

Failures of stereoscopic shape constancy over changes of viewing distance and size for bilaterally symmetric polyhedra

Ying Yu

Department of Psychology, The Ohio State University,
Columbus, OH, USA



James T. Todd

Department of Psychology, The Ohio State University,
Columbus, OH, USA



Alexander A. Petrov

Department of Psychology, The Ohio State University,
Columbus, OH, USA



Two shape matching experiments examined the effects of viewing distance and object size on observers' judgments of 3D metric shape under binocular viewing. Unlike previous studies on this topic, the stimuli were specifically designed to satisfy the minimal conditions for computing veridical shape from symmetry. Concretely, the stimuli were complex, mirror-symmetric polyhedra whose symmetry planes were oriented at an angle of 45° relative to the line of sight in a shape-matching task. Although it is mathematically possible to accurately compute the 3D shapes of these stimuli using relatively simple algorithms, the results indicated that human observers are unable to do so. Indeed, the apparent shapes of the objects were systematically expanded or compressed in depth as a function of viewing distance, in exactly the same way as has been reported for simpler stimuli that do not satisfy the minimal conditions for an accurate computational analysis. For objects presented at near distances, we also obtained statistically significant effects of object size on observers' shape judgments.

Introduction

A fundamental problem for the visual perception of 3D shape is that the patterns of visual stimulation are inherently ambiguous. One possible way of solving this problem is to combine information from multiple sources. The present article will consider the perceptual analysis of two possible sources of information from binocular disparity and bilateral symmetry, and the extent to which they can mutually constrain one another to produce accurate judgments of 3D shape.

The pattern of binocular disparity is generally recognized as one of the most powerful sources of information for estimating the 3D structure of objects

in space (Howard & Rogers, 2012). However, patterns of horizontal disparity are inherently ambiguous because of the *disparity scaling problem*. Consider two identical objects, one of which is twice the distance from the observer as the other. To a first approximation, binocular disparities scale with the inverse square of the distance. Consequently, if one object is twice the distance of another, then the range of disparities produced by the far object will be approximately one-quarter of those produced by the near one. Unless disparities are somehow rescaled as a function of viewing distance, the far object would appear compressed in depth relative to the near one (Richards, 1985). The top panel of Figure 1 shows an approximation of a one-parameter family of shapes at different distances that would all be consistent with a given pattern of horizontal disparities.

There is considerable evidence that human observers are unable to fully resolve the disparity scaling problem, so that the apparent extensions of objects in depth from binocular stereopsis can expand or compress as a function of viewing distance. For example, in a classic study by Johnston (1991) the observers had to judge whether the depth axis of an elliptical cylinder was expanded or compressed relative to a circular cylinder. When the stimuli were presented at a viewing distance of approximately 1 m, the observers' adjustments were close to veridical. However, when the viewing distance was reduced to 0.5 m the objects appeared expanded in depth, and when it was increased to 2 m they appeared compressed in depth. These findings have been replicated in numerous other studies that have examined the apparent shapes of cylinders, pyramids, or dihedral angles at varying viewing distances (e.g., Glennerster, Rogers, & Bradshaw, 1996, 1998; Hecht, Doorn, & Koenderink, 1999; Johnston, Cumming, & Landy, 1994; Scarfe & Hibbard, 2006;

Citation: Yu, Y., Todd, J. T., & Petrov, A. A. (2021). Failures of stereoscopic shape constancy over changes of viewing distance and size for bilaterally symmetric polyhedra. *Journal of Vision*, 21(6):5, 1–19, <https://doi.org/10.1167/jov.21.6.5>.



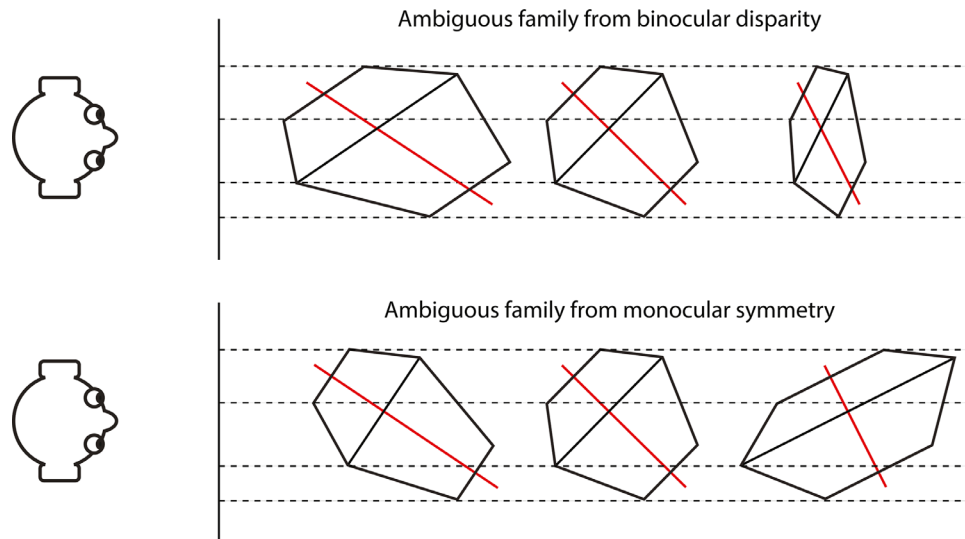


Figure 1. Schematic bird's eye view of two families of shapes that all produce the same optical projection. The shapes in the top row are all related by a stretching transformation in depth. This preserves the relative depth order of all vertices but destroys symmetry. The shapes in the bottom row are all related by a combination of shearing and stretching in depth (Li et al., 2009). This special transformation preserves the object's bilateral symmetry but alters the relative depth order of its vertices. The red line depicts the *midline* of each object—that is, the common line (or plane in 3D) through the midpoints of edges connecting corresponding points. When the midline is perpendicular to these edges, it is also the axis of bilateral symmetry. Note that the two families have only one shape in common—shown here in the middle of each row.

Todd & Norman, 2003). Similar results have also been obtained for judged intervals in depth along the ground (e.g., Baird & Biersdorf, 1967; Gilinsky, 1951; Harway, 1963; Heine, 1900; Norman, Todd, Perotti, & Tittle, 1996; Toye, 1986; Wagner, 1985) and for action-based paradigms that measure the width of the hand grip while reaching to grasp objects at different distances (Campagnoli, Croom, & Domini, 2017; Campagnoli & Domini, 2019).

Some researchers have argued that the accuracy of 3D shape judgments can be improved if binocular disparities are combined with other sources of information, such as texture or motion. For example, Richards (1985) noted that unscaled binocular disparities allow a one-parameter family of possible shape interpretations that are all related by expansions or compressions in depth. Monocularly viewed two-frame apparent motion sequences allow a different family of possible interpretations that are related by a shearing transformation in depth (Huang & Lee, 1989; Koenderink & Doorn, 1991). Thus, if both sources of information are combined, it is possible in principle to obtain a veridical estimate of 3D shape by computing the intersection of those two families. There is conflicting empirical evidence about whether human observers are able to exploit the combination of stereo and motion to obtain accurate judgments of 3D shape. One study by Johnston et al. (1994) found that they can, whereas another by Todd and Norman (2003) found that they cannot.

Another possible source of information that could be combined with stereo to obtain veridical shape estimates is provided by bilateral symmetry. Symmetry imposes a powerful constraint because one half of the object repeats the structure of the opposite half, but with complementary polarity (Franc,ois, Medioni, & Waupotitsch, 2002; Gordon, 1989; Vetter & Poggio, 1994). Thus, a single image of a symmetric object provides two distinct views of the same underlying 3D structure, except in degenerate cases (Hong, Yang, Huang, & Ma, 2004). Algorithms for multiple-view geometry (Hartley & Zisserman, 2003; Ma, Soatto, Kosecka, & Sastry, 2004) can be modified to recover 3D structure from a single image (Hong et al., 2004; Ma et al., 2004). This is an active area of research in computer vision that is sometimes referred to as *shape from symmetry* or *structure from symmetry* (e.g., Franc,ois et al., 2002; Gordon, 1989; Michaux, Kumar, Jayadevan, Delp, & Pizlo, 2017; Park et al., 2008; Sawada, Li., & Pizlo, 2011; Thrun & Wegbreit, 2005).

To take a concrete example, Li, Pizlo, and Steinman (2009) have shown that the 3D metric structure of a mirror-symmetric polyhedron can be recovered from a single orthographic image up to a one-parameter family of symmetric interpretations, which is shown in the bottom panel of Figure 1. The underconstrained parameter that generates the ambiguity is the slant of the object's symmetry plane. Note the close similarity with the ambiguity in the analogous structure-from-motion case, where the underconstrained parameter is

the angle of object rotation in depth (Huang & Lee, 1989; Koenderink & Doorn, 1991). It is important to note that there are some minimal conditions that must be satisfied for the application of the shape-from-symmetry computations: the projected image of the object must contain at least four pairs of bilaterally symmetric points (Ma et al., 2004, p. 122), and the object must be oriented so that its plane of symmetry is neither parallel nor perpendicular to the observer's line of sight.

In sum, when a bilaterally symmetric object of sufficient complexity is viewed binocularly from a nondegenerate viewpoint, the retinal inputs contain two independent and complementary sources of information. Each of these sources by itself is sufficient to recover the 3D shape of the object up to a one-parameter ambiguity. Each ambiguity is generated by a particular parameter—viewing distance for the stereoscopic source and the slant of the object's symmetry plane for the symmetry-based one. Critically, these two parameters are distinct and independent, and consequently each source can potentially be used to disambiguate the other. It can be proven mathematically that when the stimulus is in fact symmetric there is a unique interpretation that belongs to both families as illustrated in Figure 1. Furthermore, provided the aforementioned minimal conditions are satisfied, this intersection-of-constraints algorithm recovers the 3D metric structure veridically.

These mathematical results provide the theoretical foundation for computer vision algorithms that can process real images on relatively modest hardware (e.g., Park et al., 2008; Shimshoni, Moses, & Lindenbaum, 2000; Sinha, Ramnath, & Szeliski, 2012; Yang, Huang, Rao, Hong, & Ma, 2005). The models of Zygmunt Pizlo and his collaborators incorporate many of these ideas (Jayadevan, Michaux, Delp, & Pizlo, 2017; Li et al., 2009; Li, Sawada, Latecki, Steinman, & Pizlo, 2012; Pizlo, Li, Sawada, & Steinman, 2014). Two recent models in particular (Michaux, Jayadevan, Delp, & Pizlo, 2016; Michaux et al., 2017) explicitly combine monocular and binocular sources of information (see also Zabrodsky & Weinshall, 1997, for an early effort). Although these models have certain limitations and invoke additional assumptions, they serve as proof-of-concept demonstrations that the ideas illustrated in Figure 1 can be implemented in practical vision systems.

Of course, it is an altogether different question whether the human visual system can extract this information from the visual inputs and combine the two sources appropriately. The importance of stereo vision for the perception of depth and 3D structure is well documented (e.g., Howard & Rogers, 2012; Todd, 2004). Much less is known about the role of 3D

symmetry in shape perception (Treder, 2010, p. 1526). Although there are plenty of studies on symmetry perception (see Treder, 2010; Tyler, 2002, for reviews), most of them deal with 2D symmetry on the image rather than in 3D. There is also evidence that the visual system is good at symmetry *detection* in both 2D and 3D, and that it uses it to support object recognition and classification (e.g., Sawada, 2010; Vetter, Poggio, & Bülthoff, 1994). Indeed, symmetry is on a short list of *nonaccidental properties* (Biederman, 1987). However, very few studies have explicitly investigated the effects of 3D symmetry on the perception of 3D shape and/or depth (Jayadevan, Sawada, Delp, & Pizlo, 2018; Lee & Saunders, 2013; Saunders & Knill, 2001; Sawada, 2010). In our estimation, the latter studies are suggestive but inconclusive. This topic is revisited in the General discussion below.

Can human observers exploit symmetry to achieve stereoscopic shape constancy? Unfortunately, there is surprisingly little evidence to address this issue. Although symmetric stimuli have been used in most of the previous experiments that have documented the compression of apparent depth with viewing distance, none of those stimuli satisfied the minimal conditions of the structure-from-symmetry computations. They generally did not contain a sufficient number of symmetrical point pairs to perform the required computations, and/or they were most often presented at degenerate orientations for which the symmetry plane was parallel (or nearly parallel) to the line of sight. Other experiments that have satisfied those conditions (Li, Sawada, Shi, Kwon, & Pizlo, 2011; Jayadevan et al., 2018) have not compared the apparent shapes of objects presented at different viewing distances, and might be subject to other methodological concerns (see the General discussion below). Thus, the research described in the present article was designed to fill this void in the literature. Stereoscopic images of complex symmetric polyhedra were presented at nondegenerate orientations relative to the line of sight, and observers were asked to judge their apparent shapes at varying viewing distances.

Experiment 1

Methods

Observers

Twelve observers participated in the experiment, including the three authors (YY, JT, and AP) and nine others who were naïve about the purpose of the experiment. All participants reported normal or corrected-to-normal visual acuity. All gave informed consent as approved by the Institutional Review Board at the Ohio State University.

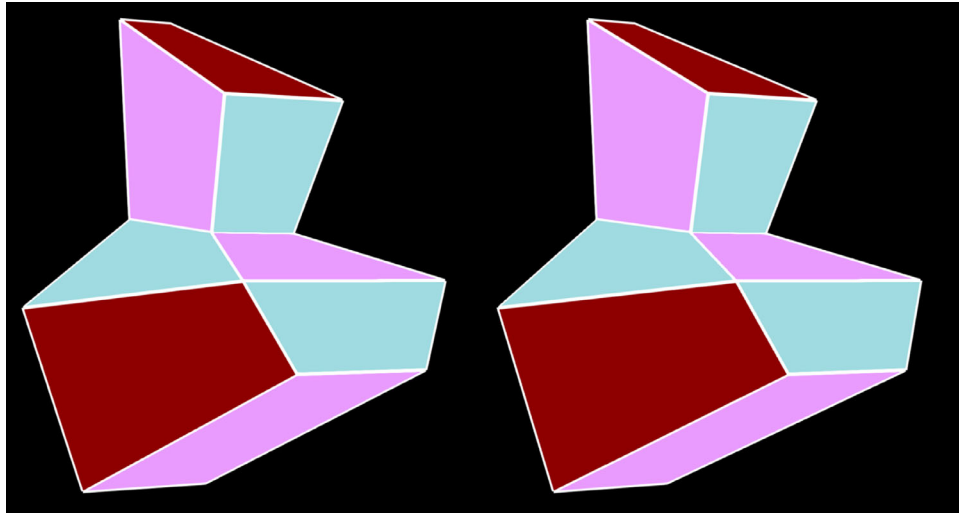


Figure 2. A stereogram of one of the 10 polyhedral objects used in the present experiments.

Stimulus displays

The stimuli consisted of 10 mirror-symmetric 3D polyhedra, which were similar to the stimuli of Li et al. (2011). Each polyhedron was composed of quadrilateral faces with one mirror symmetry plane (see Figure 2; see also Figure 11 below for another example). The objects were initially defined by a set of 16 vertices and the connections between them. These were rendered as white lines on a black background. The faces of each polyhedron were painted in three different colors such that no two adjacent faces were the same. The occluded back surface of each object consisted of three quadrilateral faces that lay in a common plane orthogonal to the symmetry plane.

Apparatus

The 3D stimuli were generated in Matlab in real time and rendered with PsychOpenGL, a set of essential functions that interfaces Psychtoolbox (Kleiner, Brainard, & Pelli, 2007; Pelli, 1997) with OpenGL. For any given stimulus, two slightly different stereoscopic perspective images were computed for observers' left and right eyes using a technique called horizontal image translation (Lipton, 1991) that horizontally shifts the viewpoint of each eye by an amount determined by the interocular distance measured for each observer. This produces an optically correct pattern of horizontal and vertical disparities. The observer viewed the stereoscopic images binocularly through LCD shutter glasses (NVIDIA 3D Vision 2) that were synchronized with the refresh rate of a mosaic display so that each eye received the appropriate image.

The mosaic display was composed of two identical LCD monitors (Dell S2716DG) placed side by side. They were synchronized into a unified and seamless

display by NVIDIA Mosaic technology and bezel correction. The refresh rate of the mosaic display was 120 Hz. Thus, the image for each eye was updated at the rate of 60 Hz, which was fast enough to avoid flicker. The mosaic display had a horizontal and vertical extent of 120×34 cm, and its spatial resolution was 5160×1440 pixels. The observers viewed the display in a darkened room at a distance of 150 cm while using a chinrest to restrict head movements.

Procedure

The basic scene geometry of the experiment is shown in Figure 3. Two polyhedra were presented side by side against a black background on the mosaic display during each trial presentation. The horizontal distance between the rightmost vertex of the left object and the leftmost vertex of the right object was approximately 9 cm and the central line between them was aligned with the chinrest. The objects were shown at eye level. The one on the right had a fixed 3D shape, and we will refer to it as the *reference* object. The one on the left could be compressed or stretched in depth by the observer, and it is referred to as the *adjustable* object. Both objects were presented in the same 3D orientation, such that their symmetry planes were at a 45° angle relative to the line of sight. This slant is the one that is most favorable for structure-from-symmetry computations.

The observers' task was to adjust the shape of the adjustable object by stretching or compressing it in depth using a handheld mouse so that it matched the apparent shape of the reference object. The reference object was always symmetric, whereas the adjustable object was generally asymmetric, except for one possible setting where it matched the shape of the reference object. The adjustment space is analogous to the depth-scaling family in the top row of Figure 1,

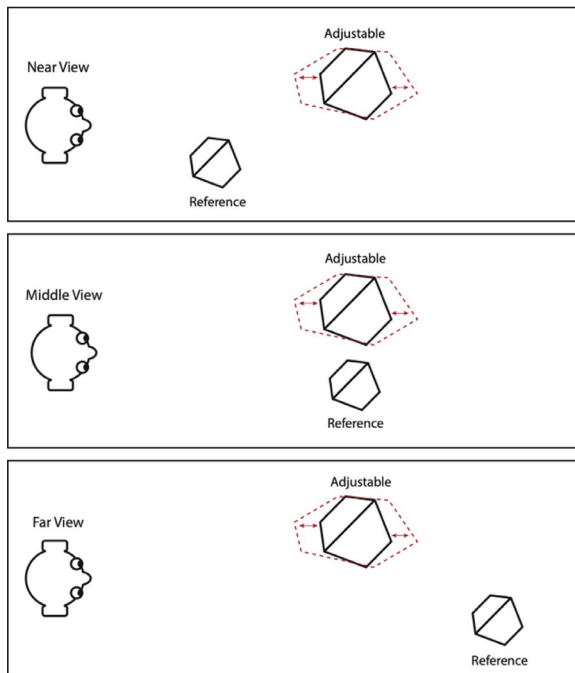


Figure 3. A bird's eye view of the viewing geometries used in the present experiments. An adjustable object was always presented in the left hemifield at a distance of 1.5 m. The reference object was always presented in the right hemifield, and its simulated viewing distance was manipulated across trials with possible values of 0.7 m, 1.5 m, and 2.3 m. Observers were required to stretch or compress the adjustable object in depth so that its apparent shape matched that of the reference object.

except that the experimental stimuli were rendered in perspective projection.

The simulated viewing distance to the adjustable object was always 150 cm, which was the same as the physical distance between the observer and the mosaic display. The simulated viewing distance of the reference object was manipulated across trials with three possible values of 70 cm, 150 cm, and 230 cm (see Figure 3). The two objects presented in each trial were rendered in different colors and sizes so that their 2D images were not identical. The size of the adjustable object was the same on every trial. The average horizontal and vertical extents of the 10 possible objects were 10.9 cm and 16.8 cm, respectively. Its extension in depth could vary based on the observers' settings. The selection of objects was constrained so that no faces appeared or disappeared from view due to the adjustment.

The size of the reference object was manipulated in two ways in separate sessions. In one session, its physical size was set at 70% of the adjustable object and was fixed across viewing distances for a given polyhedron. As a result, the size of its 2D projected image would change with the simulated viewing distance (see Table 1). In the other session, we fixed the projected size of the

reference object by changing its physical size according to the viewing distance (see Table 1 for details). On average, the reference object used in this fixed projected size session subtended 4.4° of visual angle at each of the three viewing distances.

The experiment was performed in a dark and quiet room where the display was the only source of illumination. Prior to their participation, observers were given a stereo acuity test developed by Brown et al. (2007) to confirm that they had normal stereoscopic vision. They were then asked to perform several practice trials to get familiar with the equipment and the task. During these practice sessions all the observers indicated that they could clearly perceive the compressions and expansions in depth of the adjustment object. The practice trials used a different object than the ones used in the experimental trials. At the start of each trial, the depth-to-width ratio of the adjustable object was set randomly. Observers then moved the mouse horizontally to make adjustments without time limitation.

Each observer conducted two separate sessions: one for the fixed physical size condition and one for the fixed projected size condition. The order of the two sessions was counterbalanced across observers. Within each session, three possible simulated viewing distances were presented three times each for each of the 10 polyhedral objects used in this experiment. Therefore, one session had 90 trials and was run in two blocks of 45 trials with randomized order. On average, one session took about 40 minutes.

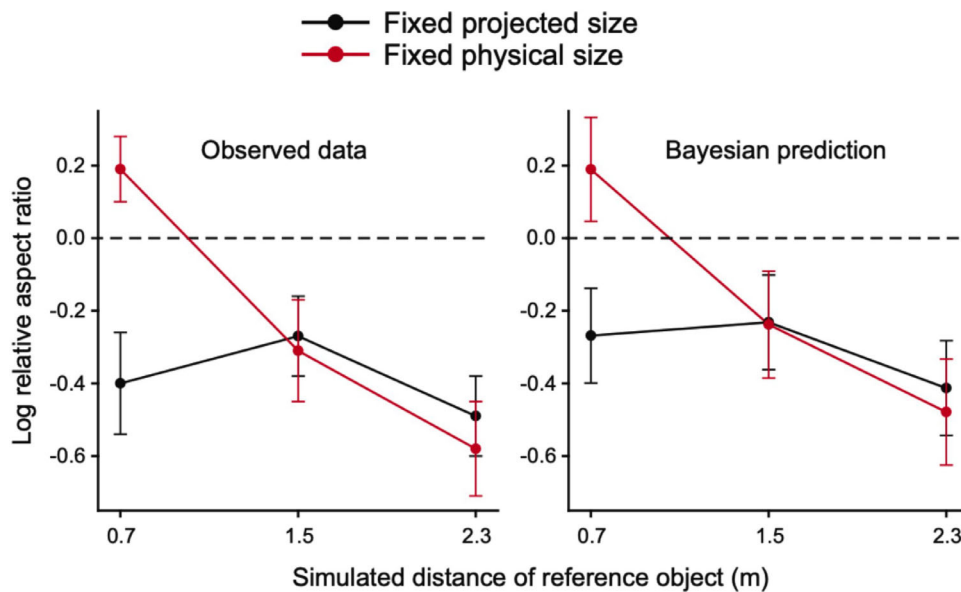
Results

During their debriefing sessions, all the observers reported that the displays produced perceptually vivid impressions of 3D structure, and that manipulations of the mouse produced clear changes in the apparent z-scaling of the adjustable object. Because the adjustable and reference objects on any given trial could have different sizes, it would not be meaningful to directly compare their relative extensions in depth. Thus, to normalize the size differences, we instead compared the depth-to-width ratio of the adjustable object relative to the depth-to-width ratio of the reference object. The dependent variable throughout our study is the *relative aspect ratio*:

$$S = \frac{z_{adj}/x_{adj}}{z_{ref}/x_{ref}} \quad (1)$$

where z_{adj} and x_{adj} represent the adjustable object's extents along the z-axis and x-axis, respectively, and z_{ref} and x_{ref} represent the extents of the reference object (see Figure 3). Note that z_{adj} is the only variable that

	Viewing distance	Physical size	Projected size
Fixed physical size	0.7 m	11.6 cm	9.5°
	1.5 m	11.6 cm	4.4°
	2.3 m	11.6 cm	3.0°
Fixed projected size	0.7 m	5.5 cm	4.4°
	1.5 m	11.6 cm	4.4°
	2.3 m	17.8 cm	4.4°
Adjustable object	1.5 m	16.6 cm	6.3°

Table 1. The distances and sizes of stimuli used in [Experiment 1](#).Figure 4. The left panel shows the average judgments of all 12 observers in [Experiment 1](#). Error bars denote ± 1 standard error of the mean (*SEM*) of the corresponding data set. The right panel shows posterior predictive fits of the hierarchical Bayesian model (see [Appendix](#)). The horizontal dashed lines represent veridical performance.

was controlled by the observers. Because this particular measure produces an asymmetry between under- and overestimates of an object's extension in depth, we transformed the scale by using the binary logarithm of S . We will refer to this measure as the *log relative aspect ratio*. $\log_2 S = 0$ indicates a perfect shape match (up to a similarity transformation). $\log_2 S > 0$ indicates the adjustable object was expanded in depth, and $\log_2 S < 0$ indicates that the adjustable object was compressed in depth relative to the reference object.

The left panel of [Figure 4](#) shows the average responses over all observers plotted as a function of the simulated viewing distance for the two size conditions. These data were analyzed in several different ways. First, we performed an analysis of variance (ANOVA) on the group data. The results revealed a significant effect of viewing distance ($F(2, 22) = 16.26, p < 0.001$), a significant effect of size ($F(1, 11) = 6.66, p < 0.05$), and a significant interaction between size

and distance ($F(2, 22) = 10.86, p < 0.001$). We also performed ANOVAs on the individual observers, whose judgments are shown in [Figure 5](#). Ten of the 12 observers showed significant effects of viewing distance, and seven of them showed significant effects of size and/or a significant size by distance interaction. It is also interesting to note in [Figure 5](#) that there were large individual differences in the magnitude and direction of constant errors, which is consistent with prior studies.

Three additional hierarchical Bayesian models of these data were implemented to quantify the overall effect sizes. The exact specification of the Bayesian models is given in the [Appendix](#). The right panel of [Figure 4](#) shows the Bayesian predictions of the observed data. [Figure 6](#) shows the posterior estimates of the effect size of the group-averaged deflection from the group mean as a function of the simulated viewing distance for the two size conditions. The shaded violins

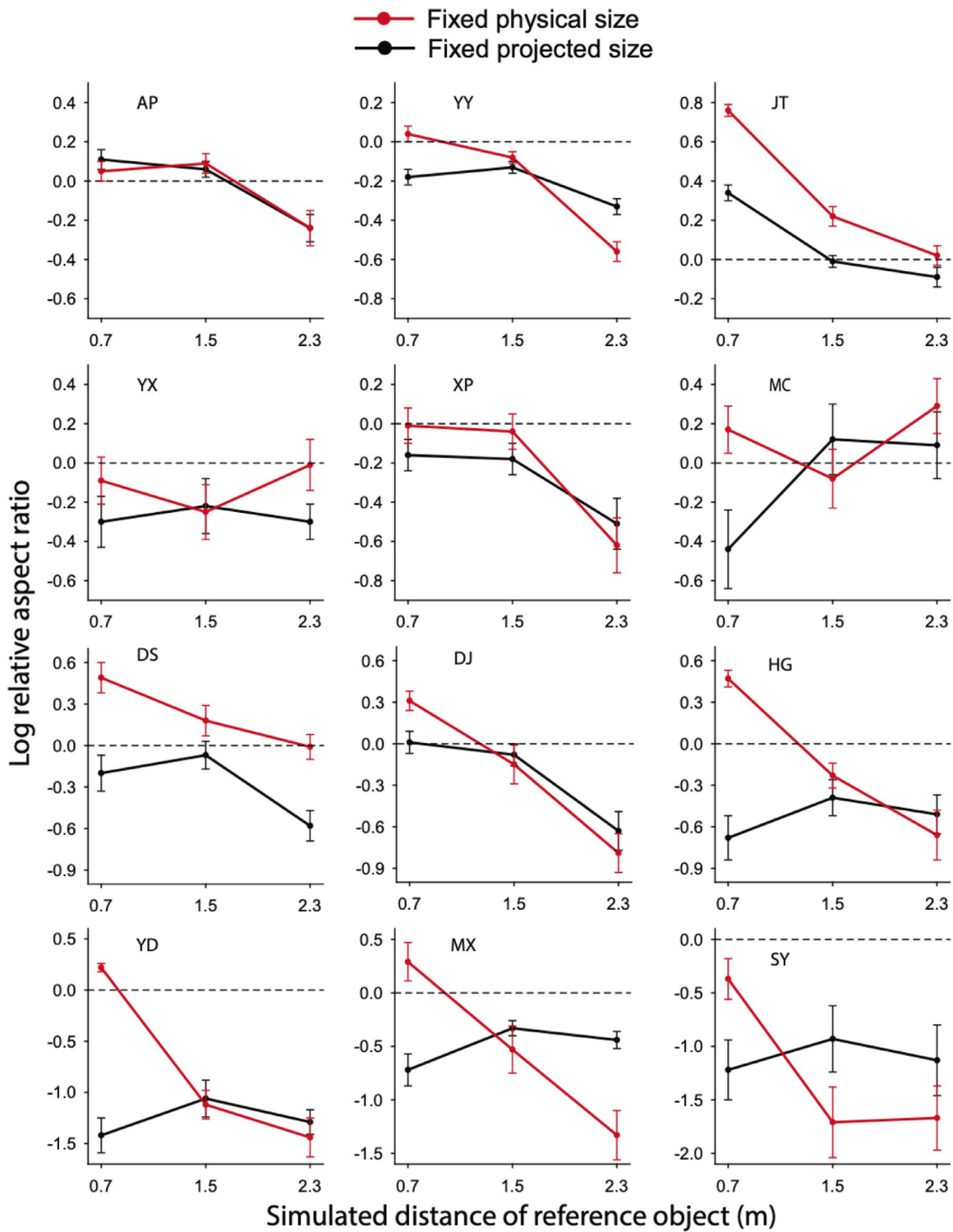


Figure 5. The average responses of each individual observer in Experiment 1. Error bars denote ± 1 standard error of the mean. Note that these graphs are plotted on different scales to accommodate the variation of scaling differences exhibited by different observers. The horizontal dashed line on each graph represents veridical performance.

denote the 95% highest density interval (HDI) of the posterior probability distribution. The point and error bars in each violin denote the mean and standard deviation of the distribution. Figure 7 shows a similar

plot of the posterior distributions for the two possible sizes at the closest viewing distance. Note that the HDIs do not overlap at all for the near and far distances, nor do they overlap for the small and large sizes at the

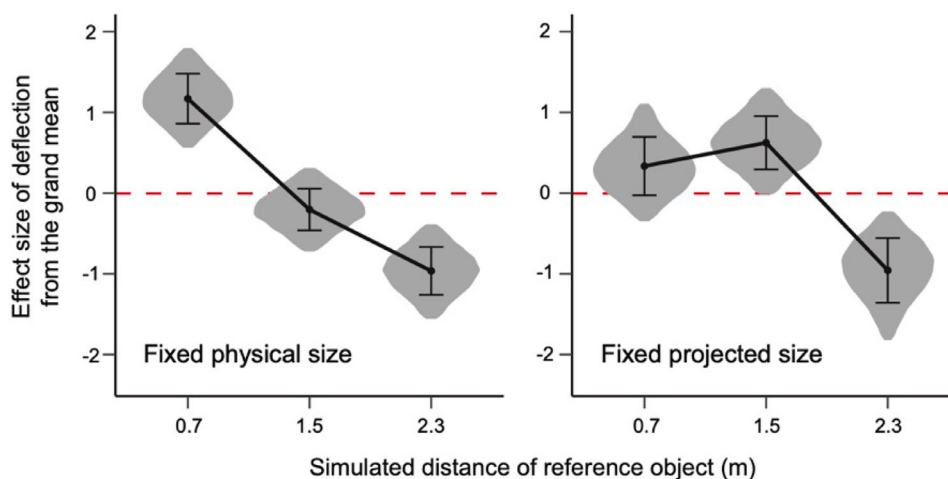


Figure 6. Posterior estimates of the effect size of the group-averaged deflection from the grand mean in [Experiment 1](#) as a function of viewing distance for the fixed physical size (Bayesian “Model-1” specified in the [Appendix](#)) and fixed projected size conditions (“Model-2”). The shaded violins denote the 95% highest density interval (HDI) of the posterior probability distribution. The point and error bars in each violin denote the mean and standard deviation of the distribution. The horizontal dashed line denotes zero deflection from the grand mean.

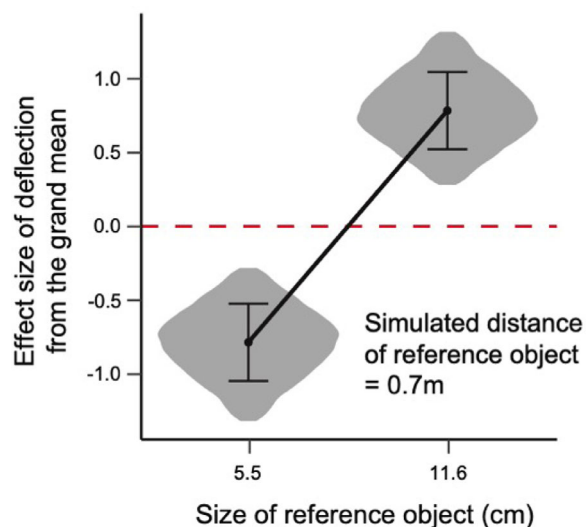


Figure 7. Posterior estimates of the effect size of the group-averaged deflection from the grand mean for the two possible physical sizes at the near distance of [Experiment 1](#). The shaded violins denote the 95% highest density interval (HDI) of the posterior probability distribution (Bayesian “Model-3” specified in the [Appendix](#)). The point and error bars in each violin denote the mean and standard deviation of the distribution. The horizontal dashed line denotes zero deflection from the grand mean.

near viewing distance. This indicates the effects of both viewing distance and object size are statistically reliable, reinforcing the ANOVA-based results above.

Discussion

There are several possible strategies by which it would be possible, in principle, to achieve accurate performance on this task. For example, if observers could accurately distinguish between symmetrical and asymmetrical objects, then they could perform the task by choosing the one setting in the adjustment space where the depicted object was perfectly symmetrical, without even bothering to compare it to the reference object (cf. top row of [Figure 1](#)). We would expect in that case that there would be no significant effects of viewing distance because the distance to the adjustment object never changed ([Figure 3](#)).

The observers could also try to scale the horizontal disparities with viewing distance, which could conceivably be measured using accommodation or convergence. Some researchers have argued that perceptual distortions can occur when using computer displays because the absence of accommodative blur provides conflicting information that the depicted objects are flat (see [Watt et al., 2005](#)). However, we would expect in that case that the conflict would be greatest at close viewing distances because that is where gradients of accommodative blur are most visible. This suggests that objects at near distances should appear compressed relative to those at far distances, which is the opposite of what is typically found in this type of experiment. Moreover, similar results have also been obtained for judgments of real objects, where there is no conflict between disparity and accommodation (e.g., see [Todd & Norman, 2003](#)).

Another possible way of scaling horizontal disparities with distance might be to exploit the vertical disparities between the projections on the two eyes. Previous mathematical analyses have shown that vertical disparities provide potential information for determining an object's distance from the observer (Koenderink & van Doorn, 1976; Petrov, 1980), and there is empirical evidence to show that human observers are able to make use of that information in at least some contexts (Rogers & Bradshaw, 1993). However, an important limitation of vertical disparities is that they become vanishingly small at small visual angles, so that their effectiveness as a source of information may be restricted to objects with relatively large angular extents. This suggests that observers' judgments of 3D shape from stereo should be most accurate for objects that are relatively large and/or relatively close to the point of observation, which is quite consistent with the pattern of results obtained in this experiment.

Given that the judgments of most of the observers exhibited large deviations from the ground truth, it is reasonable to conclude that they could not accurately implement any of the three strategies described above. Without some way of scaling disparities or exploiting monocular symmetry, observers would be forced to adopt some arbitrary standard to determine a specific depth interval that corresponds to a particular difference in disparity. Such arbitrary scaling factors are what Koenderink et al. (2001) referred to as *the observer's share*, which may become necessary for tasks that do not provide sufficient information for accurate performance. One hallmark of such tasks is that they often produce large individual differences because different observers adopt different strategies. Note that this is compatible with consistent performance across trials for each individual participant because they can apply their chosen observer's share consistently throughout the experimental session.

It is clear from Figure 5 that indeed there were considerable individual differences for this task, and these cannot be explained by the relatively minor variations among observers on the stereo acuity tests. This is especially clear in the far viewing condition, where the apparent compression of objects in depth varied among observers over a range of 0–65%. This finding is quite consistent with other studies of stereo depth scaling that have looked at individual differences. For example, Todd and Norman (2003) did a similar study with 10 observers on the apparent shapes of real cardboard pyramids presented at different distances. Nine of the 10 observers judged the far objects to be significantly more compressed in depth than the near ones. As in the present experiment, most of the variation among observers occurred in the far viewing condition, where the magnitudes of their depth scaling varied over a range of 50%.

It is important to keep in mind that the apparent compression of objects in depth with increasing viewing distance is consistent with a large number of previous studies (e.g., Baird & Biersdorf, 1967; Campagnoli et al., 2017; Campagnoli & Domini, 2019; Gilinsky, 1951; Glennerster et al., 1996, Glennerster et al., 1998; Harway, 1963; Hecht et al., 1999; Heine, 1900; Johnston, 1991; Johnston et al., 1994; Norman et al., 1996; Scarfe & Hibbard, 2006; Todd & Norman, 2003; Toye, 1986; Wagner, 1985). The primary interest of this particular replication is that the stimuli depicted moderately complex symmetric polyhedra viewed from “nondegenerate” slants. Pizlo and colleagues (Li et al., 2011; Jayadevan et al., 2018) have argued that this is a necessary condition for the veridical perception of 3D shape from stereo. Thus, the fact that the observers' perceptions of shape did not remain constant over variations in size and distance, and that they exhibited large constant errors, provides strong evidence against that hypothesis.

The effects of object size on stereoscopic shape perception have been investigated in several previous studies, but the results have been surprisingly inconsistent. For example, Bradshaw, Glennester, and Rogers (1996) obtained results similar to ours, in that the apparent depth-to-width ratios of the depicted objects increased with object size. However, Collett, Schwarz, and Sobel (1991) and Champion, Simmons, and Mamassian (2004) obtained the opposite effect, and Norman et al. (2009) obtained a negligible effect of size (see also Johnston, 1991). The discrepancies in these results are most likely due to methodological differences in the designs of the experiments. For example, the displays used by Norman et al. included gradients of shading and texture that provided salient information about 3D shape in addition to the information from binocular disparities. For Bradshaw et al. (1996) study that showed positive size effects, the outer edges of the depicted objects could be 15° or more in the periphery relative to the median plane, whereas the ones that showed negative effects used stimuli confined to a much smaller central region.

When an object's size is increased at a relatively close viewing distance, it produces perspective distortions in its optical projection. In particular, the optical projection of the front part of the object undergoes a greater expansion than the back part. Similarly, vertical disparities arise when the distance of a point to one eye is greater than its distance to the other (Rogers & Bradshaw, 1993), and this differential perspective increases systematically for points that are located farther in the periphery. This differential expansion as a function of depth is attenuated as viewing distance is increased, which could explain the interaction of size and distance. Note in Figure 4 that the difference in apparent z-scaling between the near and far viewing distances was four times greater in the fixed physical

size condition than in the fixed projected size condition. This suggests that perspective distortions or vertical disparities could have a significant influence on apparent depth-to-width ratios at relatively close viewing distances. Because the statistical significance of the size effect was relatively small due to individual differences, we decided to perform an additional replication that was focused entirely on that issue.

Experiment 2

Methods

The materials and design were the same as [Experiment 1](#), except for the manipulations of the reference object. In this experiment, the simulated viewing distance of the reference object was fixed at 0.7 m, but its size was systematically manipulated. For the largest size, the average width of the different objects was 11.6 cm (9.47°), which was identical to the fixed physical size at the closest viewing distance in [Experiment 1](#). We also included two smaller sizes with average widths of 7.8 cm (6.35°) and 5.2 cm (4.25°), respectively. As in [Experiment 1](#), the adjustable object was presented at a simulated viewing distance of 1.5 m and its physical size was always 16.6 cm (6.3°). The actual viewing distance to both objects was 1.5 m. The displays were judged by 11 of the observers who participated in [Experiment 1](#) and one additional naïve observer. Each observer participated in a single experimental session.

Results

[Figure 8](#) shows the average responses over all observers plotted as a function of object size, together with the Bayesian fits of these data. An ANOVA of the observers' judgments revealed that the apparent depth-to-width ratio of the reference object increased significantly with the size of the object ($F(2, 22) = 8.97, p < 0.001$). [Figure 9](#) shows the individual results of all 12 observers. ANOVAs for these data revealed significant effects of size for nine of the 12 observers. [Figure 10](#) shows the posterior estimates of effect size of the group-averaged deflection from the grand mean as a function of object size (see [Appendix](#) for details of the implementation of the Bayesian analyses). The key thing to note in that figure is the lack of overlap in the posterior highest density intervals for the smallest and largest objects. It thus provides evidence for a statistically reliable effect of object size on judgments of an object's 3D shape, reinforcing the ANOVA-based results above.

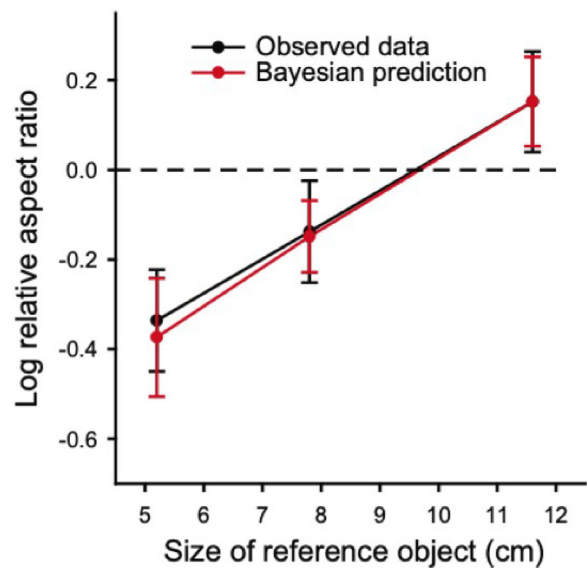


Figure 8. The average judgments of all 12 observers in [Experiment 2](#), together with posterior predictive fits of the hierarchical Bayesian model (see [Appendix](#)). The horizontal dashed lines represent veridical performance.

The most likely cause of the size effect is the nonlinear distortion of image structure that occurs due to strong perspective. As an object increases in size (or decreases in viewing distance) the front part of its optical projection expands at a greater rate than the back part, and this increases the range of both horizontal and vertical disparities in stereoscopic vision. To better appreciate how this might influence observers' perceptions it is useful to consider the three stereograms in [Figure 11](#). The one in the middle depicts a large bilaterally symmetric polyhedron. The ones above and below it depict the same object with a smaller size. One of these smaller objects has exactly the same shape as the larger one. The other has been expanded or compressed in depth by 40%. The reader is invited to judge which object has been distorted. The first thing to note when examining these stereograms is that this is not a trivial task. It is relatively easy to see that the upper object has a greater extension in depth than the lower one, but which one matches the shape of the larger object in the center? The observers in our experiment overwhelmingly chose the upper object as the best match, but the correct answer is actually the lower one.

General discussion

The apparent compression of depth intervals with increasing viewing distance is a well-documented phenomenon that was first reported by [Heine \(1900\)](#)

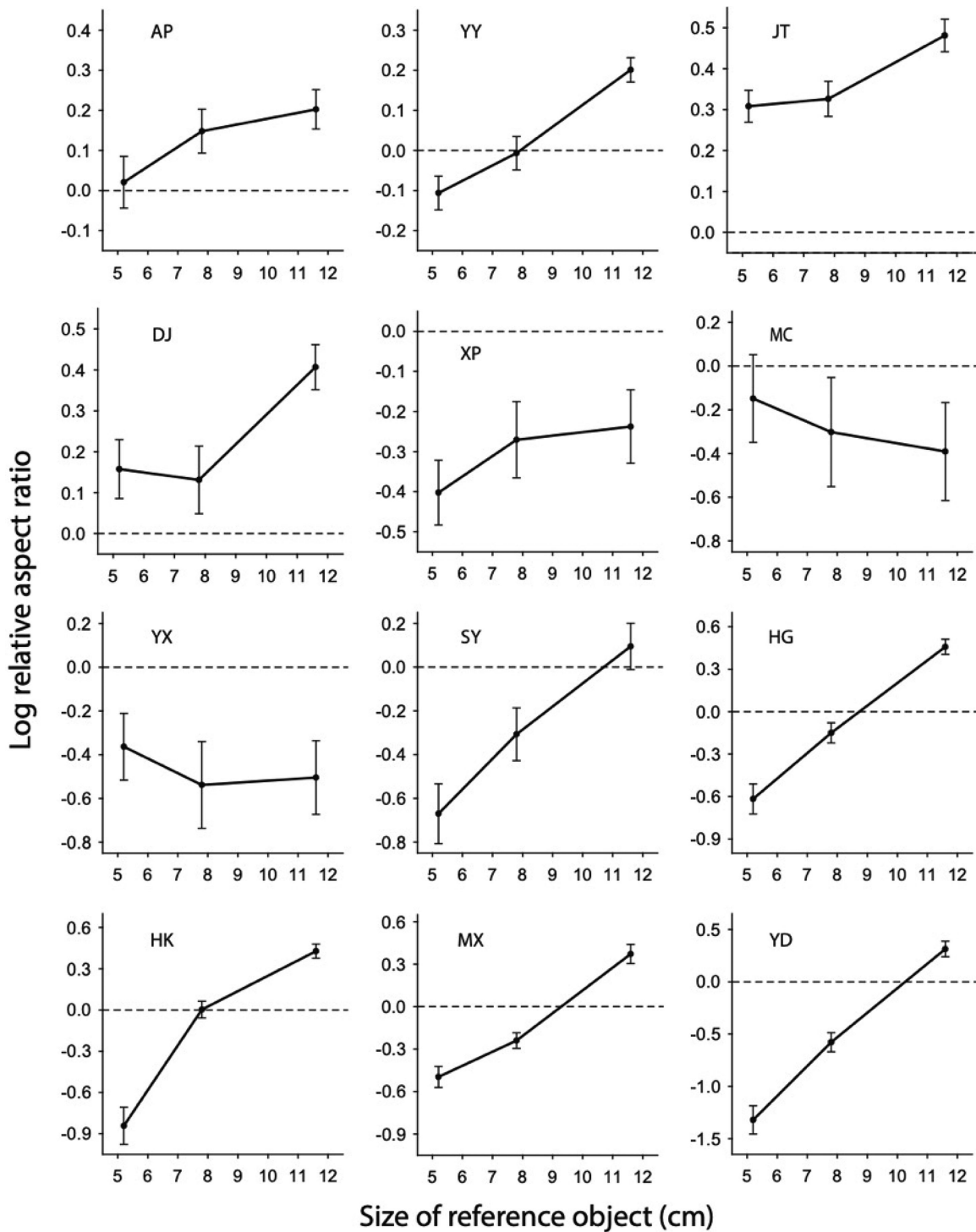


Figure 9. The average responses of each individual observer in Experiment 2. Note that these graphs are plotted on different scales to accommodate the variation of scaling differences exhibited by different observers. The horizontal dashed line on each graph represents veridical performance. Error bars denote ± 1 standard error of the mean.

over 100 years ago. A particularly good example of this can be experienced while driving on the highway. In the United States, the dashed lines that separate lanes are 10 ft long, but if observers are asked to estimate the length of a line in front of the car they are driving, the

average response is only 2 ft (Shaffer, Maynor, & Roy, 2008). Most of the early research on this topic involved judging length intervals on the ground (e.g., Baird & Biersdorf, 1967; Gilinsky, 1951; Harway, 1963; Heine, 1900; Norman et al., 1996; Toye, 1986; Wagner, 1985),

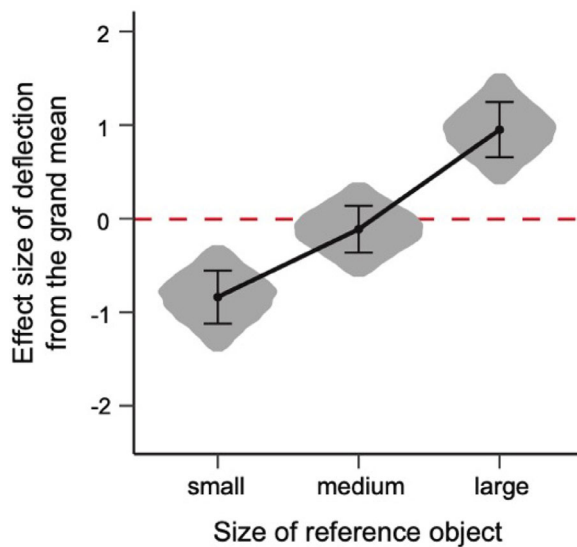


Figure 10. Posterior estimates of the effect size of the group-averaged deflection from the grand mean in [Experiment 2](#) as a function of object size. The shaded violins denote the 95% highest density interval (HDI) of the posterior probability distribution (Bayesian “Model-4” specified in the [Appendix](#)). The point and error bars in each violin denote the mean and standard deviation of the distribution. The horizontal dashed line denotes zero deflection from the grand mean.

but the same effect has been obtained when observers judge the apparent depth-to-width ratios of simple 3D objects like cylinders, square pyramids, or dihedral angles (e.g., [Campagnoli et al., 2017](#); [Campagnoli & Domini, 2019](#); [Glennester et al., 1996, 1998](#); [Hecht et al., 1999](#); [Johnston, 1991](#); [Johnston et al., 1994](#); [Scarfe & Hibbard, 2006](#); [Todd & Norman, 2003](#)).

[Pizlo, Sawada, Li, Kropatsch, and Steinman \(2010\)](#) have argued that these results are misleading because the stimuli employed did not satisfy the minimal conditions for computing veridical shape from symmetry. They did not contain four visible bilaterally symmetric point pairs, and they were most often viewed at a “degenerate” orientation for which the plane of symmetry was (nearly) parallel to the observer’s line of sight. The present experiments were designed specifically to consider whether observers can achieve stereoscopic shape constancy over variations in viewing distance and object size if the objects they are asked to judge satisfy the minimal conditions for computing shape from symmetry. The stimulus objects used in these experiments all contained more than four bilaterally symmetric point pairs, and they were also presented at “nondegenerate” orientations in which the symmetry planes were at a 45° angle relative to the line of sight. The adjustable stimuli were viewed stereoscopically at a fixed distance and orientation relative to the observer, and the experimental software allowed observers to expand or compress the extent

of the object in depth. The reference objects were presented at three possible simulated distances, and their sizes were varied as well.

Our results provide clear evidence that the apparent shapes of complex objects at “nondegenerate” orientations are systematically compressed with increasing viewing distance in exactly the same way as simple objects in “degenerate” orientations. It is important to keep in mind that these judgments could not have been achieved by simply comparing the apparent extensions in depth of the adjustable and reference objects. Because these objects always had different sizes, the observers were required to somehow normalize their judgments to compensate for that. There are several possible strategies by which this could have been achieved. For example, they could have estimated the relative depth-to-width ratio of the two objects, which is how we parameterized the results. Alternatively, they could have estimated the relative angles among different faces or the relative lengths among different edges.

As with all 3D adjustment tasks, there is no way of knowing for certain the specific object properties on which the observers based their responses, or whether they used a consistent strategy on all the trials. We suspect that variations in task strategy may have contributed to the individual differences in our data. For comparison, [Todd and Norman \(2003\)](#) asked observers to perform two different judgments on stereoscopic dihedral angles. One of the tasks required observers to adjust the angle so that its height appeared equal to its depth, and another required them to adjust the angle so that it appeared to be 90°. Judgments of the aspect ratio exhibited large compressions in depth with increasing viewing distance, whereas judgments of the angle did not (see also [Glennester et al., 1996](#); [Scarfe & Hibbard, 2013](#)).

Whatever strategy (or combination of strategies) was used to perform the adjustment task, the observed systematic deviations from veridicality indicate that our participants did not make effective use of the intersection-of-constraints algorithm outlined in the Introduction. It is important to keep in mind that that the one-parameter family of 3D interpretations for symmetric polyhedra in [Figure 1](#) assumes orthographic projection. In the case of perspective projection, there is additional information that allows a unique solution for computing 3D shape from symmetry ([Franc, ois et al., 2002](#); [Gordon, 1989](#); [Sawada et al., 2011](#)). Although the objects in the present experiment were viewed under strong perspective, the observers were unable to exploit this additional information to achieve shape constancy.

The results of the present experiments may appear at first blush to be inconsistent with an earlier study by [Li et al. \(2011\)](#). They obtained nearly veridical shape judgments using stimuli that were quite similar to the ones shown in [Figures 2 and 11](#). We suspect this is

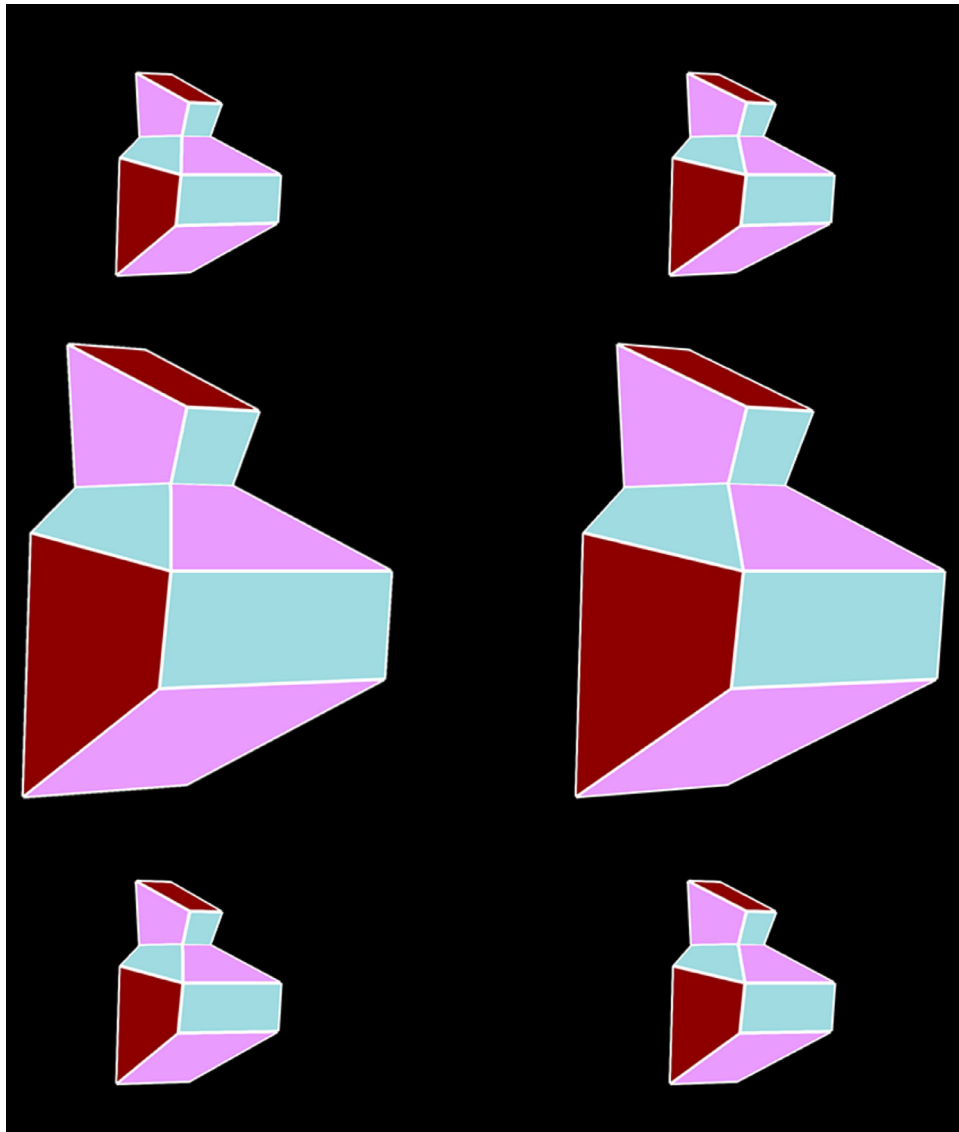


Figure 11. Three stereograms of polyhedral objects. One of the smaller ones has exactly the same shape as the large object, and the other has been expanded or compressed in depth by 40%. Can you identify which two have the same shape?

due to the unusual response task employed in their study. When asked to make judgments of 3D metric structure, observers frequently complain that this is perceptually difficult. If the specific task they are asked to perform allows some form of shortcut that makes it possible to achieve accurate performance in some other way, they will quickly learn to make use of it (see [Todd & Norman, 2003](#)). [Scarfe and Hibbard \(2006\)](#) have argued that nongeneric short cuts may be used more often than is widely appreciated. For example, the task used by [Li et al.](#) included a static, stereoscopically viewed reference object presented next to a monocularly viewed adjustable object that rotated in depth over 360° . Note that the rotating object provides much more information about 3D structure than the static object.

Indeed, it is possible to perform their task without even looking at the reference object, by exploiting the fact that the correct setting was always the one that was maximally compact.

We performed a control experiment to test the feasibility of this strategy. The rotating, adjustable object was presented by itself with no reference object to compare it to. Three observers adjusted this rotating object along the symmetry-preserving family in [Figure 1](#). This is the same adjustment space as in the experiment of [Li et al. \(2011\)](#). See their Equation B4 for details. Unlike [Li et al.](#), however, the participants in our control experiment were instructed to maximize the apparent compactness of the rotating object. The data were analyzed in terms of the dissimilarity between the

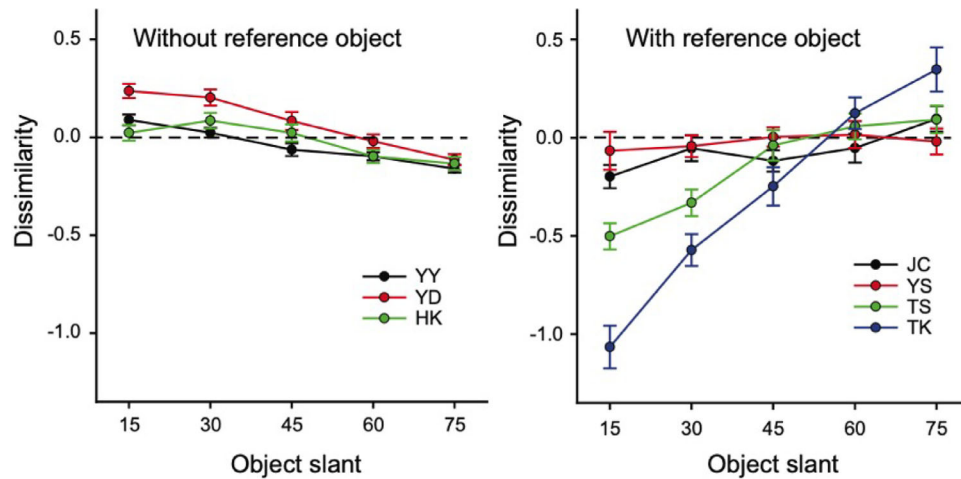


Figure 12. Dissimilarity (as per Equation B12 in Li et al., 2011) between the adjusted shape and the reference shape as a function of object slant. The task was to match the shape of the static reference object by adjusting the rotating object along the symmetry-preserving family in Figure 1. The right panel reproduces the data of Li et al. (2011, Figure 3) for stereoscopically viewed reference objects. The left panel shows the results from our control experiment, in which the adjustment family was the same as in Li et al., but the observers were told to maximize the compactness of the adjustment object in the absence of a reference object. Note that the control observers produced substantially more accurate “matches” without even seeing the reference shape. This indicates that the matching task in the experimental design of Li et al. could be performed on the basis of extraneous factors that had nothing to do with the perceived shape of the reference object. Error bars denote ± 1 standard error of the mean.

adjusted shape and the invisible reference shape. Zero dissimilarity indicates a perfect match (Equation B12 in Li et al., 2011). The left panel of Figure 12 plots our control data for five different object slants, and the right panel reproduces the results obtained by Li et al. under comparable conditions in the presence of a stereoscopic reference object. Note that the control observers produced substantially more accurate “matches” than the ones in Li et al. using a strategy that did not require them to estimate the 3D shape of the reference object. This is clearly a problem for their methodology.

Moreover, even if all observers in the Li et al. experiments had tried conscientiously to match the shape of the reference object, the adjustment space did not include any shapes that were related to that object by an expansion or compression of depth (cf. Figure 1 above). Thus, if any of the observers had misperceived the reference objects in that manner, which is a common result in the literature, the response task was incapable of detecting that. These methodological problems cast serious doubt on the interpretation of their results.

In another relevant experiment from the same lab, Jayadevan et al. (2018) employed an adjustment task that allowed objects to be expanded or compressed in depth. They concluded: “there was no systematic distortion of binocularly viewed shapes along the depth direction” (Jayadevan et al., 2018, p. 14). However, this interpretation was reached in a very informal manner—apparently by eyeballing some plots. No rigorous statistical analysis is reported in the

published article or the accompanying website. This is problematic because the unaided eye cannot distinguish unsystematic noise from systematic deviations from veridicality of the magnitude typical for our data.

In an effort to relate our results to those of Jayadevan et al., we reanalyzed our data in terms of their shape difference metric, which is defined as the normalized average of the absolute differences of corresponding angles (Equation 6 in Jayadevan et al., 2018). The black bar in Figure 13 shows the average shape difference for binocular symmetric polyhedra in their study. The red bars show the same measure applied to the data in our Experiment 1 for the three different viewing distances of the fixed-physical-size session. Note that the average error by this metric is quite comparable in both experiments, although our results have much smaller variance, most likely because our adjustment allowed fewer degrees of freedom. It should also be noted that this particular shape difference metric is incapable of detecting the clear effects of viewing distance that are evident in Figure 4. It appears, therefore, that this particular metric may be ill suited to detect systematic distortions of perceived 3D structure as opposed to random noise in the data.

The interpretation of Jayadevan et al. data is undermined further by various potential confounds. For the sake of argument, let us assume that a reanalysis of their data had revealed near-perfect match between the adjusted and reference shapes within a narrow

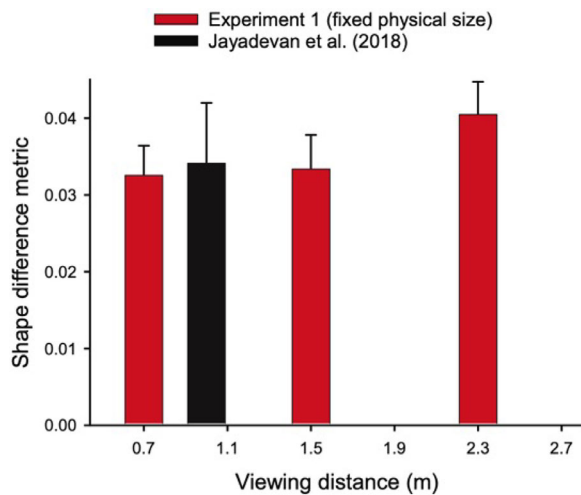


Figure 13. Shape differences between the adjustment object and the reference object as a function of viewing distance. The shape difference metric is the normalized average of absolute differences of corresponding angles as originally proposed by Jayadevan et al. (2018, Equation 6). The black bar reproduces the results for symmetric reference objects in the binocular condition of Jayadevan et al. (2018, Figure 7, averaged across their three participants). The red bars show the results from our Experiment 1 in the fixed physical size conditions. Note that this difference metric is insensitive to the large effects of viewing distance that are evident in Figure 4 above. Error bars denote ± 1 standard error of the mean.

confidence interval indicating adequate statistical power. Even this most favorable outcome still would not necessarily indicate *zero* distortion. Negligible differences in a shape matching task merely indicate *equivalent* distortions of the two objects. This is what we observed when the reference object was presented at the same simulated distance (1.5 m) as the adjustable object in our Experiment 1 (Figure 3). Our experimental design decoupled the simulated viewing distances of the two objects expressly to protect from this potential confound. In the experiment of Jayadevan et al., the two objects were presented at the same distance (1.0 m in the binocular condition). Granted, the two objects were not on a completely equal footing because one object was static while the other rotated continuously in depth, but our default expectation in the absence of explicit evidence to the contrary ought to be that *both* objects would be affected by any distortion arising within the visual system. Besides, the continuous rotation of the adjustable stimulus in this experiment raises methodological concerns in its own right as discussed in the context of Li et al. (2011) above. Finally, even if we assume—again for the sake of argument—that it were established somehow that near-veridical performance was achieved by Jayadevan et al., it would still remain an open question whether this favorable outcome

would generalize to viewing distances other than one meter. Recall that this particular distance is close to the “sweet spot” where Johnston (1991) obtained veridical performance.

What is most striking about our results is how similar they are to those obtained using simpler stimuli that do not satisfy the minimal conditions for shape-from-symmetry computations. Even though it is mathematically possible to accurately recover the 3D metric structure of our stimuli using relatively simple algorithms, the results reveal that human observers are unable to do so. These findings suggest that observers’ judgments of 3D metric structure in the present study were determined primarily by the pattern of binocular disparity magnitudes, and that the effects of symmetry on stereoscopic shape judgments are likely to be quite minimal.

Keywords: symmetry, shape constancy, veridical perception, computer vision, psychophysics

Acknowledgments

Supported by a Grant from the National Science Foundation (BCS-1849418).

Commercial relationships: none.

Corresponding author: Alexander A. Petrov.

Email: apetrov@alexpetrov.com.

Address: Department of Psychology, The Ohio State University, Columbus, OH, USA.

References

- Baird, J. C., & Biersdorf, W. R. (1967). Quantitative functions for size and distance judgments. *Perception & Psychophysics*, 2, 161–166.
- Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94(2), 115–147.
- Bradshaw, M. F., Glennerster, A., & Rogers, B. J. (1996). The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Research*, 36(9), 1255–1264.
- Brown, A.M., Lindsey, D. T., Satgunam, P., & Miracle, J. A. (2007). Critical immaturities limiting infant binocular stereopsis. *Investigative Ophthalmology & Visual Science*, 48 (3), 1424–1434.
- Campagnoli, C., Croom, S., & Domini, F. (2017). Stereovision for action reflects our perceptual experience of distance and depth. *Journal of Vision*, 17(9):21, 1–26, <https://doi.org/10.1167/17.9.21>.

- Campagnoli, C., & Domini, F. (2019). Does depth-cue combination yield identical biases in perception and grasping? *Journal of Experimental Psychology: Human Perception and Performance*, *45*(5), 659–680.
- Champion, R. A., Simmons, D. R., & Mamassian, P. (2004). The influence of object size and surface shape on shape constancy from stereo. *Perception*, *33*(2), 237–247.
- Collett, T. S., Schwarz, U., & Sobel, E. C. (1991). The interaction of oculomotor cues and stimulus size in stereoscopic depth constancy. *Perception*, *20*(6), 733–754.
- François, A. R. J., Medioni, G. G., & Waupotitsch, R. (2002). Reconstructing mirror symmetric scenes from a single view using 2-view stereo geometry. *Proceedings of the 16th IEEE International Conference on Pattern Recognition* (Vol. 4, pp. 12–16). Quebec, Canada: IEEE.
- Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Analysis*, *1*(3), 515–534.
- Gilinsky, A. S. (1951). Perceived size and distance in visual space. *Psychological Review*, *58*(6), 460–482.
- Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic depth constancy depends on the subject's task. *Vision Research*, *36*(2), 3441–3456.
- Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1998). Cues to viewing distance for stereoscopic depth constancy. *Perception*, *27*, 1357–1365.
- Gordon, G. G. (1989). Shape from symmetry. *Proceedings of the SPIE Conference: Intelligent Robots and Computer Vision VIII: Algorithms and Techniques* (Vol. 1192, pp. 297–308). Philadelphia, PA: Society of Photo-Optical Instrumentation Engineers (SPIE).
- Hartley, R., & Zisserman, A. (2003). *Multiple-view geometry in Computer Vision*. Cambridge, UK: Cambridge University Press.
- Harway, N. I. (1963). Judgment of distance in children and adults. *Journal of Experimental Psychology*, *65*, 385–390.
- Hecht, H., van Doorn, A., & Koenderink, J. J. (1999). Compression of visual space in natural scenes and in their photographic counterparts. *Perception & Psychophysics*, *61*, 1269–1286.
- Heine, L. (1900). Über Orthoskopie oder über die abhängigkeit relative entfernunggsschätzungen von der vorstellung absoluter entfernung [On “orthoscopy” or on the dependence of relative distance on the representation of absolute distance]. *Albrecht von Graefe's Archiv für Ophthalmologie*, *51*, 563–572.
- Hong, W., Yang, A. Y., Huang, K., & Ma, Y. (2004). On symmetry and multiple-view geometry: Structure, pose, and calibration from a single image. *International Journal of Computer Vision*, *60*(3), 241–265.
- Howard, I. P., & Rogers, B. J. (2012). *Perceiving in depth, volume 2: Stereoscopic vision*. New York, NY: Oxford University Press.
- Huang, T. S., & Lee, C. H. (1989). Motion and structure from orthographic projections. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, *11*, 536–540.
- Jayadevan, V., Michaux, A., Delp, E., & Pizlo, Z. (2017). 3D shape recovery from real images using a symmetry prior. *Proceedings of the IS&T International Symposium on Electronic Imaging: Computational Imaging IX* (Vol. 10, pp. 106–115). San Francisco, CA: Society for Imaging Science and Technology.
- Jayadevan, V., Sawada, T., Delp, E., & Pizlo, Z. (2018). Perception of 3D symmetrical and nearly symmetrical shapes. *Symmetry*, *10*(8), 344, 1–24.
- Johnston, E. B. (1991). Systematic distortions of shape from stereopsis. *Vision Research*, *31*(7), 1351–1360.
- Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research*, *34*(17), 2259–2275.
- Kleiner, M., Brainard, D., Pelli, D., Ingling, A., Murray, R., & Broussard, C. (2007). What's new in Psychtoolbox-3. *Perception*, *36*(14), 1–16.
- Koenderink, J. J., & van Doorn, A. J. (1976) Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, *21*, 29–35.
- Koenderink, J. J., & van Doorn, A. . (1991). Affine structure from motion. *Journal of the Optical Society of America*, *8*(2), 377–385.
- Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (2001) Ambiguity and the “Mental Eye” in pictorial relief. *Perception*, *30*, 431–448.
- Kruschke, J. K. (2015). *Doing Bayesian data analysis: A tutorial with R, JAGS, and Stan* (2nd Ed.). New York: Academic Press.
- Lee, M. D. (2008). Three case studies in the Bayesian analysis of cognitive models. *Psychonomic Bulletin & Review*, *15*(1), 1–15.
- Lee, M. D., & Wagenmakers, E. J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge, UK: Cambridge University Press.
- Lee, Y. L., & Saunders, J. A. (2013). Symmetry facilitates shape constancy for smoothly curved 3D objects. *Journal of Experimental Psychology: Human Perception & Performance*, *39*(4), 1193–1204.

- 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*(9), 979–991.
- Li, Y., Sawada, T., Latecki, L. J., Steinman, R. M., & Pizlo, Z. (2012). A tutorial explaining a machine vision model that emulates human performance when it recovers natural 3D scenes from 2D images. *Journal of Mathematical Psychology*, *56*, 217–231.
- Li, Y., Sawada, T., Shi, Y., Kwon, T., & Pizlo, Z. (2011). A Bayesian model of binocular perception of 3D mirror symmetrical polyhedra. *Journal of Vision*, *11*(4), 1–20, <https://doi.org/10.1167/11.4.11>.
- Lipton, L. (1991). *The CrystalEyes Handbook*. San Rafael, CA: StereoGraphics Corporation.
- Ma, Y., Soatto, S., Kosecka, J., & Sastry, S. S. (2004). *An invitation to 3-D vision: From images to geometric models*. New York: Springer.
- Michaux, A., Jayadevan, V., Delp, E., & Pizlo, Z. (2016). Figure-ground organization based on three-dimensional symmetry. *Journal of Electronic Imaging*, *25*(6), 061606.
- Michaux, A., Kumar, V., Jayadevan, V., Delp, E., & Pizlo, Z. (2017). Binocular 3D object recovery using a symmetry prior. *Symmetry*, *9*(5), 64.
- Norman, J. F., Swindle, J. M., Jennings, L. R., Mullins, E. M., & Beers, A. M. (2009). Stereoscopic shape discrimination is well preserved across changes in object size. *Acta Psychologica*, *131*(2), 129–135.
- 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), 173–186.
- Park, M., Lee, S., Chen, P. C., Kashyap, S., Butt, A. A., & Liu, Y. (2008). Performance evaluation of state-of-the-art discrete symmetry detection algorithms. In *IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1–8). IEEE, doi:10.1109/CVPR.2008.4587824.
- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.
- Petrov, A. P. (1980). A geometrical explanation of the induced size effect. *Vision Research*, *20*(5), 409–413.
- Pizlo, Z., Li, Y., Sawada, T., & Steinman, R. (2014). *Making a machine that sees like us*. New York: Oxford University Press.
- Pizlo, Z., Sawada, T., Li, Y., Kropatsch, W. G., & Steinman, R. M. (2010). New approach to the perception of 3D shape based on veridicality, complexity, symmetry and volume. *Vision Research*, *50*(1), 1–11.
- Plummer, M. (2003). JAGS: A program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing*. Retrieved from <http://www.ci.tuwien.ac.at/Conferences/DSC-2003/>.
- Richards, W. A. (1985). Structure from stereo and motion. *Journal of the Optical Society of America*, *2*, 343–349.
- Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature*, *361*, 253–255.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, *16*(2), 225–237.
- Saunders, J. A., & Knill, D. C. (2001). Perception of 3D surface orientation from skew symmetry. *Vision Research*, *41*(24), 3163–3183.
- Sawada, T. (2010). Visual detection of symmetry of 3D shapes. *Journal of Vision*, *10*(6), 1–22, <https://doi.org/10.1167/10.6.4>.
- Sawada, T., Li, Y., & Pizlo, Z. (2011). Any pair of 2D curves is consistent with a 3D symmetric interpretation. *Symmetry*, *3*(2), 365–388.
- Scarfe, P., & Hibbard, P. B. (2006). Disparity-defined objects moving in depth do not elicit three-dimensional shape constancy. *Vision Research*, *46*, 1599–1610.
- Scarfe, P., & Hibbard, P. B. (2013). Reverse correlation reveals how observers sample visual information when estimating three-dimensional shape. *Vision Research*, *86*, 115–127.
- Shaffer, D. M., Maynor, A. B., & Roy, W. L. (2008). The visual perception of lines on the road. *Perception & Psychophysics*, *70*(8), 1571–1580.
- Shiffrin, R. M., Lee, M. D., Kim, W., & Wagenmakers, E.-J. (2008). A survey of model evaluation approaches with a tutorial on hierarchical Bayesian methods. *Cognitive Science*, *32*(8), 1248–1284.
- Shimshoni, I., Moses, Y., & Lindenbaum, M. (2000). Shape reconstruction of 3D bilaterally symmetric surfaces. *International Journal of Computer Vision*, *39*, 97–110.
- Sinha, S., Ramnath, K., & Szeliski, R. (2012). Detecting and reconstructing 3D mirror symmetric objects. *European Conference on Computer Vision* (pp. 586–600). Berlin, Heidelberg: Springer.
- Thrun, S., & Wegbreit, B. (2005). Shape from symmetry. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1* (Vol. 2, pp. 1824–1831). IEEE, doi:10.1109/ICCV.2005.221.

- Todd, J. T. (2004). The visual perception of 3D shape. *Trends in Cognitive Sciences*, 8(3), 115–121.
- 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(1), 31–47.
- Toye, R. C. (1986). The effect of viewing position on the perceived layout of space. *Perception & Psychophysics*, 40(2), 85–92.
- Treder, M. S. (2010). Behind the looking-glass: A review on human symmetry perception. *Symmetry*, 2(3), 1510–1543.
- Tyler, C. W. (Ed.). (2002). *Human symmetry perception and its computational analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Vetter, T., Poggio, T., & Bülthoff, H. H. (1994). The importance of symmetry and virtual views in three-dimensional object recognition. *Current Biology*, 4(1), 18–23.
- Wagner, M. (1985). The metric of visual space. *Perception & Psychophysics*, 38(6), 483–495.
- Watt, S. J., Akeley, K., Ernst, M. O., & Banks, M. S. (2005). Focus cues affect perceived depth. *Journal of Vision*, 5(10), 7, 834–862, <https://doi.org/10.1167/5.10.7>.
- Yang, A., Huang, K., Rao, S., Hong, W., & Ma, Y. (2005). Symmetry-based 3D reconstruction from symmetric images. *Computer Vision and Image Understanding*, 99, 210–240.
- Zabrodsky, H., & Weinshall, D. (1997). Using bilateral symmetry to improve 3D reconstruction from image sequences. *Computer Vision and Image Understanding*, 67, 48–57.

Appendix

Hierarchical bayesian data analyses

We performed four hierarchical Bayesian data analyses: Model-1 for the fixed-physical-size session of [Experiment 1](#), Model-2 for the fixed-projected-size session of [Experiment 1](#) ([Figure 6](#)), Model-3 for the closest-distance conditions in both sessions of [Experiment 1](#) ([Figure 7](#)), and Model-4 for [Experiment 2](#) ([Figure 10](#)). Except for a minor modification involving Model-3 that will be clarified below, all four models used the same structure and the same priors as specified diagrammatically in [Figure A1](#).

It is useful to conceptualize this graphical structure as a Bayesian analog of one-way ANOVA. The experimental design matrix is counterbalanced across participants i , experimental conditions k , and replications j as indicated by the nested plates in [Figure A1](#). Each of the 10 polyhedral objects was presented three times for a total of 30 replications per condition. The four models were individuated by their experimental conditions: In Model-1 and Model-2 these were the three viewing distances for the reference object. Model-3 had only two conditions: fixed physical size ($k = 1$) versus fixed projected size ($k = 2$). Finally, the conditions in Model-4 corresponded to the three sizes of the reference object in [Experiment 2](#). One behavioral observation was collected per trial—the log relative aspect ratio (cf. [Equation 1](#) in the main text). This quantity is referred to as “adjustment” throughout this [Appendix](#). It enters the Bayesian model via the random variable y_{ijk} in the innermost plate. It is assumed that all observations for a given participant i and a given

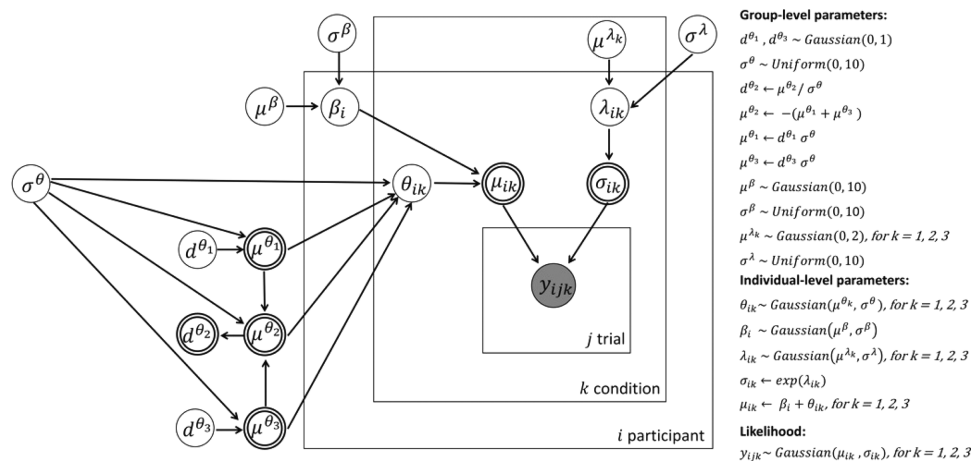


Figure A1. General structure and model specifications of the hierarchical Bayesian statistical analyses used in this study. Notational conventions: Nodes denote random variables, arrows denote dependencies, and plates denote exchangeable replications. Shaded = observable, unshaded = latent, double borders = deterministically calculated, and single borders = stochastic variables (Lee, 2008; Shiffrin, Lee, Kim, & Wagenmakers, 2008).

condition k are drawn independently from a Gaussian distribution with mean μ_{ik} and standard deviation σ_{ik} (intermediate plate in Figure A1). To accommodate individual differences, the model includes idiosyncratic means and standard deviations in each condition.

The hierarchical structure of the model is designed to accommodate both individual differences and commonalities within a single condition at the same time. The individual-level parameters that govern the distribution of the observable data reflect individual differences. They are sampled in turn from group-level distributions governed by group-level parameters, which reflect commonalities across participants within a given condition. For example, consider the individual differences in the variability of adjustments across the 30 replications in a given condition that are manifested in the unequal widths of the error bars in Figures 5 and 9. These individual differences are modeled by the random variable σ_{ik} whose natural logarithm λ_{ik} is sampled from a group-level Gaussian distribution with group-level parameters μ^{λ_k} and σ^{λ} . More importantly, there are individual differences in overall adjustment level—the individual profiles in Figures 5 and 9 “float” up and down relative to each other. The model accounts for them by partitioning each individual-level mean (μ_{ik}) into two parts:

$$\mu_{ik} = \beta_i + \theta_{ik} \quad (\text{A1})$$

where β_i is the grand mean of Participant i 's adjustments across all experimental conditions, and θ_{ik} is Participant i 's deflection in the k th condition from his/her own grand mean. The random variables β_i for participants $i = 1, 2, \dots, 12$ are sampled independently from a common group-level Gaussian distribution with group-level parameters μ^{β} and σ^{β} . The individual deflections θ_{ik} are sampled from Gaussian distributions with common standard deviation σ^{θ} . Importantly, the latter distributions have different means μ^{θ_k} that characterize the respective condition k .

The group-level parameters μ^{θ_k} are the topic of main interest in the current analyses. They are analogous to the main effect of the condition factor in traditional ANOVA. Specifically, μ^{θ_k} is the group-averaged deflection from the grand mean in experimental condition k , and thus estimates the effect of the experimental manipulation after controlling for individual differences. One technical challenge that arises at this point is to enforce the sum-to-zero

constraint. Ideally, the sum of the deflections θ_{ik} across all conditions should equal zero for any given participant i . To simplify the computation, this sum-to-zero constraint was approximated at the group level by enforcing

$$\mu^{\theta_2} = -(\mu^{\theta_1} + \mu^{\theta_3}) \quad (\text{A2})$$

where μ^{θ_1} , μ^{θ_2} , and μ^{θ_3} are the group-means of θ_{ik} for $k = 1, 2$, and 3 , respectively. In Model-3, which has only two conditions, this reduced to $\mu^{\theta_2} = -\mu^{\theta_1}$.

Instead of placing priors on μ^{θ_k} directly, we placed priors on their effect sizes, as suggested by Lee and Wagenmakers (2014). The effect sizes were denoted as d^{θ_k} and defined as $d^{\theta_k} = \mu^{\theta_k} / \sigma^{\theta}$. Following common practice (e.g., Rouder, Speckman, Sun, Morey, & Iverson, 2009), we used the standard Gaussian as the priors on effect sizes in our model. Note that because the standard deviation, σ^{θ} , is common for all conditions, the sum-to-zero constraint applies to the effect sizes d^{θ_k} as well as the group means μ^{θ_k} . For other group-level parameters that are not the focus of the current analysis, we used priors that contain very little information so that the results of the analysis would be largely driven by the data. We placed weakly informative Gaussian priors on the group-means other than μ^{θ_k} , and placed noninformative priors with large range, *Uniform* (0, 10), on the group-level standard deviations, as suggested by Gelman (2006).

The models were implemented in JAGS (Plummer, 2003). Results of each model were based on two chains of Markov chain Monte Carlo (MCMC), each consisting of 9000 samples collected following a burn-in period of 1000 samples. Convergence of the chains was confirmed by visually examining the trace plots of all the group-level parameters. Bayesian posterior predictive distributions fit well to the corresponding distributions of the observations, indicating good model fit.

In Bayesian statistics, the reliability of an effect can be evaluated by the degree of separations among distributions of the posterior estimates under different levels of the manipulation. Specifically, we used 95% highest density interval (HDI) of a distribution to characterize the range of the estimates that is credible, given the data and model assumption (Kruschke, 2015). Figures 6, 7, and 10 plot the HDIs for the group-level effect sizes d^{θ_k} estimated from the corresponding data set.