Perception & Psychophysics
2002, 64 (4), 507-520

Object pose: Perceiving 3-D shape as sticks and slabs

AUGUSTINUS H. J. OOMES and TJEERD M. H. DIJKSTRA
Ohio State University, Columbus, Ohio

Estimating the pose (three-dimensional orientation) of objects is an important aspect of 3-D shape perception. We studied the ability of observers to match the pose of the principal axes of an object with the pose of a cross consisting of three perpendicular axes. For objects, we used a long and a flat spheroid and eight symmetric objects with aspect ratios of dimensions of approximately 4:2:1. Stimulus cues were the contour and stereo for the spheroids, and contour, stereo, and shading for the symmetric objects. In addition, the spheroids were shown with or without surface texture and with or without active motion. Results show that observers can perform the task with standard deviations of a few degrees, though biases could be as large as 30º. The results can be naturally decomposed in viewer-centered coordinates, and it turns out that the estimation of orientation in the frontoparallel plane (tilt) is more precise than estimation of orientation in depth (slant, roll). A comparison of long and flat spheroids shows that sticks lead to better performance than do slabs. This can even be the case within the same object; the pose of the stick-like aspect is seen with more precision than is the pose of the slab-like aspect. The largest biases occurred when the spheroids were displayed with the binocular contour as the only cue. We can explain these biases by assuming that subjects' settings are influenced by the orientation of the rim.

Next to perceiving where objects are located in our environment, we usually want to know what their shapes are. One of the salient aspects of an object is how it is oriented relative to other objects in the environment. The description of an office scene becomes clear if we are given not only the location of the desk but also its orientation with respect to the room. In social interaction, it is important to know where the other person is, but perhaps more important to know the pose of the head and the direction of gaze in order to determine whether that person is looking at you or not. Our ability to perceive the pose of objects is the focus of this paper.

Willats (1992) argued for an unconventional explanation of our ability to perceive the global shape of a 3-D object. He classified shape into the three perceptual categories—*sticks*, *lumps*, and *slabs*—drawing on linguistic and developmental evidence (Figure 1). Sticks are objects with one long dimension and two short ones, whereas slabs are objects with one short dimension and two long ones. In the case of lumps, the three dimensions have roughly equal lengths. The analysis of their projected silhouettes reveals that, in general, sticks project to elongated silhouettes, lumps to round silhouettes, and slabs to either elongated or round silhouettes. He concluded that observers perceive elongated silhouettes as sticks and round silhouettes as lumps, whereas slabs are not recognized as such but are categorized as either sticks or lumps. What is characteristic of Willats's approach is his emphasis on the extended region in the projection (i.e., silhouette) instead of on the occluding contour only.

The crucial aspect of his analysis in the present context is that the shape of the silhouette depends on the pose (3-D orientation) of the object. This is why the projection of the slab is perceptually ambiguous; in roughly half of the possible poses, it looks elongated, and in the other poses, it looks round. Even the elongated projection of the stick is ambiguous; it could be a short stick that is seen from the side, or a long stick that makes an acute angle with the viewing direction (foreshortened view). Therefore, the perception of the global shape and the pose of an object are tightly connected.

The situation is analogous to the perception of the size and distance of an object; the estimation of the two parameters are coupled, and a misjudgment in the one parameter might lead to a misjudgment in the other. A well-known illusion for shape and pose is the *Ames window*, where a fully rotating trapezoidal window is seen as a rectangular window that is oscillating around a vertical axis (Ames, 1951). Because we assume the shape of the window to be a rectangle, we make an error in the estimation of its pose.

To simplify our discussion, we will define some common terms (see Figure 2). An object is a bounded and opaque region in 3-D space. The *pose* of the object is determined by the orientation of its principal axes relative to the environment (e.g., gravitational coordinate system: *x*-, *y*-, and *z*-

This research was partially funded by the Netherlands Organization for Scientific Research (NWO). We thank Jim Todd for comments and suggestions. Correspondence concerning this article should be addressed to T. Dijkstra, Department of Psychology, Ohio State University, 1885 Neil Avenue, Columbus, OH 43210 (e-mail: dijkstra.1@osu.edu).

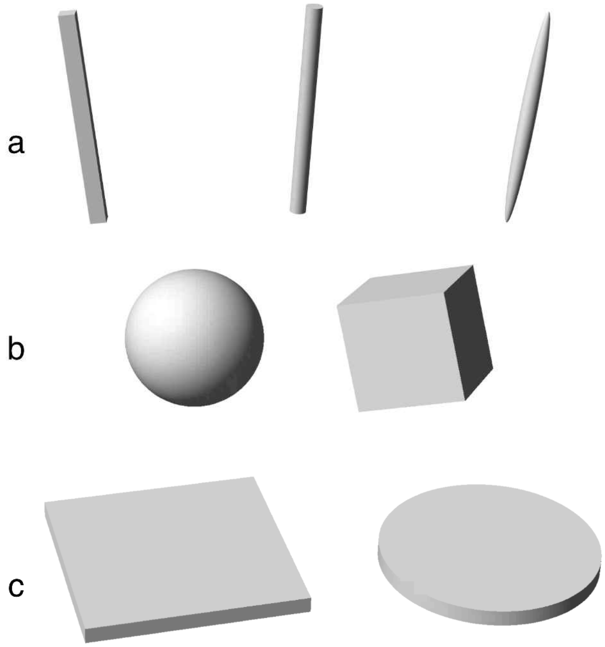

Figure 1. (a) Sticks. (b) Lumps. (c) Slabs.

axes). In turn, the principal (intrinsic) axes of the object depend on the mass distribution, and in classical mechanics, they are defined as those axes of rotation around which the moment of inertia is minimal (*major axis*), or maximal (*minor axis*), with a third (*median axis*) perpendicular to both of them. The generic view seen in Figure 2a shows the major axis, represented as a stick pinched through the long dimension of the box.

The observer's cyclopean eye determines the viewpoint. Light rays that graze the surface of the object project as the occluding *contour*. We use the term *silhouette* to refer to the contour plus the region that it encloses. The visible part of the surface of the object lies within the contour, and its surface structure can be revealed by local landmarks, texture, and reflectance (diffuse shading, specular highlights). Those points on the surface where the light rays touch form the *rim*, which is also referred to as the *contour generator*. The rim is generally a space curve, and its shape depends both on the shape of the object and the vantage point (Koenderink, 1984).

The *reference line* is defined as the connection between the viewpoint and the center of mass of the object. The *pose* of the object in viewer-centered coordinates is the orientation of the major axis of the object and can be represented by two angles: *slant* and *tilt*. The tilt denotes the angle around the line of sight, or the orientation in the (frontoparallel) plane, and we use the convention that the horizontal, pointing to the right, is 0°, and positive angles are measured counterclockwise (Figure 2b). The slant is the angle between the object axis and the line of sight and is commonly referred to as the *orientation in depth* (Figure 2c). In the case that the slant is 0°, the major axis of the object coincides with the line of sight.

Marr and Nishihara (1978) hypothesized that the perceptual representation of an object should be relative to an object-centered coordinate system. Every object defines a principal axis that indicates the global orientation: "A canonical coordinate frame must be set up within the object *before* its shape is described, and there seems to be no way of avoiding this" (Marr, 1982, p. 296). This confident message begs the question of how the principal axis is extracted from the optical input in the first place. Marr's answer is that in the case of generalized cones, the principal axis can be derived from the contour, unless the axis is too foreshortened.

This claim rests on the assumption that parts of the rim of an object have the same orientation as does the global axis. In fact, this is only true for a special type of generalized cone: one with a straight axis and a constant cross-section. For that type, the tilt of the orientation can be directly inferred from (a part of) the contour, and in the case of strong perspective, the convergence of the contours parallel to the axis is a cue to the slant. In general, however, the rim of an object is a space curve that has no such simple relation to the pose. Therefore, the contour cannot reveal the pose of the object the way Marr (1982) proposed, unless it is a member of a very restricted set of generalized cones.

Little is known about the human ability to visually estimate the pose of an object. It emerges somewhere in the first year of human development. In a study about grasping (Robinson, McKenzie, & Day, 1996), it was found that after approximately 10 months of age, infants orient their hand to the major axis of an object before grasping it. They clearly rely on optical evidence to anticipate a successful grasping movement; younger infants reach for the object and then find the proper grip by touch, before picking it up.

More quantitative evidence comes from the perception of line orientation in the frontoparallel plane. The most striking result is the oblique effect; observers are more sensitive to cardinal orientations (horizontal and vertical) than to oblique orientations (see Appelle, 1972, for a discussion of the effect in different species). The typical standard deviations for cardinal orientations are in the 0.1°−1.0° range, and they roughly double for oblique orientations. The oblique effect is also found for acuity of lines and gratings, but Westheimer and Beard (1998) found that it is most pronounced for discrimination of line orientation.

Appelle (1972) tried to explain the oblique effect by pointing to neurophysiological evidence that shows higher sensitivity of neurons for the cardinal orientations. Other authors have turned to the world instead of the brain and have pointed to the regularities in our carpentered environment to explain the large exposure to horizontal and vertical contours (Annis & Frost, 1973). However, a statistical analysis of the orientation distribution of contours in indoor and outdoor scenes has revealed that even in natural scenes, the cardinal orientations are more abundant (Coppola, Purves, McCoy, & Purves, 1998; Switkes, Mayer, & Sloan, 1978).

The first experimental treatment of 3-D object pose of which we are aware was performed by Wanger, Ferwerda,

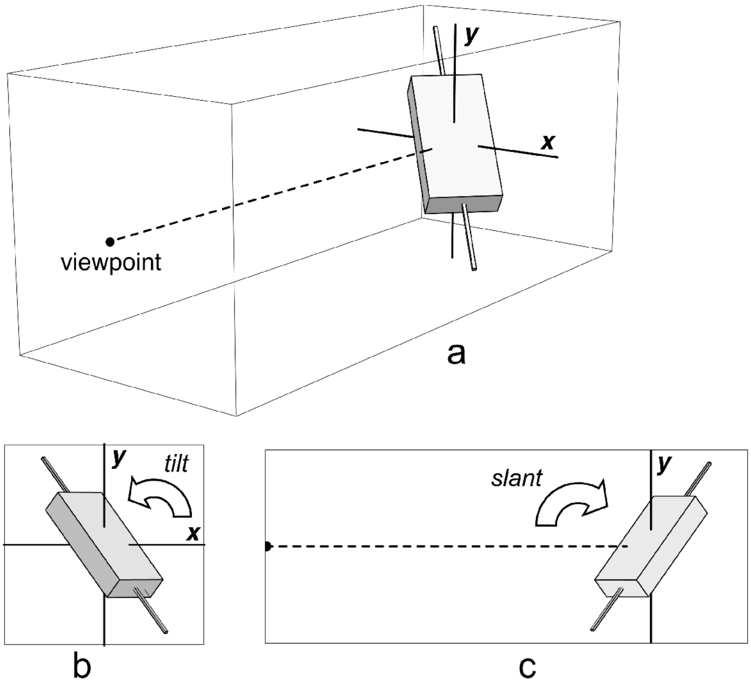

Figure 2. (a) Generic view of a rectangular box with the major axis pinched through the length dimension relative to an environmental frame of reference with the horizontal axis x and the vertical axis y and the line of sight, or reference line, as the third axis. (b) Frontal view of the box with the x- and y-axes; the tilt is defined as the counterclockwise angle between the major axis and the positive x-axis. (c) Side view of the box with y-axis and the line of sight; the slant is defined as the angle between the major axis and the line of sight.

and Greenberg (1992) and was done in the context of computer graphics research with the goal of finding simple cues for the visualization of object position, size, and orientation. In the relevant experiment, observers had to match the pose of two cubes in a simulated environment by rotating one of them. Results show a mean angular difference of 6.6° ± 3.4°. Especially the standard deviation of 3.4° is of interest here because it can be compared with the standard deviations for orientation in the plane that range from 0.1° to 1.0°.

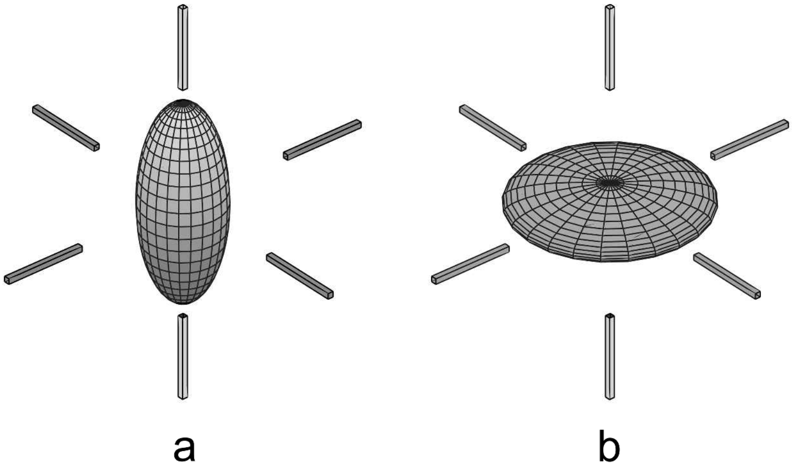

Figure 3. Objects used in Experiment 1 surrounded by three perpendicular axes with the vertical one aligned with the principal axis. A mesh is added to the surfaces for visualization purposes. (a) Long spheroid with an aspect ratio of 1.83. (b) Flat spheroid with an aspect ratio of 0.54.

Table 1
Dimensions (in Centimeters) and Aspect Ratios of the Spheroids

Spheroid	Length	Diameter	Aspect Ratio
Long	12.0	6.5	1.83
Flat	5.3	9.8	0.54

The second relevant source is Pollick, Nishida, Koike, and Kawato (1994), who were interested in the perception of the rotation axis of rotating objects. In these experiments, observers had to indicate the orientation of the axis by putting their right index finger in the proper orientation. To establish a baseline of performance, they used a static metal rod that was attached to a robot arm that could be put in the desired orientation. The overall mean angular difference was 9.0° ± 3.8°, so the precision is in the same range as that found by Wanger et al. (1992). From the presentation of their data, it can not be inferred whether it contains an oblique effect, because they did not decompose their results in slant and tilt.

The term *slant perception* usually refers to the perception of the orientation in depth of a plane, and it has been studied extensively (see Braunstein, 1976, for a review). One striking result from these experiments is that there are large differences between stationary (static) and moving (dynamic) planes. In the static case, observers usually underestimate the slant of the plane; it looks more frontoparallel than it actually is. If the edges of the plane are visible and the observer can use the perspective cue of contour convergence, this dominates over surface texture or the aspect ratio of the plane. If only a texture gradient is available, the more regular the size and shape of the texture elements, the better the performance. In the dynamic case, the performance improves dramatically and the estimation of slant is closer to veridical (Gibson & Gibson, 1957). In that case, the influence of perspective and texture is subordinate to the influence of aspect ratio. If the slant of a plane is estimated from disparity information only, the slant is again underestimated, with vertical planes leading to better performance than horizontal planes (Rogers & Graham, 1983). In all the studies reviewed so far, the global orientation of an object has been considered. As for local surface orientation, Koenderink, van Doorn, and Kappers (1992) investigated the perception of surfaces in a picture of a Brancusi sculpture with the use of a small circular probe, which subjects had to fit onto the surface. Their results showed that the scatter in the slant was much larger than the scatter in the tilt.

In summary, there seems to be no detailed theoretical proposal of the ability of human observers to estimate object pose. The only starting point is the global shape categorization of Willats (1992) that is so tightly connected to object pose. The empirical results cited above can be summarized by stating that tilt perception has typical standard deviations in the 0.1°–1° range and shows the oblique effect, whereas perception of slant is guessed to be an order of magnitude larger, in the 1°–10° range. The problem with this summary is, of course, that it combines results from different experiments that were obtained with very different stimuli (lines, cubes, rods, planes) and with different tasks (matching, pointing, discriminating). The goal of the present study was to gather more systematic data on the perception of object pose and to identify some important factors that influence performance.

EXPERIMENT 1
Spheroids

The first goal of this experiment was to establish the performance of observers in matching the poses of an object and a cross. For objects, we used a long and a flat spheroid (i.e., ellipsoids of revolution), which were stylized versions of a stick and a slab (Figure 3). Spheroids have several advantages over other objects. They have no local structure that might reveal their pose, such as the ribs of a box, and their shapes can easily be manufactured by stretching or shrinking a sphere along one dimension. Also, they are mathematically well-defined objects, which makes an analytic treatment possible. The cross consisted of three perpendicular axes with their midsections removed so that it was surrounding the spheroid. The motivation for using a cross was that it has a clearly defined orientation in all three dimensions.

The task of the subject was to align the principal axis of the spheroid with one axis of the cross. In the case of the long spheroid, the pose was determined by the orientation of the major axis, and in the case of the flat spheroid, by the minor axis, as shown by the vertical axes in Figure 3. We distinguished static and dynamic conditions; in the static condition, the spheroid had a fixed pose and the cross could be rotated around its common center of mass. In the dynamic condition, the pose of the cross was fixed and the spheroid could be rotated. Therefore, the rotation of either the cross or the spheroid was totally under the control of the subject, who determined its axis, direction, and speed.

The second goal of this experiment was to determine what factors influence pose estimation. We emphasize here that some cues were present in all condition: The subjects were always exposed to objects with a clearly resolvable contour, and the scenes were always presented stereoscopically. In the experiment, we varied the pose of the static object (spheroids or cross) and chose 12 orientations. The spheroid was shown either as a silhouette or as a textured surface. In the first case, the observer could use only the binocular contour as the cue. In the latter case, the observer could also infer the surface orientation from the texture gradient and the stereoscopic information that was carried by

Table 2
Averages and Ranges of the Spherical Mean Error and Spherical Standard Deviation

Observer	Mean Error		Standard Deviation	
B.L.	4.4°	(0.4°–23.3°)	4.7°	(1.8°–9.2°)
J.T.	5.7°	(0.2°–31.1°)	3.2°	(1.3°–8.1°)
T.D.	5.8°	(0.3°–25.1°)	4.0°	(1.7°–7.8°)

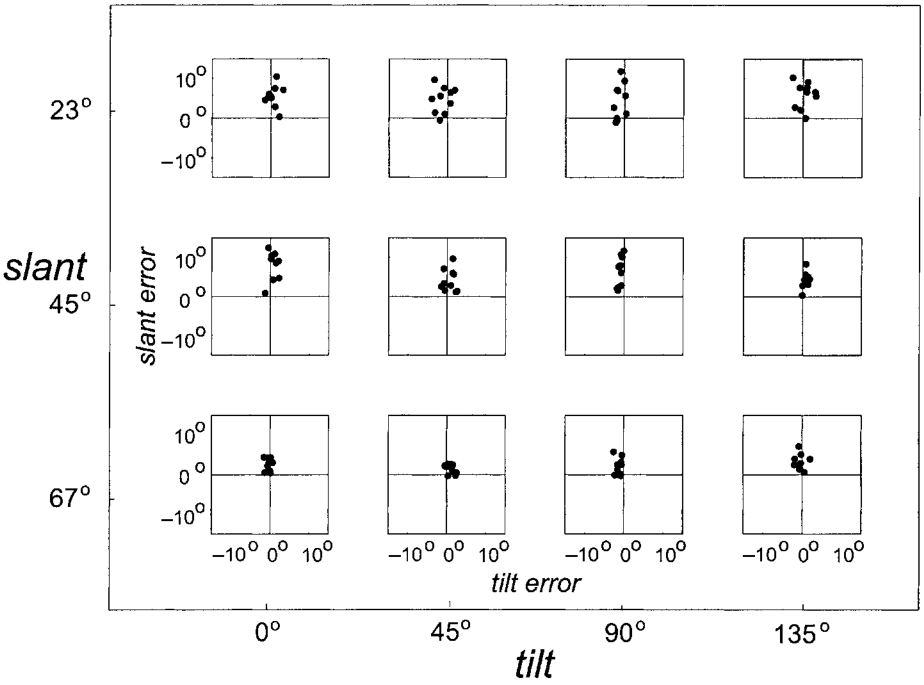

Figure 4. Raw results for Observer T.D. for the long spheroid in the dynamic silhouette condition. In each box, the tilt is along the horizontal axis, and the slant is along the vertical axis; the origin is the fiducial orientation (no error).

the texture. The spheroid was shown either statically or dynamically, and in the latter case, the motion gave extra information about its structure. Together, this led to the four cue combinations: dynamic texture, dynamic silhouette, static texture, and static silhouette.

Method

Apparatus. We used a Silicon Graphics Indigo2 with the OpenGL graphics package for the production and presentation of the stimuli. The monitor had a 120-Hz refresh rate and 1,280 × 1,024 pixel spatial resolution with an active image region of 35.4 × 27.7 cm. For stereoscopic viewing, we used ChrystalEyes liquid crystal shuttered glasses; the effective refresh rate was 60 Hz for each eye, and the effective resolution was 1,280 × 492 pixels. Head movements were restricted by a chinrest that was kept at a 105-cm distance, which resulted in a field of view of 19.5° × 15.2° visual angle.

Stimuli. The objects were one oblate and one prolate spheroid with constant volume, as shown in Figure 3. The volume was 270 cm^3. Table 1 shows the exact dimensions and aspect ratios (ratio of length to diameter) of the spheroids.

The cross consisted of three orthogonal axes with a square cross-section. Each axis had a width of 0.25 cm and a length of 29.0 cm from which the middle 15.0 cm was deleted. One axis was red, and the other two axes were blue. In the static condition, the spheroid was presented in 1 of 12 fixed poses; in the dynamic condition, the cross was put in the same poses. The slant values of these poses were 22.5°, 45.0°, and 67.5°, and the tilt values were 0.0°, 45.0°, 90.0°, and 135.0°.

The spheroid and cross were shown in perspective projection and in stereo. Their common center of mass was in the monitor plane, and the distance to the center of the spheroid was therefore also 105 cm. The distance between the eyes was taken to be 6.1 cm. A chinrest was set so that the object was at eye level and at the proper distance.

As mentioned, in half of the conditions the spheroid was shown with a textured surface (*static texture* and *dynamic texture* conditions). The texture was created by using an image with a high-density, homogeneous, and isotropic random pattern. The spheroids were built from 1,600 quadrilateral polygons that increased in size from the poles to the equator. In order to get a homogeneous and isotropic texture on the surface of the spheroid, we projected each quadrilateral on a random place in the texture image and then projected that part of the texture back onto the polygon. The scale of the texture elements was much smaller than the smallest polygon to ensure that the seams between the polygons were invisible. The surface of the textured spheroid was also shown with Gouraud shading. We used a parallel beam that frontally (perpendicular to the monitor plane) illuminated the object and the cross. The color of the light was white, and the background was set to an intermediate gray.

Procedure. Each of the four sessions consisted of 240 trials in which the cues (dynamic or static, texture or silhouette) were kept constant. The object (long or flat spheroid) was shown in 12 poses (3 slants and 4 tilts), and each condition was repeated 10 times. The order of trials with different shapes and viewpoints was randomized. There was

Table 3
Averages and Ranges of the Circular Standard Deviations of Tilt and Slant With Their Mean Ratio

Observer	Tilt		Slant		Mean Ratio
B.L.	2.1°	(0.7°–5.1°)	4.0°	(1.4°–8.8°)	2.4
J.T.	1.5°	(0.4°–3.6°)	2.7°	(0.8°–6.8°)	2.4
T.D.	1.8°	(0.4°–7.0°)	3.4°	(1.1°–7.1°)	2.5

Note—The ratios were calculated from the individual conditions, and thus the mean ratio is not equal to the ratio of the mean standard deviations of slant and tilt.

Table 4
Average Tilt Standard Deviations for the Long and Flat Spheroids and Their Ratios

Observer	Long Spheroid	Flat Spheroid	Ratio
B.L.	1.4°	2.8°	2.1
J.T.	1.1°	1.8°	1.7
T.D.	1.2°	2.3°	1.9

no time pressure, and it took the observers approximately 1 h to complete each session. No feedback about the performance was provided.

The subjects had to align the principal axis of the spheroid with the red axis of the cross. The object whose pose had to be adjusted (either spheroid or cross) was shown in a random pose at the beginning of the trial. It could be rotated around the common center of mass by holding the mouse button down and moving the mouse. The direction of the mouse movement determined the (perpendicular) direction of the rotation axis. A large mouse displacement corresponded to a large angle of rotation. If the edge of the screen was reached, the subject could release the mouse button and recenter the cursor to make further adjustments. By pushing the space bar on the keyboard, the subject could confirm the setting and was shown the next trial.

Observers. The 3 observers, B.L., J.T., and T.D., were myopic and had corrected-to-normal visual acuity. They all had extensive psychophysical experience.

Results and Discussion

The 3 observers were exposed to 96 different conditions that varied along the dimensions of tilt (4), slant (3), shape (2), and cue (4). In order to get an overview of the performance, we calculated the spherical mean error and the spherical standard deviation in each condition. The mean error is the (unsigned) difference between the fiducial orientation and the average orientation of the 10 settings. The standard deviation indicates the spread of those settings around the mean orientation. (See the Appendix for a mathematical definition of these measures). Table 2 lists the average and the range of the mean error and standard deviation for each subject.

The average performance is on the order of a few degrees, both for accuracy (mean error) and for precision (standard deviation). This is in the same ballpark as the results reported by Wanger et al. (1992) and Pollick et al. (1994). Note, however, that the mean error can be more than 30°, whereas the standard deviations will stay under 10°. Subjects can be inaccurate and have large constant errors in certain conditions, but they are also precise, in the sense that they have small relative errors.

Slant-tilt decomposition. The results in Table 2 indicate only the magnitude of the errors but not their direction. The mean error can be large, but from these numbers, we can not infer the direction of the error. Also, the standard deviation is assumed to be isotropic, although it is in fact highly directional. We have only to look at Figure 4, which shows the typical results for the long spheroid in the dynamic silhouette condition for observer T.D., to see that the distributions are not isotropic.

Figure 4 shows the detailed results for the 12 different poses in each individual box, with the tilt on the horizontal axis and the slant on the vertical axis. Notice that the scatter is generally elongated in the slant direction, which means that the variance in setting the orientation in depth is larger than setting the orientation in the frontoparallel plane. Overall, most of the bias and variance are in the slant results and little of them in the tilt. Quantitatively, this point is illustrated in Table 3, which gives the average and ranges of the circular standard deviation of the tilt and the slant settings. The decomposition in tilt and slant was calculated by projecting the vectors onto the horizontal x–z and vertical y–z planes, respectively, and then calculating the circular standard deviation in that plane (see the Appendix).

The tilt standard deviations are smaller than those for slant, and their mean ratio indicates that the standard deviations for slant are on average more than twice as large as those for tilt. These results are consistent with the difference in slant and tilt judgments in the estimation of local surface orientation as investigated by Koenderink et al. (1992) and Norman, Todd, and Phillips (1995). We analyze the tilt and slant results separately.

Tilt: Orientation in the plane. A three-way analysis of variance (ANOVA), with factors orientation (12), shape (2), and cue (4), was performed on the tilt mean error and the tilt standard deviation, separately.

The tilt mean error depended only on the orientation [$F(11,192) = 2.6$, $p < .01$],with no significant interactions. A detailed two-way ANOVA with tilt and slant as factors indicated that the tilt mean error depended only on the tilt [$F(3,24) = 8.4, p < .01$]. However, there seems to be no clear pattern in the tilt mean results that would enable us to interpret this result.

The tilt standard deviation depended on orientation [$F(11,192) = 6.8, p < .01$] and shape [$F(1,192) = 137.3$, $p < .01$], with no significant interactions. The shape had a rather large effect, and Table 4 shows that the standard deviations for the flat spheroid were roughly twice as large as those for the long spheroid. The main difference

Table 5
Average Cardinal and Oblique Standard Deviations in the Tilt Condition and Their Ratios

	Long Spheroid			Flat Spheroid		
Observer	Cardinal	Oblique	Ratio	Cardinal	Oblique	Ratio
B.L.	1.2°	1.5°	1.3	2.6°	3.1°	1.2
J.T.	0.9°	1.2°	1.3	1.4°	2.2°	1.6
T.D.	0.9°	1.5°	1.7	1.7°	2.9°	1.7

Table 6
Average Slant Standard Deviations for the Long and Flat Spheroid and Their Ratios

Observer	Long Spheroid	Flat Spheroid	Ratio
B.L.	3.4°	4.6°	1.3
J.T.	2.4°	3.0°	1.2
T.D.	3.8°	4.9°	1.3

in the task was that for the long spheroid, the subjects had to estimate the tilt of the major axis of the projected ellipse, and for the flat spheroid, they had to estimate the minor axis. This might have been the cause of the effect.

A detailed two-way ANOVA with tilt and slant as factors revealed that the tilt standard deviation depended only on tilt [$F(3,24) = 10.3, p < 0.01$] and not on slant. We then pooled the standard deviations in the cardinal (horizontal and vertical) and oblique (diagonal) orientations, as shown in Table 5, separated for the long and flat spheroid.

The standard deviations for the cardinal orientations are always higher than those for the oblique ones, as is indicated by the ratios that are larger than one. This is a clear indication of the oblique effect (Appelle, 1972), though with a smaller ratio than is usually reported. The main difference between our experiment and most of the research on this effect is that we asked the observers to judge the orientation of 3-D objects instead of 2-D patterns (lines, contours, or gratings). The ellipse that was the projection of a spheroid did not have a contour with the same tilt as the object; therefore the observer had to estimate the major or minor axis of the whole elliptical silhouette (*implicit orientation*).

Slant: Orientation in depth. Similarly, a three-way ANOVA, with the factors of orientation (12), shape (2), and cue (4), was performed on the slant mean error and slant standard deviation, separately.

The main factor for the standard deviation was shape of the spheroid [$F(1,192) = 29.1, p < .01$], although there were no interactions. Because of the influence of shape, the subjects were more precise for the long than for the flat spheroid. Table 6 shows the average standard deviations and their ratios. The ratios were not as high as for the tilt standard deviations, which were around two (Table 4), but, similar to tilt, the slant of the stick was seen with more precision than the slant of the slab.

The most interesting results are in the slant mean error. The factors that influenced performance were orientation [$F(11,192) = 20.8, p < .01$], shape [$F(1,192) = 226.0, p < .01$], and cue [$F(3,192) = 7.1, p < .01$]; all their in-

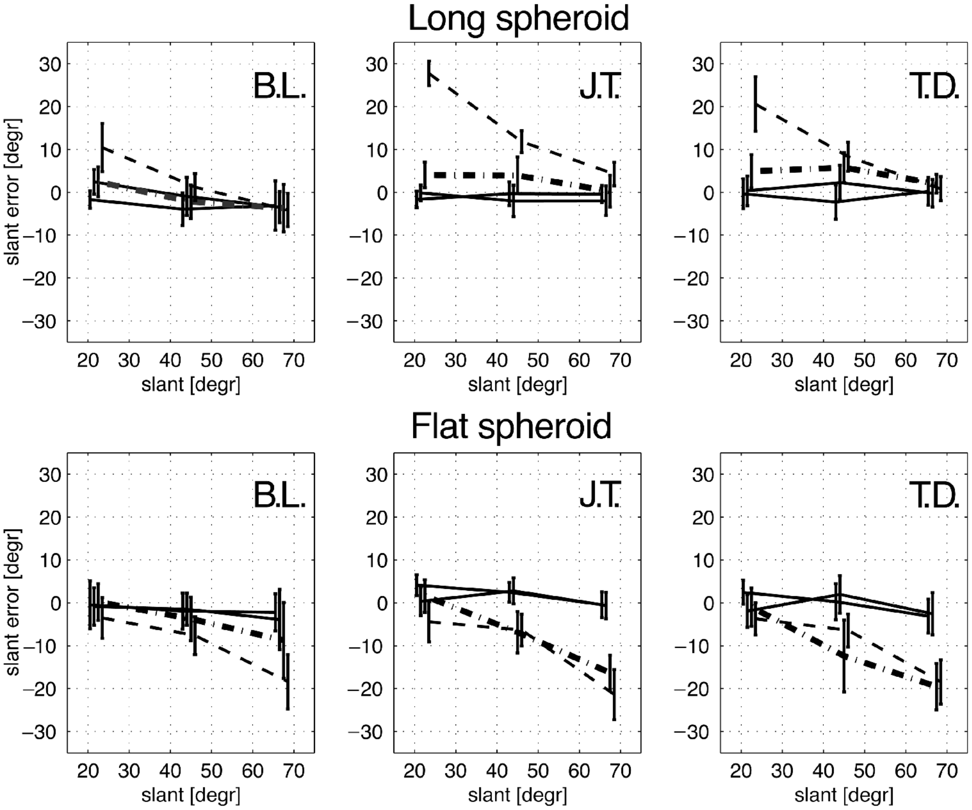

Figure 5. The results for the slant settings for the long spheroid (top row) and the flat spheroid (bottom row) for the 3 observers, B.L., J.T., and T.D. The solid lines represent the texture condition, the dashed-dotted line, the dynamic silhouette condition, and the stippled line, the static-silhouette condition. The error bars are standard deviations.

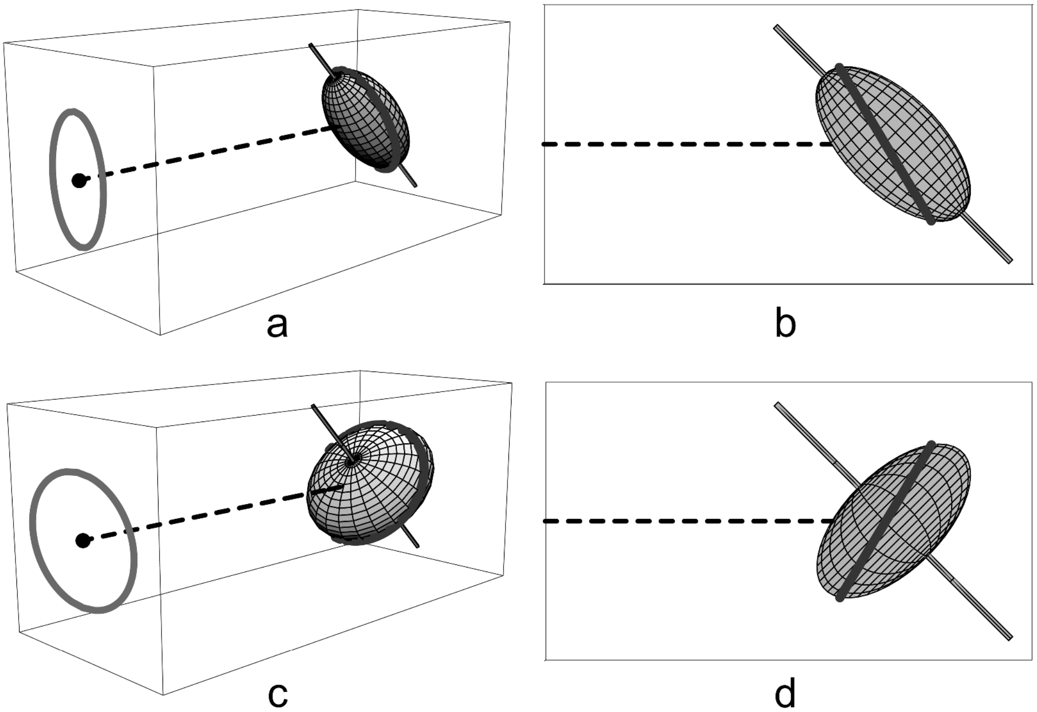

Figure 6. Representation of the long and flat spheroids with slants of 45º, and the rim represented as the thick line on the surface of the spheroid. (a) Generic view of the long spheroid. (b) Side view of long spheroid. (c) Generic view of the flat spheroid. (d) Side view of the flat spheroid.

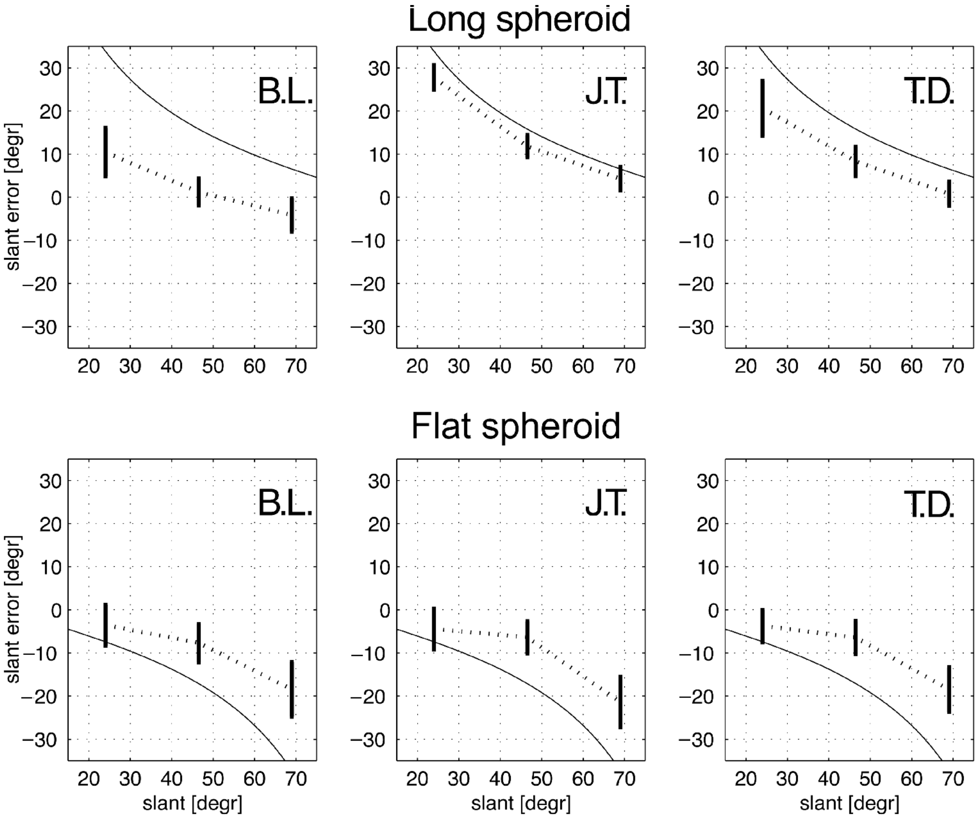

Figure 7. Replot of the results for the slant settings for the long spheroid (top row) and the flat spheroid (bottom row) in the static-silhouette condition (stippled line), with the orientation of the rim plotted as a solid line.

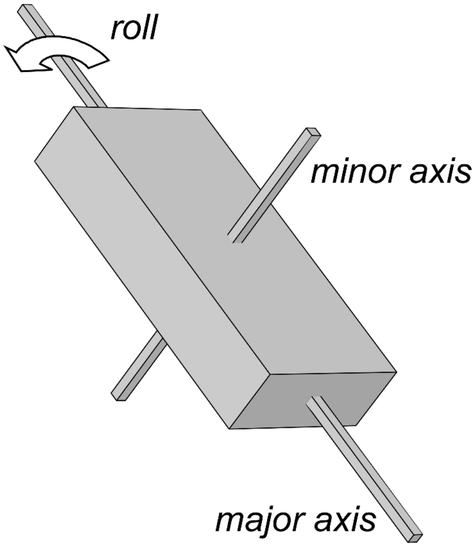

Figure 8. A rectangular box with aspect ratios of 4:2:1, with the major axis and the minor axis pinching through the length and width dimensions, respectively. The arrow denotes the angle of the minor axis (roll) around the major axis.

teractions were statistically significant. A detailed analysis revealed that tilt had no influence on slant mean error but only on the slant [$F(2,24) = 93.0, p < .01$]. We therefore averaged over all the tilts and plotted the results as the slant mean error, depending on the slant of the object for the different cue combinations (Figure 5). The results are separated for the 3 subjects (columns) and for the long and the flat spheroids (rows).

The solid lines in Figure 5 represent the results in the two textured conditions, and the biases are both indistinguishable from zero and from each other. The thick dashed-dotted line is the dynamic silhouette condition, where a difference emerges between the long and flat spheroids; for the long spheroid, the slant is slightly overestimated, whereas for the flat spheroid, the slant is underestimated, additionally following the trend that the larger the slant, the larger the error. In the static silhouette condition (thin dashed line in Figure 5), this effect is shown most dramatically, and all the observers show an overestimation for the long spheroid and an underestimation for the flat spheroid. Furthermore, there are clear trends in both cases; for the long spheroid, the smaller the slant, the larger the error, and for the flat spheroid, the larger the slant, the larger the error.

What can the explanation be for the large effect in the static silhouette condition? Let us consider the geometry of the situation in the static silhouette condition. The rim of spheroids that project to the contour of the silhouette is an ellipse (planar curve), but its orientation is neither frontoparallel nor that of the spheroid.

Figures 6a and 6b show a long spheroid from a generic (cyclopean) viewpoint and a side view, where the rim is indicated by the thick curve on the surface. The slant of the spheroid is 45°, whereas the slant of the long axis of the elliptical rim is 62° and therefore larger. Figures 6c and 6d provide the same views of a flat spheroid with a slant of 45°. In this case, the rim has a slant of 119°, but because we take the orientation of the minor axis, we take the angle perpendicular to the rim (minus 90°), which is 29° and therefore smaller than the slant of the spheroid. The orientation ρ of the rim changes with the slant, σ, of the spheroid, and the relation is

$$\tan\rho = \frac{(S^2 - 1)\sin\sigma\cos\sigma}{\cos^2\sigma + S^2\sin^2\sigma},$$

with S as the aspect ratio of the spheroid (Oomes, 1998). This dependence of rim slant on object slant and aspect ratio was calculated for the cyclopean eye. This results in a situation in which the slant of the rim deviates from the spheroid for small angles in the case of the stick and for large angles in the case of the slab. Figure 7 shows a replot of the data in the static silhouette condition (stippled lines). The added solid line is the difference between the slant of the rim and the spheroid. As can be seen, the rim orientation predicts the direction of the effect (over- or underestimation) and the direction of the trend, but not the magnitude of the effect.

In the static silhouette condition, the observers are biased toward the orientation of the rim, but they are not fooled in the sense that they confuse it with the orientation of the object itself.

EXPERIMENT 2
Symmetric Objects

Though the categorization of global shape into sticks, lumps, and slabs is very useful, it is not complete. In general, objects have different length, breadth, and width dimensions. The assumption in Willats's (1992) scheme is that at least two of them are roughly the same.

The rectangular box in Figure 8 is an example of a generic object with different aspect ratios (in this case, 4:2:1). It also shows the major and minor axes of the object as thin sticks that pinch through the length and width dimensions. These principal axes can be considered as denoting the orientations of the "stick-like" and "slab-like" aspects of the object. The pose of the object is expressed as the familiar slant and tilt of the major axis and the roll of the minor axis. The roll is the angle of the minor axis around the major axis. Together, these three angles completely describe the pose of the object.

The question was whether observers can estimate the pose of more generic objects. For simplicity, we chose a set of eight objects that had either two or three planes of bilateral symmetry. They were constructed to have aspect ratios that were roughly 4:2:1, so that they were approximately in between sticks (4:1:1) and slabs (4:4:1) on the shape continuum.

Method

Apparatus. We used the same setup as that in Experiment 1, with the only difference being that the viewing distance was 114 cm, which resulted in a field of view of 17.5° × 14.0° visual angle.

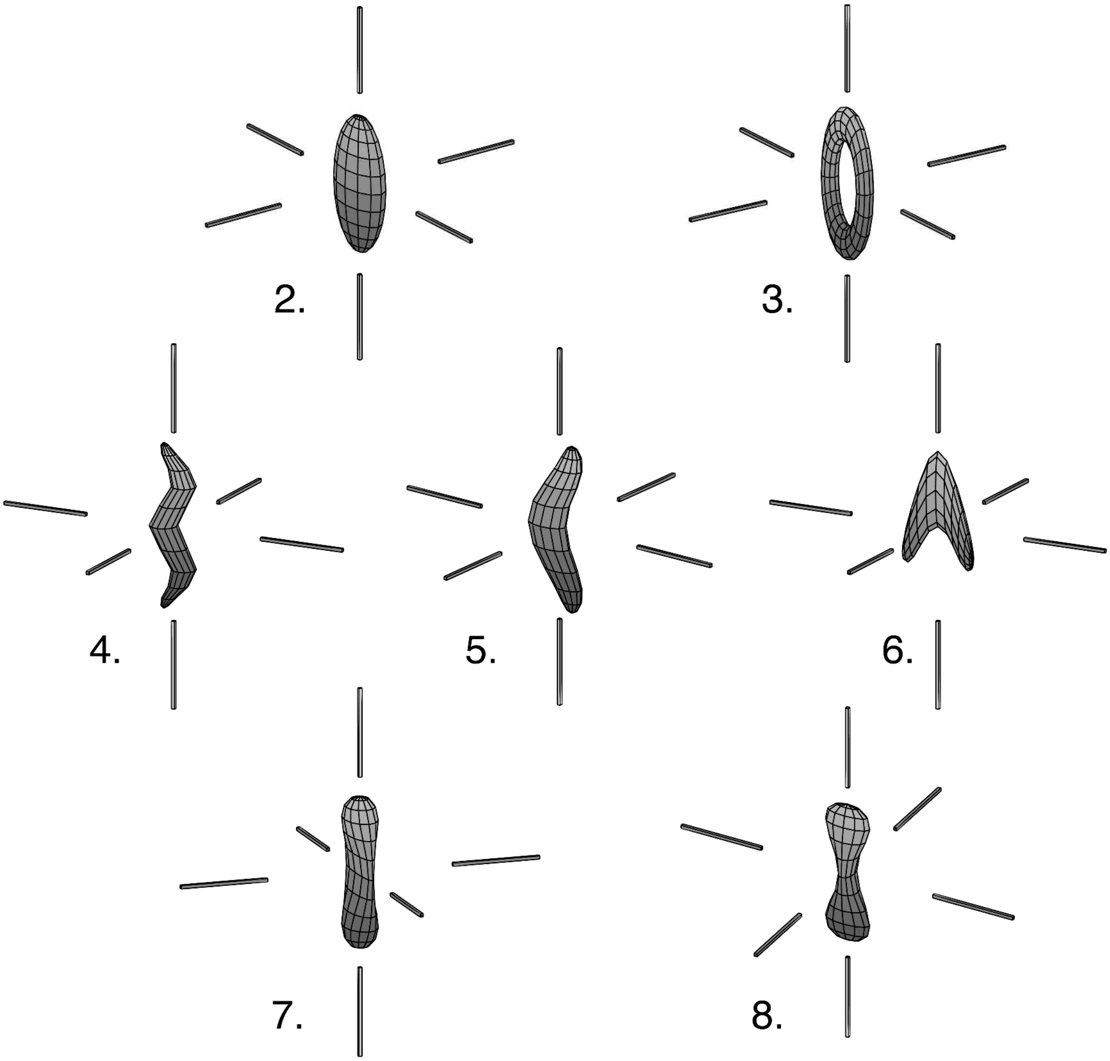

Figure 9. The other seven objects used in the experiment seen from a canonical viewpoint, surrounded by the cross (three perpendicular axes). A mesh is added to the surfaces for visualization purposes.

Stimuli. Eight objects were used in the experiment, and their shapes had either two or three planes of bilateral symmetry. Figure 8 shows the box, and Figure 9 shows the other objects, with the cross axes in the principal directions. Table 7 lists the names, dimensions, and aspect ratios. The stick-like aspect ratio is defined as $r_1 = A/\sqrt{BC}$ and the slab-like aspect ratio as $r_2 = C/\sqrt{AB}$.

The cross consisted of three orthogonal axes with a square cross-section. Each axis had a width of 0.25 cm and a length of 29.0 cm, from which the middle 15.0 cm was deleted.

All the objects were shown in 15 different orientations. The orientations of the major axis were the five (tilt, slant) combinations: (45.0°, 16.4°), (16.4°, 45.0°), (81.8°, 45.0°), (30.0°, 81.8°), and (60.0°, 81.8°). The angle around the major axis was the roll and had values of 26.7°, 86.7°, and 140°. Together, this gave 15 viewpoints of the eight objects, which resulted in 120 stimuli. In one session, the subjects were presented with all the stimuli from all viewpoints in random order. All subjects performed in all eight sessions. As an example, Figure 10 shows all presented views of the bow tie (Object 8).

The objects and axes were shown in perspective projection and in stereo. The viewing distance was 114 cm, and the distance between the eyes was taken to be 6.1 cm. For the lighting, we used a parallel beam with an oblique (upper left frontal) direction. For the light color, we picked white and for the background, an intermediate gray. The surface of the objects was visualized with Gouraud shading and was an orange-yellow, whereas the axes were colored red, cyan, and blue.

Table 7
Dimensions (in Centimeters) and Aspect Ratios of the Objects

Number	Object	A	B	C	r_1	r_2
1	box	11.2	5.6	2.8	2.8	0.35
2	ellipsoid	12.0	6.0	3.0	2.8	0.35
3	torus	12.3	6.2	1.4	4.2	0.16
4	worm	13.2	3.2	2.2	5.0	0.31
5	boomerang	13.6	4.7	1.8	4.7	0.21
6	spaceship	9.2	6.3	1.4	3.1	0.17
7	squeeze	12.0	6.0	1.9	3.6	0.26
8	bow tie	12.0	2.9	2.0	5.0	0.28

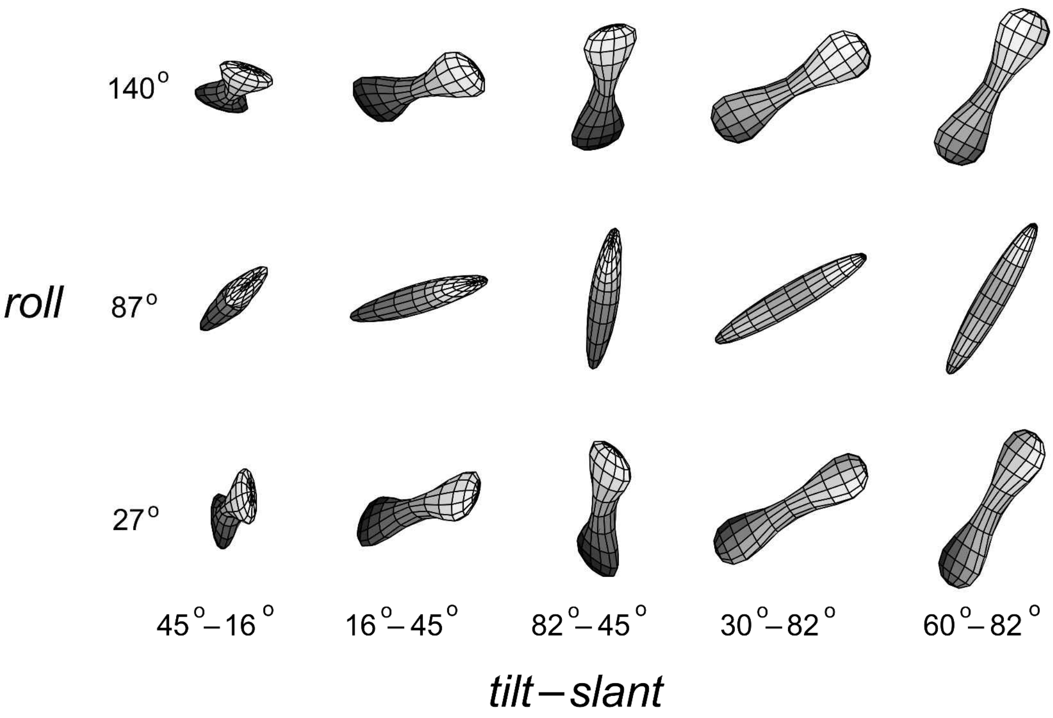

Figure 10. The bow tie as seen from all the 15 viewpoints. A mesh is added to the surfaces for visualization purposes.

The object was always shown in the center of the screen. A chinrest was placed so that the center of mass of the object was at eye level. At the bottom center of the screen, the trial number was presented. The light in the room was dimmed in order to avoid glare on the screen.

Procedure. Every session consisted of 120 trials, and the subject had to align the cross that consisted of three orthogonal axes with the principal axes of the object. The red axis was to be matched with the major axis of the object and the blue axis with the minor axis. The cross could be rotated around the origin by pressing the left mouse button and moving the mouse in the same way as described in Experiment 1. By pressing the right mouse button, the red axis stayed in the same orientation, and the other two axes could be rotated around the red one to align the minor axis. The subject was allowed to go back and forth between the two modes of operation until all the axes were aligned satisfactorily. By pushing the space bar on the keyboard, the subject confirmed the setting and was shown the next trial. There was no time pressure, and it took the observers about 45 min to complete each session. They were given one practice session in order to familiarize them with the objects and subsequently performed eight sessions.

Observers. Of the 4 observers, B.L. and S.P. had extensive psychophysical experience, whereas H.H. and T.O. had little or no experience. All observers were myopic and had corrected-to-normal visual acuity.

Results and Discussion

First, we give an overview of the data in which the responses for the different viewpoints are all rotated to the same orientation on the unit sphere. Figure 11 shows a typical cluster of results for Observer T.O. for the torus (Object 3). Figure 11a shows the results for the major axis with the tilt along the horizontal axis and the slant along the vertical axis. Consistent with Experiment 1, there was more variance in the slant than in the tilt.

Figure 11b shows the scatter for the minor axis. The variance in the vertical direction was due to a misalignment in the major axis, whereas the variance in the direction along the equator was the variance in roll proper. Note that most of the scatter is in the roll. By using the same method as that in Experiment 1, we decomposed the results in tilt, slant, and roll.

Table 8 shows that the average tilt standard deviations are smaller than the average slant standard deviations, which in turn are smaller than the average roll standard deviations. As in Experiment 1, the mean slant–tilt ratios (Table 9) are more than twice as large, and the roll–tilt ratios are even higher. The difference between tilt on the one hand, and slant and roll on the other, is the difference between the orientation in the plane and the orientation in depth. In this respect, observers behave similarly for the more generic symmetric shapes as for the spheroids. The orientation in the plane could be more reliably estimated than the orientation in depth.

Table 8
Averages and Ranges of the Circular Standard Deviations of Tilt, Slant, and Roll

Observer	Tilt	Slant	Roll
B.L.	1.3° (0.5°–3.4°)	3.0° (0.7°–8.6°)	4.1° (0.7°–12.2°)
H.H.	1.1° (0.4°–3.1°)	2.5° (0.9°–5.2°)	4.3° (0.5°–12.7°)
S.P.	0.8° (0.2°–2.4°)	1.7° (0.5°–4.6°)	2.7° (0.5°–8.3°)
T.O.	0.9° (0.3°–2.7°)	3.1° (0.6°–8.9°)	3.7° (0.3°–10.7°)

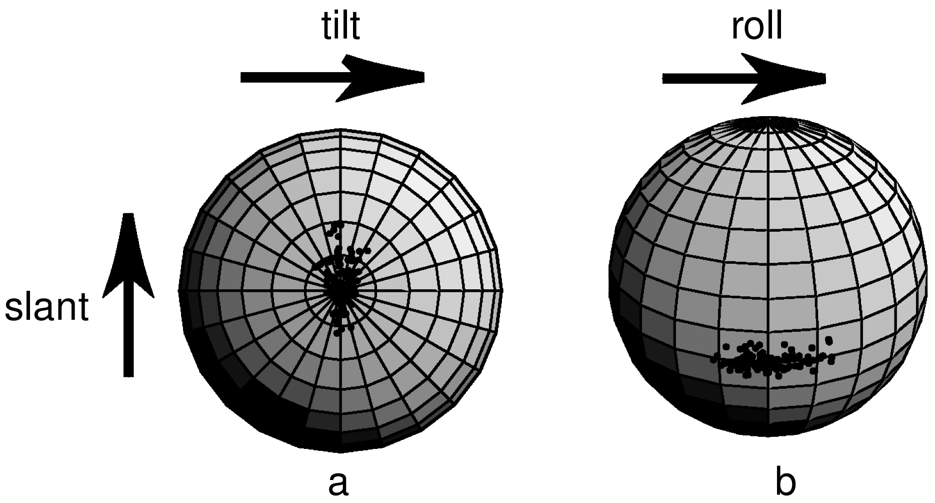

Figure 11. Responses from Observer T.O. for the torus, rotated to the same point on the unit sphere with the results for the major axis (a) and the minor axis (b).

We investigated whether the standard deviations were rank correlated with the object aspect ratios r_1 and r_2. One would expect that a more stick-like object would lead to higher precision in the slant (e.g., worm, Object 4) and that a more slab-like object (e.g., space ship, Object 6) would show higher precision in the roll. Surprisingly, we found a statistically significant Pearson rank correlation only between the stick-like aspect ratio r_1 and the tilt standard deviation [B.L., $-0.81, p < .01$; H.H., $-0.85, p < .01$; S.P., $-0.76, p < .05$; T.O., $-0.76, p < .05$]. The higher the aspect ratio, the smaller the tilt standard deviation, so the more an object looked like a stick, the higher the precision for estimating the orientation in the frontoparallel plane. That we did not find significant rank correlations for slant and roll was probably due to local shape properties that influenced the performance.

GENERAL DISCUSSION

We investigated the ability of human observers to perceive the pose of 3-D objects. They were asked to match the pose of an object with the pose of a cross, which consisted of three perpendicular axes with the middle part removed. Either the object or the cross was static; the observer could actively rotate the other shape around the common center of mass until all axes were aligned satisfactorily. For objects, we used long and flat spheroids (Experiment 1) and symmetric objects such as a box, an elongated torus, and a bow tie (Experiment 2).

The results show that the observers could do the task with a precision of a few degrees (less than 10°), though accuracy could be low, with biases as extreme as 30°. Besides establishing general performance, the goal of the these exploratory experiments was to determine some factors that influence pose estimation. We discuss these factors and the (possible) mechanisms by which they operate.

Pose

One of the main results was that precision depended on the pose of the object. The precision for the orientation in the frontoparallel plane (tilt) was consistently higher than that for the orientation in depth (slant, roll). This is analogous to the results in the perception of local surface orientation (Koenderink et al., 1992; Norman et al., 1995). As far as we know, this result has not been reported for the perception of object pose before.

It seems that the tilt could be inferred directly from the global orientation of the projected shape, although additional cues were necessary to infer slant and roll. Since, in the case of spheroids, tilt judgments were independent of the different cue combinations, a static silhouette was sufficient for performing the task. It is unclear at this point whether the observers used the contour or the silhouette to perform the tilt judgments. What is clearly the case for the tilt results is the much reported oblique effect, in which cardinal orientations (horizontal and vertical) lead to higher performance than do oblique ones. We have replicated this effect in Experiment 1. This has not been reported for 3-D objects before. In contrast to orientation in the plane, the orientation in depth was much harder to estimate, and performance highly depended on the available cues. We discuss the influence of shape, motion, shading, and texture in more detail below.

Another way to look at the decomposition in the frontoparallel and depth dimensions is to do so in terms of frames

Table 9
Mean Ratios of the Tilt, Slant, and Roll Standard Deviations

Observer	Slant–Tilt Ratio	Roll–Tilt Ratio	Roll–Slant Ratio
B.L.	2.7	3.5	1.7
H.H.	2.7	4.6	1.9
S.P.	2.8	3.7	1.9
T.O.	3.9	4.6	1.5

of reference. The world-centered frame is aligned with the gravitational vertical; the object-centered frame is aligned with the principal axes; and the viewer-centered frame is aligned with the observer's head or direction of gaze. By design, the world- and head-centered frames were aligned in these experiments, so we cannot determine whether the decomposition in slant and tilt was caused entirely by the viewpoint or was partly due to the orientation relative to the world (room, monitor). There is some evidence that the latter was influential. Cuijpers, Kappers, and Koenderink (2001) reported that in setting two stick-like probes parallel that are at different positions in space, the orientations aligned with the room were much easier than other orientations. This *slant oblique effect* resulted in biases around 2° for orientations parallel to the walls and could go up to 45° for oblique orientations. Unlike the objects used here, which showed the same center of mass, those in Cuijpers et al. were some distance apart.

Shape

The shape of the object had a clear influence on the precision of pose estimation. According to the categorization of Willats (1992), the pose of sticks is easier to estimate than the pose of slabs. This can even be true within one object that has both a stick-like orientation in depth (slant) and a slab-like orientation in depth (roll). Willats's claim that slabs are not recognized from silhouettes was not applicable in our experiments. Our observers always had information in stereo and never confused long and flat spheroids, which might have shown up in the results as a misalignment in the tilt of around 90°.

Motion

The dynamic silhouette condition in which the motion was under the control of the observer led to almost veridical results for the long spheroid and to an underestimation of the slant of the flat spheroids. Especially surprising was that the slant of the long spheroid could be seen from the moving silhouette. The observers themselves were in control of the motion, and this active vision loop provided enough information to do the task. It is not clear why the same cues were not helpful in the case of flat spheroids. Present theories of structure from motion cannot explain these results, because there were no identifiable points on the surface of the spheroid that could be tracked (Norman, Dawson, & Raines, 2000). The rim slid over the surface when the spheroid was rotated in depth, so the points on the contour could not be used to establish correspondence. This is a major challenge for future theories of object structure from motion.

Estimating the orientation of rotation axes results in a much lower performance (Norman & Todd, 1994; Pollick et al., 1994) than does judging orientation of object axes. It seems that orientation axes are more fundamental than rotation axes. It might be the case that the visual system uses the estimation of the orientation of a rotating object to infer the rotation axis.

Shading and Texture

Spheroids that were visualized with a textured surface led to the highest performance, whether the object was moving or not. Texture was apparently a rich source of information for the task of pose estimation. It helps to form a percept of the surface shape by being a carrier for the stereo information, which is sufficient to infer the pose of the object. It is clear that without the diffuse shading on the objects in Experiment 2, the subjects would not have been able to do the task with the same performance they had shown. The static silhouette condition in Experiment 1 gives ample evidence of this fact. In this condition, we found large biases, which we could explain by assuming that the subjects were biased by the orientation of the rim of the spheroid. This suggests that in the static case, some surface information in the form of shading and/or texture was necessary.

The sources in the literature on object-pose estimation are very sparse, but the present exploratory experiments have shed some light on the phenomenon. Some factors have been identified, though much detailed work remains to be done to understand the exact mechanisms involved. A challenge has emerged: to develop a theory of object perception that can explain the process of pose estimation in a way that is consistent with the present results.

REFERENCES

Ames, A., Jr. (1951). Visual perception and the rotating trapezoidal window. *Psychological Monographs*, **65**(7), 1-32.

Annis, R. C., & Frost, B. (1973). Human visual ecology and orientation anisotropies in acuity. *Science*, **182**, 729-731.

Appelle, S. (1972). Perception and discrimination as a function of stimulus orientation: The "oblique effect" in man and animals. *Psychological Bulletin*, **78**, 266-278.

Braunstein, M. L. (1976). *Depth perception through motion*. New York: Academic Press.

Coppola, D. M., Purves, H. R., McCoy, A. N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences*, **95**, 4002-4006.

Cuijpers, R. H., Kappers, A. M. L., & Koenderink, J. J. (2001). On the role of external reference frames on visual judgements of parallelity. *Acta Psychologica*, **108**, 283-302.

Fisher, N. I. (1993). *Statistical analysis of circular data*. Cambridge: Cambridge University Press.

Gibson, J. J., & Gibson, E. J. (1957). Continuous perspective transformations and the perception of rigid motion. *Journal of Experimental Psychology*, **54**, 129-138.

Koenderink, J. J. (1984). What does the occluding contour tell us about solid shape? *Perception*, **13**, 321-330.

Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. L. (1992). Surface perception in pictures. *Perception & Psychophysics*, **52**, 487-496.

Mardia, K. V., & Jupp, P. E. (2000). *Directional statistics*. Chichester, U.K.: Wiley.

Marr, D. (1982). *Vision*. New York: W. H. Freeman.

Marr, D., & Nishihara, H. K. (1978). Representation and recognition of the spatial organization of three-dimensional shapes. *Proceedings of the Royal Society of London: Series B*, **200**, 269-294.

Norman, J. F., Dawson, T. E., & Raines, S. R. (2000). The perception and recognition of natural object shape from deforming and static shadows. *Perception*, **29**, 135-148.

Norman, J. F., & Todd, J. T. (1994). Perception of rigid motion in depth from the optical deformations of shadows and occlusion boundaries. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 343-356.

Norman, J. F., Todd, J. T., & Phillips, F. (1995). The perception of surface orientation from multiple sources of optical information. *Perception & Psychophysics*, **57**, 629-636.

Oomes, A. H. J. (1998). *Human visual perception of spatial structure: Symmetry, orientation, and attitude.* Unpublished doctoral dissertation, University of Nijmegen.

Pollick, F. E., Nishida, S., Koike, Y., & Kawato, M. (1994). Perceived motion in structure from motion: Pointing responses to the axis of rotation. *Perception & Psychophysics*, **56**, 91-109.

Robinson, J. A., McKenzie, B. E., & Day, R. H. (1996). Anticipatory reaching by infants and adults: The effect of object features and apertures in opaque and transparent screens. *Child Development*, **67**, 2641-2657.

Rogers, B. J., & Graham, M. (1983). Anisotropies in the perception of three-dimensional surfaces. *Science*, **221**, 1409-1411.

Switkes, E., Mayer, M. J., & Sloan, J. A. (1978). Spatial frequency analysis of the visual environment: Anisotropy and the carpentered environment hypothesis. *Vision Research*, **18**, 1393-1399.

Wanger, L. R., Ferwerda, J. A., & Greenberg, D. P. (1992). Perceiving spatial relationships in computer-generated images. *IEEE Computer Graphics & Applications*, **12**, 44-58.

Westheimer, G., & Beard, B. L. (1998). Orientation dependency for foveal line stimuli: Detection and intensity discrimination, orientation discrimination and Vernier acuity. *Vision Research*, **38**, 1097-1103.

Willats, J. (1992). Seeing lumps, sticks, and slabs in silhouettes. *Perception*, **21**, 481-496.

APPENDIX

We will briefly describe some descriptive measures in spherical and circular statistics. Details can be found in Fisher (1993) and Mardia and Jupp (2000).

A. Spherical Statistics

For n vectors with coordinates (x_i, y_i, z_i) the mean vector is

$$\bar{\boldsymbol{x}} = (\bar{x}, \bar{y}, \bar{z}) = \left(\frac{1}{n}\sum_{i=1}^{n} x_i, \frac{1}{n}\sum_{i=1}^{n} y_i, \frac{1}{n}\sum_{i=1}^{n} z_i \right),$$

with its length

$$R = |\bar{\boldsymbol{x}}| = \sqrt{\bar{x}^2 + \bar{y}^2 + \bar{z}^2}.$$

The normalized vector $\bar{\mathbf{x}}/R$ is the mean direction on the unit sphere. The mean angle with the z-axis is

$$\bar{\theta} = \arccos\left(\frac{\bar{z}}{R}\right).$$

The spherical standard deviation is

$$v = \sqrt{2(1-R)}\ .$$

B. Circular Statistics

For the slant, tilt, and roll, we modified the vectorial equations into equations applicable to our axial data. For n measurements of angle θ_i, the mean cosine and sine factors are

$$C = \frac{1}{n}\sum_{i=1}^{n} \cos 2\theta_i,$$

$$S = \frac{1}{n}\sum_{i=1}^{n} \sin 2\theta_i\,,$$

with length

$$R = \sqrt{C^2 + S^2}\ .$$

The factor 2 allows the orientations to be distributed over the entire circle. The mean of the angles is

$$\bar{\theta} = \frac{1}{2}\arctan\left(\frac{S}{C}\right),$$

and the axial circular standard deviation is

$$v = \sqrt{-\frac{\log R}{2}}\ .$$

(Manuscript received March 3, 2000;
revision accepted for publication July 18, 2001.)