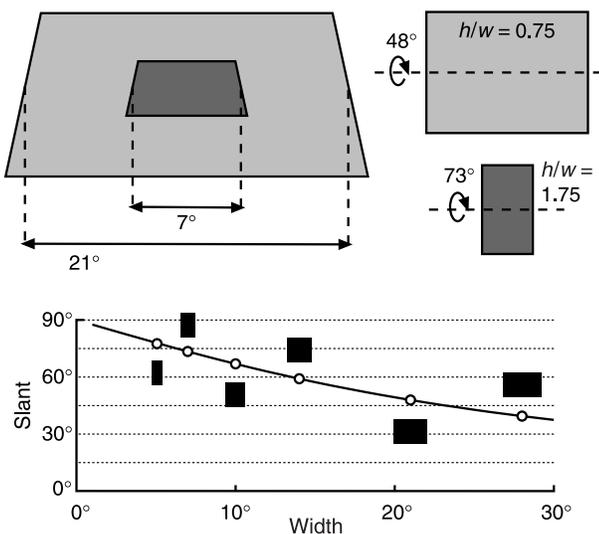


Figure 3. Illustration of how the 3D interpretation of perspective information depends on projected size. The upper left panel shows two trapezoid-shaped projected contours that have different angular sizes but identical shapes. If these contours are projections of rectangles with parallel sides, their slants can be determined (see Figure 2), which in turn specifies the dimensions of the rectangular object. For the large size, the 3D interpretation is a wide rectangle at intermediate slant (upper right), whereas for the small size, the 3D interpretation is a long rectangle with high slant (middle right). The lower graph plots the slant specified by perspective for this example contour shape across a range of projected sizes. The small black rectangles illustrate object dimensions for some sample points on the graph.

it should appear less slanted. The experiments presented here test this prediction, as a means to identify the contribution of perspective convergence.

Previous work on scaled images

In the context of picture perception, other researchers have pointed out that the 3D interpretation of perspective information is changed when an image is scaled (e.g., Farber & Rosinski, 1978; Nichols & Kennedy, 1993; Sedgwick, 1980, 1991). This is sometimes described as viewing an image from the wrong distance, rather than viewing a scaled image, although these are geometrically equivalent. A perspective picture or photograph is only geometrically accurate when viewed from a particular vantage point, corresponding to the optical center of projection of the image. If an observer moves further from a picture, its projected image subtends a smaller visual angle but remains otherwise similar; hence, its 3D interpretation has greater depth relative to width and height (as in Figure 3). Thus, the problem of interpreting a perspective picture viewed from the wrong distance is equivalent to the problem of interpreting a perspective picture when the size of an image has been rescaled, as in our experiments.

Previous studies have tested whether perceived depth in photographs is affected by rescaling an image or by changing viewing distance (Bengston, Stergios, Ward, & Jester, 1980; Lumsden, 1983; Smith, 1958a, 1958b; Smith & Gruber, 1958). Most of these studies compared scaled and unscaled photographs with matching perspective information. Bengston et al. (1980) found close agreement between judgments for scaled and unscaled photos that would be expected to appear equivalent based on perspective geometry. Other experiments found more limited correspondence (Lumsden, 1983; Smith, 1958a, 1958b). Regardless of the degree of correspondence, one cannot infer from these results whether perceived depth in scaled images changes accurately because any biases in the interpretation of perspective information would have affected the perception of *both* scaled and unscaled images. Smith and Gruber (1958) compared depth judgments based on either photographs or actual views of a corridor. To the extent that perspective information contributes to perceived depth of the real scene, however, this paradigm has the same limitation as the other studies.

To our knowledge, Smith (1967) is the only previous study that directly tested whether perceived depth from perspective is accurately modulated by projected size. Smith compared slant judgments for stimuli with similar shapes but varied projected sizes. He found that slant estimates changed depending on size in the expected direction but by a much smaller magnitude than predicted by perspective geometry, in agreement with our results here.

Size-invariant monocular cues

Perspective convergence differs from other pictorial cues in its dependence on size. One size-invariant cue is *foreshortening* or *compression* of a projected contour. If the overall shape of a planar object is assumed to be *isotropic* (Garding, 1993; Witkin, 1981), then the aspect ratio of its projected contour provides a cue to its slant. Figure 4a illustrates this cue for the case of trapezoidal projected contours.

Another size-invariant cue is skew symmetry (Kanade, 1981; Perkins, 1976; Saunders & Knill, 2001). If a pair of axes (or edges) that intersect at a skewed angle in an image is assumed to intersect at a right angle in the world, then the skewed axes provide a cue to 3D orientation. As illustrated in Figure 4b, the skew of a set of axes can provide a slant cue that is independent of projected size. In scaled images, these size-invariant image cues will conflict with perspective convergence and are expected to reduce the effect of image size on the perceived slants of surfaces in the depicted scene.

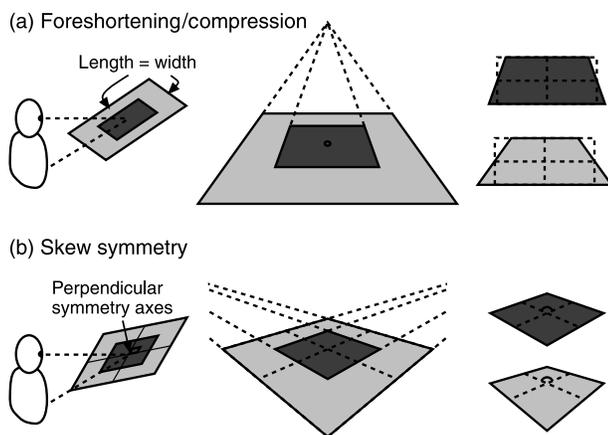


Figure 4. Two monocular depth cues that provide information that does not depend on absolute projected size: foreshortening (a) and skew symmetry (b). The top center figure shows the projected outline contours of two slanted squares with the same slant but different sizes. These trapezoids have different amounts of convergence, but their mean aspect ratios are the same (upper right), equal to the cosine of surface slant. If objects are assumed to be isotropic, slant could be recovered from projected aspect ratio even if size were not known. A size-invariant measure of foreshortening can alternatively be formulated in terms of the slope of the line connecting the geometric center of the trapezoid to one of its corners (Braunstein & Payne, 1969). The bottom middle figure shows the projected contours of two squares with different sizes that have been rotated by 45 deg within their plane. Size affects the amount of convergence but not the angle formed between symmetry axes at the geometric center (bottom right). If objects are assumed to be mirror symmetric, rotated in depth around a horizontal axis, slant could be recovered from the skew angle without knowledge of size.

There is evidence that observers are sensitive to differences between the 3D structure specified by perspective and by size-invariant cues. Nichols and Kennedy (1993) presented subjects with line drawings of cubes viewed from their corners, some of which were generated by rescaling a perspective image. A given drawing was judged to be a good cube over a range of sizes, but the size that was geometrically accurate for a cube was rated best. Yang and Kubovy (1999) performed a similar experiment, presenting cube-like line drawings with varying relationships between projected size and amount of perspective and asking subjects to identify the best cube. They also observed a weak preference for geometrically accurate images, consistent with the work of Nichols and Kennedy. In both studies, a wide range of stimuli were rated to be good cubes, which is consistent with reliance on size-invariant cues.

We wanted to isolate perspective convergence from other pictorial cues. We therefore used trapezoidal projected images like those of Figure 3, where the side edges are symmetric and where the top and bottom edges are parallel in the image. This image is a special case: Assuming right-angle intersections does not provide the normal size-invariant slant cue. One way to understand this is that when viewing a real rectangle, the vanishing points for opposite edge pairs have visual directions that are separated by 90 deg. This angle normally becomes smaller or larger when the image of the rectangle is scaled. For a trapezoidal image, however, this angle remains 90 deg (because one “vanishing point” is at infinity) and, therefore, provides no information about whether the image has been scaled.

Although the use of trapezoid-shaped stimuli eliminates the skew symmetry cue, isotropy is still a potentially useful constraint: Observers could assume that the images are the projections of square objects in 3D, which would generally lead to conflict with perspective convergence. In Experiment 1, the task was to judge whether the object was taller versus wider than a square; thus, heavy reliance on an assumption of isotropy would cause judgments to be unreliable but would not by itself lead to bias. In Experiment 2, subjects matched perceived length in depth to the height of a frontally oriented bar, and projected aspect ratio was an independent variable, which allowed us to assess its effect.

Experiment 1

In this experiment, stimuli were monocular images like those shown in Figure 5, and subjects judged whether the objects appeared to be taller versus wider than a square, when considered as objects in 3D. There is a correct answer in this task based on the perspective convergence cue. We tested two different side angles, 25 and 7.125

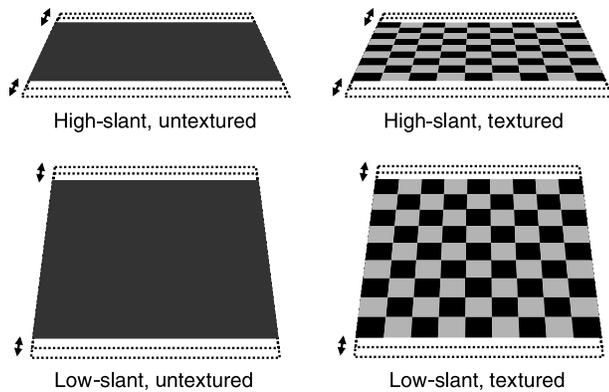


Figure 5. The four classes of perspective images used in Experiment 1. Images were consistent with a perspective view of planar rectangular objects, slanted in depth. Objects were either of uniform color (untextured conditions, left) or covered with a checkerboard pattern (textured conditions, right). The convergence angle of the sides of the trapezoids was either 25 deg (high-slant conditions, top) or 7.125 deg (low-slant conditions, bottom). Each base image was presented at different sizes, with the width of the centerline subtending 7, 14, or 21 deg (not shown in this figure). Here, the trapezoids are shown as dark on a white background, but in the actual displays, the contrast was reversed.

deg, which we will refer to as the *high-slant* and *low-slant* conditions, respectively (Figure 5, top and bottom rows). Both shapes were presented with a range of projected sizes. If subjects use perspective convergence correctly at a given projected size, one would expect a large difference across size conditions in the projected shape that is judged to be square, with the largest projected size having to be the tallest to appear square. Figure 6 shows the contour shapes that accurately correspond to the projections of a square, for each of the six size and slant conditions. In addition to overall shape, we also varied the presence or absence of an internal grid texture (Figure 5, right vs. left). The textured stimuli contained an additional cue to depth: the gradient of compression of vertical spacing. This cue is effective in its own right (Andersen et al., 1998). As with perspective convergence, the 3D interpretation of the compression cue depends on projected size; hence, varying size might have a larger effect for textured than untextured rectangles.

Methods

Apparatus and display

Stimuli were rear projected from an InFocus LP350 projector, with $1,024 \times 768$ resolution, onto a 166×125 cm region of a large screen positioned 2 m from the observer. The rectangular projected region subtended 45 deg horizontally and 35 deg vertically, and its boundaries were dimly visible. Subjects wore a patch over their left eye throughout the experiment. Subjects were seated on a stool

and were instructed to remain stationary during judgments but were not otherwise restricted (no chin rest or bite bar was used). Images were grayscale and antialiased, rendered using OpenGL on a workstation with Nvidia Quadro FX 1000 graphics board.

Stimuli simulated perspective views of slanted rectangular surfaces, filled either with uniform gray (untextured) or with a 9×9 checkerboard pattern (textured), on a black background. Slant was always around a horizontal axis (i.e., the tilt direction was vertical). The left and right sides of the projected contours had 2D orientations that were either ± 25 or ± 7.125 deg relative to vertical in the high- and low-slant conditions, respectively. Both shapes were presented at three different sizes, with widths of 25 cm (7.1 deg), 50 cm (14 deg), or 75 cm (20.6 deg), as measured horizontally through the geometric center of the projected contours. By “center,” we mean the intersection point of the trapezoid diagonals, which is also the projected location of the center of the original 3D rectangle. The screen location of the center of the trapezoid was the same for all stimuli. The slants specified by perspective convergence were as follows: 82 and 63 deg (small), 75 and 45 deg (medium), or 68 and 34 deg (large), respectively.

Procedure

Subjects made forced-choice judgments whether the simulated 3D object was longer versus wider than a

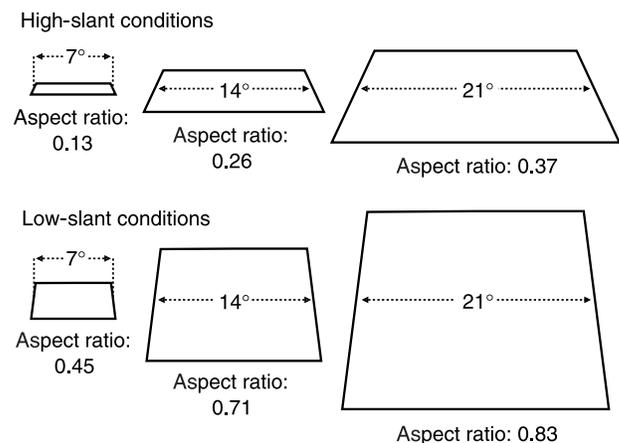


Figure 6. Predicted results if performance were veridical. Each trapezoid is an accurate projection of a square, for various widths and slants. The convergence angles are matched across the three high-slant conditions (top row) and the three low-slant conditions (bottom row), but the slants that these correspond to differ depending on size. In the high-slant conditions, the slants specified by perspective convergence under an assumption of parallelism are 82, 75, and 68 deg (from left to right). In the low-slant conditions, the slants specified by perspective convergence are 63, 45, and 34 deg. The mean aspect ratios of the projected trapezoids are equal to the cosine of the slant.

square. They were instructed to base their judgments on the perceived shape of the slanted 3D object, not on the screen projection. In the case of textured stimuli, subjects were told to base their responses on the rectangle as a whole, not component rectangles. Trials were self-paced, and subjects received no feedback.

The aspect ratios of projected contours were varied across trials using a new adaptive method (see [Appendix A](#)). The set of judgments from each condition was fit to a cumulative Gaussian psychometric function, using maximum-likelihood criteria. The mean of the best fitting function was taken as the point of subjective equality (PSE), which, for this task, was the aspect ratio of the image. Discrimination ability was measured as the difference between the PSE and the 75% points, divided by the PSE. This is a Weber fraction that corresponds to the change in aspect ratio required for 75% discrimination.

We define the projected aspect ratio of a trapezoid to be $(w_{\text{bottom}} + w_{\text{top}})/(2h)$, where w_{bottom} and w_{top} are the widths of its bottom and top edges and h is its projected height. This ratio is related to the aspect ratio of the corresponding 3D rectangle by a factor of the cosine of slant (Braunstein & Payne, 1969). Thus, the same Weber fraction describes discriminability for the 3D rectangles and for their projections. Note that the “mean width” used to compute aspect ratio for the Weber fraction is not same as width measured through the center of the trapezoid.

Textured and untextured stimuli were presented in separate blocks, whereas size and slant were randomized within blocks. The experiment consisted of two 1-hr sessions, each with one block of textured stimuli and one block of untextured stimuli, with order randomized across subjects and sessions. Each block contained 300 trials, yielding a total of 100 trials for each of the 12 conditions.

In an additional control condition, separate subjects judged the aspect ratios of 2D projected contours. The stimuli were either trapezoidal or rectangular contours, and subjects reported whether the contour’s height was greater or less than half the central width of the contour. No feedback was given. The trapezoidal contours used in the control experiment were the same as in the untextured, high-slant condition of the main experiment. The rectangular contours were presented at the same three sizes as the trapezoidal contours. Contour height was varied across trials using the same adaptive procedure, and the PSEs and 75% thresholds for the 2D task were computed as before. Subjects performed two blocks of 300 trials each, and both contour shape and size were randomized within blocks.

Subjects

Twelve paid subjects participated in the main experiment. All had normal or corrected-to-normal vision and were naive to the purposes of the experiment. Five additional naive subjects, along with the first author, participated in the control experiment using the 2D task.

Subjects gave informed consent in accordance with a protocol approved by the IRB panel at the University of Pennsylvania.

Results

[Figure 7](#) shows mean contour aspect ratios derived from subjects’ judgments, averaged across subjects, as a function of projected size. The two graphs plot results for the high- and low-slant conditions, respectively, and the two solid lines on each graph correspond to the textured and untextured conditions. In all cases, aspect ratios were larger for larger projected sizes, which is in the predicted direction, ANOVA for high slant: $F(2,55) = 19.33$, $p < .001$; low slant: $F(2,55) = 39.53$, $p < .001$. There was no evidence that the presence or absence of texture made a difference in the effect of size, high slant: $F(2,55) = 0.628$, $p = .47$ (*ns*); low slant: $F(2,55) = 0.57$, $p = .57$ (*ns*). Eleven of 12 subjects showed the size effect. There were also small but significant main effects of texture for both high- and low-slant conditions, high slant: $F(1,55) = 6.119$, $p = .016$; low slant: $F(1,55) = 10.96$, $p = .002$.

Although the effect of projected size was reliable, it was much smaller in magnitude than would be expected based on perspective geometry. The dashed lines on each plot show the ratios for veridical use of this cue. For the high-slant conditions ([Figure 7](#), left), the observed height-to-width ratios for the largest and smallest trapezoids differed by 19% for textured stimuli and 26% for untextured stimuli, whereas the predicted difference is 180%. For the low-slant conditions ([Figure 7](#), right), PSEs were closer to veridical, but the effect of size was still compressed (12% and 8% for the textured and untextured conditions vs. a predicted difference of 86%). In all cases, the projections of apparently square rectangles were taller than the projections of actual squares.

[Figure 8](#) shows the mean 75% discrimination thresholds for the same conditions as in [Figure 7](#). Discrimination thresholds were relatively low, indicating that although judgments were biased, they were reliable. For the high-slant conditions, textured stimuli produced significantly lower thresholds, $F(1,55) = 8.7$, $p = .005$, whereas for the low-slant conditions, there was no reliable difference for texture, $F(1,55) = 1.5$, $p = .23$ (*ns*). Threshold decreased as projected size increased in the high-slant conditions, $F(2,55) = 12.6$, $p < .001$; this trend was not significant in the low-slant conditions, $F(2,55) = 1.4$, $p = .25$ (*ns*). There was no interaction between size and the presence of texture for either slant condition, high slant: $F(2,55) = 0.36$, $p = .70$ (*ns*); low slant: $F(2,55) = 0.14$, $p = .87$ (*ns*).

Discussion

Consistency of depth judgments

Our results demonstrate that subjects can make consistent judgments about 3D lengths from monocular

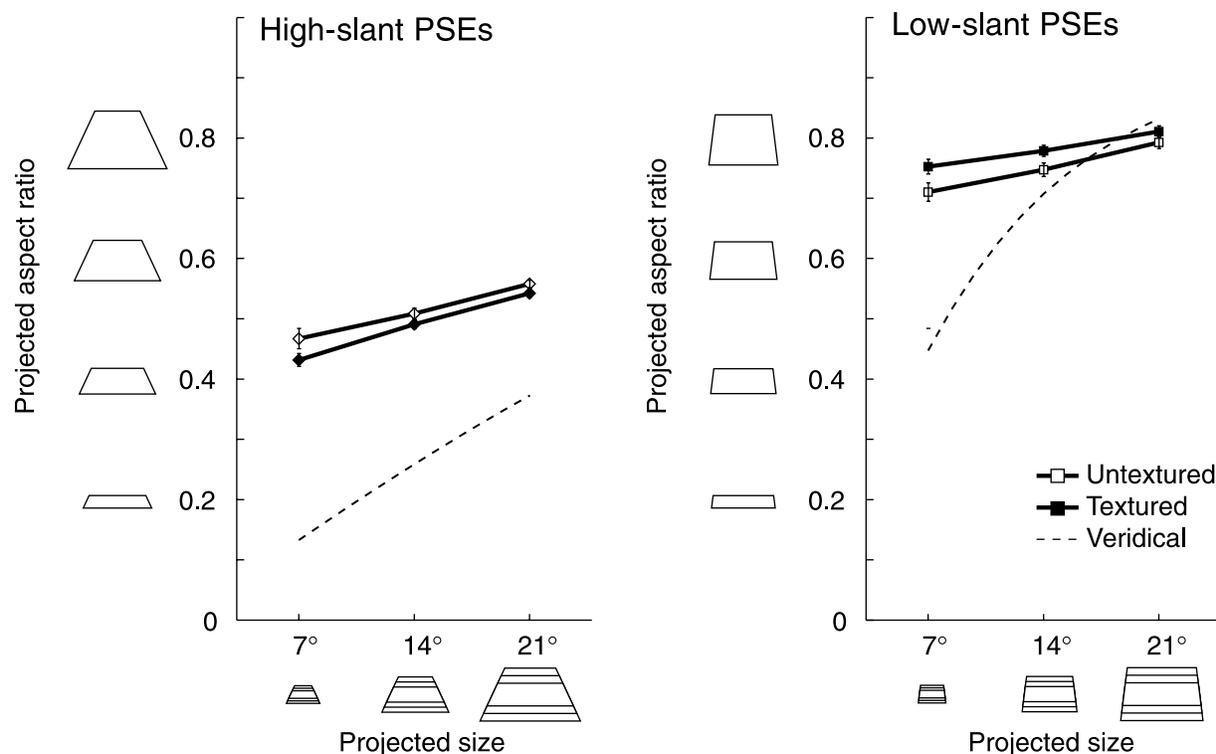


Figure 7. Mean PSEs from the data of Experiment 1. The graphs plot the mean projected aspect ratios of projected contours that would be perceived as images of square objects, as a function of projected size. The left and right graphs show results for the high- and low-slant conditions, respectively; closed and open symbols correspond to textured and untextured conditions, respectively. The dashed lines plot veridical responses based on accurate use of perspective convergence.

images containing perspective convergence. At the largest projected sizes, the average discrimination thresholds for untextured stimuli were only 6% for the low-slant condition and 9% for the high-slant condition.

The presence of internal texture resulted in lower discrimination thresholds. This is not surprising, as there are a number of ways that the texture could have served to improve performance. First, texture provides additional independent sources of information about depth, including the following: the compression of individual texture elements, the gradient of texture compression, and the gradients of texture size and spacing. The contributions of any of these texture cues, which have all been observed to contribute in at least some conditions (e.g., Knill, 1998b; Rosenholtz & Malik, 1997), could have improved the precision of depth estimates. Second, the textured stimuli contained multiple converging lines (the columns of the texture), which could have improved image measurements of perspective convergence and thereby improved the precision of depth estimates.

In the case of the textured stimuli, improvement in discrimination performance at large projected sizes could potentially be explained by better estimates of the size and shapes of texture elements rather than by better use of information from perspective convergence. However, thresholds for the untextured stimuli improved with projected size by as much as or more than those for the

textured stimuli; hence, any effect of size on the reliability of texture information must have been comparatively small.

Perceptual biases

Although judgments were relatively reliable, they showed large biases relative to veridical, especially for the high-slant conditions. In the worst case—high slant and small width—the contours that appeared to be views of square objects were, on average, over three times the height of an actual projection of a square.

Some component of the overall bias could be related to a known bias in perception of 2D dimensions: the horizontal/vertical illusion. For example, subjects might have been comparing perceived dimensions of the slanted 3D object to some biased internal standard of a square. However, the magnitude of the horizontal/vertical bias has been measured to be around 4% in the case of 2D shapes (Henriques, Flanders, & Soechting, 2005); thus, it could not fully account for the much larger biases we observed.

The presence of the checkerboard texture had only a small effect on biases, despite the fact that the textured stimuli, in principle, provide more information and are subjectively more compelling. This suggests that texture cues added little depth information beyond perspective convergence for our displays. This is in partial agreement

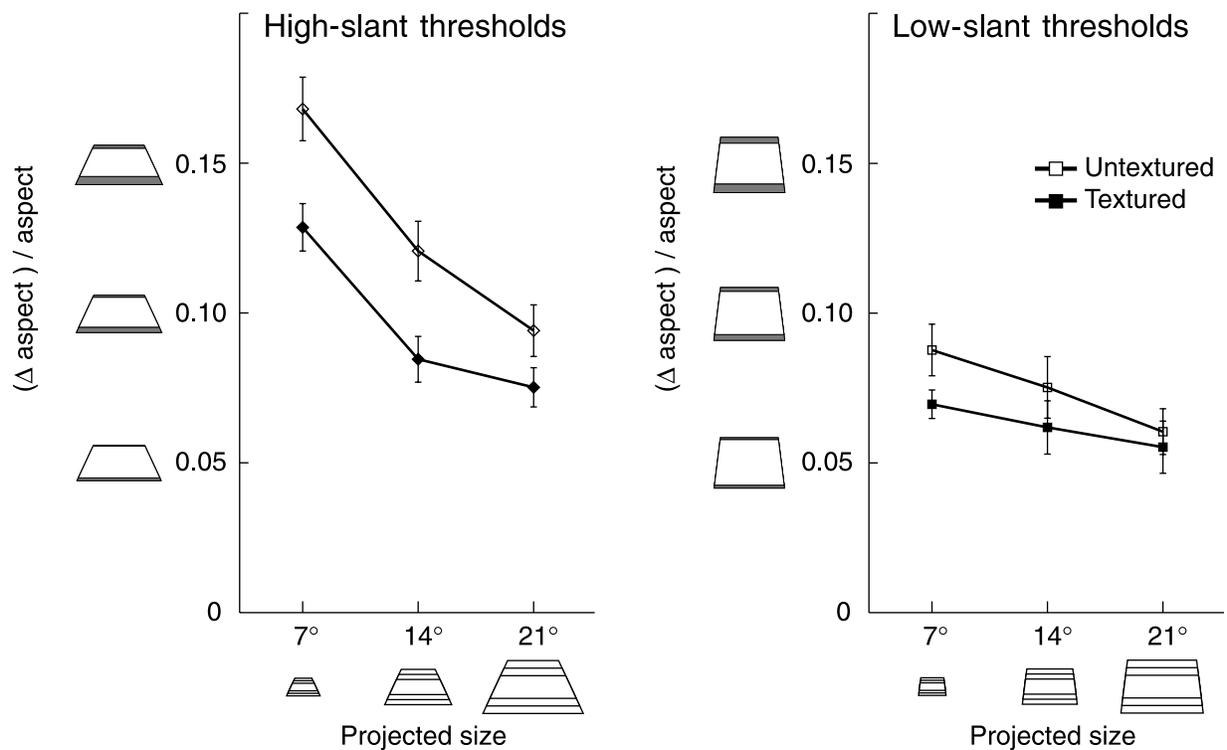


Figure 8. Mean discrimination thresholds from the data of Experiment 1. Thresholds are expressed as Weber fractions, computed by taking the difference between aspect ratios at the PSE and 75% points and dividing by the PSE aspect ratios. The left and right graphs show results for the high- and low-slant conditions, respectively; closed and open symbols correspond to textured and untextured conditions, respectively. The small figures beside the y-axis labels are graphical representations of the range of aspect ratios corresponding to a given threshold value, with the shaded regions depicting ± 1 threshold units around the mean.

with previous studies. Todd et al. (2005) measured judgments of the dihedral angle formed by two textured surfaces and found a reliable but small difference between ruled textures, which isolated perspective convergence information, and plaid textures. Andersen et al. (1998) also report modest differences in judgments of dihedral angles for ruled surfaces and grids, although they found a comparatively larger advantage for grid textures when subjects judged the slant of a single surface.

Effect of projected size

The contours that appeared to be square objects, in addition to being biased overall, also did not change with projected size as much as predicted by perspective geometry. This is in agreement with the results of Smith (1967). Smith measured direct estimates of slant for trapezoidal contour stimuli, similar to our untextured stimuli, and included conditions in which contour shape was matched across different sizes. Smith also found that judgments were dependent on size but that this effect was much smaller than predicted.

Although its effect was modest, projected size did reliably modulate subjects' judgments in our experiment, and the magnitude of the effect exceeded discrimination

thresholds. Thus, the smaller-than-expected size modulation cannot be attributed to simply poor sensory measurement of projected size.

One depth cue that we were not able to isolate, even in the untextured condition, was the overall foreshortening of the projected contour. However, any bias caused by the foreshortening cue would be in the direction of an isotropic interpretation; that is, the 3D interpretation of the object would always be closer to being square. For the task we used, this might make *discrimination* more difficult, but it could not directly account for size-dependent *biases* in what projected shapes appear to be squares.

Control for 2D strategy

One potential concern is that subjects might not have based their responses on a 3D percept. Rather, they might have made 2D shape judgments, comparing the projected aspect ratio of a contour with some imagined 2D standard. If subjects were using 2D projected shapes directly, rather than the perceived shapes of slanted 3D objects, then our experiments would have little relevance to the question of how the visual system uses perspective convergence to perceive 3D slant and shape.

There are several reasons to believe that subjects used 3D shape. First, for a cognitive 2D strategy, it is not obvious why judgments would be modulated by size at all. One would have to assume that the internal 2D standards used for comparison had varying aspect ratios depending on projected size. Second, if subjects were judging 2D shape, thresholds would be expected to show the same general pattern as previously observed for 2D aspect ratio judgments. However, aspect ratio discrimination for 2D shapes is unaffected by large variations in overall scale (Regan & Hamstra, 1992). In our experiments, thresholds decreased as image size increased.

Our additional control experiment also addresses this issue. Instead of judging the perceived shape of a slanted 3D object, subjects were instructed to judge the aspect ratio of the 2D trapezoidal contour: whether the contour's height was greater or less than half of its central width. For comparison, we also had subjects perform the same task for rectangular 2D contours with the same sizes. The results are shown in Figure 9. Thresholds were higher overall for the trapezoids relative to the rectangles, but projected size had no effect. Comparing these thresholds to those obtained using a 3D task (Figure 8, high-slant conditions), it is clear that the 2D shape task produces qualitatively different results. The PSEs from our control experiment also did not depend on projected size.

Given that thresholds and PSEs both varied with size in our 3D task but not in our 2D task and that performance was worse when subjects were instructed to use the 3D

strategy, we conclude that subjects' judgments in the 3D task were, in fact, based on 3D percepts. In that case, our results indicate that perceived depth from perspective convergence is both more reliable and more accurate for large images. As we will describe later, an ideal observer exhibits similar performance.

Perceptual compression of depth

The direction of the observed biases is consistent with underestimation of the perceived slant of the rectangles, which has been observed in other studies that isolated perspective information (Andersen et al., 1998; Smith, 1967; Todd et al., 2005). Consider, for example, the contour shown in the upper left panel of Figure 6. The slant implied by perspective convergence is very high in this case (82 deg). Suppose that observers tended to see the figure as being less slanted. The resulting perceived 3D object would be compressed in length relative to veridical. To be perceived as a 3D square, it would then have to be stretched vertically, as in our results.

This interpretation agrees with our phenomenal impressions of the stimuli: They appear much less slanted than they should, based on perspective convergence. Another aspect of the phenomenal appearance that is also consistent with underestimation of slant is that, for the textured surfaces, individual texture elements did not appear to have uniform aspect ratios along the 3D surface. Rather, the upper elements appeared to be more compressed in length. If perceived slant were underestimated, then the compression gradient in a projected image would be greater than would be expected for a homogeneous surface with the (biased) perceived slant, which could account for the inhomogeneous appearance of the textured stimuli.

There are a number of factors that could have reduced perceived slant in our stimuli. Real slanted surfaces produce an accommodative gradient, which is known to contribute to slant perception (Watt, Akeley, Ernst, & Banks, 2005). The absence of an accommodative gradient in our displays indicated a frontal surface. Another factor is the frame provided by the projection screen. Although the images were subjectively compelling as 3D objects, subjects were aware that they were looking at projected images, and the boundaries of the screen were visible. This could also have acted as conflicting information or "cross talk" specifying a frontal surface, as suggested by Sedgwick (1991). Eliminating cues that specify a flat pictorial surface has been observed to enhance perception of depth (Koenderink, van Doorn, & Kappers, 1994). The visual system may also have an a priori bias toward seeing uniform depth when 3D cues are weak or absent (Gogel, 1965). All of these factors would have the effect of compressing perceived depth toward the frontal plane.

If perceived slant were some blend between the slant specified by perspective and the slant indicated by conflicting cues and any prior assumptions, then one could

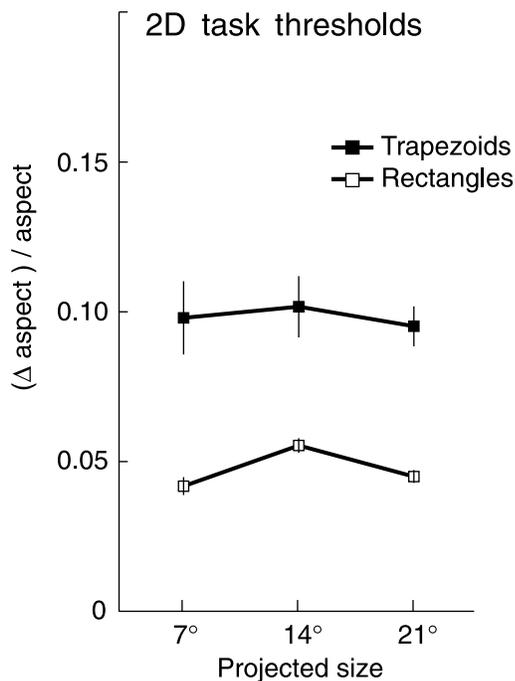


Figure 9. Mean 75% discrimination thresholds for 2D aspect ratio judgment task.

also explain the smaller-than-expected effect of projected size. That is because, to the extent that the perspective cue is only partially “weighted,” changes in the perspective information would have less influence on the final percept. We will discuss this possibility in more detail in a later section.

It is also possible that the visual system has simply learned an incorrect mapping between rectangular 3D objects and their perspective projections. Rectangular or square planar objects are common in normal environments; thus, the visual system would have ample exposure to this class of objects. However, there may normally be no cost to underestimating the slant or length in depth of rectangles. Informal observation suggests that real squares (e.g., on sidewalks) may also be misperceived as being relatively wider than a square, consistent with compression of perceived depth even under full-cue conditions.

Experiment 2

Figure 10a illustrates what PSEs from Experiment 1 represent: the shapes of projected contours that are perceived to be square 3D objects, for various projected

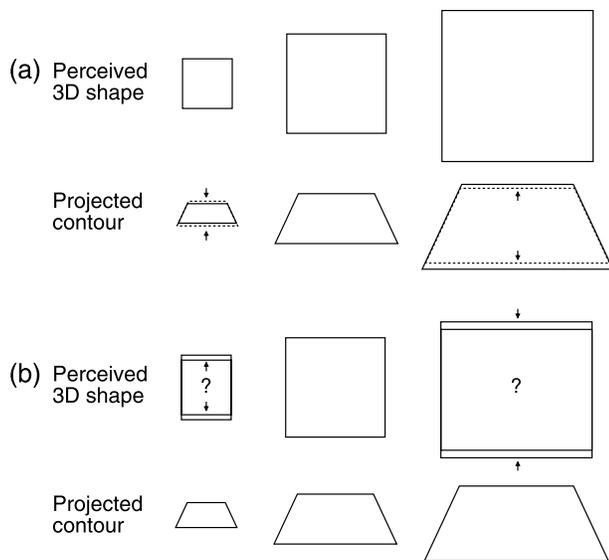


Figure 10. (a) Interpretation of the PSEs from Experiment 1. For each projected size, the contour that is perceived as a square 3D object has a different shape. The contours shown are consistent with the mean PSEs in the high-slant condition. (b) A related question: for a given contour shape, what is the perceived 3D object shape across different projected sizes? The degree to which the perceived 3D object is stretched in depth at small sizes (left) or compressed at large sizes (right) cannot be determined from Experiment 1 alone.

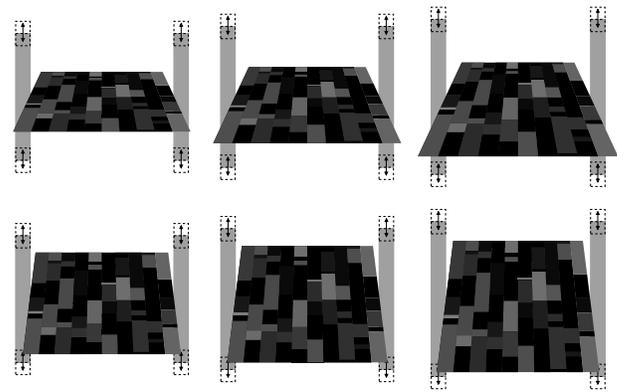


Figure 11. The stimuli and length-matching task used in Experiment 2. Subjects adjusted a pair of vertical bars to appear the same length as the slanted 3D rectangle. Surfaces were textured with random planks (see text). Convergence angles and projected sizes were the same as in Experiment 1. In the high-slant conditions, the trapezoids had aspect ratios of 0.4, 0.5, or 0.6 (top, left to right). In the low-slant conditions, the aspect ratios were 0.7, 0.8, or 0.9 (bottom, left to right). From the results of Experiment 1, the short trapezoids (left) would be expected to appear as wide rectangles and the taller trapezoids (right) as long rectangles, with the intermediate trapezoids appearing close to square.

sizes. One could also ask: How does perceived 3D object shape vary across different sized contours with the same shape (Figure 10b)? From the data from Experiment 1, we can infer that a contour that appears to be a square object at an intermediate size would appear as an elongated rectangle when presented at smaller sizes and as a shortened object at larger sizes. However, the extent to which the perceived 3D objects appear elongated or shortened cannot necessarily be determined.

In particular, the aspect ratio of a projected contour might interact with perspective information in determining perceived 3D slant and shape. For example, if foreshortening contributed as a slant cue, varying projected aspect ratio would change both perceived slant and perceived 3D shape. This would tend toward making the 3D objects appear closer to being square and thereby reduce the effect of projected size on perceived depth. It is also possible that projected aspect ratio interacts in the opposite direction; for example, tall projected contours might be perceived as more slanted than shorter contours with the same perspective convergence.

These possibilities are tested in Experiment 2. Figure 11 depicts the stimuli and task. Subjects viewed images with trapezoid-shaped figures, which appeared as 3D rectangles, and adjusted the length of vertical (frontal) comparison rectangles until its length matched the apparent length of the 3D rectangle. This task allows one to compare perceived shape for different-sized but identically shaped contours.

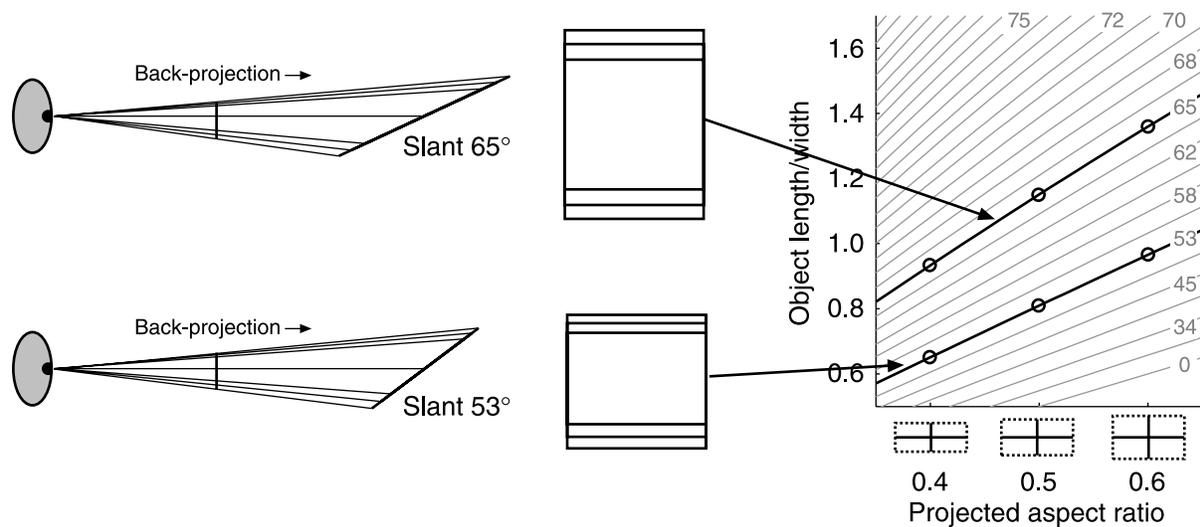


Figure 12. Predicted results for [Experiment 2](#) if judgments were consistent with some constant planar surface across changes in contour aspect ratio. When a projected contour is back projected onto a slanted surface (left), the length of the resulting 3D object (middle) depends on both surface slant and contour height. The rightmost graph plots the length-to-width ratio of a 3D object (y-axis) as a function of the aspect ratio of its projection contour (x-axis), for various possible surface slants (curve labels). If perceived slant depends on convergence but not contour height (or aspect ratio), judgments of 3D object dimensions should lie along one of these lines. The two example cases shown on the left have been highlighted on the graph.

The slant of a rectangular object is determined by the width and side angles of its projected contour. The contour's aspect ratio also factors into the 3D shape of the rectangle but not into its slant. In the extreme, if perceived slant were determined solely by perspective convergence, then trapezoidal projected contours with the same width and side angles would be perceived to have the same slant, regardless of the aspect ratio of the projected contour. This would in turn imply a relationship between the aspect ratio of a projected contour and its 3D interpretation. [Figure 12](#) illustrates this mapping, for a range of possible slants.

Methods

Apparatus and display

The display apparatus was the same as in [Experiment 1](#). Stimuli consisted of a textured trapezoid and an adjoining pair of (identical) variable-height comparison rectangles. [Figure 11](#) shows the six trapezoidal shapes that were used as projected contours. Two convergence angles were tested (same as [Experiment 1](#)): 25 and 7.125 deg. For each convergence angle, three projected aspect ratios were tested: 0.3, 0.4, and 0.5 for the high-slant condition (top row, left to right) and 0.7, 0.8, and 0.9 for the low-slant condition (bottom row, left to right). These aspect ratios were chosen because they spanned the range of shapes that were judged to be the projections of squares, based on the mean data from [Experiment 1](#). That is, we would

expect, on average, the leftmost stimuli to be seen as wider than longer and the rightmost stimuli as longer than wider. Each projected shape was presented at three different sizes, with horizontal widths of 7.1, 14, or 20.6 deg (same as [Experiment 1](#)).

In [Experiment 1](#), judgments were more reliable (lower thresholds) for the checkerboard textured surface than for the blank surface, although PSEs were similar. One possibility is that the texture helped stabilize the interpretation of the 3D object. To achieve this potential benefit without introducing a strong cue based on the gradient of vertical spacing, we used a random plank surface texture (illustrated in [Figure 11](#)) similar to that of Andersen et al. (1998). This texture tiled the surface with 10 columns of rectangles with uniform width but variable length. Each column was subdivided into separate tiles at 10 locations, chosen randomly from a uniform distribution. A different random pattern of planks and brightness values was generated for each trial. The comparison rectangles were simulated to have the same width as the rectangle, with height being controlled by the subject.

Procedure

Subjects' task was to adjust the height of the comparison rectangles until they appeared to match the length of the slanted rectangular object. The initial comparison height on a trial was chosen randomly to be between 70% and 140% of the contour's projected width, and subjects

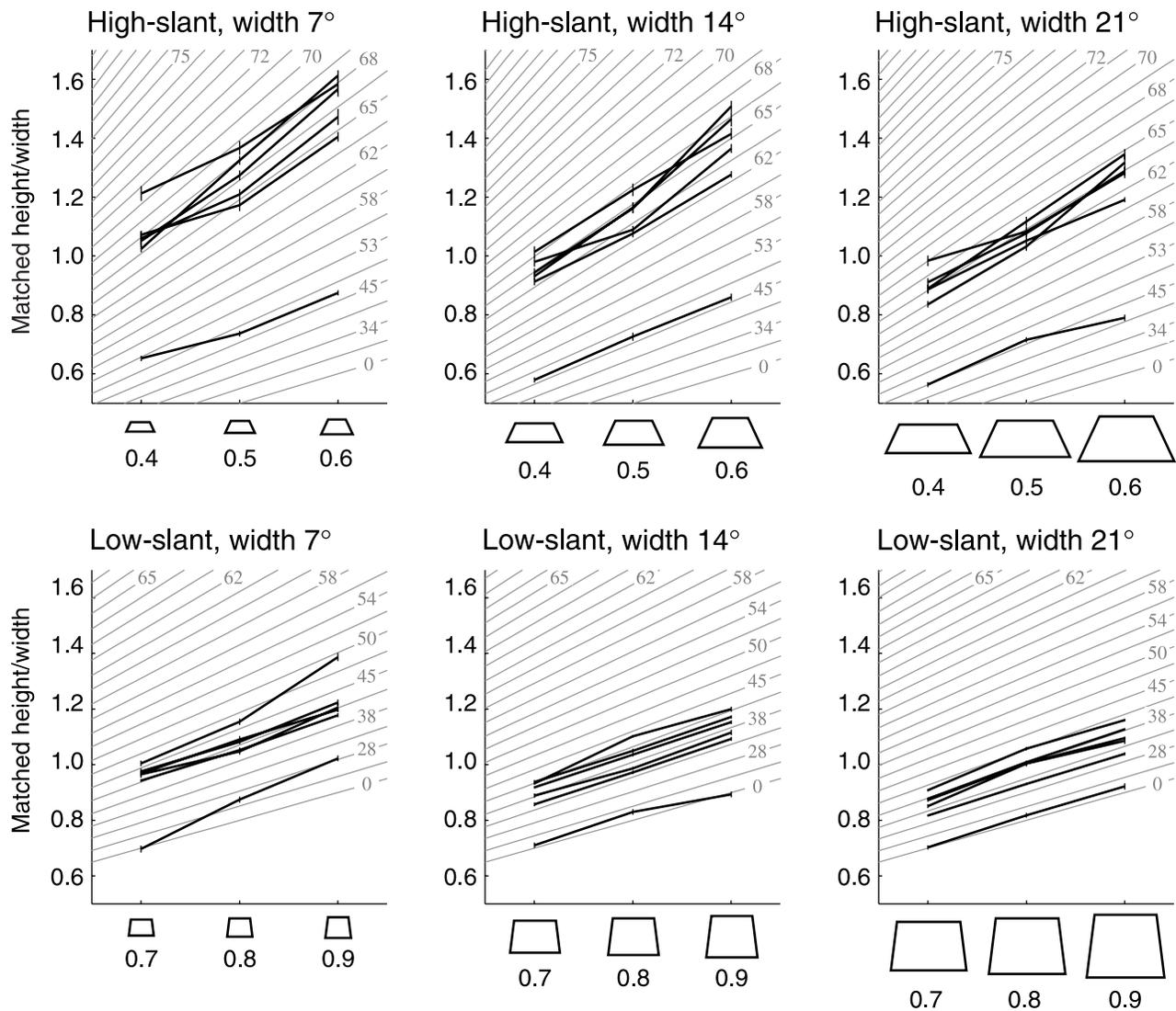


Figure 13. Results of Experiment 2. The black lines plot mean matching length settings for individual subjects as a function of projected aspect ratio. The six graphs correspond to high- and low-slant conditions (top and bottom) and to the different size conditions (left to right). The gray lines depict the patterns of responses that would be consistent with a constant surface orientation for various possible slants (see Figure 10 for description). Overall, the data plots were close to being aligned with predicted curves, indicating that subjects' judgments changed as a function of aspect ratio at a rate that was consistent with their mean bias, as would be expected if they perceived the objects as having a constant (but biased) surface orientation. Note that the one low outlier on each graph is the same subject.

increased or decreased the height in 1.4% steps using a keyboard. Trials were self-paced with no feedback. The experiment consisted of two blocks of 180 trials in a single 1-hr session, yielding 20 trials for each of 18 conditions (3 sizes \times 2 slants \times 3 aspect ratios) per subject. Conditions were randomized within blocks.

Subjects

Six subjects participated in Experiment 2. One of the subjects was the first author. The others were naive to the purposes of the experiment and were paid for participating. All subjects had normal or corrected-to-normal

vision. Subjects gave informed consent in accordance to a protocol approved by the IRB panel of University of Pennsylvania.

Results

Height settings were averaged across trials in a condition for a given subject, then divided by projected width to normalize for scale. The resulting measure represents the length-to-width ratio of the perceived 3D rectangle. Figure 13 plots mean length-to-width ratios for individual subjects as a function of projected aspect ratio, for each of the three

sizes (left to right) and two side angles (top and bottom). The labeled gray lines show how length-to-width ratio varies for a given slant as the contour's aspect ratio changes. Across conditions with the same size and side angle, the mean length-to-width settings for a given observer lie along one these curves, which means that their settings were consistent with perceiving the same 3D slant regardless of the contour aspect ratio.

As in [Experiment 1](#), the effect of projected size was significant but smaller than expected based on perspective geometry. If responses were veridical, length-to-width ratios would lie along the curves corresponding to slants of 82, 75, and 68 deg for the high-slant conditions (small to large widths) and slants of 63, 45, and 34 deg for the low-slant conditions. As can be seen in [Figure 13](#), subjects' judgments corresponded to surface slants that had much less size modulation. In all subjects, the slants implied by judgments decreased with projected size, but this decrease was, on average, only 3.8 deg for the high-slant conditions and 6.4 deg for the low-slant conditions, corresponding to gains of 27% and 34%, respectively. One outlier subject made responses that were consistent with lower overall perceived slants but similarly exhibited the smaller-than-expected size modulation.

[Figure 14](#) plots inferred slant as a function of contour side angle (high slant vs. low slant) and projected size, combining data across projected aspect ratio conditions. The filled circles plot the mean inferred slants, averaged across the six subjects. The dashed lines depict predicted results based on accurate use of perspective convergence. Relative to veridical, subjects' perceived slants are biased

toward zero (frontal) and show less change as a function of projected size, consistent with the results of [Experiment 1](#). For comparison, we computed inferred perceived slants based on the PSEs from [Experiment 1](#), which are also plotted in [Figure 14](#) (open circles). The biases are similar across the two experiments.

Discussion

The results of [Experiment 2](#) demonstrate that, for our stimuli, small changes in the aspect ratio of a projected contour have little effect on perceived slant in depth. Across conditions with different aspect ratios but with the same perspective convergence and projected size, judgments were consistent with a constant slant.

This finding is consistent with earlier results of Braunstein and Payne (1969). In their study, slant information from perspective convergence and from foreshortening were independently varied by manipulating the horizontal and vertical spacing of a texture composed of a grid of dots. Braunstein and Payne similarly observed that convergence was the dominant factor in determining perceived slant. The trapezoidal contour stimuli used by Smith (1967) also varied aspect ratio independently of side angle. Unfortunately, the conditions cannot be easily compared to ours. Smith varied the width rather than the height of projected contours, which changes both foreshortening and perspective information.

Although our results suggest that perceived slant in depth was determined primarily by perspective convergence, the data do not allow a strong test of this hypothesis. It remains possible that projected aspect ratio does influence perceived slant for our class of stimuli but that its effect is too small to be observed across the modest range of aspect ratios we tested.

What we can conclude is that any effect of contour aspect ratio is not sufficient to account for the smaller-than-expected effect of projected size on judgments of length in depth. The projected aspect ratio for which the object appears square in 3D (as was measured in [Experiment 1](#)) can be inferred from the data in [Experiment 2](#) by looking at the point where perceived length-to-width ratio equals 1 (i.e., where $y = 1$ on the plots in [Figure 13](#)). If there was a significant interaction between perspective convergence and aspect ratio, these points might be similar across projected size conditions, even if the perceived length-to-width ratio for a given aspect ratio changed by a large amount. This is clearly not the case.

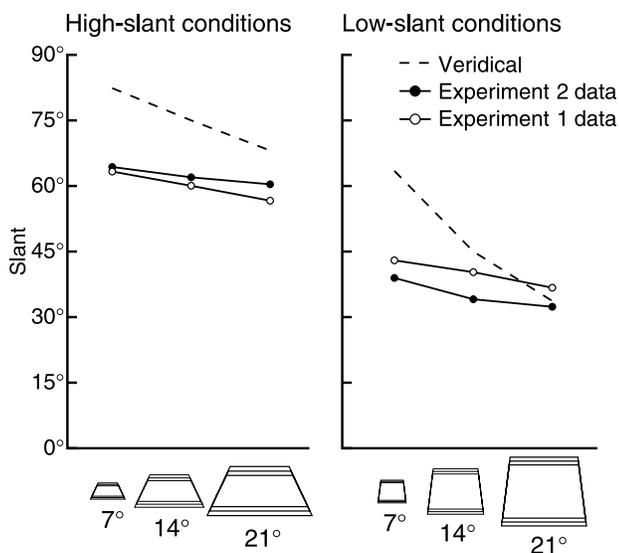


Figure 14. Replotting of data as inferred perceived slants. Open circles and filled circles show slants inferred from [Experiments 1](#) and [2](#), respectively. Dashed lines plot the veridical slants assuming parallel sides.

Ideal observer model

In both experiments, subjects' judgments were biased relative to veridical. We hypothesized that a model that

correctly internalized perspective geometry might still show the observed pattern of biases, if the perspective cue was not strong enough, by itself, to overcome conflicting information and prior assumptions.

As described earlier, a bias toward perceiving depth as flattened toward the frontal plane would explain why image trapezoids would have to be taller, relative to veridical, to be perceived as 3D squares. Such an overall bias might be due to absent or conflicting cues, awareness of the pictorial surface, or some prior assumption of constant depth. Additionally, if perspective information were only partially weighted relative to conflicting information, this could also explain why projected size had less effect on subjects' judgments than predicted by perspective geometry.

The simplest variant of this cue conflict explanation would be if perceived slant were some constant weighted average of the slant from perspective and the slant specified by other information. Because the latter slant is zero, perceived slant would be linearly related to perspective slant, by a constant factor. This clearly does not fit our data; the discrepancy between subjects' judgments and veridical performance varies greatly depending on both contour size and shape.

A constant-weight linear model is also overly simplistic in principle because it ignores how the information specified by perspective varies across conditions. First, sensory measures of contour shape and size would have varying amounts of noise. For example, it is known that discrimination of line orientations depends on both line length and overall orientation (Heeley, Buchanan-Smith, Cromwell, & Wright, 1997; Regan & Price, 1986; Snippe & Koenderink, 1994) and that discrimination of angles depends on the reference angle (Chen & Levi, 1996; Heeley & Buchanan-Smith, 1996; Regan, Gray, & Hamstra, 1996). Second, noise in image measurements propagates to different amounts of uncertainty in slant. Finally, because the slant from perspective depends on projected size, the amount of conflict between the perspective cue and other information effectively varies with size as well.

To incorporate such factors in our hypothesized explanation, we simulated the performance of a nonlinear Bayesian ideal observer for the task and stimuli we used. A bias toward perceptual compression of depth was modeled as a prior distribution over the set of possible surface slants, which was integrated with perspective information.

The task of the ideal observer model was to estimate the 3D shape and slant of planar object, based on a trapezoid-shaped projected contour. The projected contour was specified by its projected width (w), aspect ratio (r_{proj}), and side angles (a_{proj}). We assumed that slant was around a horizontal axis (vertical tilt direction); thus, the 3D object would be a symmetric trapezoid as well. The 3D object was also specified by its size, length-to-width ratio (r_{obj}), and the angle of its sides relative to its midline

(a_{obj}). There is an unavoidable ambiguity with respect to the overall size and distance of the 3D object; hence, without loss of generality, we ignore its size parameter. Thus, in terms of the defined parameters, the model's task was to estimate the slant (s) and shape (r_{obj} , a_{obj}) of the 3D object, from a projected contour with a given projected width (w) and shape (r_{proj} , a_{proj}). The results of [Experiment 2](#) suggest that perceived slant does not depend on projected aspect ratio; thus, we initially estimated s and a_{obj} based solely on w and a_{proj} . The estimate of slant was combined with r_{proj} to determine r_{obj} , which was what the model (and subjects) judged.

The model's estimate of s and a_{obj} was the combination that maximized the posterior probability function $P(s, a_{\text{obj}} | a_{\text{proj}}, w)$. By applying Bayes' rule, this can be expressed in terms of the likelihood function $P(a_{\text{proj}} | s, a_{\text{obj}}, w)$ and the priors on s and a_{obj} :

$$P(s, a_{\text{obj}} | a_{\text{proj}}, w) \sim P(a_{\text{proj}} | s, a_{\text{obj}}, w) P(s) P(a_{\text{obj}}). \quad (2)$$

To compute the likelihood function $P(a_{\text{proj}} | s, a_{\text{obj}}, w)$, we assumed that the image measures of w and a_{proj} were unbiased but corrupted by noise and then marginalized over the possible true values. Details of the noise model are given in [Appendix B](#).

The bottom left panel of [Figure 15](#) shows the likelihood function $P(a_{\text{proj}} | s, a_{\text{obj}}, w)$ computed for an example image trapezoid (top left). There is a range of 3D interpretations with high likelihood lying along a curve. Two particular points along the curve are marked for illustration. One is the zero slant interpretation; in this case, the 3D object has the same trapezoid shape as the projected contour. The other special case marked in the figure is where the high likelihood curve intersects the axis $a_{\text{obj}} = 0$, which corresponds to the 3D interpretation assuming parallel sides. The other points with high likelihood are intermediate cases, where the 3D shape is a trapezoid with less steeped sides than the 2D projected contour and is less slanted than the parallel-sides interpretation.

The middle and right panels in the bottom part of [Figure 15](#) show the result of combining the likelihood function $P(a_{\text{proj}} | s, a_{\text{obj}}, w)$ with the different priors for s and a_{obj} . The top middle panel shows $P(s, a_{\text{obj}})$ assuming a uniform prior for s and a Gaussian prior for the shape parameter a_{obj} , centered around zero with standard deviation $\sigma_{\text{persp}} = 6$ deg. This prior assigns higher likelihood to interpretations for which the object's sides are near parallel. The bottom middle panel shows the result of multiplying these priors with $P(a_{\text{proj}} | s, a_{\text{obj}}, w)$ to obtain the posterior $P(s, a_{\text{obj}} | a_{\text{proj}}, w)$. As might be expected, the maximum of this function is very close to the parallel-sides interpretation. The top right panel shows a different set of priors. The prior on a_{obj} is the same, but the prior on s is weighted toward zero, $P(s) = \cos(s)$. This particular prior has been suggested by Hillis et al. (2004), who point out that it

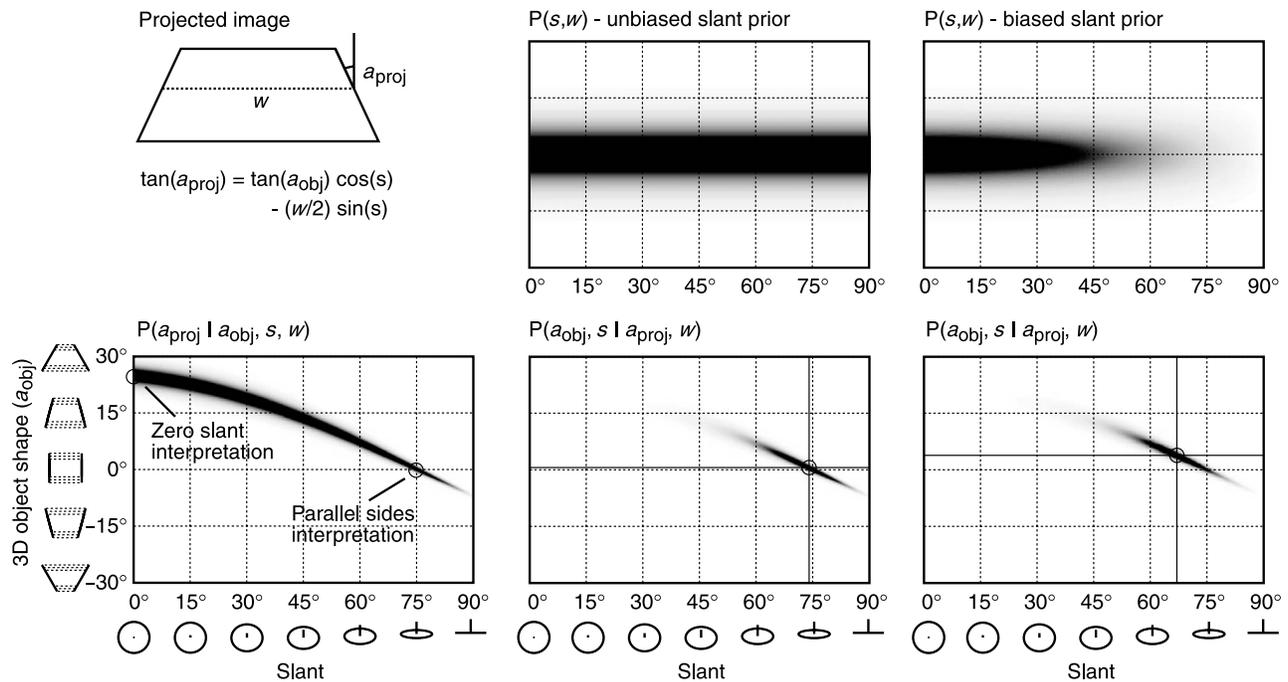


Figure 15. Illustration of ideal observer computation (see text).

describes the distribution of viewer-relative slants in an environment where all 3D surface orientations are equally likely. When this biased slant prior is integrated with perspective information, the peak of $P(s, a_{\text{obj}} | a_{\text{proj}}, w)$ is shifted away from the correct parallel interpretation to a point with lower slant (bottom right).

The final step in our model simulations was to convert slant estimates into simulated performance in the 3D dimension judgment task performed by subjects. Given a projected contour and its estimated slant, the shape of the corresponding 3D object can be determined by simply back projecting the contour. We assumed no decision noise; hence, judgments were directly determined by the length-to-width ratio of the back-projected shape. On any given trial, however, image measures of the width and shape of a projected contour would be perturbed by noise. The same noise models used in deriving the likelihood functions served as generative models for trial-to-trial variability. By integrating model judgments over the possible perturbations, we computed expected psychometric functions for a given stimulus.

Figure 16 plots the simulated performance of the model with a biased cosine prior on slant, for the conditions tested in Experiment 1. The model's PSEs (left) exhibit both an overall bias and reduced modulation by projected size, as in the human data. The model's discrimination thresholds (right) are lower overall than the human results but otherwise show a similar pattern. This agreement provides confirmation that our choice of noise parameters was reasonable. The fact that these parameters also result in a good fit to PSE data supports our hypothesis, demonstrat-

ing that a general bias toward underestimating depth could account for much of the deviations from veridicality observed in our results.

On the other hand, the quantitative agreement is clearly imperfect and is not easily improved by choice of parameters within our simple formulation. With a fixed slant prior and noise parameters determined from psychophysical data, the only remaining freedom is in the choice of the parallelism prior, $P(a_{\text{obj}})$, which in our model is specified by the single parameter σ_{persp} . Setting this parameter much higher than in our simulations can lead to qualitative discrepancies. In addition, there is one discrepancy between model performance and human data that cannot be explained within our formulation, regardless of parameters. In the 21 deg, low-slant condition, subjects' judgments on average were close to veridical, and for some subjects, they were biased slightly in the opposite direction as in the other conditions. In our model, the prior toward low slants is the only factor that produces deviations from veridical performance; hence, one would never expect biases in the direction opposite to perceptual compression of depth.

Thus, there must be some factors, other than conflicting depth cues or priors, contributing to biases in subjects' performance. We have argued that foreshortening information could not account for the smaller-than-predicted size modulation. However, an influence of this cue might be sufficient to shift overall biases. The possibility that there are simply errors in the learned mapping from perspective convergence to slant or depth also remains. In our formulation, this would correspond to $f(a_{\text{obj}}, s, w')$

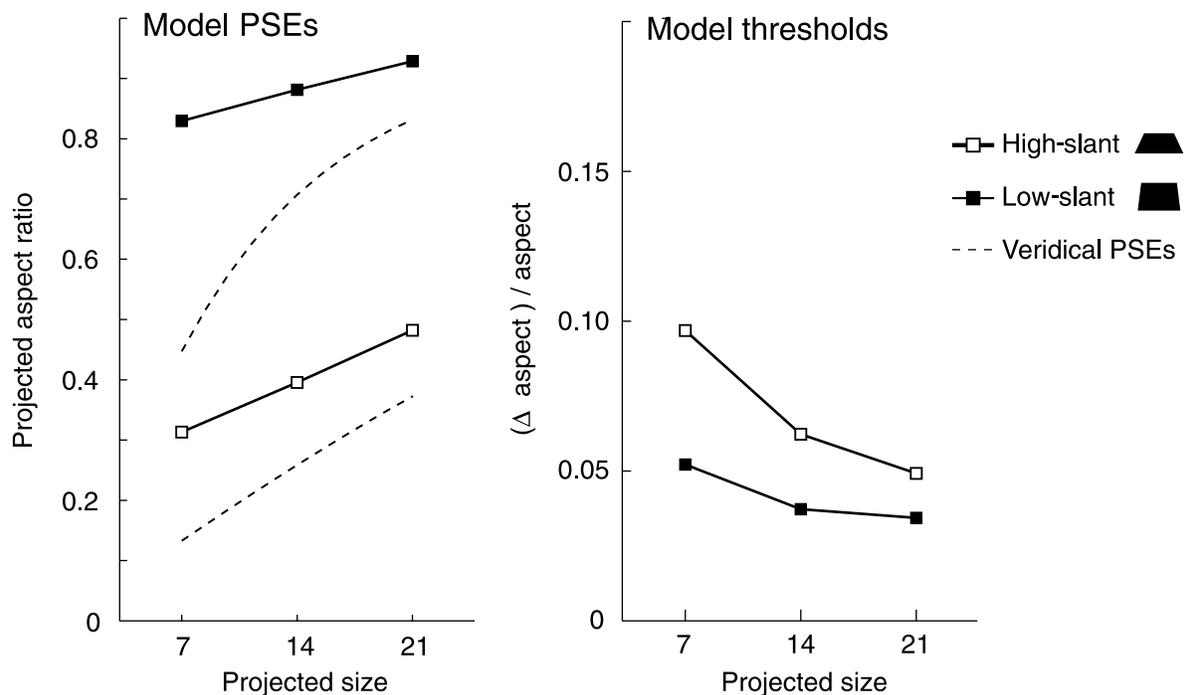


Figure 16. PSEs and 75% thresholds from the model simulations (see text).

being inaccurate. Perceptual distortions have been observed even for a very simple shape matching task (Henriques et al., 2005); hence, the possibility of systematic distortions in 3D interpretations cannot be discounted.

Although the model fit is not perfect, the simulations demonstrate that much of the observed pattern of biases could potentially be explained by a general perceptual bias toward an absence of depth, which has been observed in many settings. Thus, despite the fact that subjects' judgments were not veridical, it remains possible that the visual system was interpreting perspective information in a geometrically appropriate way but that the isolated perspective cue is relatively weak compared to conflicting information.

General discussion

Our results demonstrate that subjects are able to make reliable judgments of length in depth based on relatively impoverished images that isolated perspective convergence as 3D information. Judgments showed large biases relative to veridical but had relatively low variability. The perceived length of slanted objects, as indicated by observers' judgments, was larger (relative to width) for small images than large images, as predicted by perspective geometry. The only monocular depth cue that specified nonzero slant, besides perspective convergence, was foreshortening of the projected contour (i.e., its projected aspect ratio). The slant implied by this cue does

not vary with overall image size, and therefore could not explain the size effect observed in the data. Moreover, the results of Experiment 2 suggest that foreshortening had little effect on judgments in our task and conditions. Thus, our results clearly implicate perspective convergence as the basis for perceived 3D structure in our stimuli.

An earlier attempt by Freeman (1966a) to measure the ability to discriminate slant from perspective was criticized on the grounds that the task could have been performed as 2D shape comparisons between sequential presentations (see Flock, 1965, or Smith, 1967). Our method avoids this problem because judgments were relative to a fixed, familiar standard. To account for our results in terms of a 2D shape-based strategy, one would have to assume that the shape standard used for comparison varied systematically depending on projected size. In addition, we demonstrated with a control experiment that 2D shape discrimination judgments for our stimuli are invariant to projected size, whereas performance in the 3D judgment task showed significant improvement with increased size. We conclude that our experiment was effective in probing perception of 3D shape from perspective.

One distinguishing characteristic of our paradigm is that absolute accuracy can be assessed because there was always a correct response based solely on the monocular information, which depended on a well-learned standard (a square). We found that, across conditions, judgments showed large overall biases in the direction consistent with perceptual compression of depth. This is in agreement with results from other studies using direct report or probe-matching tasks, which have consistently found

underestimation of slant and depth based on perspective convergence (Andersen et al., 1998; Rosinski et al., 1980; Smith, 1967; Todd et al., 2005).

Judgments were also inaccurate in that the effect of image size was not as large as would be expected based on perspective geometry. A similar finding was reported by Smith (1967) for the case of slant judgments for simple contour stimuli. The fact that some size modulation was consistently observed indicates that this was not simply due to poor sensitivity to image size, and the results of [Experiment 2](#) rule out the foreshortening cue as a primary causal factor.

We hypothesized that both the overall bias in perceived depth and the smaller-than-expected effect of size were due to a general perceptual compression of depth. Perceptual compression of depth has been observed under various conditions for a variety of measures (for a review, see Todd & Norman, 2003). In our experiment, perceptual compression could be due to some a priori bias and/or the influence of conflicting information specifying flatness, such as lack of accommodation, influence of the screen frame, and so forth.

To test the viability of this explanation, we simulated the performance of a Bayesian ideal observer that incorporated a probabilistic version of a parallelism assumption. With a uniform prior for slant, average performance of the model was veridical, but with a prior that is weighted toward low slants (i.e., biased toward an absence of depth), estimates exhibited biases that were qualitatively similar to the human data. Thus, the deviations from veridicality we observed could arise even for a model that accurately internalized perspective geometry, if one assumes that perspective information by itself is not “strong” enough (i.e., a weak parallelism assumption) to fully counteract either a priori assumptions or conflicting information indicating an absence of depth. We cannot rule out the possibility that the visual system has simply instantiated a biased model of perspective projection. However, the simulations demonstrate that such errors need not be assumed to account for most of the pattern of our results, particularly given that there is prior reason to expect some overall bias toward compression of depth.

Some previous studies have compared slant judgments for slanted rectangles with different sizes but with the same slant (Freeman, 1966b; Stavrianos, 1945). This is different from what we tested, because when the slant is fixed, scaling a rectangular object increases the amount of convergence along with projected size. In contrast, we varied projected size while keeping convergence constant. However, from our data, we can infer how perceived slant would change if the object size was varied while holding slant constant. As can be seen in [Figure 7](#), the slant specified by perspective (dashed line) was similar for the small contours with low convergence and the large contours with high convergence. The contours judged to be squares, however, were much taller for the small stimuli, implying that they were perceived as less slanted.

Extrapolating from our data, this difference would be expected to remain even if slant from perspective was exactly matched. Thus, our results indicate that a large rectangle would be perceived as more slanted than a smaller rectangle with the same slant, which is consistent with earlier reports (Freeman, 1966b; Stavrianos, 1945).

Our results are in partial conflict with those of Nichols and Kennedy (1993) and Yang and Kubovy (1999). In both studies, subjects rated images as most cube-like when their projected size was consistent with perspective information, across a range of sizes. One interpretation is that observers interpret perspective information in a way that accurately depends on projected size. For example, subjects might have used perspective convergence to perceive extent in depth and then based their judgments on whether the corners appeared stretched or compressed relative to that of a cube. By this interpretation, a preference for geometrically consistent sizes conflicts with our results because we observed that perceived depth was compressed overall and did not scale with projected size by the predicted amount. In Yang and Kubovy’s experiment, biases could have been obscured by the relatively coarse sampling of sizes. However, given the large overall biases we observed relative to veridical, some detectable effect would be expected.

A significant difference between our stimuli and those used by both Nichols and Kennedy and Yang and Kubovy is that their stimuli contained size-independent depth cues. If the faces of the cube-like object were implicitly assumed to have right-angle corners or if they were assumed to be isotropic, then a unique 3D interpretation is possible even if projected size were unknown (see [Figure 4](#)). Thus, the stimuli where the perspective cue was inconsistent with projected size were cue conflict situations, with the task being to judge the degree of conflict. In contrast, we chose restricted conditions to isolate the information available from perspective convergence. In particular, our stimuli lacked a skew symmetry cue, which is scale invariant and which has been shown to be an effective cue to slant (Saunders & Knill, 2001). For trapezoid-shaped contours such as the ones we used, an assumption of parallel sides implies right-angle corners and vice versa; hence, there is no conflict between perspective and skew symmetry information. In a follow-up study, we are exploring whether perceived depth from perspective is affected by the presence of scale-invariant information from skew symmetry.

Implications for picture perception

Finally, we consider the implications of our results regarding the perception of 3D structure in perspective pictures and photographs. As many others have noted, observers are not typically positioned at the correct center of projection when viewing pictures. Consequently, the pictorial cues presented to the observer would generally

specify a 3D structure that is distorted relative to the depicted scene.

Much of the previous work on perception of 3D structure in pictures has focused on the effect of viewing pictures from an angle, which can induce a variety of perceptual distortions (Goldstein, 1987, 1988; Halloran, 1993; Koenderink, van Doorn, Kappers, & Todd, 2004; Perkins, 1974). In this situation, the visual system could potentially use information about the 3D orientation of the pictorial surface to compensate for a slanted viewpoint. There is considerable debate both as to the amount of robustness in perceived 3D structure and the extent that pictorial surface cues contribute to compensation (Farber & Rosinski, 1978; Goldstein, 1979, 1987, 1988; Halloran, 1993; Koenderink et al., 2004; Kubovy, 1986; Perkins, 1974; Rosinski & Farber, 1980; Rosinski et al., 1980; Vishwanath, Girshick, & Banks, 2005; Wallach & Marshall, 1986). Although this is an interesting issue, it is not directly relevant to the present experiments because only projected size was varied. When viewing pictures from the wrong distance, rather than from an angle, knowledge about the pictorial surface no longer provides useful information for perceptual compensation. Projected size would still be required to interpret depth from perspective convergence, regardless of the distance of the picture.

In normal viewing of pictures, inconsistencies in pictorial information due to incorrect scaling may be larger and more common than inconsistencies due to viewing angle. When allowed, observers will likely choose viewing positions that are roughly normal to the picture surface (e.g., sitting in the center of a theater, holding a snapshot frontally). In contrast, large inconsistencies in projected size would likely remain. The relationship between angular field of view depicted in a picture and the angle that it subtends when viewed depends on many factors, such as the power of the camera lens (i.e., telephoto or wide angle), the physical size of the picture, and the distance of the observer from the picture.

Based on our results, the perceived 3D structure of a scene based on a photograph would indeed be expected to change as a function of its magnification. The effect of projected size we observed was smaller than that predicted by accurate use of perspective information, corresponding to a gain of 0.2–0.3. This size dependence, while limited, would still produce significant distortions in perceived depth when viewing photographs with varying magnification.

Our finding that size modulation is partial might appear to conflict with the results of some previous studies, which have reported that 3D judgments from magnified or minified images are consistent with geometric predictions (e.g., Bengston et al., 1980; Smith & Gruber, 1958). However, the notion of geometric consistency tested by these experiments is very different, and there is no actual conflict with our results. Previous studies measured whether scaled and unscaled stimuli *with identical perspective information* were judged to have equivalent depth structure.

This is essentially a cue conflict paradigm. In the case of the experiment by Smith and Gruber (1958), which compared judgments for photos and actual scenes, the conflict would consist of any cues that differ between views of an actual scene and a photograph. In other studies, scaled and unscaled photos with matching perspective convergence were compared (Bengston et al., 1980; Lumsden, 1983; Smith, 1958a, 1958b). In this case, conflicts would arise from size-invariant monocular depth cues, such as texture compression or familiar size. One can imagine an analogous variant of our experiment, in which judgments were based on either rendered stimuli or monocular views of actual checkerboard surfaces, constructed to have matching projected images. If results were similar, it would not imply that perspective convergence was interpreted in an accurate, scale-dependent way. Rather, it would imply that perspective convergence dominated other 3D cues. Similarly, to the extent that results differed, it would imply that other cues influenced judgments. Thus, this type of design addresses the question of how much perspective contributes relative to other depth information. This is very different from asking, as in our experiment, whether perceived depth from convergence changes in a geometrically accurate way when projected size is varied.

If perception of 3D structure from perspective does depend on projected size, as suggested by our results, there remains a question as to why perception of pictures appears “robust”—that is, why we do not more frequently experience noticeable distortions. One possibility is that size-dependent distortions are tolerated because the (incorrectly) perceived scene is also plausible. This sort of explanation is discussed in Koenderink et al. (2004). For example, in the special case of rectangular objects that are partially aligned with the picture plane (as tested here), distortions due to image magnification preserve properties like mirror symmetry and right-angle corners (i.e., rectangles remain rectangular) and, therefore, might not interfere with the ability to perceive the general structure of the scene. More generally, perception of 3D structure might be based on intrinsic geometric relations that are invariant to the sorts of distortions caused by changes in viewpoint (Busey, Brady, & Cutting, 1990; Gibson, 1950; Perkins, 1972; Sedgwick, 1983, 1991).

It is also possible that the size dependence we observed was due to the degenerate nature of our stimuli. Real-world scenes provide a variety of monocular depth cues besides perspective convergence, some of which provide size-invariant information (e.g., skew symmetry). Presenting an image at the wrong projected size (as when viewing a picture from the wrong distance) generally introduces a conflict between the depth specified by perspective and size-invariant cues. However, in our stimuli, this conflict was intentionally minimized. A natural question for further work is whether perceived depth from perspective is modulated by image size in a similar way when other monocular cues are available.

Appendix A: Minimized expected entropy staircase method

In choosing the stimulus aspect ratios to test on each trial, we used a new adaptive procedure, which we term *minimized expected entropy staircase method*.

For a given trial, the probe ratios and responses from previous trials in the same condition, $\{x_k, r_k\}$, were used to estimate a posterior probability distribution $P(\mu, \sigma | x_1, r_1, x_2, r_2, \dots, x_n, r_n)$, where μ is the PSE and σ is the difference between the PSE and the 75% point. The next probe x_{n+1} was chosen to minimize the expected entropy, $-p \log(p)$, of the posttrial posterior function, $P(\mu, \sigma | x_1, r_1, x_2, r_2, \dots, x_{n+1}, r_{n+1})$. The entropy cost function rewards probes that would be expected to result in a more peaked and concentrated posterior distribution over the space of possible combinations of μ and σ , consistent with the goal of estimating μ and σ with minimal bounds of uncertainty.

There are only two possibilities for the next response, 0 or 1, and for each of these possibilities, one can compute what the new postresponse likelihood distribution would be, as well as its entropy. The expected value of entropy is simply a weighted average of the two possible results, where weights are proportional to their probabilities, $P(r_{n+1} = 0 | x_{n+1})$ and $P(r_{n+1} = 1 | x_{n+1})$. If μ and σ were known, these probabilities would be directly determined by the model psychometric function. Thus, to estimate $P(r_{n+1} | x_{n+1})$, we marginalized over μ and σ , using the posterior distribution computed from previous response history as an estimate of $P(\mu, \sigma)$:

$$P(r_{n+1} | x_{n+1}) \approx \sum_{\mu, \sigma} P(r_{n+1} | x_{n+1}, \mu, \sigma) \times P(\mu, \sigma | x_1, r_1, \dots, x_n, r_n). \quad (\text{A1})$$

In our implementation, we used a logistic function to model the psychometric function $P(r_{n+1} | x_{n+1}, \mu, \sigma)$, rather than a more standard cumulative Gaussian, to simplify computation during probe selection. Also, the function was scaled to range from 0.025 to 0.975 rather than from 0 to 1, to reduce the effect of lapses of attention and guessing on the probe selection. The space of possible bias and threshold values was discretely sampled to carry out marginalization, with σ sampled linearly from the set $\{0.05, 0.1, \dots, 0.8\}$ and μ sampled exponentially from the set $\{0.26, 0.274, \dots, 3.87\}$ for low-slant conditions and from the set $\{0.094, 0.100, \dots, 1.42\}$ for high-slant conditions.

Our staircase method is a greedy algorithm, in that it minimizes the expected entropy only after the succeeding trial, not for the whole future sequence. We do not yet know how much this greedy method diverges from a full

optimization. However, informal testing of the procedure revealed it to be highly efficient and robust.

One aspect of the method's behavior that could be problematic in practice is that, once the estimates of μ and σ have converged, the probe choices tend to oscillate between two values, symmetric around the PSE, and be anticorrelated with the previous response. This occurs because the expected entropy function at this point has two local minima that are very similar, such that a single response switches their relative depths. Consequently, probe values would tend to alternate, which could influence a subject's behavior. In the experiment reported here, there were many interleaved conditions in each block and there were a modest number of trials per staircase; hence, temporal correlations were not a concern. However, in a design with few conditions and many trials per staircase, this would be a more serious problem. A simple solution is to use a random subset of the response history to estimate the posterior function, rather than the whole history, once a sufficient number of trials are recorded. Because the method converges to a rough estimate quickly (within 15–20 trials), excluding a subset of trials has little effect on the final distribution of probe samples. Note that, with this modification, there is no need to run multiple interleaved staircases using our method, as is commonly done when using standard staircases.

Appendix B: Measurement noise for the ideal observer model

In this appendix, we describe how we modeled noise in image measurements for our ideal observer simulations.

We modeled the noise in the shape parameter a_{proj} as being Gaussian, with a width parameter σ_a that was set based on previous psychophysical measures of 2D orientation discrimination. Discrimination of 2D orientations exhibits an oblique effect: Thresholds are higher away from the horizontal and vertical axes (Heeley et al., 1997; Regan & Price, 1986; Snippe & Koenderink, 1994). In the case of our stimuli, this would imply that orientations are encoded less reliably for our high-slant conditions than for our low-slant conditions. Uncertainty in shape measurement could alternatively be modeled as a function of corner angles of the projected figure, as opposed to the orientations of its side edges. Thresholds for 2D angle discrimination follow an m-shaped function of base angle, with a local minimum at 90 deg (Chen and Levi, 1996; Heeley and Buchanan-Smith, 1996; Regan, Gray, et al., 1996); thus, one would similarly expect greater noise for the high-slant conditions. On the basis of these various results, we estimated that the effective orientation/angle noise for the high-slant condition would be about twice as high as for the low-slant condition, with all other factors equal. Orientation discrimination for 2D

lines has also been found to strongly depend on line length. For extended lines, thresholds decrease roughly with the square root of length (Heeley & Buchanan-Smith, 1998; Orban, Vandenbussche, and Vogels, 1984). Incorporating this length dependence, our model for noise in projected shape was $2.5 \text{ deg}/\sqrt{L}$ for the low-slant conditions and $5 \text{ deg}/\sqrt{L}$ for the high-slant conditions, where L is the length of the side edges (in degrees of visual angle).

For uncertainty in measurement of the width of projected figures, we assumed proportional noise around the correct value, such that $\log(w)$ is a Gaussian with deviation $\sigma_w = 0.05 \text{ log units}$. This would be consistent with results from studies of interval length discrimination, which have found that thresholds increase proportionally with length, with Weber fraction of approximately 0.05, across a range of conditions (Burbeck, 1987; Toet, van Eekhout, Simons, & Koenderink, 1987; Whitaker & Latham, 1997). We found that this noise parameter could be varied somewhat without affecting the qualitative performance of the model, provided that it remains small compared with the uncertainty introduced by orientation noise.

For any combination of slant and 3D object shape, the true projected shape parameter a'_{proj} is determined by perspective geometry; hence, only the projected width parameter w' needs to be explicitly marginalized. Using the noise models, the desired likelihood function becomes:

$$P(a_{\text{proj}}|s, a_{\text{obj}}, w) \sim \int Z([a_{\text{proj}} - f(a_{\text{obj}}, s, w')]/\sigma_a) \times Z([w - w']/\sigma_w)dw', \quad (\text{B1})$$

where Z is a standard Gaussian distribution and f is the projection function mapping a_{obj} to a_{proj} for a given slant and width.

The final step of the modeling was to simulate the effect of measurement noise on performance of our experimental task (see main text). For this, it was necessary to assume a noise model for the measurement of contour height, in addition to noise in measurement of width and side angles. We assumed proportional noise for contour height, with the same Weber fraction, 5%, as for image measurement of contour width.

Acknowledgment

This research was supported by NIH Grant EY-013988.

Commercial relationships: none.

Corresponding author: Jeffrey A. Saunders.

Email: jeffrey_a_saunders@yahoo.com.

Address: 3401 Walnut Street, Philadelphia, PA 19104, USA.

References

- Andersen, G. J., Braunstein, M. L., & Saidpour, A. (1998). The perception of depth and slant from texture in three-dimensional scenes. *Perception*, *27*, 1087–1106. [PubMed]
- Attneave, F., & Olson, R. K. (1966). Inferences about visual mechanisms from monocular depth effects. *Psychonomic Science*, *4*, 133–134.
- Banks, M. S., & Backus, B. T. (1998). Extra-retinal and perspective cues cause the small range of the induced effect. *Vision Research*, *38*, 187–194. [PubMed]
- Bengston, J. K., Stergios, J. C., Ward, J. L., & Jester, R. E. (1980). Optic array determinants of apparent distance and size in pictures. *Journal of Experimental Psychology: Human Perception and Performance*, *6*, 751–759. [PubMed]
- Braunstein, M. L., & Payne, J. W. (1969). Perspective and form ratio as determinants of relative slant judgments. *Journal of Experimental Psychology*, *81*, 584–590.
- Burbeck, C. A. (1987). Position and spatial frequency in large-scale localization judgments. *Vision Research*, *27*, 417–427. [PubMed]
- Busey, T. A., Brady, N. P., & Cutting, J. E. (1990). Compensation is unnecessary for perception of faces in slanted pictures. *Perception & Psychophysics*, *48*, 1–11. [PubMed]
- Chen, S., & Levi, D. M. (1996). Angle judgement: Is the whole the sum of its parts? *Vision Research*, *36*, 1721–1735. [PubMed]
- Clark, W. C., Smith, A. H., & Rabe, A. (1955). Retinal gradients of outline as a stimulus for slant. *Canadian Journal of Psychology*, *9*, 247–253. [PubMed]
- Clark, W. C., Smith, A. H., & Rabe, A. (1956). The interaction of surface texture, outline gradient, and ground in the perception of slant. *Canadian Journal of Psychology*, *10*, 1–8. [PubMed]
- Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The interaction, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Perception of space and motion* (pp. 69–117). San Diego, CA: Academic Press.
- Farber, J., & Rosinski, R. R. (1978). Geometric transformations of pictured space. *Perception*, *7*, 269–282. [PubMed]
- Flock, H. R. (1965). Optical texture and linear perspective as stimuli for slant perception. *Psychological Review*, *72*, 505–514. [PubMed]
- Freeman, R. B., Jr. (1966a). Absolute threshold for visual slant: The effect of stimulus size and retinal perspective.

- Journal of Experimental Psychology*, 71, 170–176. [PubMed]
- Freeman, R. B., Jr. (1966b). Function of cues in the perceptual learning of visual slant: An experimental and theoretical analysis. *Psychological Monographs*, 80, 1–29. [PubMed]
- Garding, J. (1993). Shape from texture and contour by weak isotropy. *Artificial Intelligence*, 64, 243–297.
- Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.
- Gillam, B. J. (1968). Perception of slant when perspective and stereopsis conflict: Experiments with aniseikonic lenses. *Journal of Experimental Psychology*, 78, 299–305. [PubMed]
- Gogel, W. C. (1965). Equidistance tendency and its consequences. *Psychological Bulletin*, 64, 153–163. [PubMed]
- Goldstein, E. B. (1979). Rotation of objects in pictures viewed at an angle: Evidence for different properties of two types of pictorial space. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 78–87. [PubMed]
- Goldstein, E. B. (1987). Spatial layout, orientation relative to the observer, and perceived projection in pictures viewed at an angle. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 256–266. [PubMed]
- Goldstein, E. B. (1988). Geometry or not geometry? Perceived orientation and spatial layout in pictures viewed at an angle. *Journal of Experimental Psychology: Human Perception and Performance*, 14, 312–314. [PubMed]
- Halloran, T. O. (1993). The frame turns also: Factors in differential rotation in pictures. *Perception and Psychophysics*, 54, 496–508. [PubMed]
- Heeley, D. W., & Buchanan-Smith, H. M. (1996). Mechanisms specialized for the perception of image geometry. *Vision Research*, 36, 3607–3627. [PubMed]
- Heeley, D. W., & Buchanan-Smith, H. M. (1998). The influence of stimulus shape on orientation acuity. *Experimental Brain Research*, 120, 217–222. [PubMed]
- Heeley, D. W., Buchanan-Smith, H. M., Cromwell, J. A., & Wright, J. S. (1997). The oblique effect in orientation acuity. *Vision Research*, 37, 235–242. [PubMed]
- Henriques, D. Y., Flanders, M., & Soechting, J. F. (2005). Distortions in the visual perception of shape. *Experimental Brain Research*, 160, 384–393. [PubMed]
- Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: optimal cue combination. *Journal of Vision*, 4(12), 967–992, <http://journalofvision.org/4/12/1/>, doi:10.1167/4.12.1. [PubMed] [Article]
- Kanade, T. (1981). Recovery of the three-dimensional shape of an object from a single view. *Artificial Intelligence*, 17, 409–460.
- Knill, D. C. (1998a). Discrimination of planar surface slant from texture: Human and ideal observers compared. *Vision Research*, 38, 1683–1711. [PubMed]
- Knill, D. C. (1998b). Ideal observer perturbation analysis reveals human strategies for inferring surface orientation from texture. *Vision Research*, 38, 2635–2656. [PubMed]
- Knill, D. C., & Saunders, J. A. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, 43, 2539–2558. [PubMed]
- Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. (1994). On so-called paradoxical monocular stereoscopy. *Perception*, 23, 583–594. [PubMed]
- Koenderink, J. J., van Doorn, A. J., Kappers, A. M., & Todd, J. T. (2004). Pointing out of the picture. *Perception*, 33, 513–530. [PubMed]
- Kubovy, M. (1986). *The psychology of perspective and renaissance art*. Cambridge: Cambridge University Press.
- Li, A., & Zaidi, Q. (2000). Perception of three-dimensional shape from texture is based on patterns of oriented energy. *Vision Research*, 40, 217–242. [PubMed]
- Lumsden, E. A. (1983). Perception of radial distance as a function of magnification and truncation of depicted spatial layout. *Perception & Psychophysics*, 33, 177–182. [PubMed]
- Nichols, A. L., & Kennedy, J. M. (1993). Angular subtense effects on perception of polar and parallel projections of cubes. *Perception & Psychophysics*, 54, 763–772. [PubMed]
- Orban, G. A., Vandenbussche, E., & Vogels, R. (1984). Human orientation discrimination tested with long stimuli. *Vision Research*, 24, 121–128. [PubMed]
- Perkins, D. N. (1972). Visual discrimination between rectangular and nonrectangular parallelepipeds. *Perception & Psychophysics*, 12, 396–400.
- Perkins, D. N. (1974). Compensation for distortion in viewing pictures obliquely. *Perception & Psychophysics*, 14, 13–18.
- Perkins, D. N. (1976). How good a bet is good form? *Perception*, 5, 393–406. [PubMed]
- Regan, D., Gray, R., & Hamstra, S. J. (1996). Evidence for a neural mechanism that encodes angles. *Vision Research*, 36, 323–330. [PubMed]

- Regan, D., Hajdur, L. V., & Hong, H. (1996). Two-dimensional aspect ratio discrimination for shape defined by orientation texture. *Vision Research*, *36*, 3695–3702. [PubMed]
- Regan, D., & Hamstra, S. J. (1992). Shape discrimination and the judgment of perfect symmetry: Dissociation of shape from size. *Vision Research*, *32*, 1845–1864. [PubMed]
- Regan, D., & Price, P. (1986). Periodicity in orientation discrimination and the unconfounding of visual information. *Vision Research*, *26*, 1299–1302. [PubMed]
- Rosas, P., Wichmann, F. A., & Wagemans, J. (2004). Some observations on the effects of slant and texture type on slant-from-texture. *Vision Research*, *44*, 1511–1535. [PubMed]
- Rosenholtz, R., & Malik, J. (1997). Surface orientation from texture: Isotropy or homogeneity (or both)? *Vision Research*, *37*, 2283–2293. [PubMed]
- Rosinski, R. R., & Farber, J. (1980). Compensation for viewing point in the perception of pictured space. In M. A. Hagen (Ed.), *The perception of pictures* (Vol. 1, pp. 137–176). New York: Academic Press.
- Rosinski, R. R., Mulholland, T., Degelman, D., & Farber, J. (1980). Picture perception: An analysis of visual compensation. *Perception & Psychophysics*, *28*, 521–526. [PubMed]
- Saunders, J. A., & Backus, B. T. (2006). Perception of slant from oriented textures. *Journal of Vision*, *6*(9), 882–897, <http://journalofvision.org/6/9/3/>, doi:10.1167/6.9.3. [PubMed] [Article]
- Saunders, J. A., & Knill, D. C. (2001). Perception of 3D surface orientation from skew symmetry. *Vision Research*, *41*, 3163–3183. [PubMed]
- Sedgwick, H. A. (1980). The geometry of spatial layout in pictorial representation. In M. A. Hagen (Ed.), *The perception of pictures* (Vol. 1, pp. 33–90). New York: Academic Press.
- Sedgwick, H. A. (1983). Environment-centered representation of spatial layout: Available visual information from texture and perspective. In J. Beck, B. Hope, & A. Rosenfield (Eds.), *Human and machine vision* (pp. 425–458). New York: Academic Press.
- Sedgwick, H. A. (1991). The effect of viewpoint on the virtual space of pictures. In S. R. Ellis, M. K. Kaiser, & A. C. Grunwald (Eds.), *Pictorial communication in virtual and real environments* (pp. 460–469). London: Taylor & Francis.
- Smith, A. H. (1967). Perceived slant as a function of stimulus contour and vertical dimension. *Perceptual and Motor Skills*, *24*, 167–173.
- Smith, O. W. (1958a). Comparison of apparent depth in a photograph viewed from two distances. *Perceptual and Motor Skills*, *8*, 79–81.
- Smith, O. W. (1958b). Judgments of size and distance in photographs. *American Journal of Psychology*, *71*, 529–538. [PubMed]
- Smith, O. W., & Gruber, H. (1958). Perception of depth in photographs. *Perceptual and Motor Skills*, *8*, 307–313.
- Snippe, H. P., & Koenderink, J. J. (1994). Discrimination of geometric angle in the fronto-parallel plane. *Spatial Vision*, *8*, 309–328. [PubMed]
- Stavrianos, B. K. (1945). The relation of shape perception to explicit judgments of inclination. *Archives of Psychology*, No. 296.
- Stevens, K. A. (1983). Slant-tilt: The visual encoding of surface orientation. *Biological Cybernetics*, *46*, 183–195. [PubMed]
- Todd, J. T., & Norman, J. F. (2003). The visual perception of 3-D shape from multiple cues: Are observers capable of perceiving metric structure? *Perception & Psychophysics*, *65*, 31–47. [PubMed]
- Todd, J. T., Thaler, L., & Dijkstra, T. M. (2005). The effects of field of view on the perception of 3D slant from texture. *Vision Research*, *45*, 1501–1517. [PubMed]
- Toet, A., van Eekhout, M. P., Simons, H. L., & Koenderink, J. J. (1987). Scale invariant features of differential spatial displacement discrimination. *Vision Research*, *27*, 441–451. [PubMed]
- Vishwanath, D., Girshick, A. R., & Banks, M. S. (2005). Why pictures look right when viewed from the wrong place. *Nature Neuroscience*, *8*, 1401–1410. [PubMed]
- Wallach, H., & Marshall, F. J. (1986). Shape constancy in pictorial representation. *Perception & Psychophysics*, *39*, 233–235. [PubMed]
- Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision*, *5*(10), 834–862, <http://journalofvision.org/5/10/7/>, doi:10.1167/5.10.7. [PubMed] [Article]
- Whitaker, D., & Latham, K. (1997). Disentangling the role of spatial scale, separation and eccentricity in Weber's law for position. *Vision Research*, *37*, 515–524. [PubMed]
- Witkin, A. P. (1981). Recovering surface shape and orientation from texture. *Artificial Intelligence*, *17*, 17–45.
- Yang, T., & Kubovy, M. (1999). Weakening the robustness of perspective: Evidence for a modified theory of compensation in picture perception. *Perception & Psychophysics*, *61*, 456–467. [PubMed]
- Zanker, J. M., & Quenzer, T. (1999). How to tell circles from ellipses: Perceiving the regularity of simple shapes. *Naturwissenschaften*, *86*, 492–495. [PubMed]