Shape constancy and depth-order violations in structure from motion: A look at non-frontoparallel axes of rotation

Julian M. Fernandez   Institute for Sensory Research, Syracuse University, Syracuse, NY, USA

Bart Farell   Institute for Sensory Research, Syracuse University, Syracuse, NY, USA

Humans can recover the structure of a 3D object from motion cues alone. Recovery of structure from motion (SFM) from the projected 2D motion field of a rotating object has been studied almost exclusively in one particular condition, that in which the axis of rotation lies in the frontoparallel plane. Here, we assess the ability of humans to recover SFM in the general case, where the axis of rotation may be slanted out of the frontoparallel plane. Using elliptical cylinders whose cross section was constant along the axis of rotation, we find that, across a range of parameters, subjects accurately matched the simulated shape of the cylinder regardless of how much the axis of rotation is inclined away from the frontoparallel plane. Yet, we also find that subjects do not perceive the inclination of the axis of rotation veridically. This combination of results violates a relationship between perceived angle of inclination and perceived shape that must hold if SFM is to be recovered from the instantaneous velocity field. The contradiction can be resolved if the angular speed of rotation is not consistently estimated from the instantaneous velocity field. This, in turn, predicts that variation in object size along the axis of rotation can cause depth-order violations along the line of sight. This prediction was verified using rotating circular cones as stimuli. Thus, as the axis of rotation changes its inclination, shape constancy is maintained through a trade-off. Humans perceive the structure of the object relative to a changing axis of rotation as unchanging by introducing an inconsistency between the perceived speed of rotation and the first-order optic flow. The observed depth-order violations are the cost of the trade-off.

Keywords: structure from motion, kinetic depth effect, affine structure, shape perception

Introduction

Humans can recover 3D structure from the projected 2D motion field of a rotating object. There are many ways to rotate an object, but structure from motion (SFM) has been studied almost exclusively only when the axis of rotation lies in the frontoparallel plane. Perceived shape in this condition is usually nonveridical. Previous studies indicate that SFM is recovered from the first-order velocity field, which implies that shape is recoverable only up to a scaling factor in depth (Norman & Todd, 1993; Todd, 1998; Todd & Norman, 1991; Werkhoven & van Veen, 1995). Hence, accuracy is low for judgments requiring veridical perception of Euclidean metric structure, such as judgments of lengths or angles (Braunstein, Liter, & Tittle, 1993; Cornilleau-Peres & Droulez, 1989; Eagle & Blake, 1995; Hogervorst, Kappers, & Koenderink, 1993; Liter, Braunstein, & Hoffman, 1993; Norman & Lappin, 1992; Todd & Bressan, 1990). Conversely, accuracy is high for judgments of an object's affine structure, such as depth order between pairs of points, parallelism between lines defined by pair of points on the surface, and coplanarity

among points (Braunstein et al., 1993; Eagle & Blake, 1995; Hogervorst et al., 1993; Liter et al., 1993; Tittle, Todd, Perotti, & Norman, 1995; Todd & Bressan, 1990). However, violations of affine structure have also been reported (Domini & Braunstein, 1998).

The trajectories of points on a surface rotating about a frontoparallel axis are linear and parallel to one another under orthographic projection. However, the trajectories of points on a surface rotating about an axis slanted out of the frontoparallel plane are curved and do not bear a simple geometrical relationship to each other. (We measure slant as the angle of a line or plane relative to the frontoparallel plane. In what follows, we use the words "slant" and "inclination" interchangeably.) This additional complexity is perhaps one of the reasons why this latter condition is rarely studied. One exception is the work of Loomis and Eby (1988), who studied this condition using elongated ellipsoids. They found that perceived depth becomes a smaller fraction of the simulated depth as the angle of rotation deviates from the vertical.

In this article, we investigate in more detail the ability of humans to recover SFM for an object rotating about an axis slanted out of the frontoparallel plane. In Experiment 1, we

measure subjects' recovery of the depth structure of a rotating cylinder. In Experiment 2, we measure subjects' recovery of the angle of inclination of the axis of rotation. From these two experiments, we find that the perceived structure is not affine. In Experiment 3, we measured departures from affinity using an alternative method that is independent from that used in the first two experiments. Also, we show that for shapes satisfying specific conditions, this violation of affinity results in perceived depth-order violations.

Results

Experiment 1

Subjects viewed a motion-defined elliptical cylinder textured with random dots (Figure 1a). They then viewed a cylindrical cross section, which they adjusted to match the cross-sectional profile of the motion-defined cylinder through a plane normal to the rotational axis (Figure 1b).

We define curvature (C) as the cylinder's depth-to-width ratio (i.e., $C = 1$ for a circular cylinder). Figure 2 shows how the subjects' perceived curvature (C_{obs}), measured by the adjustment procedure, compares with the simulated curvature (C_{sim}). The ratio C_{obs}/C_{sim} does not change significantly with the simulated angle of inclination θ_{sim} (Figure 2a). The average ratio for the four subjects fluctuates narrowly and unsystematically between 0.83 and 0.88. However, C_{obs}/C_{sim} does depend on the simulated curvature (Figure 2b). In all cases, cylinders were perceived as less curved than simulated, but the

Figure 1. Experiments 1 and 2. (a) Subjects viewed a motion-defined elliptical cylinder textured with random dots. (b) Experiment 1: Subjects adjusted a cylindrical cross section to match the profile of the motion-defined cylinder (a). (c) Experiment 2: Subjects adjusted the inclination of a bar to match the inclination of the motion-defined cylinder (a).

difference was greatest for flattened cylinders ($C_{sim} < 1$). A two-way repeated measures ANOVA revealed a statistically significant effect of the simulated curvature on C_{obs}/C_{sim}, $F(4, 60) = 15.87$, $p < .0001$. The test revealed no statistically significant effect of the simulated angle of inclination on C_{obs}/C_{sim} or of the interaction between simulated angle of inclination and simulated curvature ($p > .05$ in both cases). Thus, the perceived shape of a simulated object relative to the axis of rotation does not change with the simulated inclination of the axis of rotation. However, this perceived shape is, in general, nonveridical and the extent of the difference between the perceived and the simulated shapes is a function of the simulated shape itself.

Results shown in Figures 2a and 2b are averages of four different conditions combining short and sustained dot lifetimes and occluded and unoccluded surface boundaries. The short dot lifetime eliminates potential dot-density cues, and the occluder eliminates potential boundary cues. Dot lifetime had little effect on performance (Figure 2c). The presence of an occluder had a small but noticeable effect, which might be due to the greater uncertainty about the cylinder's width when an occluder is present. Thus, if subjects judge the cylinder to be wider when its boundaries are occluded, then the cylinder's width-to-depth ratio will be increased and C_{obs}/C_{sim} will be reduced. However, the effect of the occluder is small. A two-way repeated measures ANOVA revealed a statistically significant effect of the occluder on C_{obs}/C_{sim}, $F(1, 6) = 45.02$, $p = .00053$, but no significant effect of dot lifetime or of interactions between dot lifetime and occluder.

The important conclusion from Figure 2c is that subjects' performance reflects pure motion cues, not dot-density or boundary cues. The use of short dot lifetimes eliminates density cues but can, in principle, add a new cue. The number of dots appearing and disappearing at each new frame due to the finite lifetime of the dots is the same for all frames, but different regions of the projected image have different ratios of appearing to disappearing dots, and these ratios change across frames. In principle, then, these ratios could provide an indication of the density variation that would be present in a sustained dot lifetime condition. However, we believe it to be unlikely that subjects could make use of this new cue: The number of appearing and disappearing dots at different positions in the image plane is perceptually unrelated to changes in density from subjects' previous experience with physically 3D objects, a consideration that makes both prior learning of this cue and evolutionary pressure to take advantage of it improbable.

For rotations around an axis in the frontoparallel plane, our result that perceived shape is not veridical is in agreement with results of previous studies: Shape was recoverable only up to a scaling factor in depth. It is currently unknown how this scaling factor is computed. We will come back to this issue in the Discussion section. For now, we just want to mention that in pilot experiments (data not shown) we found that a change in the simulated
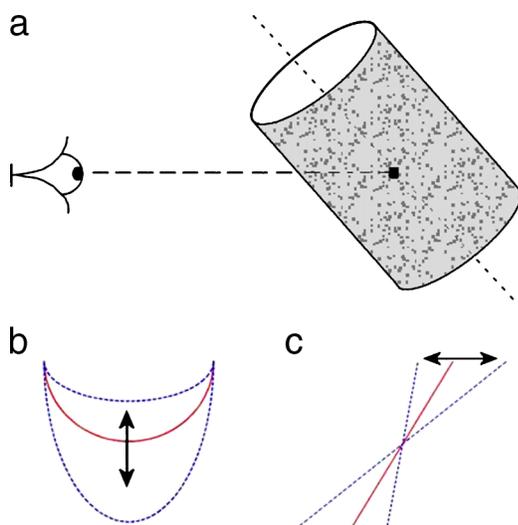
speed of rotation does not produce a change in the adjusted shape. The minimum velocity was always zero; hence, curvature was not estimated from a difference in the ratio of maximum and minimum speeds. But the change over time in the optic-flow field does provide cues about shape. For instance, one cue is the direction of the change in position of the point of maximum projected

velocity: If this point moves in the same direction as the cylinder, then the object has $C > 1$; if it moves in the opposite direction, then the object has $C < 1$. However, we were unable to find a variable obtained from a simple combination of optic-flow parameters that could predict the observed change in the scaling factor with curvature.

Our data show that recovered depth relative to the axis of rotation does not depend on the angle of inclination (Figure 2a). Thus, the ratio between perceived and simulated depth *along the line of sight* does change with the angle of inclination (see Equation 1). This is consistent with the results of Loomis and Eby (1988). Recovering the depth along the line of sight is the same as recovering the distance of the surface relative to the axis of rotation only when this axis is frontoparallel. To account for depth recovery in the general case, we need first to assess the perceived inclination of the angle of rotation. This will be done in the next experiment.

Experiment 2

In a second experiment, we assessed subjects' ability to recover the inclination of the axis of rotation. Subjects viewed circular cylinders with axes of rotation inclined over the range from 0° to 85° relative to the frontoparallel plane (Figure 1a). They then adjusted the orientation of a line to match the inclination of the perceived axis of rotation (Figure 1c).

Figure 3a shows the perceived angle as a function of the simulated angle. The inclination was always misperceived, the perceived angle being much smaller than the simulated angle. The relationship between perceived and simulated inclination is linear, with the slope varying across subjects. This result agrees with previous findings (Pollick, Nishida, Koike, & Kawato, 1994; see also Caudek & Domini, 1998), although our slopes are somewhat smaller, presumably due to differences in stimuli and procedure.

Now, the ratio between perceived and simulated depth d along the line of sight can be computed from our data as (see Equation A3)

$$d = (C_{obs}/C_{sim})(\cos\theta_{sim}/\cos\theta_{obs}), \tag{1}$$

where θ_{sim} and θ_{obs} are the simulated and perceived angle of inclination of the axis of rotation, respectively. Figure 3b shows this ratio, d, for the three subjects, as a
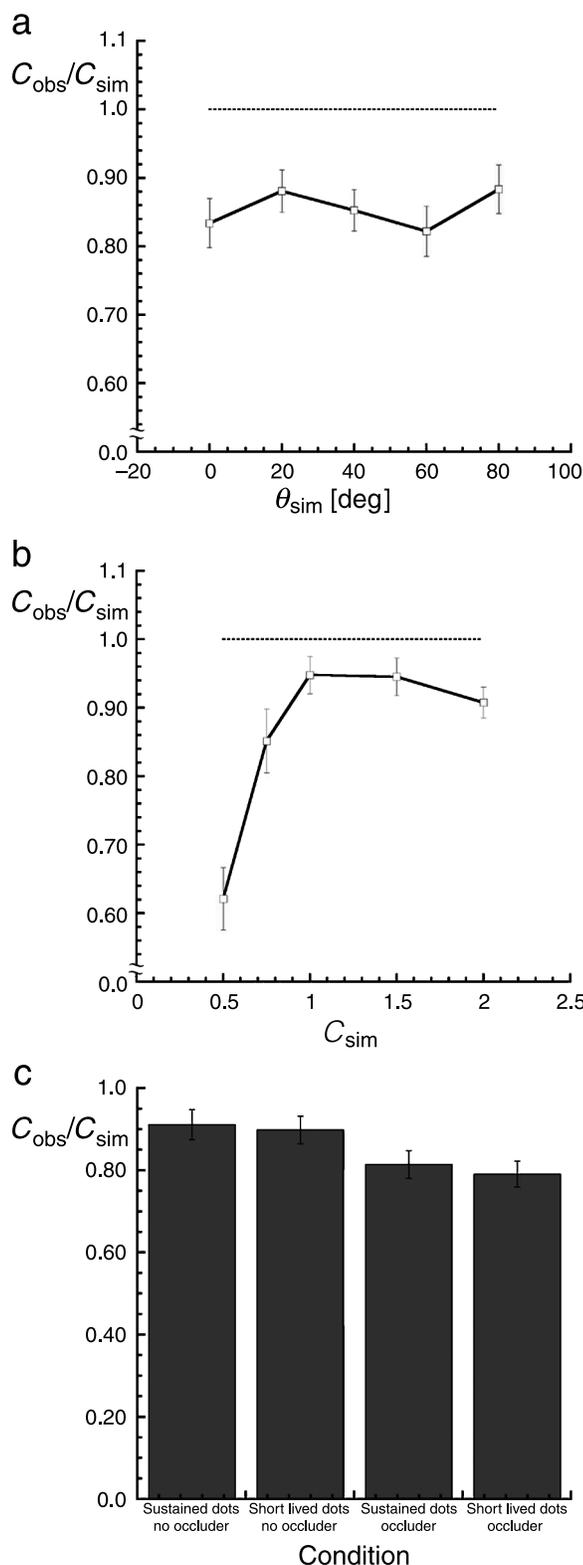

Figure 2. Recovered depth. (a) The ratio C_{obs}/C_{sim} between perceived and simulated curvature, averaged across the four subjects and across the different simulated curvatures, does not change with the simulated angle of inclination θ_{sim}. (b) C_{obs}/C_{sim}, averaged across the four subjects and across simulated angles, does depend on the simulated curvature. (c) C_{obs}/C_{sim} as a function of viewing condition (dot lifetime and presence of an occluder). In all panels, error bars are individual subject's standard error of the mean averaged across subjects and conditions.
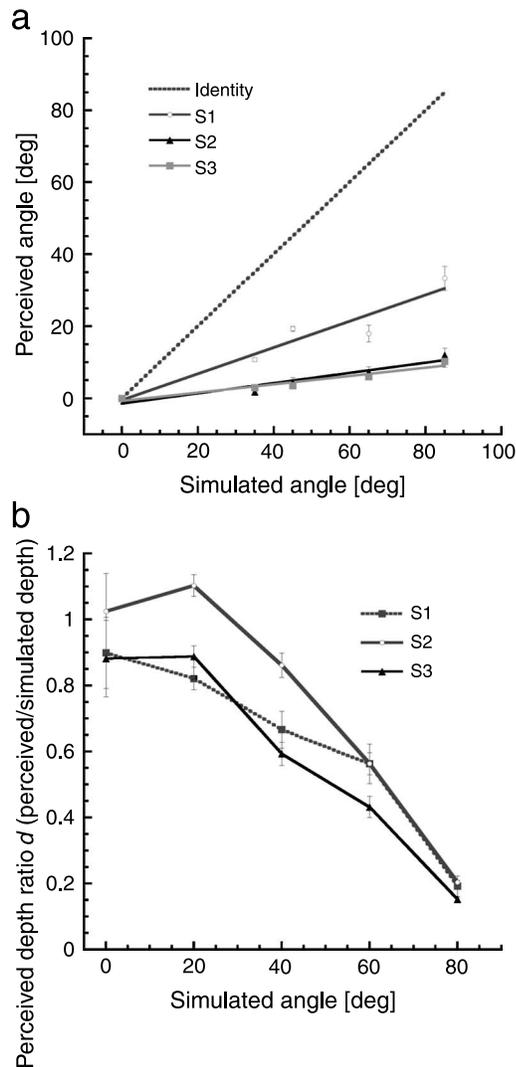
Figure 3. Effect of inclination of the axis of rotation. (a) Perceived inclination as a function of the simulated inclination. The relationship is linear, with perceived angle substantially lower than the simulated angle. (b) Ratio between perceived and simulated depth along the line of sight (*d*), as a function of the simulated inclination of the axis of rotation. In both panels, error bars are standard error of the mean.

function of θ_{sim}. Perceived depth becomes a smaller fraction of the simulated depth as the inclination of the axis of rotation deviates from the vertical. The results agree well with those from a previous study (Loomis & Eby, 1988) using elongated ellipsoids.

There is a major difference between rotations around a vertical axis and rotations around an axis inclined toward the subject. Vertical rotations can be considered a degenerate case of general rotations. They satisfy, under orthographic projection, the following relationship between relative depth and relative speeds (Fernandez, Watson, & Qian, 2002):

$$\Delta v = -\Omega\,\Delta Z/Z_0, \tag{2}$$

where Ω is the angular speed of rotation, Z_0 is the distance of the subject to the axis of rotation, and Δv and ΔZ are the differences in retinal velocity and depth, respectively, between any two points on the object. The instantaneous velocity field yields no information about Ω and, as a consequence, object shape is recoverable only up to a scale factor in depth. The visual system is thus free to set Ω, which, in general, it does nonveridically, resulting in the nonveridical recovery of the object's shape.

When the axis of rotation is not in the frontoparallel plane, motion is no longer rectilinear, and the 3D structure of the object under orthographic projection must satisfy the following two relations:

$$\Delta Z = Z_0\Delta v_x/(\Omega\cos\theta) + Z_0\Delta y \tan\theta \tag{3}$$

and

$$\Omega = -\partial_x v_y/\sin\theta, \tag{4}$$

so that

$$\Delta Z = (-\Delta v_x/\partial_x v_y + \Delta y)Z_0\tan\theta, \tag{5}$$

where, as shown in Figure 4a, Z_0 is the distance to the object, Ω is the angular speed of rotation, θ is the
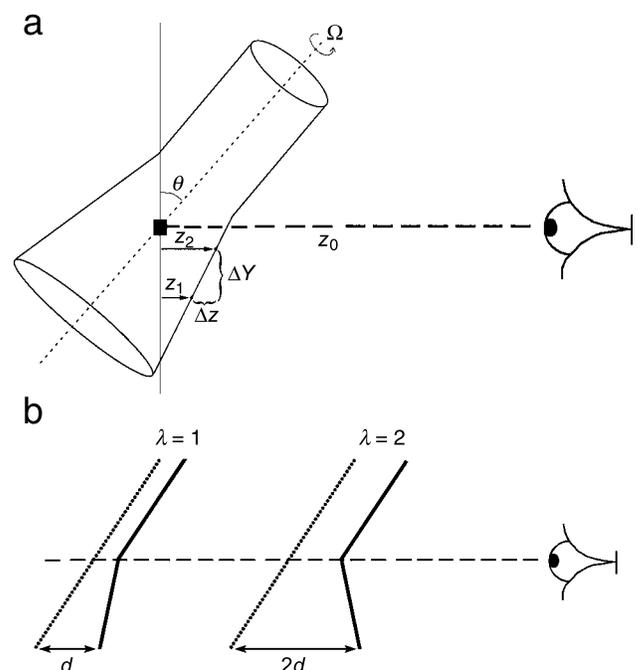

Figure 4. Nonaffine structure. (a) Schema showing the nomenclature utilized. (b) The effect of changing λ for a constant inclination of the axis of rotation. A change in λ results in a depth-order reversal of the bottom half of the object, as all the distances from the object to the axis of rotation (dotted line) double.

perceived angle of inclination of the axis of rotation ($0° < \theta < 90°$) from the frontoparallel plane, and $\partial_x v_y$ is the derivative with respect to x (horizontal direction) of the vertical component of the retinal velocity, v_y. It is assumed here that the axis of rotation lies within the sagittal plane passing through the eye. Equation 5, which implicitly assumes rigidity, gives the difference in depth ΔZ between any two points on the object from their differences in horizontal retinal velocity Δv_x and angular vertical position Δy.

Neither the inclination of the axis of rotation, θ, nor the speed of rotation, Ω, can be recovered from the instantaneous velocity field. However, using Equation 4, we can reduce the recovered shape from Equation 3 into a one-parameter family of recoverable shapes that differ by a scaling factor in depth (Equation 5). Thus, for inclined rotations, θ plays the role of the scaling parameter, in the same way as Ω played this role for vertical rotations.

Thus, our result showing that subjects do not perceive θ veridically is expected if they recover SFM only from the instantaneous velocity field. However, this failure to perceive θ veridically, together with the fact that perceived curvature was close to veridical for $C_{sim} \geq 1$, violates Equations 3 and 4. This can be seen intuitively by considering that if shape relative to the axis of rotation is recovered almost veridically (for $C_{sim} \geq 1$), then the angle of inclination should have been recovered almost veridically, too. Thus, neither the Euclidean nor the affine structure was recovered, although the instantaneous velocity field does carry information about affine structure. The object's shape along the line of sight was recovered with distortions beyond a simple scaling in depth.

The violation of Equations 3 and 4 suggests that Ω, the angular speed of rotation, is not estimated from the instantaneous velocity field (Equation 4). Let us introduce a factor λ into Equation 3 so as to quantify the nonveridical Ω that is estimated by whatever heuristic subjects use in place of the instantaneous velocity field. Then, from Equation 5, the recovered structure becomes

$$\Delta Z = (-\lambda \Delta v_x / \partial_x v_y + \Delta y) Z_0 \tan \theta. \qquad (6)$$

The factor λ can be seen as a scaling factor for depth relative to the slanted plane that contains the axis of rotation and is normal to the sagittal plane. Thus, if $\lambda = 2$, all distances to this plane are doubled (see Figure 4b). From Figure 4b, it is easy to see that for $\lambda \neq 1$, violations of depth order should be observed: The relative depth of any two points on the bottom half of the object will be perceived as reversed, whereas pairs of points in the top half, where the cylindrical radius is constant, will be perceived with the correct depth order. Thus, a uniform distortion relative to the axis of rotation results in a nonuniform distortion across the image plane, in this case, along the vertical direction. This stands in contrast to the

uniform distortions due to depth scaling found in the case of a frontoparallel axis of rotation, where relief structure is always conserved. A depth-order violation when $\lambda \neq 1$ can only occur when the shape of the object changes along the axis of rotation, as in the examples shown in Figure 4b. In fact, it can be shown that any value of $\lambda \neq 1$ will result in a depth-order violation for any pair of points satisfying certain simple geometric relations.

Although this deviation from affinity produces depth-order violations, it does not produce violations of parallelism. The structure defined by Equation 6 is affine to the simulated structure (Equation 5) within any plane in which y = constant (i.e., any plane parallel to the xz—horizontal—plane). Thus, parallelism is not violated on these planes. On the other hand, affinity is violated on planes parallel to the yz—vertical—plane, but here, the deformation is such that two lines that have the same inclination in the simulated object will also have the same inclination in the perceived object; the line inclinations in the simulated and in the perceived objects will, in general, differ from each other (see Appendix A).

Thus, for an inclined axis of rotation, we have two independent free parameters, θ and λ. A value of $\lambda \neq 1$ implies that the recovered depth structure is not affine. For a frontoparallel axis of rotation, by contrast, we have only one free parameter, namely, angular speed Ω in Equation 2; thus, depth structure can be recovered up to an affine transformation in depth.

From Experiments 1 and 2, it is possible to estimate λ (see Equation A13). Values obtained, shown in Figure 5a, depart greatly from 1.0 for all subjects tested. The minimum value in Figure 5a is 1.99. This result implies large departures from affinity in the recovered depth structure along the line of sight. However, the way in which λ was estimated for Figure 5a is susceptible to any systematic bias that subjects bring to their estimation of θ_{obs}. A bias, if present, would be strongly amplified when estimating λ because $\sin\theta_{obs}$, which is usually a small quantity, is inversely proportional to λ (see Equation A13).

Regardless of potential biases that might have corrupted our crude estimations of λ, our finding that perceived curvature does not change with inclination clearly suggests that λ values differ from 1.0. A nonunity value of λ is necessary for the object—that is, the perceived structure relative to the axis of rotation—to be independent of the inclination (see Figure 5b). $\lambda = 1$ would guarantee affinity but carries the expense of a change in the perceived shape of the object as the axis of rotation changes its inclination. Thus, if subjects perceive the axis of rotation as more vertical than it is (as seems to be the case), then they would perceive the object as flatter than it is.

Note that a value can be obtained for the perceived speed of rotation that is consistent with the perceived inclination and with the first-order optic flow (see Equation 4). By definition, this implies $\lambda = 1$, so that the affine structure of the object is recovered. Thus, using a consistent value of Ω would result in a perceived change
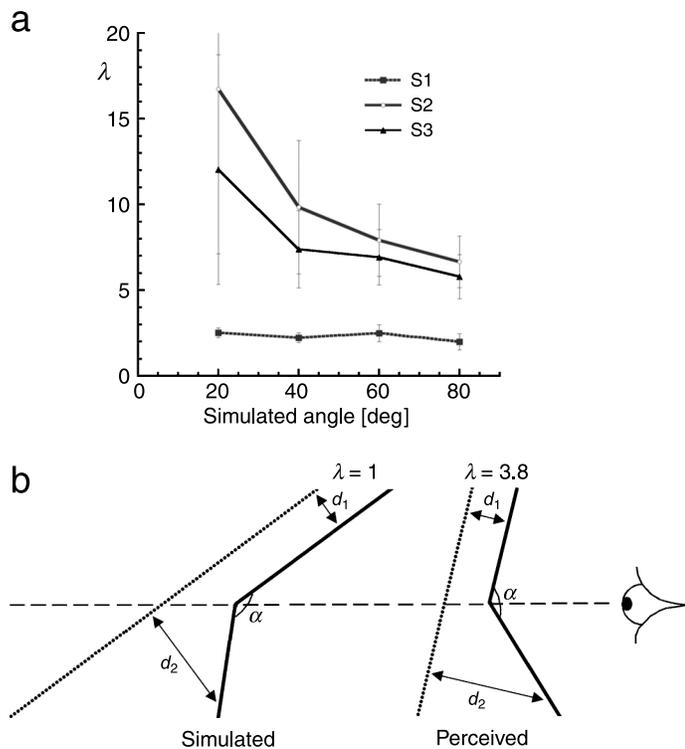
a

b

Figure 5. Nonaffine structure. (a) Variation of λ with the simulated angle of inclination. Error bars are standard error of the mean. (b) Same as Figure 4b, but this time showing the difference between the perceived and simulated object. The shape of the perceived object can be made independent of the inclination of the axis of rotation by using an adequate nonunity value of λ. It can lead to depth-order violations, as shown here. By contrast, a value of $\lambda = 1$, that is, an affine structure, would preserve depth order but result in changes in d_1, d_2, and α.

in object shape when the axis inclination changes. Avoiding changes in perceived shape when inclination changes means using a value of Ω that is inconsistent with the first-order optic flow; that is, $\lambda \neq 1$. This comes at the cost of depth-order violations along the line of sight. Figure 5b shows the relationship between the perceived and the simulated object in such a case.

It is important to stress (see Equation A14) that λ can be computed from first-order optic-flow quantities. This guarantees that the task of computing λ is something the visual system can, in principle, do.

Experiment 3

There is another way to estimate λ that is independent of any measurement of the angle of inclination and more precise than the method used earlier. If subjects perceive surface slant without bias, then, for $\lambda = 1$, a surface simulated as vertical must be perceived as vertical

regardless of the perceived value of the inclination of the axis of rotation. However, for $\lambda \neq 1$, a surface simulated as vertical will be perceived slanted in depth, which implies a depth-order violation (i.e., there exist near-vertical surfaces that would be perceived with the reversed slant). Thus, we can estimate the nonunity value of λ from the slant of the nonvertical simulated surface that appears vertical. Control experiments are needed to measure any intrinsic bias subjects may have in perceiving surface slant.

Subjects viewed a simulated rotating cone attached to a circular cylinder, with an axis of rotation inclined by 45° from the frontoparallel plane. Subjects' task was to indicate the slant of the upper cone's surface relative to the frontoparallel plane (Figure 6a). Subjects were offered a two-alternative forced choice: top of the cone near versus far with respect to the bottom of the cone. Control experiments (Figure 6a) assessed subjects' capacity to perform the task and their bias in perceiving slant. Control surfaces rotated about a vertical axis and, thus, were not subject to errors due to nonaffinity when $\lambda \neq 1$ (Equation 6). Any misperception of the surface's slant in this case can be attributed to the subject's intrinsic bias. (Of course, there is still the possibility that the vertical axis condition represents a special case of unbiased perception of surface inclination, with the bias confined to inclined-axis conditions. Thus, the absence of bias in the inclined-axis condition cannot be completely ruled out.)

The gray bars of Figure 6b show that three of the four subjects do not show a statistically significant bias for control stimuli. They perceive the surface as vertical when it is simulated to be so. One of the subjects does have a statistically significant bias (S3: $p = .0049$), but its direction is opposite that needed to explain the experimental result. Obtained λ values, shown as black bars in Figure 6b, differ significantly from 1.0 ($p < .0001$) for all four subjects, implying depth-order violations. The direction of this difference—λ exceeds 1.0 in all cases—implies that subjects see the top of the cone as closer than the bottom (i.e., $\gamma > 0$). This perceived shape conflicts with the physical shape of the simulated cone, which had the bottom closer. In the control condition, one subject shows a bias to perceive the top of the cone as far relative to the bottom (i.e., $\gamma < 0$) and the others show a slight tendency in that direction. Thus, the λ values obtained are not an artifact of the subjects' intrinsic biases. Values for λ obtained from this experiment are smaller than those obtained earlier from the inclination matching task (Figure 5a), although their rank across subjects is conserved (i.e., subject S2 has the largest λ, etc.). Besides differences in task and stimuli, another reason for the smaller λ values is subjects' bias to perceive inclination in the opposite direction to that of the depth-order violation. In addition, as already mentioned, any systematic bias that could be present in the estimation of θ_{obs} from Experiment 2 would be strongly amplified in estimates of λ in Figure 5a.
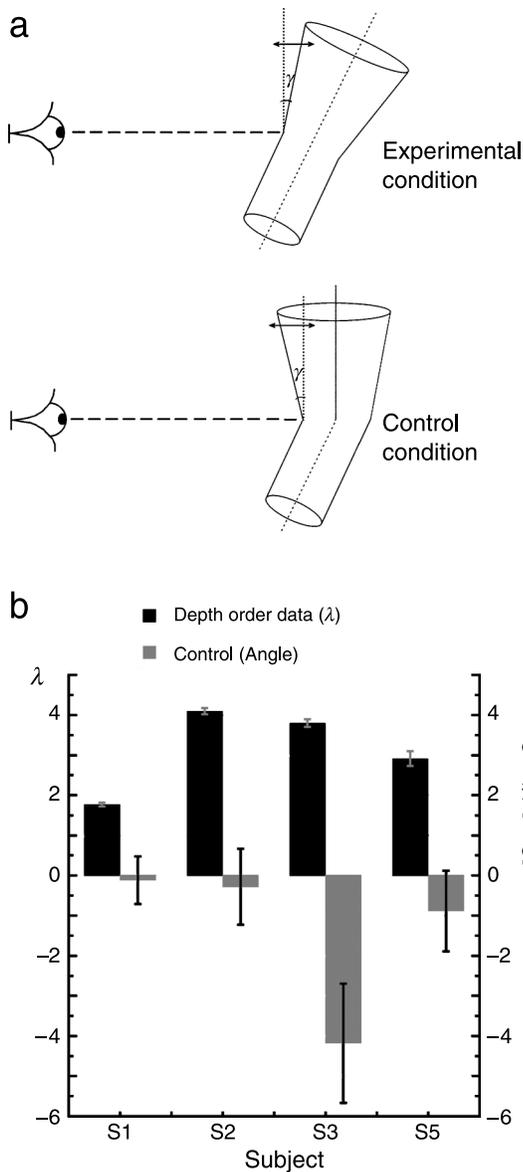
Figure 6. Depth-order violation. (a) Experimental and control stimuli used in depth-order violation experiments. γ is shown as negative in the experimental example and positive in the control example. Drawings are not to scale. (b) λ recovered from depth-order experiments (black bars) and subjects' intrinsic bias (gray bars). Error bars are standard error of the mean.

This confirmation of nonunitary values of λ and of the predicted depth-order violations shows that recovery of SFM from the general case of a non-frontoparallel axis of rotation is not even affine.

Discussion

We studied the ability of humans to recover SFM for an object rotating about an axis slanted out of the frontoparallel plane. We found, using cylinders, that the recovered structure is not affine. We proposed a simple model to account for deviations from veridical in the perceived shape. This model also predicted that for specific shapes, such as cones, the violation of affinity could result in depth-order violations. This prediction was then confirmed in our third experiment.

Structure recovered from rotations around an axis in the frontoparallel plane differs from the more general case of structure recovered from rotations around an arbitrary axis. It is still unknown whether the former preserves affine structure; the latter, in general, does not. While the nonaffine recovery of depth structure can generate depth-order violations along the line of sight, it yields shape constancy relative to the axis of rotation. Thus, the perceived object shape does not change as the axis of rotation slants toward or away from the frontoparallel plane. To maintain this shape constancy, however, the perceived speed of rotation must be inconsistent with the first-order optic flow.

Curved versus flat surfaces

Experiment 1 included conditions in which the simulated cylinder rotated around an axis in the frontoparallel plane. In agreement with previous studies using a frontoparallel axis of rotation, we found that the perceived shape was not veridical. Shape was recoverable only up to a scaling factor in depth. It is currently unknown what heuristics might be used to compute this scaling factor. It has been proposed (see, e.g., Domini & Braunstein, 1998) that the scaling factor is a monotonic function of a first-order property of the optic flow called def (see Appendix B for a formal definition of def). The ratio between perceived and simulated depth is proposed to be

$$d = \Omega \cos(\theta_{\text{sim}}) f(\text{def}), \tag{7}$$

where Ω is the angular speed of rotation. Domini, Caudek, and Richman (1998) did not give an explicit form for the function $f(\text{def})$, but they showed that, to be consistent with their psychophysical data, it must be a decreasing function of def.

In what follows, for the sake of brevity, we will refer to the model developed in Domini and Braunstein (1998) and Domini et al. (1998) as the Domini et al. model. This model seems to work well for planar surfaces. But what works for planar surfaces might not generalize to curved surfaces. Their model was formally developed for perceived surfaces that were planar, and hence, def was constant across each surface. It is not clear how to generalize the model so that it would apply to curved surfaces: Is def averaged across the whole surface and then used to compute a single scaling factor, or is it computed locally and used to derive a scaling factor that varies across the surface?

Caudek and Domini ([1998](#)) suggest that an averaged def is applied to the whole surface. For our cylinders, the average def equals zero because left and right halves have the same def magnitude but opposite signs. We will assume that the visual system computes an average def for each half of the cylinder. The value of def averaged across each half-cylinder can be reasonably approximated (see [Appendix B](#)) as

$$\langle \text{def} \rangle \simeq \Omega \sqrt{\sin^2 \theta_{\text{sim}} + a C_{\text{sim}}^2}, \tag{8}$$

where $a \simeq 5.39$. For $\theta_{\text{sim}} = 0$, [Equation 7](#) becomes

$$\frac{C_{\text{obs}}}{C_{\text{sim}}} = \Omega f(\text{def}). \tag{9}$$

However, the heuristic that seems to work for planar surfaces seems not to work for ellipsoidal cylinders. From [Equations 8](#) and [9](#), we can see that our results do not give support to the notion proposed by Domini et al. that $f(\text{def})$ is a decreasing function of $\langle \text{def} \rangle$. If it were a decreasing function of $\langle \text{def} \rangle$, then the curve in [Figure 2b](#) would be a decreasing function of C_{sim}, which is the opposite of what we found.

Caudek and Domini ([1998](#)) proposed that the perceived slant of the axis of rotation is also a function of def:

$$\tan \theta_{\text{obs}} = \Omega \sin \theta_{\text{sim}} / g(\text{def}), \tag{10}$$

where $g(\text{def})$ is a monotonic increasing function of def. Using [Equation 8](#) for $\langle \text{def} \rangle$, and assuming $g(\langle \text{def} \rangle) = a\sqrt{\langle \text{def} \rangle}$, where a is a constant that depends on the subject, we find a good match (not shown) between our results and the Caudek and Domini prediction. Notice that θ_{obs} can be obtained exclusively from the optic flow because the numerator in [Equation 10](#) can be (see [Equation 4](#)).

Affinity and consistency

We studied the ability of humans to recover SFM for an object rotating about an axis slanted away from the frontoparallel plane. We found, using cylinders, that the perceived structure was not affine. We proposed a simple model to account for the observed data. This model also predicted that for specific shapes, such as cones, the violation of affinity could result in depth-order violations. This prediction was then tested and confirmed. Structure recovered from rotations around an axis in the frontoparallel plane differs from the more general case of structure recovered from rotations around an arbitrary axis. It is still unknown if the former case preserves affine structure (for

a discussion, see [Appendix B](#)). However, we have shown that the latter case, in general, does not.

We have already shown that our results do not support the general claim that, in SFM, the perceived depth between two points is a decreasing function of def (Domini et al., [1998](#)). We have also pointed out that the Domini et al. model seems to work reasonably well for planar surfaces undergoing frontoparallel rotations. This is true even when two such surfaces are connected by a smoothly curved surface. In addition to predicting the ratio between perceived and simulated depth, Domini and Braunstein ([1998](#)) and Domini et al. ([1998](#)) claim that their model predicts violations of affine structure, including the appearance of depth-order violations and internal inconsistencies in the recovered 3D structure. In the following sections, we rework this proposed model, generalize it to curved surfaces, and give an alternative interpretation to their results.

The Domini et al. model

The Domini et al. model arose from the observation that the visual system makes use only of first-order optic-flow information. Domini, Caudek, and Proffitt ([1997](#)) showed that (a) the magnitude of the frontoparallel component of the angular velocity of rotation is misperceived in SFM because it is derived as a monotonically increasing function of def and (b) misperceptions of angular velocities lead to misperceptions of rigidity. Thus, a single planar surface with one value of def was seen as rotating rigidly, whereas two simultaneously presented planar surfaces, each with a different value of def, were perceived as rotating nonrigidly. The same hypothesis led to the prediction that the perceived slant of the axis of rotation would also be a function of def and would, thus, also be misperceived. Caudek and Domini ([1998](#)) confirmed this prediction.

In subsequent work (Domini & Braunstein, [1998](#); Domini et al., [1998](#)), it was proposed that the perceived slant of a planar surface is also a monotonically increasing function of def (the function was assumed to be sublinear, for consistency with their results). This function predicted that the depth separation between two points in a planar surface would be a decreasing function of def. But in this version of the model, the perceived angular speed was assumed not to be a function of def.

The hypothesis that the perceived slant is a function of def generates a radical prediction: The recovered depth structure of the object would be internally inconsistent. That is, the integral of the signed depths across a closed path would not sum to zero. This inconsistency has some similarity (although it is actually worse) to that found in the "ever-ascending" staircase of Maurits Escher. Also predicted are distortions of depth-order relations and parallelism, so that the recovered structure is neither affine nor Euclidean.

In a series of experiments, Domini and Braunstein (1998) and Domini et al. (1998) found supporting evidence for these predictions. Can we then conclude from their results that there is no globally consistent interpretation of an SFM shape? We do not think so.

All the experiments they preformed can be reinterpreted in the light of an alternative theory, one based on their original hypothesis that the angular velocity is misperceived in SFM. As mentioned, this hypothesis was successfully tested (Caudek & Domini, 1998; Domini et al., 1997) but was replaced by their later hypothesis that perceived slant is a function of def, which makes very different predictions about the shape recovered in SFM.

Our theory, based on a further elaboration of the original work of Domini et al., has two advantages. First, it resolves the contradiction between their former and later hypotheses. Second, it avoids the radical view that the recovered structure from SFM is internally inconsistent.

An alternative interpretation

Previous work has suggested that the depth structure recovered from the projection of rotational motion is marked by violations of affinity and consistency. We take up each of these issues in turn.

Affinity

As previously mentioned, Domini et al. (1997) suggested that the projection of the angular speed of rotation onto the frontoparallel plane is derived as a function of def. The surfaces used in their experiments were always planar. Hence, def had a constant value across the surface, and, as predicted, subjects observed individual surfaces rotating at a single speed. When two intersecting planar surfaces with differing values of def were presented, each surface was perceived, again as expected, as moving with a different speed of rotation, resulting in nonrigid motion. However, when testing the hypothesis that perceived slant is a function of def (Domini & Braunstein, 1998; Domini et al., 1998), it was assumed that perceived speed of rotation did not vary across surfaces, even for intersecting planar surfaces, although there was no test of perceived rigidity in those studies.

We assume that each planar surface was perceived as rotating at a different speed and that motion was indeed perceived nonrigidly in the Domini and Braunstein (1998) and Domini et al. (1998) experiments. The recovered structure of each independently moving surface could, thus, be affine. We show in Appendix B that for planar surfaces, this hypothesis predicts the same measurements as obtained by Domini and Braunstein and Domini et al. in their experimental settings. For these stimuli, apparent violations of depth-order relations and of parallelism are explained without recourse to violations of affinity.

Consistency

In an experiment that explicitly tested for the internal consistency of the recovered structure, Domini and Braunstein (1998) used a surface consisting of two planar patches smoothly joined by a curved surface. In their experiment, there were four probe dots presented as two dot pairs. One of the pairs was on one planar surface; another was on another such surface; and the two planar surfaces were linked by a curved surface so as to predict the same perceived depth difference between the two near dots and the two far dots across pairs. Figure 7 shows a schematic top view of the four dots. Probe Dots 1 and 2 rest on a planar surface with a given slant, and Probe Dots 3 and 4 rest on another planar surface with a different slant. The separation in depth between Dots 1 and 2 is the same as that between Dots 3 and 4. The curved surface (not shown) joining the two planar surfaces is identical between Dots 2 and 3 and between Dots 1 and 4. The hypothesis of Domini et al. predicts that the integral of the signed depths across the closed path 1–2–3–4–1 does not sum to zero. The reason is that $\Delta Z_{2-3} = \Delta Z_{1-4}$ because they are joined by similarly shaped surfaces, but $\Delta Z_{1-2} < \Delta Z_{4-3}$ because they are located in planar patches of different slant.

Our interpretation is different. Let us assume that for a curved surface, the recovered speed of rotation is a function of def and varies locally at each point in the surface. It can be shown (Appendix B) that under these assumptions, the integral of the signed depths across any closed path is always predicted to equal zero. Thus, the surface depth structure is always predicted to be internally
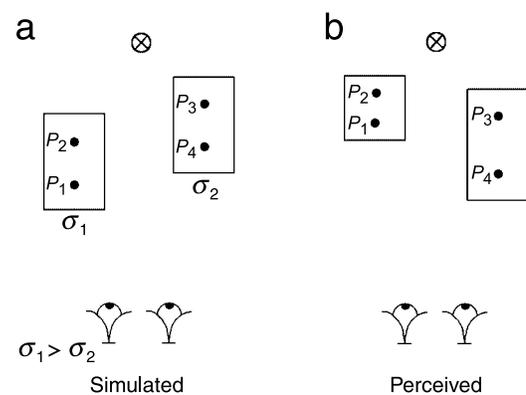

Figure 7. (a) Bird's-eye view of the setup of Domini and Braunstein (1998) for testing internal consistency of perceived shape. Four probe dots (P_1 to P_4) were located in a surface (also defined by moving dots) consisting of two planar patches of different slants (σ_1 and σ_2). The two patches were united by a smooth curved surface (not shown). The circle with the cross inside represents the axis of rotation. The distance between dots P_1 and P_2 is the same as that between P_3 and P_4. (b) Perceived configuration predicted by our model of the simulated structure shown in Panel a.

consistent. However, when the axis of rotation lies behind the surfaces, it predicts not only that $\Delta Z_{1-2} < \Delta Z_{4-3}$ but also that $\Delta Z_{2-3} < \Delta Z_{1-4}$. Evidence consistent with this latter prediction appears in the data of the Domini and Braunstein (1998). In addition, they found that judged separations differed between axis-in-front and axis-behind conditions. To explain this, they introduced an additional dependence of perceived depth on the velocity ratio between dots. Our model predicts this difference without added assumptions because, for a constant speed of rotation, the depth difference is a function of both relative and absolute retinal speeds (see Equations B12 and B13) and these speeds vary with axis position. In the Domini et al. model, the depth difference is a function of def, which is not a function of axis position.

Our model predicts a nonzero closed-path integral when it is computed as Domini and Braunstein (1998) did. They assumed that the perceived relative positions of the dots would be the same as the simulated positions (Figure 7a). Hence, they computed the "integral" I across a closed path as $I = \Delta Z_{1-2} + \Delta Z_{2-3} - \Delta Z_{4-3} - \Delta Z_{1-4}$, where all ΔZ values refer to their absolute (positive) values. Our model predicts that the perceived relative positions of the dots will be, instead, as shown in Figure 7b. This predicts that the integral $I' = \Delta Z_{1-2} - \Delta Z_{2-3} - \Delta Z_{4-3} + \Delta Z_{1-4}$ will add up to zero. Our model predicts (substituting $I' = 0$ into the definition of I used by Domini & Braunstein, 1998) that the measured value will be $I = 2(\Delta Z_{2-3} - \Delta Z_{1-4}) \neq 0$.

Thus, the hypothesis that the recovered structure is internally inconsistent is not supported. Our model shows that an alternative internally consistent interpretation is possible.

Conclusions

Current data from frontoparallel rotations—including those of Domini and Braunstein (1998) and Domini et al. (1998)—are inconclusive as to whether the recovered structure is affine or not. Our results using a slanted axis of rotation show conclusive evidence of nonaffinity and depth-order violations in the recovered structure. This violation to affinity, which occurs when the axis of rotation is slanted out of the frontoparallel plane, can result in depth-order violations. In principle, a single angular speed and depth-scaling factor could be used to recover the structure of the whole object without violations of internal consistency.

Although the recovered depth structure along the line of sight is nonaffine for inclined axes of rotation, this structure is closely related to the structure recovered from a frontoparallel rotation. As we have seen, the depth structure perpendicular to the axis of rotation does not change with the inclination of the axis of rotation. Thus,

recovering the depth structure in the general case can be decomposed into two stages. In the first stage, the frontoparallel structure is computed: The information about depth structure relative to the axis of rotation is available from the optic-flow field in the projected component of the speeds perpendicular to the axis of rotation (but, as mentioned before, we still lack an algorithm that allows us to predict the recovered depth structure for the frontoparallel case). In the second stage, the parameter λ is computed. This gives the depth structure for an arbitrary axis' inclination.

Our results show that the second stage introduces nonaffine distortions; it is still unknown whether the first stage is affine. If future experiments show that the recovered structure in the frontoparallel rotation condition is also not affine, this class of nonaffinity must be added to the nonaffinity found in this article. The two would have different origins, arise in different computational stages, and act independently of each other. According to the two-stage formulation, nonaffinity in the first stage would be inherited by the second stage, which would add its own nonaffine contribution. In this case, the two nonaffine distortions would act in superposition.

Methods

General methods

The stimuli consisted of moving dark "dots" ($3.4' \times 3.4'$ squares) on a moderately bright background. An attenuator used to boost luminance resolution to approximately 12 bits drove only the monitor's green gun. Thus, the color of both the dots and the background was green. Dot and background luminances were 2 and 29 cd/m^2, respectively. The motion was shown at the refresh rate of the monitor (75 Hz). Stimuli were monocularly viewed at an optical distance of 94 cm, using a chin rest to stabilize head position. No separate fixation point was required for the task, but one was present as a $10' \times 10'$ square between trials to guide fixation to the center of the display. Binocular viewing was also tested in pilot experiments and produced no significant differences from monocular viewing.

Perceived depth from SFM displays often increases to an asymptotic value with increases in parameters such as dot density, dot lifetime, stimulus duration, range of oscillation, and speed of rotation. Pilot experiments were performed to verify that the parameters used in our experiments resulted in perceived depth reaching asymptotic values.

One experienced subject (S1), and three naïve inexperienced subjects (S2 to S4) were used in Experiment 1. Three subjects (S1 to S3) were used in Experiment 2, and four subjects (two experienced [S1 and S5] and two naïve

and inexperienced [S2 and S3]) were used in Experiment 3. Subject numbering reflects subject identity across experiments. All subjects gave written informed consent for participation in the study, which was approved by the Institutional Review Board of Syracuse University.

Experiment 1

We created 28-frame (373 ms) movies of individual opaque elliptical rotating cylinders defined by moving dots ($n = 300$). Cylinders rotated at a constant angular speed (135 deg/s). The axis of rotation corresponded with the cylinder's longitudinal axis. Rotation was confined to an arc of 50°; one cycle of rotation brought the cylinder to its starting position after rotating a half-cycle of 50° in one direction and another half-cycle of 50° in the opposite direction. Test stimuli were presented in 1.49-s movies. Each movie displayed two cycles of rotation, each cycle formed by playing in forward and reversed order the same 28-frame animation sequence. The axis of rotation could be slanted away from the frontoparallel plane by an angle of 0°, 20°, 40°, 60°, or 80°. Regardless of the slant, the axis of rotation was always within the subjects' midsagittal plane; that is, its projection onto the frontoparallel plane was vertical and centered on the subject's line of sight. The cylinders' cross section was elliptical. The cylinders' rotational path was such that one of the main axes of the ellipse crossed the midsagittal plane halfway through each of the cylinders' half-cycles of rotation (i.e., the starting phase was −25°). Cylinders' curvature was defined as the ratio between the two main axes of the ellipse. In computing the ratios, the axis used in the numerator was the axis that crossed the midsagittal plane during the rotation.

Curvature values of 0.5, 0.75, 1, 1.5, and 2 were used. Initial position of the dots defining the cylinder was random in the projected view and, therefore was not random on the 3D cylinder's surface. To eliminate density cues, which become more important the more elongated the cylinders are, we tested two different dot conditions: sustained dots versus finite-lifetime dots. A sustained dot remained attached to the cylinder's surface during the whole movie. A finite-lifetime dot disappeared and was replaced by a new dot at a random position in the projected view. Dot lifetimes were distributed as follows. Let us define, for a given movie, the maximum dot lifetime, lt_{max}. Then, at every frame in the movie, we replaced n/lt_{max} dots (rounded to the nearest integer) with new randomly positioned dots, where $n = 300$ is the total number of dots. A different set of dots was replaced on each frame until all dots were used, and the process was started again in the same order until the end of the movie. To keep dot lifetimes inside the range of asymptotically perceived depth, lt_{max} covaried with curvature, being 24, 18, 12, 9, and 6 frames for curvatures of 0.5, 0.75, 1, 1.5, and 2, respectively. These values were low enough at each

curvature value to also ensure that no dot-density cues were present, as assessed by judging individual frames, which did not allow the identification of the cylinder's shape.

Cylinders were occluded at the top and bottom by dark rectangular maskers (2 cd/m^2, 4° [H] × 2° [V]) to cover these borders. The projected vertical cylinder height between the masks was 4°. Cylinders' horizontal size was 4° when located midway through the half-cycle. In side-masking conditions, the cylinders' left and right borders were occluded by maskers whose horizontal separation matched the minimal lateral extent of the cylinders during rotation. This kept the visible portion of the cylinder constant, rather than expanding or contracting laterally during rotation. Maskers' size, thus, never exceeded 10% of the cylinders' width.

Subjects' task was to view the rotating cylinder and then to adjust a cylindrical cross section to match the profile of the cylinder previously seen in the movie. Subjects could repeat the 1.49-s movie as many times as they wanted during the adjustment procedure. The cylindrical cross section had the same horizontal angular size as the moving cylinder; subjects adjusted the cross section by clicking with a mouse on "+" and "−" symbols located on the screen below the cross section. Clicking on "++" and "−−" symbols adjusted the cross section more coarsely. Each run consisted of 20 trials obtained by randomly selecting without replacement 1 of the 100 different stimuli obtained by combining the five curvatures, the five axis inclinations, the two masking conditions, and the two dot lifetime conditions. Each subject completed 40 runs and, thus, adjusted each stimulus eight times.

Experiment 2

Experiment 2 differed from Experiment 1 in three ways: (a) Only the circular cylinder was used. (b) The tested inclinations of the axis of rotation were 0°, 35°, 45°, 65°, and 85°. (c) Subjects were required to adjust the orientation of a line appearing on the screen to match the inclination of the axis of rotation.

Experiment 3

We created 112-frame (1.49 s) movies of rotating geometric structures each defined by 1,200 moving dots. The structures were opaque and rotated at a constant angular speed of 33.75 deg/s about an axis of rotation along their longitudinal axis. The angular speed of rotation was made slower than in the two previous experiments to keep retinal speeds in the same range, as the stimulus was now larger. Rotation was in both directions and confined to a 50° arc. Test stimuli consisted of a 5.96-s movie, obtained by running the 1.49-ms movie forward and in reverse, twice. The axis of rotation was

slanted 45° away from vertical, with the upper part of the axis slanted away from the subject. Initial positions of the dots defining the structure were random in the projected view. Maximum dot lifetime, lt_{max}, was 18 frames, with a vanished dot replaced by a new dot placed at random in the projected view, using the procedure explained in Experiment 1. The structures were viewed through an $8° \times 8°$ square window. The structure's cross section was circular, with its bottom half forming a cylinder and its top half forming a cone. For test stimuli, the radius of the cylinder was 6°, and that of the cone increased from 6° at a linear rate such that the cone's surface within the midsagittal plane was inclined with angle γ relative to the vertical (Figure 6a). For control stimuli, the only difference was in the top half. Here, both the axis of rotation and the longitudinal axis of the cone were vertical; hence, the bottom and top halves had different axes of rotation. The radius along the vertical axis was variable, increasing or decreasing linearly from 6° so that the cylinder's surface within the midsagittal plane was inclined with angle γ relative to the vertical (Figure 6a). For both test and control stimuli, all four boundaries were occluded from view by maskers.

We used a two-part hybrid object (i.e., cylinder plus cone), instead of only a single cone (which, in principle, ought to be sufficient for our purpose), because we observed that a single cone viewed through a window is often perceived as a 2D deforming flow of dots (a valid, or rather a veridical, interpretation of the optic-flow field) rather than as a 3D rotating object. Adding a cylinder seems to avoid this problem.

For test stimuli, the psychometric functions recording the proportion of trials in which subjects reported the bottom of the cone as "near", as a function of the simulated surface's inclination, γ_{sim}, were obtained using the method of constant stimuli. From γ_{sim} we can obtain β (see Appendix A, Equation A15), and thus, $\lambda = 1/\beta$. The five γ_{sim} values were individually selected for each subject based on pilot data to optimize the range of the psychometric function (in the pilot experiments, $\gamma_{sim} = 0°, \pm10°, \pm20°,$ and $\pm30°$ were used for all subjects; in the main experiments, we used $\gamma_{sim} = -15°, -20°, -25, -30°,$ and $-35°$ for S1 and $\gamma_{sim} = -5°, -15°, -25, -35°,$ and $-45°$ for the rest of the subjects). A cumulative normal was fit to the psychometric functions by probit analysis, from which the points of perceived verticality were obtained. This point was defined as the point at which the response rate for bottom-seen-as-near was 50%.

Subjects' task was to indicate whether the top or the bottom of the cone appeared as near. To ease this task, we placed two static square markers ($10' \times 10'$, 2 cd/m²) at the upper and lower extremes of the line defined by the intersection of the cone's surface and the midsagittal plane. It was this line, at the horizontal midpoint of the cone, that subjects were to judge. Each run consisted of 50 trials obtained by randomly choosing one of the five different stimuli that differed in γ_{sim}. Each subject totaled

at least two runs so as to judge each stimulus at least 20 times. The same procedure was followed for control stimuli ($\gamma_{sim} = 0°, \pm3.5°,$ and $\pm7°$ for S1 and $\gamma_{sim} = 0°, \pm10°,$ and $\pm20°$ for the rest of the subjects).

To test for the statistical significance of $\lambda \neq 1$ for test stimuli, or of $\gamma \neq 0$ for control stimuli, we obtained precise estimates of the standard deviation of the 50% thresholds using the bootstrap method described by Foster and Bischof (1997), which allows the use of a normal distribution to compute probabilities. Fifty percent threshold values (i.e., point of perceived verticality) and their standard deviations obtained from the bootstrap method were virtually identical to those obtained using probit.

Appendix A

Derivation of Equations 3 and 4

To derive Equation 3, observe that, for an object rotating around a vertical axis, the trajectories of points are horizontal and the distance between a given point on the object and the frontoparallel plane that includes the axis of rotation is (under orthographic projection)

$$r_{axis} = Z_0 \, v/\Omega, \qquad (A1)$$

where Z_0 is the distance between the object and the subject, v is the retinal speed of the given point, and Ω is the angular speed of rotation (for a derivation, see Fernandez et al., 2002). In what follows, it is assumed that the object is distant enough from the subject so that Z_0 approximates the distances of all points on the object.

Let us consider now the same object when the axis of rotation makes an angle θ with respect to the frontoparallel plane. The distance between a given point on the object and an arbitrary reference frontoparallel plane will be (see Figure A1)

$$Z = Z_1 + Z_2. \qquad (A2)$$

From simple trigonometry, we have (for $0° < \theta < 90°$)

$$Z_2 = r_{axis}/\cos\theta \qquad (A3)$$

and

$$Z_1 = Y\tan\theta, \qquad (A4)$$

where Y is the distance between the given point and a horizontal plane passing through the line of sight.

In the case of an inclined axis of rotation, the trajectories of points on the surface are no longer horizontal but curved. Let (X', Y', Z') be the coordinates of a given point in a coordinate system in which the axis Y' coincides with the axis of rotation (Figure A2). Let us assume that the axis of rotation is inclined by θ relative to
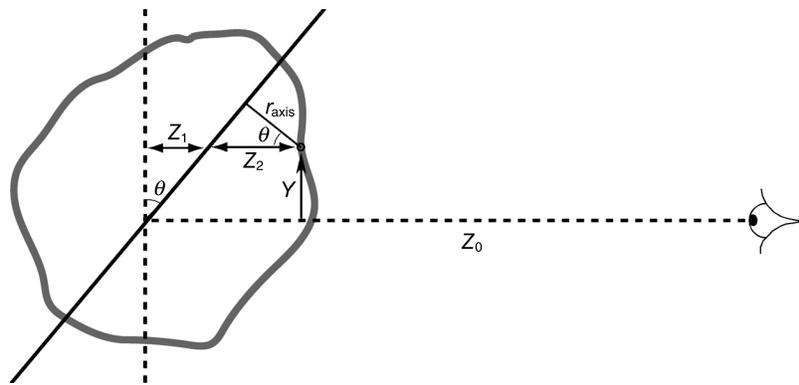
Figure A1. Viewing geometry. See text for details.

the frontoparallel plane, and let (X, Y, Z) be the coordinates in our canonical reference frame in which Y is vertical (Figure A2)—note that, for simplicity, we use the same labels for axes and the coordinates of a point relative to those axes. Then, a simple rotation of coordinates gives

$$
\begin{aligned}
X &= X' \\
Y &= Y'\cos\theta - Z'\sin\theta \\
Z &= Y'\sin\theta + Z'\cos\theta.
\end{aligned}
\tag{A5}
$$

Only the first two equations need to be considered for our purposes. If we differentiate both sides of these equations with respect to time, and then divide by Z_0 (and noticing that $v_{y'} = 0$ for rotations about axis Y'), we obtain

$$
\begin{aligned}
v_x &= v_{x'} \\
v_y &= -v_{z'}\sin\theta.
\end{aligned}
\tag{A6}
$$

Taking into account that $v_{x'} = v$ (Equation A1) and $v_{z'} = \Omega x' = \Omega x$ (lower caps refer to angular variables):

$$
v_x = v
\tag{A7}
$$

$$
v_y = -\Omega x \sin\theta.
\tag{A8}
$$

From Equation A8, after differentiation with respect to x, we obtain Equation 4. Substituting from Equations A1, A3, A4, and A7, Equation A2 becomes

$$
Z = Z_0 v_x / (\Omega\cos\theta) + Z_0 y\tan\theta,
\tag{A9}
$$

where $y = Y/Z_0$ is the angular distance between the given point and a horizontal plane passing through the line of sight. Taking differences between any two points using Equation A9 yields Equation 3.

Obtaining λ from Experiments 1 and 2

Let us consider two points belonging to the rotating object, one obtained from the intersection of the line of sight and the object's surface (P_2), and the other from the intersection of the line of sight and the axis of rotation (P_1; Figure A3).

Let us write Equations 5 and 6 for these two points using the simulated axis of rotation and again using the perceived axis of rotation. By definition, $\Delta y = 0$; hence, we have

$$
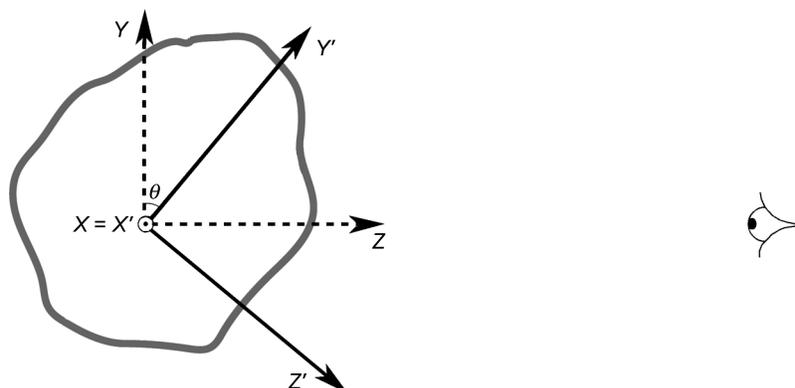\Delta Z_{\text{sim}} = -Z_0\tan\theta_{\text{sim}}\Delta v_x / \nabla_x v_y
\tag{A10}
$$

Figure A2. Viewing geometry. See text for details.

for the simulated axis and

$$\Delta Z_{\mathrm{obs}} = -\lambda Z_0 \tan\theta_{\mathrm{obs}} \Delta v_x / \nabla_x v_y \qquad (A11)$$

for the perceived axis.

From Equations A10 and A11, we obtain λ:

$$\lambda = \tan\theta_{\mathrm{sim}} \Delta Z_{\mathrm{obs}} / (\tan\theta_{\mathrm{obs}} \Delta Z_{\mathrm{sim}}). \qquad (A12)$$

Using the relationship $r = \Delta Z \cos\theta$ (Figure A3) for simulated and perceived values of θ, and using $C_{\mathrm{obs}}/C_{\mathrm{sim}} = r_{\mathrm{obs}}/r_{\mathrm{sim}}$, Equation A12 becomes:

$$\lambda = (C_{\mathrm{obs}}/C_{\mathrm{sim}})(\sin\theta_{\mathrm{sim}}/\sin\theta_{\mathrm{obs}}). \qquad (A13)$$

We can also express λ as a function of optic-flow properties, which is important to guarantee that the task of computing λ is something the visual system can actually carry out. Using Equations A1, A13, 4, and 10, we obtain, after a lengthy calculation,

$$\lambda = \frac{r_{\mathrm{obs}}\sqrt{[g(\mathrm{def})]^2 + (\partial_x v_y)^2}}{Z_0 v_x}, \qquad (A14)$$

where r_{obs} and v_x are the perceived distance to the axis of rotation and the horizontal angular speed, respectively, of any point on the object. It is not known how r_{obs} is obtained from the optic-flow field—as we already mentioned, the heuristics proposed by Caudek and Domini (1998) seem not to be valid for cylinders. Two points deserve to be stressed here. First, to perceive a constant shape independent of the inclination of the axis of rotation, it is necessary and sufficient that r_{obs} be a

function only of v_x. Second, notice that λ is a function not only of def but also of more basic optic-flow properties, such as the gradients (like $\partial_x v_y$ and possibly also $\partial_x v_x$ through r_{obs}) and angular speeds.

Obtaining λ from Experiment 3

Let us consider two points, P_1 and P_2, on the rotating object, located on the same sagittal plane, but having different depths and heights relative to the subject (Figure A4). The line joining the two points meets with the vertical to form a perceived angle γ given by

$$\tan\gamma = -\Delta Z/\Delta Y. \qquad (A15)$$

Using Equation 6, this becomes

$$\tan\gamma = \tan\theta(\lambda\beta - 1), \qquad (A16)$$

where $\beta = \Delta v_x/(\nabla_x v_y \Delta y)$ is a quantity that depends only on the physical parameters of the stimulus and, thus, can be set by the experimenter.

By definition, the surface is perceived as vertical when $\gamma = 0$. In this case, we can obtain λ as

$$\lambda = 1/\beta. \qquad (A17)$$

We can obtain β (and thus λ) at the point of perceived verticality from a psychometric function in which β is the independent variable, as described in the Methods section. Note that the value of λ obtained in this way is independent of θ and γ and does not require their measurement.
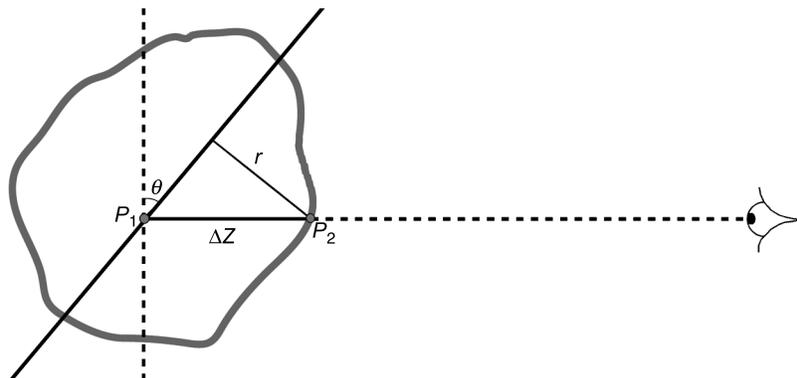

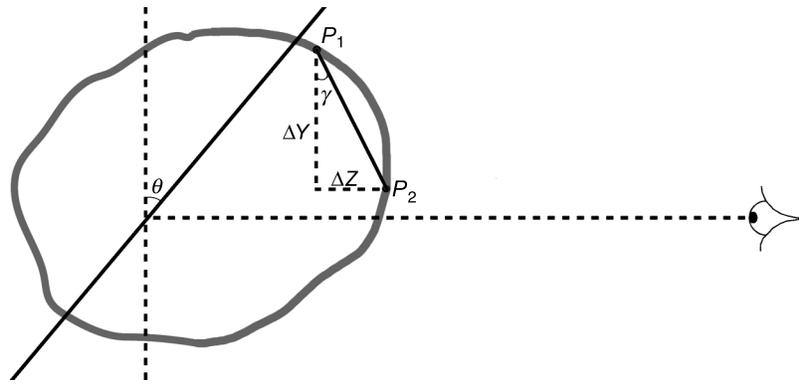Figure A3. Viewing geometry. See text for details.

Figure A4. Viewing geometry. See text for details.

A depth-order violation occurs when the sign of the perceived γ is opposite the sign of the simulated γ:

$$\text{sign}(\gamma_{\text{per}}) \neq \text{sign}(\gamma_{\text{sim}}). \tag{A18}$$

By definition, $\lambda = 1$ for the simulated object; thus, Equation A18 is equivalent to

$$\text{sign}(\lambda\beta - 1) \neq \text{sign}(\beta - 1). \tag{A19}$$

This inequality gives three cases that result in depth-order violations:

Case 1. $\lambda < 1/\beta$ and $\beta < 0$
2. $\lambda > 1/\beta$ and $0 < \beta < 1$ $\qquad$ (A20)
3. $\lambda < 1/\beta$ and $1 < \beta$

Thus, any $\lambda \neq 1$ will result in a depth-order violation for any pair of points satisfying Equation A19.

Also notice that any line simulated with a given inclination γ_{sim} will result in a similarly inclined perceived line γ_{per}. This is easily seen from Equation A16. Equation A16 is valid for either the simulated or the perceived object; in the latter case, we must use $\lambda = 1$. First, note that Equation A16 implies that for $\gamma_{\text{sim}} = $ constant, then $\beta = $ constant (assuming that $\theta_{\text{sim}} = $ constant). Thus, two parallel lines must have the same value of β. Using Equation A16 again, this time for γ_{per}, we obtain that $\gamma_{\text{per}} = $ constant, because λ is also a constant (and also assuming that $\theta_{\text{obs}} = $ constant).

Appendix B

Calculation of def

Def is defined as

$$\text{def} = \sqrt{\text{def}_1^2 + \text{def}_2^2}, \tag{B1}$$

where

$$\begin{aligned}\text{def}_1 &= \partial_x v_x - \partial_y v_y \\ \text{def}_2 &= \partial_y v_x - \partial_x v_y.\end{aligned} \tag{B2}$$

Here, $\partial_x(\partial_y)$ represent the partial derivative with respect to $x(y)$ and $v_{x,y}$ are the horizontal and vertical components of the retinal speed. For the cylinders of Experiment 1, we have $\partial_y v_y = \partial_y v_x = 0$; thus, we obtain

$$\text{def}^2 = (\partial_x v_x)^2 + (\partial_x v_y)^2. \tag{B3}$$

We can easily evaluate Equation B3 when the apex of the cylinder crosses the line of sight. In this case, we have

$$z' = C_{\text{sim}}\sqrt{1 - x^2}, \tag{B4}$$

which is the equation of the ellipse in the primed axis system (without losing generality, we are assuming cylinders of horizontal radius equal to one). Using the fact that

$$v_x = v_x' = -\Omega z' \tag{B5}$$

in Equation B4, and calculating the derivative, we obtain

$$\partial_x v_x = \frac{\Omega \, C_{\text{sim}} x}{\sqrt{1 - x^2}}. \tag{B6}$$

Using Equation A8, we obtain

$$\partial_x v_y = -\Omega \sin\theta. \tag{B7}$$

Finally, using Equations B6 and B7 into B3, we obtain

$$\text{def}^2 = \Omega^2 \left(\sin^2\theta + \frac{C_{\text{sim}}^2 x^2}{1 - x^2} \right). \tag{B8}$$

The value of def averaged across each side of the cylinder ($\langle\mathrm{def}\rangle$) can be obtained as a reasonable approximation as

$$\langle\mathrm{def}\rangle \simeq \sqrt{\langle\mathrm{def}^2\rangle} = \Omega\sqrt{\sin^2\theta + aC_{\mathrm{sim}}^2}, \tag{B9}$$

where $a \simeq 5.39$.

For vertical cylinders, $\theta = v_y = 0$; hence, we can compute def exactly (i.e., without using the approximation given by Equation B9) as

$$\langle\mathrm{def}\rangle = \Omega C_{\mathrm{sim}}. \tag{B10}$$

Speed of rotation as a function of def

For an object rotating about an axis in the frontoparallel plane, the distance z of a point on the object to the frontoparallel plane containing the axis of rotation is (from Equation 2)

$$z = \frac{Z_0 v}{\Omega}, \tag{B11}$$

where Ω is the angular speed of rotation, Z_0 is the distance of the subject to the axis of rotation, and v is the retinal velocity of the point. Notice that in Equation 2, Z refers to the distance to the subject, and here, z refers to the distance from the frontoparallel plane (positive toward the subject) containing the axis of rotation.

Assuming that the perceived speed of rotation, Ω_{per}, is a function of def, then, for a planar surface, Ω_{per} will be constant across the surface. Thus, the perceived depth difference between any two points on a planar surface can be obtained from Equation B11 as

$$\Delta z = \frac{Z_0 \Delta v}{\Omega_{\mathrm{per}}}. \tag{B12}$$

Two planar surfaces of different slant have different values of def; therefore, by assumption, they will be perceived as rotating with different speeds. If the two planar surfaces intersect at an edge (open book), Equation B12 prescribes that they will not be perceived joining at an edge, but rather a gap in depth will be perceived between them (Figure A5). The perceived depth difference between surfaces at this edge will be

$$\Delta z = \frac{Z_0 v}{\Omega_{\mathrm{per1}}} - \frac{Z_0 v}{\Omega_{\mathrm{per2}}} = -Z_0 v \frac{\Delta\Omega_{\mathrm{per}}}{\Omega_{\mathrm{per1}}\Omega_{\mathrm{per2}}}, \tag{B13}$$

where v is the same for both edges.

The Domini et al. (1998) prediction is that the perceived distance between P_2 and P_3 will be larger than that between P_1 and P_3 due to the difference in the slants of

the surfaces these pair of points belong to. Thus, in their experiment, subjects indicated which of the two points, P_1 or P_2, is closer in depth. They found that P_2 is seen as closer, but the same prediction is made by our model (see Figure A5).

In the same way, the results of all of the experiments using planar or piecewise planar surfaces can be explained by our hypothesis. This includes their "ever-ascending" crown. Basically, the main perceptual difference between their predictions and ours is that they assume that the simulated edge where two planes intersect is perceived as an edge, whereas our model predicts that there will be a depth separation between the perceived planes at the simulated edge. Another difference, already mentioned, is that they assume that all surfaces will be perceived as rotating with the same angular speed, whereas we predict that they will be perceived as rotating at different speeds.

To understand the experiments that used curved surfaces, we need a more detailed formalism, which will be developed below—a detailed look at one of the experiments involving curved surfaces can be found in the Discussion section.

Let us assume that, for a curved surface, it is still valid that the perceived speed of rotation varies locally as a function of def. We can approximate a smooth surface as a collection of infinitely small planes and then use Equations B12 and B13 in their differential form to calculate the depth difference between two neighboring points as

$$dz = dz_1 + dz_2 = \frac{Z_0}{\Omega_{\mathrm{per}}}dv - \frac{Z_0 v}{\Omega_{\mathrm{per}}^2}d\Omega_{\mathrm{per}}, \tag{B14}$$

where dz_1 is the depth difference inside the differential plane (from Equation B12) and dz_2 is the depth difference between contiguous differential planes (from Equation B13).
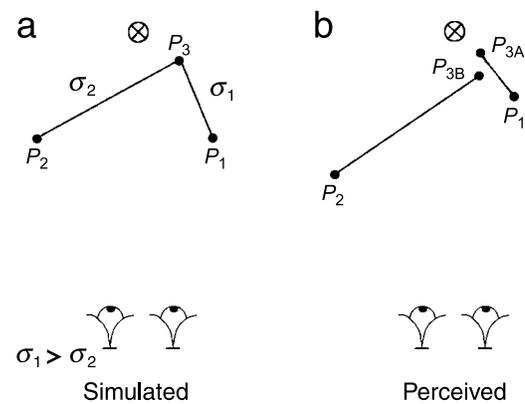

Figure A5. (a) Bird's-eye view of the setup of Domini et al. (1998). Probe dots P_1 and P_2 belong to planar patches that have different slants (σ_1 and σ_2). The distance between P_1 and P_3 is the same as that between P_2 and P_3. (b) Perceived configuration predicted by our model of the simulated structure shown in Panel a.

Both v and Ω_{per} are a function of position (x, y) on the frontal plane; thus, we have

$$
\begin{aligned}
\mathrm{d}v &= \partial_x v \, \mathrm{d}x + \partial_y v \, \mathrm{d}y \\
\mathrm{d}\Omega_{\mathrm{per}} &= \partial_x \Omega_{\mathrm{per}} \, \mathrm{d}x + \partial_y \Omega_{\mathrm{per}} \, \mathrm{d}y.
\end{aligned} \tag{B15}
$$

Replacing Equation B15 into Equation B14, we obtain

$$
\begin{aligned}
\mathrm{d}z &= Z_0 \left[\left(\frac{\partial_x v}{\Omega_{\mathrm{per}}} - \frac{v \partial_x \Omega_{\mathrm{per}}}{\Omega_{\mathrm{per}}^2} \right) \mathrm{d}x + \left(\frac{\partial_y v}{\Omega_{\mathrm{per}}} - \frac{v \partial_y \Omega_{\mathrm{per}}}{\Omega_{\mathrm{per}}^2} \right) \mathrm{d}y \right] \\
&= A \, \mathrm{d}x + B \, \mathrm{d}y.
\end{aligned} \tag{B16}
$$

A shorter but less instructive way of obtaining Equation B16 is to directly compute $\mathrm{d}z$ as

$$
\mathrm{d}z = \partial_x z \, \mathrm{d}x + \partial_y z \, \mathrm{d}y, \tag{B17}
$$

where $\partial_x z$ and $\partial_y z$ are obtained from Equation B11 assuming v and Ω as functions of position (x, y).

By integrating Equation B16, we can calculate the depth difference between any two points in the surface

$$
\Delta z_{12} = \int_{P_1}^{P_2} (A \, \mathrm{d}x + B \, \mathrm{d}y). \tag{B18}
$$

This is a line integral (also known as path integral). In general, its value is a function of the path chosen to go from P_1 to P_2. To have a consistent object, the value of the integral in Equation B18 must be independent of the path. This is also equivalent to stating that the integral of any closed path will give a value of zero. Using Stokes' theorem (Kaplan, 1952), it is easy to show that this happens if and only if $\partial_y A = \partial_x B$. An easy but lengthy calculation shows that this is indeed the case. Thus, our alternative to the Domini et al. hypothesis results in an internally consistent recovered depth structure. This structure, though, is not affine.

We can apply the same procedure to the Domini et al. hypothesis. Now, instead of Equation B11 we have (Domini & Braunstein, 1998)

$$
z = \frac{\sigma}{\sqrt{1 + \tau^2}} (x + \tau y), \tag{B19}
$$

where σ and τ are the slant and tilt of the surface, respectively. Using Equation B17 and assuming σ and τ as functions of position (x, y), we get, after a lengthy calculation, an equation similar to Equation B18. In this

case, however, $\partial_y A \neq \partial_x B$, which indicates that the surface is internally inconsistent, as expected.

Model variations

Our model of locally computed angular speeds agrees with the model of Domini et al. that SFM is not affine. Both models are consistent with psychophysical data for which the stimuli were piecewise planar surfaces and surfaces composed of planar patches smoothly joined (i.e., without edges) by a curved surface. But for fully curved surfaces, such as cylinders, both theories predict that $C_{\mathrm{obs}}/C_{\mathrm{sim}}$ should decrease with C_{sim}, contrary to our results (Figure 2b). There are a few ways to deal with this discrepancy. In the Domini et al. model, the dependence of perceived slant on def might differ between planes and strongly curved surfaces. Rather than being a decreasing function of def, it could be an increasing one for curved surfaces. In our version of the model, perceived angular speed of rotation could be taken as a decreasing function of def, rather than an increasing one. This change could be made without fundamentally altering other model predictions.

Another way to deal with the discrepancy between cylinders and piecewise planar surfaces is to resort in an alternative version of our model. In this version, which we call the nondifferential version, the angular speed of rotation is computed as an average across differentiable surfaces, rather than locally. The logic here is that some objects, such as those composed of multiple planar surfaces, might be perceptually segmented, with each patch possessing a different angular speed (or, in the Domini et al. model, slant). At the limit, a continuously differentiable surface, such as the surface of a cylinder, would be a single perceptual unit with a single angular speed. But the average def would have no role here, because averaged across the cylinder def is zero. An important aim of future research will be to uncover a general algorithm, one that applies not only to planar surfaces but also to any type of surface. Currently, no such algorithm exists. Yet, regardless of how the speed of rotation of a given surface patch is estimated, it is clear that the recovered structure will be affine for each patch, and each patch will be perceived as moving with its own angular speed, making the object nonrigid in the general case.

A similar distinction between differential and discrete interpretations arises for curved surfaces within the Domini and Braunstein (1998) framework. The depth difference between two points could be obtained either from the average slant across the surface between the points or by integrating the depth between the points as a function of the local slant at each intermediary point. Current data are compatible with either the differential or the nondifferential versions of both the Domini et al. model and the model presented here.

Acknowledgments

Commercial relationships: none.
Corresponding author: Julian M. Fernandez.
Email: julian_fernandez@isr.syr.edu.
Address: Institute for Sensory Research, Syracuse University, 621 Skytop Rd., Syracuse, NY 13224, USA.

References

Braunstein, M. L., Liter, J. C., & Tittle, J. S. (1993). Recovering three-dimensional shape from perspective translations and orthographic rotations. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 598–614. [PubMed]

Caudek, C., & Domini, F. (1998). Perceived orientation of axis of rotation in structure-from-motion. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 609–621. [PubMed]

Cornilleau-Peres, V., & Droulez, J. (1989). Visual perception of surface curvature: Psychophysics of curvature detection induced by motion parallax. *Perception & Psychophysics, 46,* 351–364. [PubMed]

Domini, F., & Braunstein, M. L. (1998). Recovery of 3-D structure from motion is neither Euclidean nor affine. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 1273–1295.

Domini, F., Caudek, C., & Proffitt, D. R. (1997). Misperceptions of angular velocities influence the perception of rigidity in the kinetic depth effect. *Journal of Experimental Psychology: Human Perception and Performance, 23,* 1111–1129. [PubMed]

Domini, F., Caudek, C., & Richman, S. (1998). Distortions of depth-order relations and parallelism in structure from motion. *Perception & Psychophysics, 60,* 1164–1174. [PubMed]

Eagle, R. A., & Blake, A. (1995). Two-dimensional constraints on three-dimensional structure from motion tasks. *Vision Research, 35,* 2927–2941. [PubMed]

Fernandez, J. M., Watson, B., & Qian, N. (2002). Computing relief structure from motion with a distributed velocity and disparity representation. *Vision Research, 42,* 883–898. [PubMed]

Foster, D. H., & Bischof, W. F. (1997). Bootstrap estimates of the statistical accuracy of thresholds obtained from psychometric functions. *Spatial Vision, 11,* 135–139.

Hogervorst, M., Kappers, A. M. L., & Koenderink, J. J. (1993). Perception of metric depth from motion parallax. *Perception, 22,* 101.

Kaplan, W. (1952). *Advanced calculus.* Reading, MA: Addison-Wesley Publishing Company, Inc.

Liter, J. C., Braunstein, M. L., & Hoffman, D. D. (1993). Inferring structure from motion in two-view and multi-view displays. *Perception, 22,* 1441–1465. [PubMed]

Loomis, J. M., & Eby, D. W. (1988). Perceiving structure from motion: Failure of shape constancy. In *Proceedings of Second International Conference on Computer Vision* (pp. 383–391). Washington, DC: Computer Society of the IEEE.

Norman, J. F., & Lappin, J. S. (1992). The detection of surface curvatures defined by optical motion. *Perception & Psychophysics, 51,* 386–396. [PubMed]

Norman, J. F., & Todd, J. T. (1993). The perceptual analysis of structure from motion for rotating objects undergoing affine stretching transformations. *Perception & Psychophysics, 53,* 279–291. [PubMed]

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. [PubMed]

Tittle, J. S., Todd, J. T., Perotti, V. J., & Norman, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 663–678. [PubMed]

Todd, J. T. (1998). Theoretical and biological limitations on the visual perception of 3D structure from motion. In T. Watanabe (Ed.), *High-level motion processing–Computational, neurophysiological and psychophysical perspectives* (pp. 359–380). Cambridge, MA: MIT Press.

Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics, 48,* 419–430. [PubMed]

Todd, J. T., & Norman, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception & Psychophysics, 50,* 509–523. [PubMed]

Werkhoven, P., & van Veen, H. A. (1995). Extraction of relief from visual motion. *Perception & Psychophysics, 57,* 645–656. [PubMed]