Mental geometry of three-dimensional size perception

Akihito Maruya Graduate Center for Vision Research, State University of New York, New York, NY

Qasim Zaidi Graduate Center for Vision Research, State University of New York, New York, NY

Judging the poses, sizes, and shapes of objects accurately is necessary for organisms and machines to operate successfully in the world. Retinal images of three-dimensional objects are mapped by the rules of projective geometry and preserve the invariants of that geometry. Since Plato, it has been debated whether geometry is innate to the human brain, and Poincare and Einstein thought it worth examining whether formal geometry arises from experience with the world. We examine if humans have learned to exploit projective geometry to estimate sizes and aspects of three-dimensional shape that are related to relative lengths and aspect ratios.

Numerous studies have examined size invariance as a function of physical distance, which changes scale on the retina. However, it is surprising that possible constancy or inconstancy of relative size seems not to have been investigated for object pose, which changes retinal image size differently along different axes. We show systematic underestimation of length for extents pointing toward or away from the observer, both for static objects and dynamically rotating objects. Observers do correct for projected shortening according to the optimal back-transform, obtained by inverting the projection function, but the correction is inadequate by a multiplicative factor. The clue is provided by the greater underestimation for longer objects, and the observation that they seem to be more slanted toward the observer. Adding a multiplicative factor for perceived slant in the back-transform model provides good fits to the corrections used by observers. We quantify the slant illusion with two different slant matching measurements, and use a dynamic demonstration to show that the slant illusion perceptually dominates length nonrigidity.

In biological and mechanical objects, distortions of shape are manifold, and changes in aspect ratio and relative limb sizes are functionally important. Our model shows that observers try to retain invariance of these aspects of shape to three-dimensional rotation by correcting retinal image distortions due to perspective projection, but the corrections can fall short. We discuss how these results imply that humans have internalized particular aspects of projective geometry through evolution or learning, and if humans assume that images are preserving the continuity, collinearity, and convergence invariances of projective geometry, that would simply explain why illusions such as Ames' chair appear cohesive despite being a projection of disjointed elements, and thus supplement the generic viewpoint assumption.

Introduction

A biological or machine visual system can successfully operate in the world only by accurately judging poses, sizes, and shapes of objects. In eyes with lenses, projective geometry describes the mapping of three-dimensional (3D) scenes to retinal images, so in understanding what they are viewing, animals and humans have constant experience with the consequences of this geometry. Not surprisingly, whether geometrical operations are innately embedded in the human mind has been debated since antiquity, for example, Plato's Meno (Cooper, 2002). In an interesting development, Poincare (2017) and Einstein (1921) expanded the debate to whether formal geometry arises from everyday experience. Consequently, if we can show that human estimates of object size and shape are based on geometric knowledge, that could provide answers to age-old questions about the links between geometry and experience.

A 3D object seen from different views forms quite different retinal images, and many different objects can form identical retinal images (Ittelson, 1952), so 3D inferences based solely on monocular two-dimensional (2D) retinal information are underspecified. However, the frequently occurring projection of reflections from objects on the ground to retinal images is a 2D-to-2D mapping, described by an invertible trigonometric function. So in this case, the back-transform derived by inverting the projection function could be used to make veridical inferences from retinal images. We showed previously that human observers consistently apply

Citation: Maruya, A., & Zaidi, Q. (2020). Mental geometry of three-dimensional size perception. *Journal of Vision*, *20*(8):14, 1–16, https://doi.org/10.1167/jov.20.8.14.

https://doi.org/10.1167/jov.20.8.14 Received September 13, 2019; published August 7, 2020 ISSN 1534-7362

the optimal observer-oriented back-transform for pose inferences in 3D scenes and in pictures of 3D scenes (Koch et al., 2018). This leads to veridical estimates for real 3D scenes, albeit with a systematic frontoparallel bias, and to an illusory rotation toward the observer in obliquely viewed pictured scenes. In the images we used for pose estimation, we noticed that the perceived sizes of objects vary with pose, and shapes of objects seem to be distorted, especially those aspects of shape that depend on relative sizes in different directions. Here we tackle the mental geometry of estimating relative sizes, using graphic displays calibrated against real objects. Variations in these size estimates across different poses also provide information about perceived shape distortions, such as aspect ratios and relative sizes of limbs.

Mathematically defined, shape is the geometrical attribute of an object that is invariant to translation, rotation, and scale effects (Kendall et al., 2009), but whether these invariances hold for perceptions of particular solid 3D shapes is an empirical question. Invariance to translation and rotation are properties of Euclidean spaces, whereas retinal image formation is described by perspective projection, which does not allow for these invariances. Perceptual invariance would thus require neural processes that overcome distortions created by the retinal projection, so the first step is to quantify perceived aspects of 3D shape under different views.

There is a long tradition of research on shape and size constancy, but surprisingly some important aspects of this issue have been almost ignored, because almost all experiments have examined the perceptual effects of placing identical objects at different distances from the observer, which scales the size of the retinal image, but does not test for rotation or translation invariance. For example, Gibson (1950) examined size invariance as a function of physical distance and maintained that we see approximately the veridical size by making use of inverse projection to recover the structure of the environment from the structure in the optic array. A number of similar studies have shown that humans are able to compensate partially for the retinal projection, so estimated sizes are closer to the physical size of the object (Gibson & Cornsweet, 1952; Joynson & Newson, 1962; Kaiser, 1967; Purdy, 1960; Sedgwick, 1986; Wallach & Moore, 1962). However, studies that asked observers to match a width to a depth found inconstancy of perceived 3D relative size, as depth had to be set 1.5 to 5.0 times the width to be judged as equal over different distances (Beusmans, 1998; Loomis, Da Silva, Fujita, & Fukushima, 1992). In scene perception, a similar systematic perceptual anisotropy of depth versus width has been found (Baird & Biersdorf, 1967; Levin & Haber, 1993; Loomis & Philbeck, 1999; Norman et al, 1996; Philbeck, 2000; Ribeiro, Fukusima, & Da Silva, 1995; Toye, 1986; Wagner, 1985), suggesting a common cause, possibly insufficient correction of image compression caused by perspective projection. This perceptual anisotropy is not found when distances are estimated by nonvisual motor tasks such as blind walking (Elliott, 1986, 1987; Fukusima, Loomis, & Da Silva, 1997; Loomis et al., 1992; Loomis, Klatzky, Philbeck, & Golledge, 1998; Philbeck & Loomis, 1997; Philbeck, Loomis & Beall, 1997; Rieser, Ashmead, Talor, & Youngquist, 1990; Sinai, Ooi, & He, 1998; Steenhuis & Goodale, 1988; Thomson, 1983), thus characterizing the anisotropy as related to the greater compression ratios for depths in retinal images as compared with widths.

Object poses change retinal image size differently along different axes, but constancy of relative size seems not to have been investigated for these conditions, so we address that lacuna in this study. The view from the top in the movie in Figure 1 shows a rigid 3D object with two physically equal limbs at a right angle, rotating on the ground. In the view from the front looking down a 15° angle, most observers see the limb passing through poses pointing at or away from them as transitorily shorter than the orthogonal limb. Using quantitative measurements, we show that, for estimating relative sizes, observers generally correct sufficiently for projective distortions in accordance with the optimal back-transform, except for poses close to the line of sight. Size underestimation increases with object length, which we show is due to a slant illusion. The illusion occurs because increased length and increased slant both increase projections along the same axis, so in the absence of stereo cues, the visual system is unable to disambiguate the two causes. We discuss how results showing performance determined by the optimal geometric back-transform imply that humans have internalized particular aspects of projective geometry through evolution or learning.

Methods

Experiment 1: Size estimates of 3D objects

The first question we address is how well observers can estimate 3D sizes across different poses of the same object lying on the ground, because projections lead to different amounts of retinal image compression depending on pose angle (Figure A1). Using Blender, we created a blue rectangular 3D parallelepiped (test stick) lying on the center of a dark ground, and a yellow vertical 3D cylinder (measuring stick) standing on the test stick. Blue parallelepipeds were presented in one of 16 poses from 0° to 360° every 22.5° (Figure 2a). Their length was equal to 10, 8, or 6 cm with a 3 × 3 cm cross-section (Figure 2b). Observers estimated physical lengths by adjusting the height of the vertical

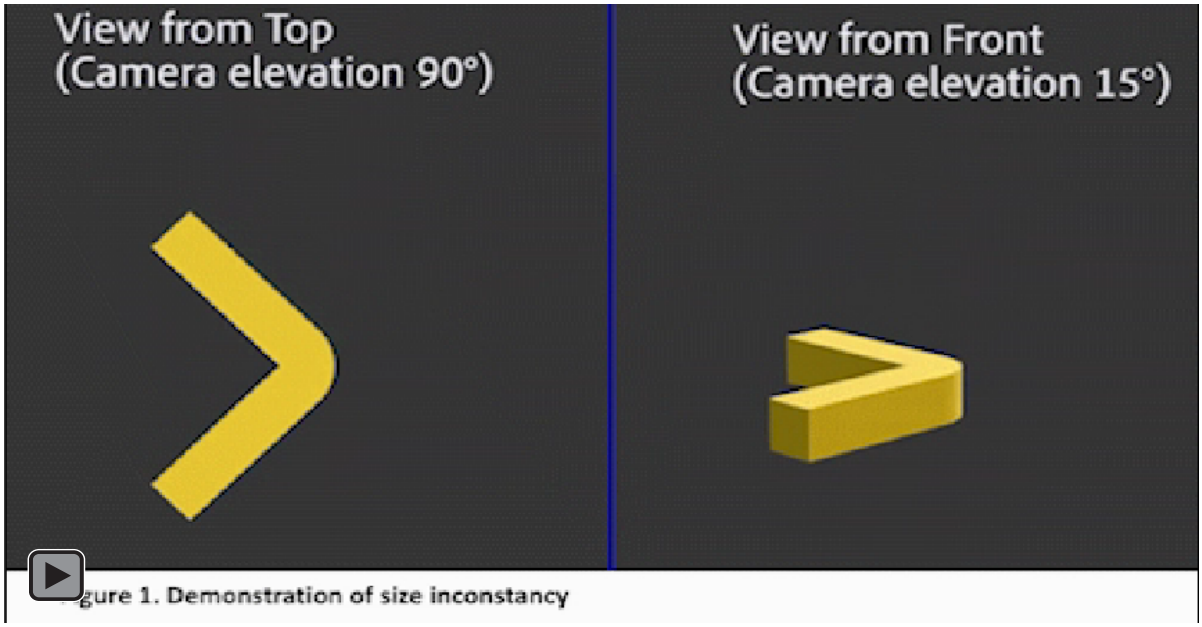

Figure 1. Demonstration of size inconstancy. A rigid object with two physically equal limbs at a right angle is rotating on the ground, as shown in the view from the top. When viewed from the front at a 15° elevation, the limb pointing at or away from the observer appears transitorily shorter than the orthogonal limb.

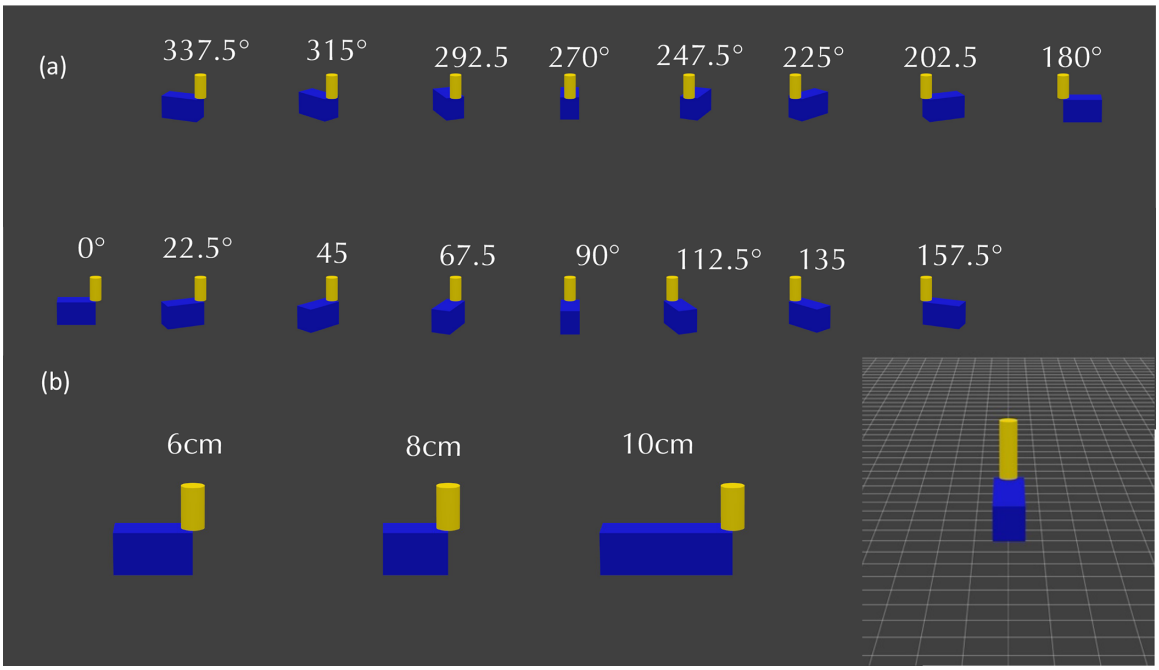

Figure 2. Stimuli for experiment 1. (a) Experiment 1a: Blue test stick lying on dark ground in 16 different poses, from 0° to 337.5° at every 22.5°. (b) Experiment 1a: Three lengths, 6, 8, 10 cm, of blue stick were presented in each pose, with adjustable yellow measuring stick. (c) Experiment 1b: Same as Experiment 1a, except for a grid making the ground plane explicit.

measuring stick between 2.75 and 12 cm, until it seemed to be the same 3D length as the horizontal test stick. In Experiment 1a, the object was presented on a dark ground plane and in Experiment 1b a regular white grid was superimposed on the ground plane to make it explicit (Figure 2c). The dark ground was used so observers could not use a ground grid to estimate size.

Stimuli and methods

The observer's viewing position was stabilized by using a chinrest so that the image was viewed on the monitor with an elevation angle of 15° at a distance of 1.0 m, matching the rendering parameters of the camera in Blender. Displayed sizes in the Blender-rendered images were calibrated against exact geometrical derivations to ensure accuracy of the simulations (Appendix). The retinal image was thus identical to that from the 3D object, except for the absence of stereo disparity. Observers were instructed to equate the physical lengths of the two limbs by pushing buttons to adjust the height of the measuring stick. There were no time limits. Randomly ordered trials were repeated in three sets, and observers were allowed to take a break between sets. The line of sight through the center of the ground was designated the 90° to 270° axis and the line orthogonal to it, the 0° to 180° axis. Images were displayed on a 22-inch DELL SP2309W Display. Matlab and PsychoToolbox were used to display the stimulus, run the experiment, and analyze the data for all the experiments. Six observers with normal or corrected vision participated. Viewing was binocular because it was more natural for the observers, and Koch et al. (2018) had not found any difference between monocular and binocular viewing for pose estimation in similar conditions. All experiments presented in this article were conducted in compliance with the protocol approved by the institutional review board at SUNY College of Optometry and the Declaration of Helsinki, and observers gave written informed consent.

Results

Figure 3a shows perceived 3D lengths as a function of 3D pose, averaged over three repeats each for six observers. Dashed lines represent the physical length. Two trends are salient: there is greater underestimation of length for poses pointing toward or away from the observer, and there is greater underestimation of length for longer objects. Individual data show both trends for every observer (Figure A2). Underestimation for different object lengths can be compared on the same relative scale in Figure 3b, where the logarithm of the ratio of perceived length over physical length is plotted against the 3D poses. This figure confirms the two trends. The first factor we ruled out for the underestimation is that the dark ground made it ambiguous whether the object was lying on the ground, by rerunning all the conditions of Experiment 1a on a white grid drawn on the ground (Experiment 1b). When the average lengths perceived on the white grid are plotted against average lengths perceived on the dark ground, most symbols fall close to the unit diagonal

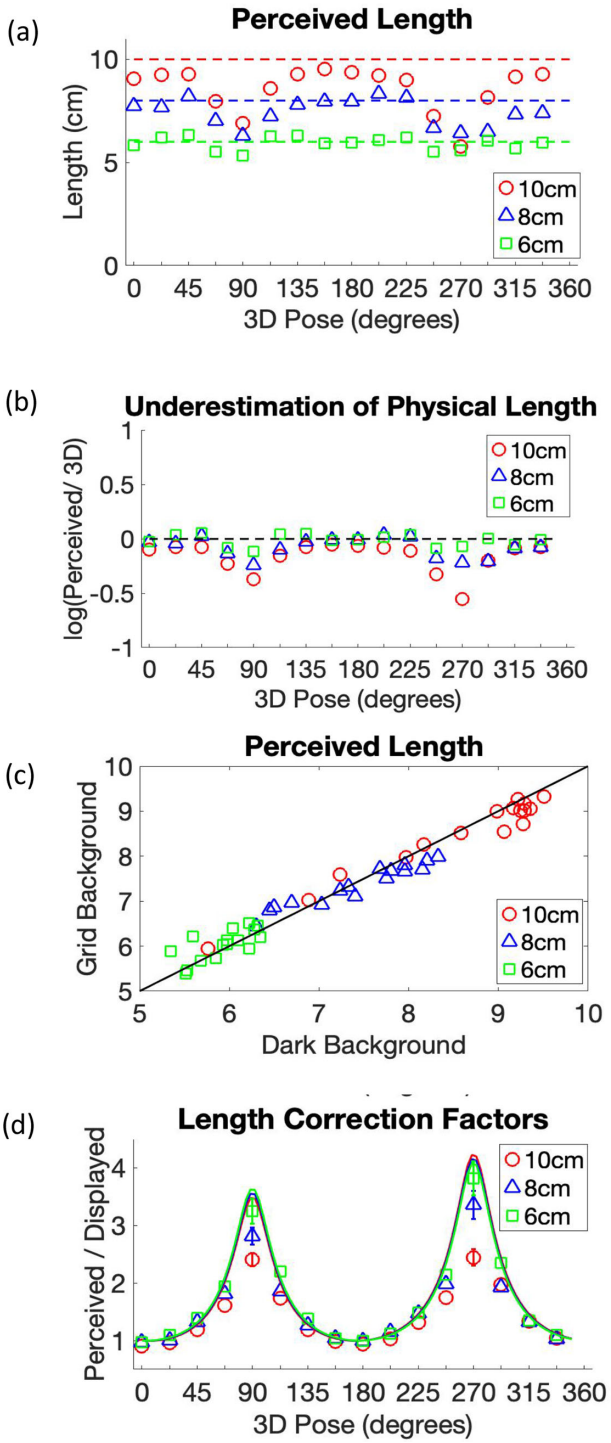

Figure 3. Perceived 3D lengths in experiment 1. Symbols represent parallelepipeds with lengths of 10 cm (red circle), 8 cm (blue triangle), and 6 cm (Green square). (a) Average perceived length across 3D pose (six observers.) Dashed lines indicate physical 3D lengths. Perceived lengths are underestimated as a systematic function of pose. (b) Logarithm of average perceived 3D length divided by physical length across different 3D poses. Length estimates of the test sticks were close to veridical for frontoparallel poses but were seriously underestimated for poses pointing at or away from

→

(Figure 3c), showing that the estimation of 3D length is very similar on the two ground planes.

Comparing empirical with optimal size estimation

We now try to understand estimation of 3D length, especially the underestimation for poses toward and away from the observer, and the greater underestimation for longer lengths, by deriving the geometrical information available to observers. A schematic diagram of the projection (Figure A1), and a mathematical derivation of Equation 1, are included in the Appendix. For a parallelepiped of length (L_{3D}), the projected length (L_c) changes with pose Ω as a distorted sinusoid (viewing elevation = Φ_c, focal length of the camera = f_c, and distance from the object = d_c):

$$L_c = \frac{L_{3D} \cdot f_c \cdot \sqrt{cos^2(\Omega) + sin^2(\Omega) \cdot sin^2(\Phi_c)}}{d_c - L_{3D} \cdot \sin(\Omega) \cdot \cos(\Phi_c)} \quad (1)$$

However, the projected length of the vertically oriented cylinder L_m (where the physical length is L_{3DM}), stays invariant with pose because the object is rotated around the axis of the cylinder:

$$L_m = \frac{L_{3DM} \cdot f_c \cdot \cos(\Phi_c)}{d_c - L_{3DM} \cdot \sin(\Phi_c)} \quad (2)$$

Given the projected lengths on the display, the projected lengths on the retina, L_r, would be (focal length of the eye = f_r, and distance from the display = d_r):

$$L_r = \frac{f_r}{d_r} \cdot L_c \quad (3)$$

On the dark ground, the retinal image of the object contains the only information available for doing size estimation and our model predicts 3D size estimates solely from retinal projections of the objects. Koch et al. (2018) showed that humans are excellent at inferring 3D pose of objects on the ground, and their estimations closely match predictions from the geometrical back-projection from retina to the ground plane. In principle, an observer could similarly make veridical estimates of 3D sizes by using the geometrical back-projection from the retinal image, which is given by substituting the expression for L_c from Eq. 1 into Eq. 3, and then manipulating the equation to get an expression for the optimal estimated 3D length $\hat{L}_{3D}$:

$$\hat{L}_{3D} = \frac{L_r \cdot d_r \cdot d_c}{L_r \cdot d_r \cdot \sin(\Omega) \cdot \cos(\Phi_c) + f_c \cdot f_r \cdot \sqrt{cos^2(\Omega) + sin^2(\Omega) \cdot sin^2(\Phi_c)}} \quad (4)$$

By dividing the physical length by the projected length, we obtain the optimal length correction index for each pose. From Equation 1, the projected length is almost a linear function of physical length, with a slight acceleration, so the optimal length correction varies little for the three sizes when plotted as a function of 3D pose, as shown by the overlap of the three solid line curves in Figure 3d. The symbols in Figure 3d plot the ratio of perceived length from Figure 3a over projected length (empirical length correction), and show that for poses other than frontoparallel, observers estimate 3D lengths to be longer than projected lengths (empirical length correction of > 1.0). The greatest correction takes place for poses pointing toward or away from the observer, but that is still less than what is required for veridical estimates, especially for the longer lengths. The general form of empirical length correction as a function of 3D pose is similar to the optimal length correction curve, suggesting that observers may be using the optimal back-transform, but with an additional multiplicative factor leading to the suboptimality.

Geometric model for 3D size estimation

A strong clue to the multiplicative factor is revealed by inspection of the stimuli in Figure 4a. When a 6-cm and a 10-cm parallelepiped are placed on the ground, the upper surface of the longer stick looks more slanted down. The genesis of this illusion is that, if the 6-cm stick were increased in length to 10 cm, its projection would come further down along the vertical axis of the image. A similar increase in vertical coordinates in the projected image would happen if the slant of the 6-cm stick was increased or equivalently if the object was pictured from a higher camera elevation. The visual system is thus faced with deciding between the quantitative increase in slant versus the quantitative increase in length. Figure 4b (left) shows that increasing the camera elevation changes the aspect ratio of a 10 cm × 3 cm top surface at 90° pose by a factor of 5.65 from

←

the viewer (around 90° and 270°.) The underestimation ratio increased with the physical length of the test stick. (c) Perceived length on grid background (experiment 1b) plotted against perceived length on dark background (experiment 1a), showing points falling close to the unit diagonal (solid line), indicating that they are similar. (d) Optimal correction factor (*solid line*) and empirical correction factor (*symbols*) across 3D pose. The empirical correction factor is close to one for the front-parallel poses, which is optimal. The empirical correction factor is greater than the one near the line of sight (90° and 270°), but significantly lower than the optimal for the longer sticks. Bars on symbols are 95% confidence intervals.

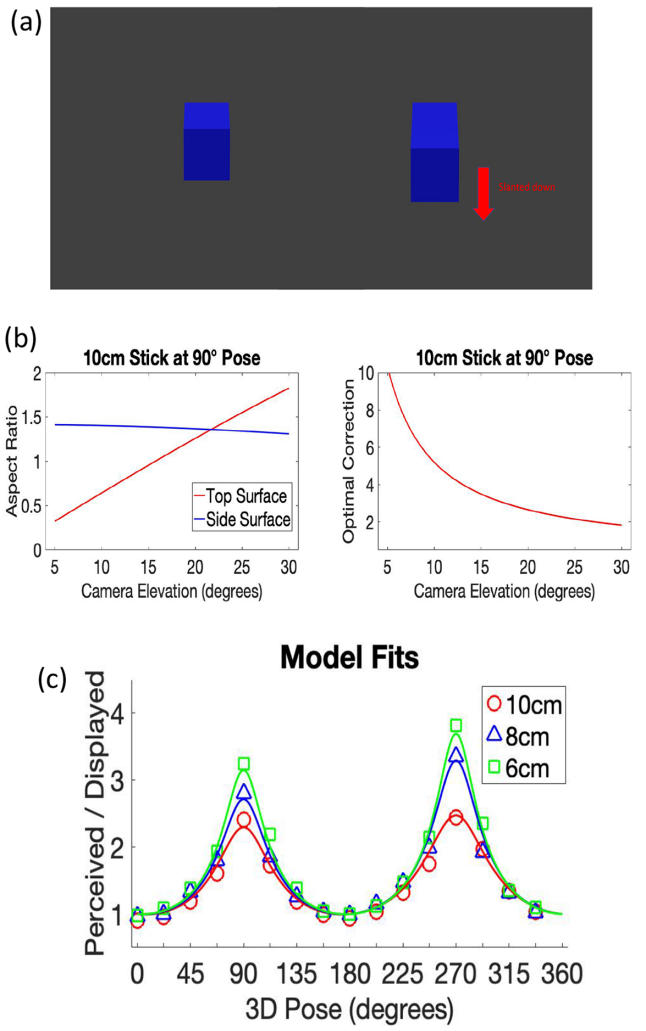

Figure 4. Geometry model for 3D size estimation. (a) The 6- and 10-cm test sticks are lying on ground, pictured with 15° camera elevation. The longer stick is seen as more slanted down toward the viewer, equivalent to an increase in viewing camera elevation. The effect is the same with a grid on the ground. (b) (*left*) Aspect ratio of the top surface of 10 × 3 cm (*red*) and front surface of 3 × 3 cm (*blue*) of stick lying on ground in a 90° pose against camera elevation. The main change is the projection of the length of the top surface. (*right*) Optimal correction factor for the stick as a function of camera elevation. The correction factor decreases with increasing camera elevation. (c) A model using the optimal geometrical back-transform but with overestimation of viewing elevation fits the underestimates of object length.

0.3227 (at 5°) to 1.8247 (at 30°), but barely changes the aspect ratio of the 3 cm × 3 cm front surface from 1.4120 at 5° to 1.3096 at 30°, a factor of 1.08. The main change is due to the shortening of the projected length of the top surface. The front surface changes are thus not a strong enough cue to discern change of camera elevation, or equivalently the slant of the object. In the absence of other cues, it seems that the visual system hedges its bets and the physically longer stick is seen as both longer and more slanted than the physically shorter stick. The illusion is powerful enough to be visible, even when three different length parallelepipeds are joined together and placed on a gridded ground (Figure A4). If a visual system assumes that an object is more slanted or pictured from a higher elevation, it will apply a smaller correction factor. Figure 4b (right) shows that the optimal length correction decreases by almost a factor of five as the cameral elevation increases from 5° to 30°; thus, a misperception of increased slant would lead to a smaller correction factor applied to the projected length. Note that the multiplicative factor will change the estimated length most for poses close to 90°, and very little for poses close to 0°.

Based on the results and the visual observations, we formulated the hypothesis that observers are using an optimal back-transform, but overestimating the slant of the object (or equivalently the camera elevation), thus correcting less than required for the shortening. Therefore, we tested whether adding a multiplier $k > 1$ to the viewing elevation in the optimal geometrical back-transform function could provide good fits to the empirical estimates for different physical lengths:

$$\hat{L}_{3D} = \frac{L_r \cdot d_r \cdot d_c}{L_r \cdot d_r \cdot \sin(\Omega) \cdot \cos(k \cdot \phi_c) + f_c \cdot f_r \cdot \sqrt{cos^2(\Omega) + sin^2(\Omega) \cdot sin^2(k \cdot \phi_c)}} \quad (5)$$

Figure 4c replots the empirical corrections from Figure 3d for the three lengths of test sticks. For each length, we found the k for which the optimal correction factor curve best fits the empirical correction factors, and shows that the model fits the results for all three physical lengths well, with just one free parameter that increases perceived camera elevation. because the perceived slant can be slightly different for 90° and 270° poses, we allowed k to be different values for 0° to 180° and 180° to 360°. Based on the best fitting k values, the estimated camera elevations were around 16° for the 6-cm stick, 20° for the 8-cm stick, and 25° for the 10-cm stick. Parenthetically, we also tested whether putting a multiplier on distance, to simulate misperceived distance (Sedgwick, 1989), could explain the data, but that was not successful. Consequently, we tested whether the slant illusion would hold up to quantitative measurements.

Experiment 2: Slant misestimation as a factor for suboptimal length correction

Perceived slant is affected by length of object and viewing elevation, so there is no way to measure

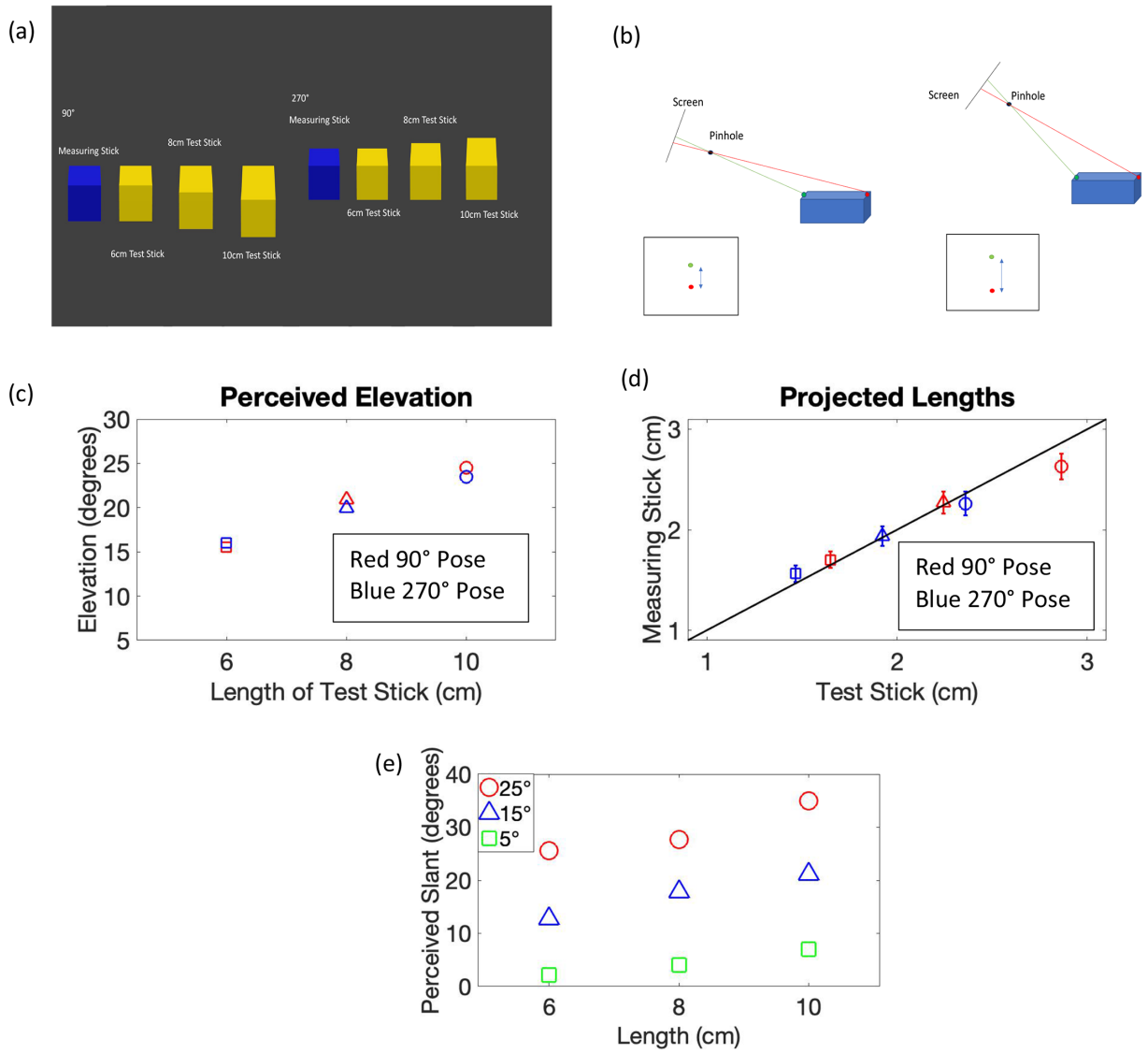

Figure 5. Slant misestimation as a function of length. Experiment 2a: (a) Blue adjustable measuring stick (6 cm) was paired with one yellow test stick (10, 8, or 6 cm) lying on dark ground in in the same pose (90° or 270°). (b) Observers were asked to match slant, but they were actually changing camera elevation on just the measuring stick (note the change in angle between camera screen and top surface of object). (c) Relative perceived camera elevation for each length of fixed stick (*circle*, 10 cm; *triangle*, 8 cm; *square*, 6 cm), separately for the two poses (*red*, 90°; *blue*, 270°). Observers' overestimation of camera elevation increases with the length of test stick, which corresponds with an increase in perceived slant. (d) At the slant match, projected lengths (vertical extents in image) for the two sticks are roughly equal (bars indicate 95% confidence intervals). (e) Experiment 2b: Directly matched perceived slant as a function of camera elevation. Consistent with experiment 2a, the perceived slant increases with the physical size of the test stick. Average perceived slants are: 2.16°, 4.02°, and 6.98° for a 5° camera elevation (the *green square*), 12.71°, 17.80°, and 21.11° for 15° camera elevation (the *blue triangle*), and 25.62°, 27.73°, and 34.95° for 25° camera elevation (the *red circle*).

absolute perceived slants with good precision. Instead we checked whether relative perceived slants across stick lengths followed the trend predicted by the best-estimated camera elevations in the model. The biggest size undercorrection was for poses along the line of sight, and the optimal back-transform predicts that overestimating slant will have little effect on poses close to frontoparallel, so we made slant measurements only for poses pointing toward or away from the observer. In Experiment 2a, we compared fixed perceived slants of 6-, 8-, and 10-cm sticks to the adjustable perceived slant of a 6-cm stick, when both sticks were either at 90° poses or at 270° (Figure 5a). In Experiment 2b, a single parallelepiped was presented on the screen at 90° or 270°, and observers adjusted a vector to match its perceived orientation.

Methods

The observers from Experiment 1 also participated in Experiment 2. In Experiment 2a, Blender was used to make two parallelepipeds, a yellow stick image was rendered from a fixed 15° camera elevation, and a blue stick with camera elevations that could be adjusted from 0° to 30° at every 1° step. The other properties of the rendering were the same as Experiment 1, as were the display monitor and observer viewpoint settings. Observers were asked to adjust the slant of the blue stick to match that of the paired yellow fixed stick by pushing buttons, without time limits. Unknown to them, observers were actually adjusting the camera elevation on the rendering of the blue stick through the 31 possible settings. Figure 5b shows that this manipulation changes the angle between camera screen and top surface of the object, so it has the same effect as changing the slant of the object. Because the aspect ratio of the front surface barely changes in these settings (Figure 4b left), the front surface provides almost no clue to the relative slant for these adjustments. The two sticks were randomly assigned to left and right on each trial. Three separate sets contained random assignments of every condition (2 poses of the pair × 3 lengths of fixed stick). Observers were allowed to take a break between sets.

In Experiment 2b, nine images were made using Blender, three lengths of the stick (6, 8, and 10 cm) rendered from three camera elevations (5°, 15°, and 25°) to vary the perceived slant for each length. Observers viewed the image of a single stick with the same geometry as Experiments 1 and 2a, that is, from a viewing angle matched to the 15° rendering. They were asked to adjust the orientation of a vector around a circle to match the slant of the stick. The vector was displayed on a 12.9-inch iPad hanging on the wall next to the observer placed vertically at the same height as the stick, orthogonal to the main display screen, so the vector orientation could be matched to the object's slant without requiring mental rotations. Three sets each contained a random arrangement of each image. Observers were allowed to take a break between sets.

Results

The main result (Figure 5c) is that observers perceived camera elevations as higher for the longer fixed sticks, despite all sticks lying flat on the same ground plane. In Experiment 2a, when both sticks were posed at 90°, the average camera elevations were and 15.50°, 20.94°, and 24.50° for the 6-cm, 8-cm, and 10-cm sticks, respectively. For the 270° poses, the averages were 16.00°, 19.94°, and 23.44°, respectively. Individual results are shown in Figure A5. These values are close to but not exactly the same as we estimated for the best fits of the model, because we measured relative slants and the model incorporates absolute perceived slants. The results are compatible with our hypothesis that observers may be applying a smaller correction factor to longer sticks because they see them as more slanted. It is interesting that, when equating slant, observers end up also equating projected lengths of the two top surfaces, that is, the excursion along the vertical axis of the image (Figure 5d), thus corroborating our conjecture that the increased vertical extent as a function of length is the cause of the slant illusion. Because the observers equated the projected lengths of the measuring stick and the test stick in Experiment 2a, to rule out that observers simply matched lengths rather than perceived slants, we did direct measurements of slant in Experiment 2b. The results are plotted in Figure 5e (individual results in Figure A6). The 5° and 25° camera elevations were used to create a variation in the perceived slants separate from variations in lengths. Estimated slants increase systematically with camera elevation of rendering, validating this method as measuring perceived slant. The results relating to the size measurements are the ones at 15° camera elevation: 12.71° for the 6-cm stick, 17.80° for the 8-cm stick, and 21.11° for the 10-cm stick. It is difficult to compare absolute numbers for the two different slant-matching techniques for a number of reasons. One is that the ground rises to meet the horizon, so it itself would be matched to different slants, depending on viewing elevation and degrees of visual angle. The perceived ground plane slant would need to be factored out in unknown ways from the absolute slant estimates in Experiment 2b, but not from the equated slants in Experiment 2a. It is worth noting that observers found it much easier to equate two slants in Experiment 2a than to match absolute slant with a vector in Experiment 2b, and this finding was reflected in greater variance for the vector settings across the three repeats for every observer. The important point is that Experiment 2b confirmed with independent measurements that perceived slant increases as a function of increased physical length.

Slant illusion demonstration

That the illusion of increased slant is perceptually compelling is further demonstrated by the movie in Figure 6. The view from the top in the movie in Figure 6 shows same rotating object as in Figure 1, but with dynamically changing lengths of the limbs, with the limb passing through poses pointing at or away from the observer made transitorily longer than the orthogonal limb. The adjustment was based on estimates from the fitted model in Figure 4c to so that the limbs seem to be approximately perceptually equal

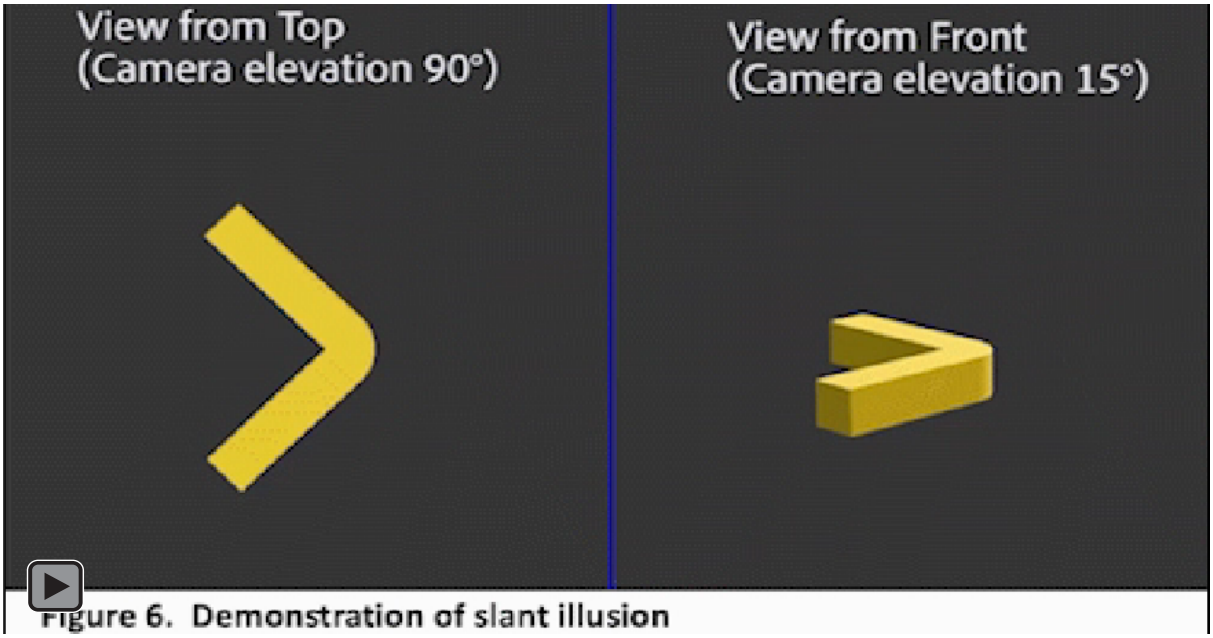

Figure 6. Demonstration of slant illusion. A rigid object with two limbs at a right angle is rotating on the ground. The length of the limb passing through the line of sight is lengthened and then shortened, as shown in the view from the top. This makes the limbs seem to be equal in all poses in the view from the 15° elevation. The percept is maintained by increasing the length of the limb passing through the line of sight according to the average size estimate in experiment 1. Instead of seeing the physical length changes, each limb seems to bounce up and down when it faces toward or away from the observer, because the slant illusion dominates the percept.

in all poses when viewed from the front at 15° elevation. As the object rotates, the limbs do not seem to change much in length, unlike in Figure 1, thus validating the adjustment. Instead, the limb passing through poses pointing at or away from the observer seems to bounce up and down because of a slant illusion similar to the static case, which seems to dominate the percept of changing length. In the dynamic case, the perceptual domination of slant changes could be related to the possibility that articulated objects are more common in the world then are objects that increase or decrease in length over a short period. The same appeal to natural statistics could be invoked to explain perceiving expansion or contraction in depth of solid objects is difficult (Jain & Zaidi, 2011; Jansson & Johansson, 1973; Johansson, 1964), while other deformations are easy to discern, even for rotating and flowing shapes (Cohen, Jain & Zaidi, 2010; Fantoni, Caudek & Domini, 2014; Bates et al, 2019). However, the slant illusion is just as compelling in the static case, so a major factor could be biases in perceiving depth versus extent from retinal images (Cohen & Zaidi, 2007; Jain & Zaidi, 2013; Kim & Burge 2018). Although not a part of this study, we want to note that perceived slant is also affected by other factors, such as object shape. For example, 6-, 8-, and 10-cm cubes are much more similar in apparent slant than the three sizes of parallelepipeds used in this study, and informal manipulations suggest that length to height ratio is also a factor.

Discussion

The first empirical contribution of this study is to measure size estimates of 3D objects as they are rotated on a ground plane into different poses, which is equivalent to changing viewpoints around the object. Invariance to rotation is one of the mathematical properties of shape, so it is surprising that no previous studies looked at the constancy or inconstancy of relative sizes and aspect ratios of shapes across poses. We found systematic and repeatable distortions of perceived relative size across poses even for simple symmetric and regular objects. Estimates of length were close to veridical for frontoparallel poses, but were seriously underestimated for poses pointing at or away from the observer. Interestingly, the underestimation increased with the physical length of the parallelepiped. Slant matching measurements revealed that longer objects were seen as slanted down, equivalent to an increase in viewing elevation. The second empirical contribution of the study is the slant illusion when slanting or lengthening an object lead to very similar

images, and the video demonstration that the visual system seems to report a dynamic illusory slant instead of a physical length change.

The main theoretical contribution of this study is to link 3D size estimates to the mental use of projective geometry. Size estimates as a function of pose form a curve that has the same shape as the optimal back-transform, and the back-transform curve fits the estimates with one free multiplicative parameter. Thus, our model that incorporates observers' misestimates of object slant in the optimal geometrical back-transform equation can explain the inconstancy of relative size for different poses, and for different object sizes. Thus, size inconstancy results despite observers using the correct geometric back-transform, if retinal images evoke misestimates of slant. Remarkably, relative size estimates as a function of object pose were similar across our observers, suggesting that the mental subconscious use of this geometrical knowledge is common among all observers.

Although we have expressed our measurements in terms of length, our measurements also reveal one type of inconstancy of shape perception across poses, or equivalently across different views of one static object. Consider the dual-limbed object formed by the parallelepiped and cylinder together in Figure 2. Because the perceived length of the parallelepiped changes in different poses relative to the perceived length of the cylinder, the object is not perceived as constant in shape. A dynamic example of perceived shape changes, because of relative limb lengths not being perceived as constant, is provided by the movie in Figure 1. In addition, our measurements predict that a solid 3D object will undergo perceived aspect ratio changes in different poses, as perceived size along the axis pointing toward the observer will be underestimated compared with the two orthogonal axes. Our results suggest that observers do try to correct for the projected shortening, using knowledge of projective geometry, but it is not enough to achieve shape constancy. The inconstancy of the estimated depth relative to width, found by previous studies on perceived object shapes and distances in perceived scenes, may also be explained by using the slant parameter in our geometrical back-transform model.

It is theoretically interesting that we show that humans do not completely discount the distortions created by perspective projection, despite possibly using the optimal geometric back transform. It is worth considering what else could be explained by assuming mental knowledge of projective geometry. Projective geometry preserves continuity, collinearity, and convergence. If a visual system assumes that, when collinear edges and intersections between edges occur in an image, it is generally because the perspective projection is preserving continuous edges and corners from the 3D world, then some objects separated into disjointed parts, for example, Ames chair (Ittelson, 1952), would be seen as unbroken and cohesive from the proper viewpoint. This explanation of cohesion is necessary before invoking stronger assumptions of simplicity or regularity in reconstructing the 3D world (Attneave & Frost, 1969; Leclerc & Fischler, 1992; Li & Pizlo, 2011; Marill, 1991). The "generic viewpoint assumption" (Freeman, 1994) thus implicitly contains a "generic projective geometry assumption," and this can be made more concrete in object and scene perception by incorporating priors that assume the invariants of projective geometry, especially when extracting 3D shapes from contours (Elder, 2018; Li, Pizlo & Steinman, 2009; Sugihara, 1986; Wang et al., 2018). The fact that slant is ambiguous in perspective projection is compatible with some real-world illusions of slant and nonrigidity (Griffiths & Zaidi, 1998, 2000). Animals and humans have constant exposure to perspective projection through image-forming eyes. Therefore, it has been an open question whether brains have learned to exploit projective geometry to understand 3D scenes. Our results imply that human brains use embedded knowledge of projective geometry to estimate 3D sizes and shapes from their retinal images. Combining the new results to our previous results that humans use optimal projective geometry back-transforms to estimate 3D poses in real 3D scenes and their pictures (Koch et al., 2018), strongly suggests that human brains have internalized particular aspects of projective geometry through evolution or learning.

Keywords: 3D size, 3D shape, 3D pose, projective geometry, inverse optics

Acknowledgments

Supported by National Institutes of Health Grants EY13312 and EY07556. We are grateful to Erin Koch for many discussions about all aspects of this study.

Author Contributions: AM and QZ designed the study. AM programmed and ran the experiments. AM and QZ analyzed and modeled the results. AM and QZ wrote the paper.

Commercial relationships: none.
Corresponding author: Qasim Zaidi.
Email: qz@sunyopt.edu.
Address: Graduate Center for Vision Research, State University of New York, 33 West 42nd St, New York, NY 10036.

References

Attneave, F., & Frost, R. (1969). The determination of perceived tridimensional orientation by minimum criteria. *Perception & Psychophysics, 6*(6), 391–396.

Baird, J. C., & Biersdorf, W. R. (1967). Quantitative functions for size and distance judgments. *Perception & Psychophysics, 2*(4), 161–166.

Bates, C. J., Yildirim, I., Tenenbaum, J. B., & Battaglia, P. (2019). Modeling human intuitions about liquid flow with particle-based simulation. *PLoS Computational Biology, 15*(7), e1007210.

Beusmans, J. M. (1998). Optic flow and the metric of the visual ground plane. *Vision Research, 38*(8), 1153–1170.

Cohen, E. H., & Zaidi, Q. (2007). Fundamental failures of shape constancy resulting from cortical anisotropy. *Journal of Neuroscience, 27*(46), 12540–12545.

Cohen, E. H., Jain, A., & Zaidi, Q. (2010). The utility of shape attributes in deciphering movements of non-rigid objects. *Journal of Vision, 10*(11), 29–29.

Cooper, J. M. (2002). *Plato: Five Dialogues: Euthyphro, Apology, Crito, Meno, Phaedo*. Cambridge, MA: Hackett Publishing.

Einstein, A (1921) Geometry and experience. *Science Studies, 3*(4), 665–675.

Elder, J. H. (2018). Shape from contour: Computation and representation. *Annual Review of Vision Science, 4*, 423–450.

Elliott, D. (1986). Continuous visual information may be important after all: A failure to replicate Thomson (1983). *Journal of Experimental Psychology. Human Perception and Performance, 12*(3), 388–391.

Elliott, D. (1987). The influence of walking speed and prior practice on locomotor distance estimation. *Journal of Motor Behavior, 19*(4), 476–485.

Fantoni, C., & Caudek, C. (2014). Domini F misperception of rigidity from actively generated optic flow. *Journal of Vision, 14*(3), 10.

Freeman, W. T. (1994). The generic viewpoint assumption in a framework for visual perception. *Nature, 368*(6471), 542.

Fukusima, S. S., Loomis, J. M., & Da Silva, J. A. (1997). Visual perception of egocentric distance as assessed by triangulation. *Journal of Experimental Psychology: Human Perception and Performance, 23*(1), 86.

Gibson, J. J. (1950). The perception of the visual world. *American Journal of Psychology, 63*(3), 367–384.

Gibson, J. J., & Cornsweet, J. (1952). The perceived slant of visual surfaces—optical and geographical. *Journal of Experimental Psychology, 44*(1), 11.

Griffiths, A. F., & Zaidi, Q. (1998). Rigid objects that appear to bend. *Perception, 27*(7), 799–802.

Griffiths, A. F., & Zaidi, Q. (2000). Perceptual assumptions and projective distortions in a three-dimensional shape illusion. *Perception, 29*(2), 171–200.

Ittelson, WH (1952) *Ames Demonstrations in Perception: A Guide to their Construction and use*. Princeton: Princeton University Press.

Jain, A., & Zaidi, Q. (2011). Discerning nonrigid 3D shapes from motion cues. *Proceedings of the National Academy of Sciences of the United State of America, 108*(4), 1663–1668.

Jain, A., & Zaidi, Q. (2013). Efficiency of extracting stereo-driven object motions. *Journal of Vision, 13*(1), 18–18.

Jansson, G., & Johansson, G. (1973). Visual perception of bending motion. *Perception, 2*(3), 321–326.

Johansson, G. (1964). Perception of motion and changing form: A study of visual perception from continuous transformations of a solid angle of light at the eye. *Scandinavian Journal of Psychology, 5*(1), 181–208.

Joynson, R. B., & Newson, L. J. (1962). The perception of shape as a function of inclination. *British Journal of Psychology, 53*(1), 1–15.

Kaiser, P. K. (1967). Perceived shape and its dependency on perceived slant. *Journal of Experimental Psychology, 75*(3), 345.

Kendall, D. G., Barden, D., Carne, T. K., & Le, H. (2009). *Shape and shape theory (Vol. 500)*. John Wiley & Sons.

Kim, S., & Burge, J. (2018). The lawful imprecision of human surface tilt estimation in natural scenes. *eLife, 7*, e31448.

Koch, E., Baig, F., & Zaidi, Q. (2018). Picture perception reveals mental geometry of 3D scene inferences. *Proceedings of the National Academy of Sciences of the United States of American, 115*(30), 7807–7812.

Leclerc, Y. G., & Fischler, M. A. (1992). An optimization-based approach to the interpretation of single line drawings as 3D wire frames. *International Journal of Computer Vision, 9*(2), 113–136.

Levin, C. A., & Haber, R. N. (1993). Visual angle as a determinant of perceived interobject distance. *Perception & Psychophysics, 54*(2), 250–259.

Li, Y., Pizlo, Z., & Steinman, R.M. (2009) A computational model that recovers the 3D shape of

an object from a single 2D retinal representation. *Vision Research, 49*, 979–991.

Li, Y., & Pizlo, Z. (2011) Depth cues vs. simplicity principle in 3D shape perception. *Topics in Cognitive Science, 3*, 667–685.

Loomis, J. M., & Philbeck, J. W. (1999). Is the anisotropy of perceived 3-D shape invariant across scale? *Perception & Psychophysics, 61*(3), 397–402.

Loomis, J. M., Da Silva, J. A., Fujita, N., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance, 18*, 906–921.

Loomis, J. M., Klatzky, R. L., Philbeck, J. W., & Golledge, R. G. (1998). Assessing auditory distance perception using perceptually directed action. *Perception & Psychophysics, 60*(6), 966–980.

Marill, T. (1991). Emulating the human interpretation of line-drawings as three-dimensional objects. *International Journal of Computer Vision, 6*(2), 147–161.

Norman, J. F., Todd, J. T., Perotti, V. J., & Tittle, J. S. (1996). The visual perception of three-dimensional length. *Journal of Experimental Psychology: Human Perception and Performance, 22*(1), 173.

Philbeck, J. W. (2000). Visually directed walking to briefly glimpsed targets is not biased toward fixation location. *Perception, 29*(3), 259–272.

Philbeck, J. W., & Loomis, J. M. (1997). Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human Perception and Performance, 23*(1), 72.

Philbeck, J. W., Loomis, J. M., & Beall, A. C. (1997). Visually perceived location is an invariant in the control of action. *Perception & Psychophysics, 59*(4), 601–612.

Poincaré, H. (2017). *Science and Hypothesis: The complete text*. Bloomsbury Publishing.

Purdy, W. C. (1960). *The Hypothesis of Psychophysical Correspondence (General Electric Tech. Rep. No. R60ELC56)*. New York: General Electric.

Ribeiro, N. P., Fukusima, S. S., & Da Silva, J. A. (1995, November). Size and distance perception in an environmental layout. *Paper presented at the Meeting of the Psychonomic Society,* Los Angeles, CA.

Rieser, J. J., Ashmead, D. H., Talor, C. R., & Youngquist, G. A. (1990). Visual perception and the guidance of locomotion without vision to previously seen targets. *Perception, 19*(5), 675–689.

Sedgwick, H. A. (1986). Space perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of Perception and Human Performance: Vol. 1. Sensory processes and perception* (pp. 21.1–21.57). New York: Wiley.

Sedgwick, H. A. (1989) The effects of viewpoint on the virtual space of pictures. Available at: https://ntrs.nasa.gov/search.jsp?R=19900013616. Accessed July 25, 2020.

Sinai, M. J., Ooi, T. L., & He, Z. J. (1998). Terrain influences the accurate judgement of distance. *Nature, 395*(6701), 497.

Steenhuis, R. E., & Goodale, M. A. (1988). The effects of time and distance on accuracy of target-directed locomotion: Does an accurate short-term memory for spatial location exist?. *Journal of Motor Behavior, 20*(4), 399–415.

Sugihara, K. (1986). *Machine Interpretation of Line Drawings* (*Vol. 1*). Cambridge, MA: MIT Press.

Thomson, J. A. (1983). Is continuous visual monitoring necessary in visually guided locomotion? *Journal of Experimental Psychology: Human Perception and Performance, 9*(3), 427.

Toye, R. C. (1986). The effect of viewing position on the perceived layout of space. *Perception & Psychophysics, 40*(2), 85–92.

Wagner, M. (1985). The metric of visual space. *Perception & Psychophysics, 38*(6), 483–495.

Wallach, H., & Moore, M. E. (1962). The role of slant in the perception of shape. *American Journal of Psychology, 75*, 289–293.

Wang, S., Wu, J., Sun, X., Yuan, W., Freeman, W. T., Tenenbaum, J. B., . . . Adelson, E. H. (2018, October). 3d shape perception from monocular vision, touch, and shape priors. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (pp. 1606–1613). New York: IEEE.

Appendix

Derivation of Equation 1 for projected length given physical length and 3D pose

For a parallelepiped of length (L_{3D}), the projected length (L_c) changes with pose Ω as a distorted sinusoid (viewing elevation = Φ_c,).

Derivation of 2D projected lengths from lengths of 3D objects at any pose

Consider a line in R^3 with physical length L_{3D}, and pose Ω, lying on the X-Z ground plane, and extending from the center of the plane (0, 0, 0) to (x, 0, z):

$$(x, 0, z) = (L_{3D}\cos(\Omega), 0, L_{3D}\sin(\Omega)) \quad \text{(A1)}$$

If the line is viewed with the camera elevation Φ_c from the Z-axis (Figure A1), for the camera screen that is equivalent to a rotation of the coordinates around the X-axis by Φ_c. The center point stays at (0, 0, 0), and the endpoint coordinates (x', y', z') are given by:

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\Phi_c) & -\sin(\Phi_c) \\ 0 & \sin(\Phi_c) & \cos(\Phi_c) \end{pmatrix} \begin{pmatrix} x \\ 0 \\ z \end{pmatrix} \quad \text{(A2)}$$

simplifying to:

$$\begin{aligned} x' &= x \\ y' &= -z\sin(\Phi_c) \\ z' &= z\cos(\Phi_c) \end{aligned} \quad \text{(A3)}$$

For the focal length of the camera $= f_c$ and the distance from the object $= d_c$, the projection of the center (0, 0, 0) in the 2D U-V picture plane is (0, 0). The projection of (x', y', z') to (u, v) is given by:

$$\begin{aligned} u &= \frac{x'}{d_c - z'} \cdot f_c \\ v &= \frac{y'}{d_c - z'} \cdot f_c \end{aligned} \quad \text{(A4)}$$

Thus, the projected length of the stick L_c, is given by:

$$L_c = \sqrt{u^2 + v^2} \quad \text{(A5)}$$

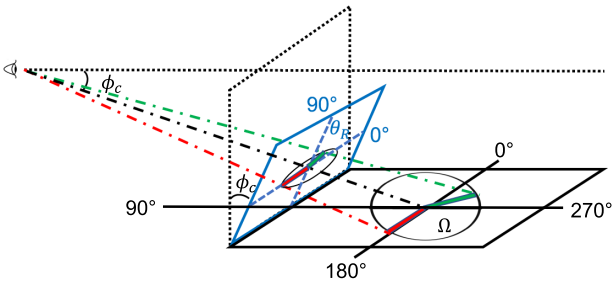

Figure A1. Schematic diagram of the projection of two sticks centered on the ground, viewed from an elevation of Φ_c.

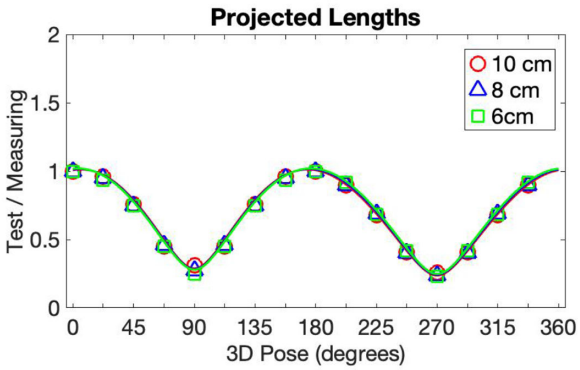

Figure A2. Calibrating rendered images for relative size. Projected length of the test stick divided by projected length of an equal sized Measuring Stick. There is no difference in the derived ratios for the three physical lengths shown as solid line curves. Points represent physical measurements of lengths of objects rendered by Blender on display screen.

Substituting Equations (A1), (A3), and (A4) in (A5):

$$\begin{aligned} L_c &= \sqrt{\left(\frac{x'}{d_c - z'} \cdot f_c\right)^2 + \left(\frac{y'}{d_c - z'} \cdot f_c\right)^2} \\ &= \sqrt{\left(x^2 + z^2 \cdot sin^2(\Phi_c)\right)\left(\frac{f_c}{d_c - z \cdot cos(\Phi_c)}\right)^2} \\ &= \frac{\sqrt{(L_{3D} \cdot \cos(\Omega))^2 + (L_{3D} \cdot \sin(\Omega))^2 \cdot sin^2(\Phi_c)}}{d_c - L_{3D} \cdot \sin(\Omega) \cdot cos(\Phi_c)} f_c \\ &= \frac{L_{3D} \cdot f_c \cdot \sqrt{cos^2(\Omega) + sin^2(\Omega) \cdot sin^2(\Phi_c)}}{d_c - L_{3D} \cdot \sin(\Omega) \cdot \cos(\Phi_c)} \end{aligned}$$

Calibrating sizes rendered by Blender

Theoretical projected lengths calculated from Equations 1 and 2 were compared with the lengths of the sticks rendered by Blender on the display screen. Because our main concern in Experiment 1 was with the relative lengths of the horizontal and vertical sticks in the 3D scenes, we calculated the ratios of the projected lengths of the parallelepiped to the orthogonally attached cylinders of the same physical length and plotted them as a function of 3D pose. Blender and geometrically derived ratios both followed a distorted and asymmetric sinusoidal curve, and had very similar values (Figure A2). Because the derived projection of the measuring stick is always the same length, this curve also describes

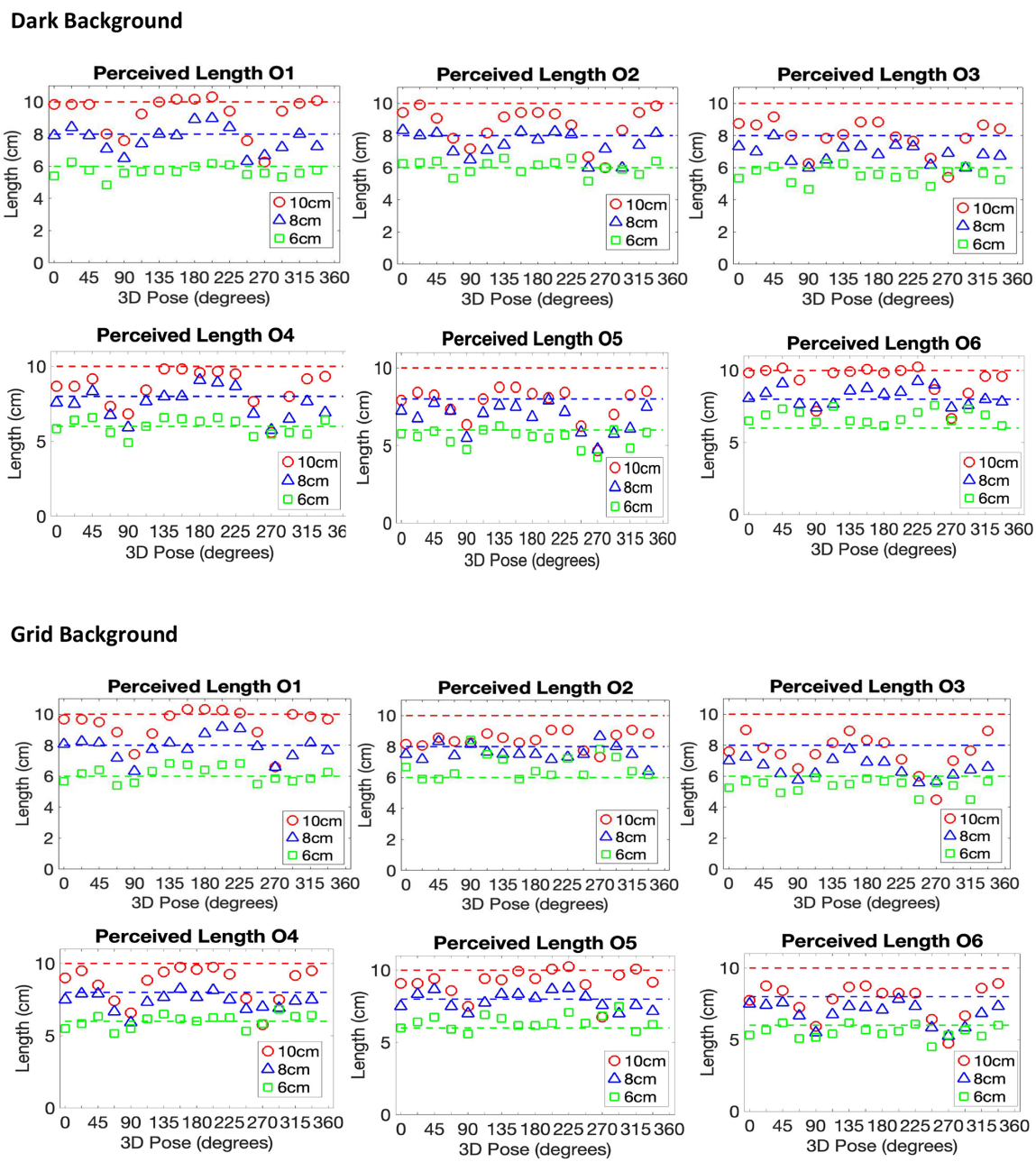

Figure A3. Perceived 3D length for each observer. Symbols represent parallelepipeds with lengths 10 cm (*red circle*), 8 cm (*blue triangle*), and 6 cm (*green square*). Average perceived length across 3D pose (six observers.) *Dashed lines* indicate physical 3D lengths. Perceived lengths are underestimated as a systematic function of pose. The systematic patterns are similar for all observers.

the projected length of the test stick as a function of pose.

Lengths estimated by individual observers

Individual observer's length estimates corresponding to Figure 3a are shown in Figure A3.

Slant illusion for conjoined objects on gridded ground

Misestimates of slant by individual observers

Individual observer relative slant estimates corresponding to Figure 5c are shown in Figure A4, and absolute slant matches corresponding to Figure 5e in Figure A5

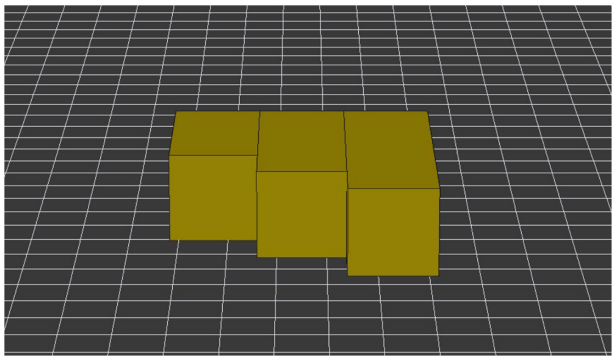

Figure A4. Slant illusion for conjoined objects on gridded ground: Front surfaces look identical, but the top surfaces look more slanted for the longer solids.

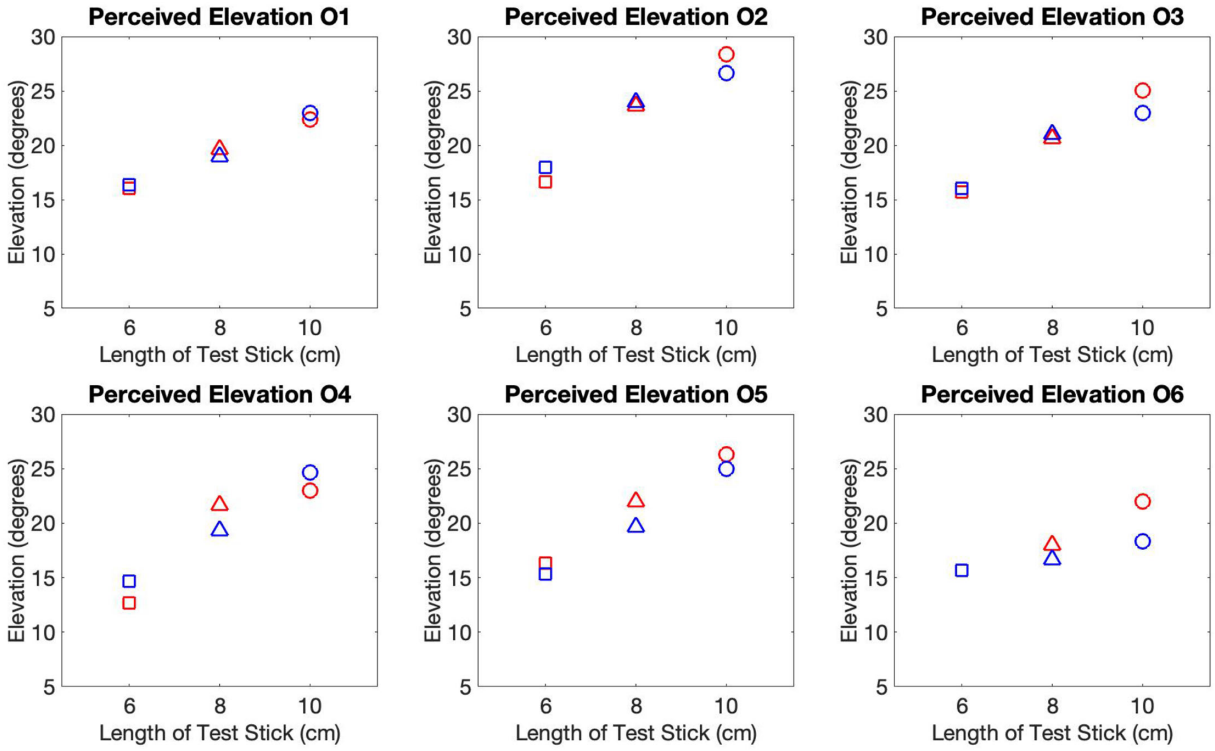

Figure A5. Equal slants across object lengths for each observer. Relative perceived camera elevation for each length of fixed stick (*circle*, 10 cm; *triangle*, 8 cm; *square*, 6 cm), separately for the two poses (*red*, 90°; *blue*, 270°) for each observer.

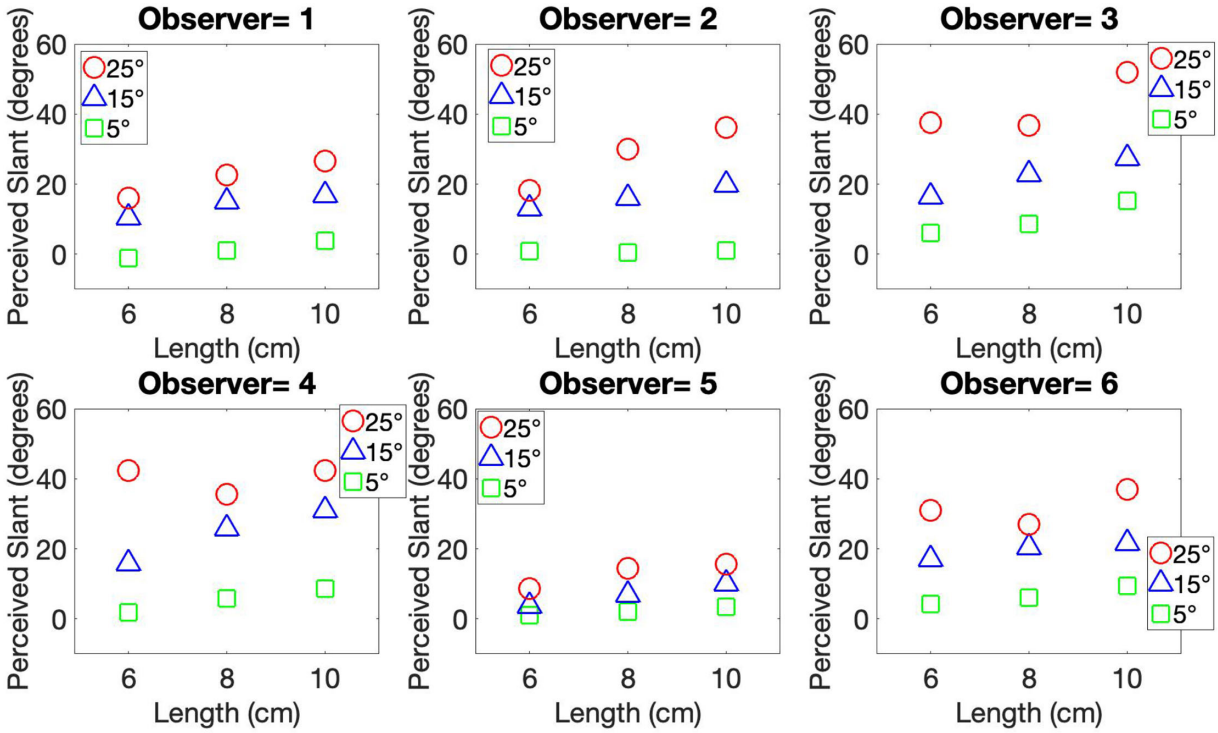

Figure A6. Matched slants across object lengths for each observer. Directly matched perceived slant as a function of camera elevation for each observer.