

# Planar motion permits perception of metric structure in stereopsis

JOSEPH S. LAPPIN and STEVEN R. LOVE  
*Vanderbilt University, Nashville, Tennessee*

A fundamental problem in the study of spatial perception concerns whether and how vision might acquire information about the metric structure of surfaces in three-dimensional space from motion and from stereopsis. Theoretical analyses have indicated that stereoscopic perceptions of metric relations in depth require additional information about egocentric viewing distance; and recent experiments by James Todd and his colleagues have indicated that vision acquires only affine but not metric structure from motion—that is, spatial relations ambiguous with regard to scale in depth. The purpose of the present study was to determine whether the metric shape of planar stereoscopic forms might be perceived from congruence under planar rotation. In Experiment 1, observers discriminated between similar planar shapes (ellipses) rotating in a plane with varying slant from the frontal-parallel plane. Experimental conditions varied the presence versus absence of binocular disparities, magnification of the disparity scale, and moving versus stationary patterns. Shape discriminations were accurate in all conditions with moving patterns and were near chance in conditions with stationary patterns; neither the presence nor the magnification of binocular disparities had any reliable effect. In Experiment 2, accuracy decreased as the range of rotation decreased from 80° to 10°. In Experiment 3, small deviations from planarity of the motion produced large decrements in accuracy. In contrast with the critical role of motion in shape discrimination, motion hindered discriminations of the binocular disparity scale in Experiment 4. In general, planar motion provides an intrinsic metric scale that is independent of slant in depth and of the scale of binocular disparities. Vision is sensitive to this intrinsic optical metric.

Everyday human experience and performance in perceiving constant object shapes from changing perspectives in varying contexts indicate that vision acquires information about the spatial structure of environmental objects with impressive speed, reliability, precision, and generality. Despite the familiarity and functional importance of these visual achievements, we lack a theoretical understanding of the optical information and visual mechanisms on which they are based.

Indeed, the specific geometric relations that might be used by vision to perceive shape have seldom been directly evaluated by psychophysical techniques. Sperling, Landy, Doshier, and Perkins (1989) recently criticized the literature on perceiving three-dimensional (3-D) structure from motion for having failed to provide such evidence. The apparent 3-D shape constancy of perceived environmental objects, however, suggests that vision usually provides information about the metric structure of environmental

surfaces; and the precision of visual-motor coordination exhibited in athletics seems to indicate that vision also provides metric information about the locations and trajectories of environmental objects relative to the observer's position.<sup>1</sup>

Recent experiments on this issue, however, have shown that metric relations in 3-D are not perceived in many cases. Todd and Reichel (1989) found that depth relations (perpendicular to the frontal-parallel plane) among neighboring points on curved surfaces, seen in 3-D space by virtue of shading and texture gradients, were discriminated as if the perceived depth relations in empty Euclidean 3-D space ( $E^3$ ) were only ordinal and not metric. In more recent experiments on the kinetic depth effect (KDE), Todd and Bressan (1990) and Todd and Norman (1991) have found that observers are unable to discriminate relative lengths of line segments or to discriminate differences in the depth of smooth surfaces. They have concluded that vision probably obtains information only about affine and not about metric structure from KDE patterns—that is, that the relative scale of the depth axis is visually indeterminate in KDE patterns.<sup>2</sup>

Evidence suggesting that observers can accurately discriminate the curvature and depth of rotating surfaces of revolution was obtained by Todd (1984) in experiments done with cylindrical surfaces with varying radii of curvature rotating around their central axes in the frontal-parallel plane. As Todd pointed out, however, these discriminations could have been based on differences in ve-

---

This research was supported in part by NIH Grants EY05926 and P30EY08126. Experiments 1-3 were conducted by S. R. Love as part of the honors program in psychology at Vanderbilt. We are grateful to Farley Norman and John Kuster for their help in conducting Experiment 4, to Warren Craft and Rebecca Fisher for help in preparing several figures, to Randolph Blake, Myron Braunstein, James Todd, and an anonymous reviewer for comments on an earlier version of this paper, and to James Todd and Farley Norman for numerous discussions of the theoretical issues. Correspondence should be addressed to Joseph S. Lappin, Department of Psychology, Vanderbilt University, Nashville, TN 37240.

locity rather than differences in depth or curvature per se. Indeed, in subsequent experiments, Todd (personal communication, November 1990) found that these surfaces cannot be discriminated if the rotational velocities are randomly varied. Thus, the metric structure of these surfaces seems to be indiscriminable.

Can vision discriminate metric structure from motion under any conditions? In contrast with the results of Todd and his colleagues (Todd & Bressan, 1990; Todd & Norman, 1991), Lappin and Fuqua (1983) found accurate bisection discriminations of 3-D distances in kinetic depth displays of three collinear points rotating in a slanted plane. The perceived spatial relations in these experiments were not merely affine, because (1) the patterns in most conditions were displayed with exaggerated perspective rather than by orthographic projection, so that the relative spacing among the points in the image plane differed from that in  $E^3$  and changed with the angle of rotation; (2) the three collinear points appeared subjectively to move rigidly in depth, with the distances between points appearing to remain constant under changing directions in space; and (3) observers were inaccurate in bisecting projected lengths in the two-dimensional (2-D) image plane. The discrepancy between these results and those of Todd and Bressan (1990) and Todd and Norman (1991) may well have derived from differences in differential structure of the retinal image motions in the two sets of studies: The projected velocity fields in Lappin and Fuqua's (1983) displays remained constant over time, corresponding to rotation within a slanted plane, whereas those of Todd and Bressan (1990) and Todd and Norman (1991) changed over time at any given position in the image, corresponding to depth rotation of volumetric and solid objects. Thus, Lappin and Fuqua's (1983) patterns satisfied the "planarity constraint" that Hoffman and Flinchbaugh (1982) have shown is theoretically sufficient for metric structure to be determined from two successive views, whereas those of Todd and Bressan (1990) and Todd and Norman (1991) did not satisfy this constraint.

The perceptual utility of this planarity constraint was studied more extensively in the present experiments. A main objective was to determine whether such planar rotation in depth was sufficient to provide accurate perception of metric structure in stereoscopic patterns. To anticipate, we found that the metric shape of a planar form rotating in a slanted plane could be accurately discriminated, even when the binocular disparities were exaggerated and unreliable! Without such rotation, however, stereopsis permitted only poor discrimination of the shapes of stationary forms in depth.

**Stereoscopic Depth Constancy**

This investigation was motivated in part by the geometric fact that the depth separation between two environmental points depends not only on the binocular disparity between their two monocular half-images but also on their egocentric distance from the observer. Small changes in viewing distance yield large changes in depth. The prob-

lem of stereoscopic depth constancy is the problem of how stereopsis might provide constant information about spatial relations in depth, independently of changes in viewing distance (Cormack, 1984; Fox, Cormack, & Norman, 1987; Mauk, Crews, & Fox, 1987; Ono & Comerford, 1977). (A schematic illustration of this relationship between binocular disparity, viewing distance, and depth is given in Figure 1.)

When two points lie within a few degrees of the perpendicular bisector of the line between the nodal points of the two eyes, the depth separation,  $z$ , between the two points is given by

$$z = (I/2)\{\tan[\tan^{-1}(2D/I) + \delta/2]\} - D, \quad (1)$$

where  $I$  is the interocular distance,  $D$  is the distance to the nearer point, and  $\delta$  is the angular binocular disparity. If  $D$  is the viewing distance to the farther of the two points,  $\delta$  and  $z$  are negative. A more general equation for the case in which the two points are not necessarily centered between the eyes is given by Cormack and Fox (1985).

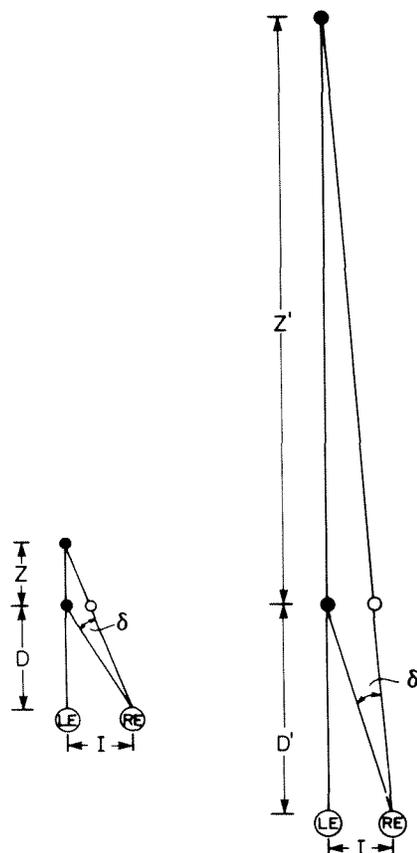


Figure 1. An illustration of the relationship between viewing distance,  $D$ , and depth,  $Z$ , for a constant binocular disparity  $\delta$ . [From "Perceiving the metric structure of environmental objects from motion, self-motion, and stereopsis," by J. S. Lappin, 1990, in R. Warren & A. H. Wertheim (Eds.), *Perception and control of self-motion* (pp. 541-578). Hillsdale, NJ: Erlbaum. Copyright 1990 by Lawrence Erlbaum Associates. Reprinted by permission.]

Equation 1 describes a distinctly nonlinear relation: When the depth is small relative to the viewing distance, then, for a constant binocular disparity, the depth separation is roughly proportional to the *square* of the viewing distance.

If the relationships illustrated in Figure 1 adequately describe the geometry of stereopsis, uncertainties about viewing distance should have the following effect on the stereoscopically perceived shapes of solid objects: A variety of different potential shapes might be perceived, depending on the estimated viewing distance. A convex spherical surface, for example, would appear spherical only when its distance was correctly estimated; if the estimated distance was too small, the curvature of the surface would appear reduced, yielding an ellipsoid flattened in the direction of gaze; and if the estimated distance was too great, the surface would appear elongated like a football in the direction of gaze. Whenever the estimated viewing distance was incorrect, the perceived length of a contour would vary with its orientation, stretching and compressing as the object rotated in depth.

One might guess that the viewing distance and thus the depth could be obtained by triangulation, using the vergence angle between the two eyes to the point of fixation. Although psychophysical evidence indicates that variations in vergence angle do influence perceived depths (Foley, 1967), changes in viewing distance produce only small changes in vergence angle but large changes in depth. That is, the partial derivative of depth relative to vergence angle is very large and nonlinear. Therefore, vergence angle is a poor source of information with which to scale stereoscopic depth. Moreover, for viewing distances beyond about 2 m, the vergence angle is essentially constant, but the apparent depth of stereoscopic afterimages (i.e., with fixed disparity) increases with viewing distances on the order of thousands of meters (Cormack, 1984).

Thus, the binocular disparities in a single pair of monocular half-images of a stationary object do not determine the metric structure of the object (cf. Foley, 1980; Koenderink & van Doorn, 1991); the scaling of distances in depth relative to distances in the frontal-parallel plane is ambiguous. The KDE information provided by two successive views of a moving object is no less ambiguous. Moreover, the results of Todd and Bressan (1990) and Todd and Norman (1991) suggest that vision is insensitive to the added geometric constraints associated with three or more views and that only affine spatial structure can be perceived from motion.

Because of this inherent ambiguity about the scaling of distances in depth, many investigators (e.g., Longuet-Higgins, 1986; Marr, 1982; Mayhew & Longuet-Higgins, 1982) have concluded that the retinal optical information from stereopsis and KDE must be supplemented by non-retinal information. The present study, however, demonstrates that retinal patterns of planar rotation can be sufficient for the perception of metric structure.

### Spatial Structure from Congruence under Rotation

Two successive (monocular) views of a rotating object or two simultaneous stereoscopic views of a stationary object permit a one-parameter family of perceived structures (Bennett, Hoffman, Nicola, & Prakash, 1989). With orthographic projections, these potential alternative structures differ from one another by affine transformations of the scale of depth relative to the frontal-parallel plane (Koenderink & van Doorn, 1991; Todd & Bressan, 1990); and with perspective projections, when the viewing distance is small or when viewing is stereoscopic, the family of potential alternative structures includes additional variations in shape associated with variations in viewing distance. Additional views of a rotating object produce additional image transformations, and the scaling parameters derived from these multiple transformations must agree if the object remains isometric under these motions. Thus, the rigidity or congruence of moving surfaces and contours might be used as the basis for scaling distances in  $E^3$ .<sup>3</sup>

The logic of this idea was cogently expressed by Killing (1892). Two spaces are congruent if and only if they can be covered by the same object; two objects are congruent if and only if they can cover the same space. Thus, objects and spaces constitute mutually constraining relational structures. The structure of both may be derived from congruence under motion.

Under what conditions might vision be sensitive to this potential geometric information? Different groups of motions produce projected images that differ in the potential visibility of information about 3-D spatial structure. One noteworthy distinction involves translation versus rotation in depth. Although translations and rotations may produce image motions that are practically indistinguishable locally, rotations are globally simpler because they can be characterized by a single free parameter corresponding to the angular extent of rotation, whereas translations require more parameters. Moreover, "motion parallax" patterns arising from lateral translation often appear to be rotations (Braunstein & Andersen, 1981; Lappin, 1991). Thus, vision seems to represent differential local image motions by 3-D rotations.

Other potentially important distinctions concern (1) the complexity of the moving surface structure (e.g., a single smooth surface vs. multiple separate surfaces or disconnected features), (2) opaque versus transparent surfaces, (3) the complexity of the surface's trajectory in space-time (e.g., a surface of revolution vs. a volume of space-time), (4) the angle between the surface and the axis of rotation, and (5) the angle between the axis of rotation and the projected image. All of these characteristics affect the complexity of the differential structure of the projected image motions and thus affect the potential visibility of the underlying spatial structure. The quantitative effects of these conditions on the visibility of 3-D spatial

structure have not yet been clearly established by psychophysical research, however.

The present experiments examined a particularly simple case, in which planar shapes were rotated in a slanted plane. In this case, the sequential positions of the moving object formed a planar surface in space-time—a 2-D manifold. Additionally, the monocular half-images were effectively orthographic projections, with negligible perspective. Thus, the projective mapping of the moving form onto its monocular images could be described by a single affine coordinate transform. Likewise, the metric tensor parameters for embedding the monocular images of the form into  $E^3$  were also constant over retinal positions and constant over time (see the Appendix). This condition satisfies what Hoffman and Flinchbaugh (1982) termed the *planarity constraint*. They proved that, under this condition, two views of three or more points (which need be only piecewise rigidly connected) are sufficient to determine the 3-D positions and motions of the points up to a reflection about the frontal-parallel plane. A different proof of the sufficiency of this condition for recovering metric structure and motion is given in the Appendix. The hypothesis that human vision can exploit this planarity constraint, however, has not previously been tested experimentally. The present study provides such a test, and it demonstrates that metric structure can indeed be accurately perceived in this case.

The hypothesis that *stereoscopic* perception of 3-D spatial structure may be derived from congruence under motion has not been tested directly, but indirect evidence is provided by several experiments by Wallach and his colleagues (Wallach & Karsh, 1963a, 1963b; Wallach, Moore, & Davidson, 1963) and by Fisher and Ebenholtz (1986). Using a telestereoscope that magnified binocular disparities by increasing the optical distance between the eyes, observers viewed wire forms rotating around a vertical axis. Because of the exaggerated disparities that expanded the depth axis, these forms appeared to deform as they rotated. Wallach & Karsh (1963a, 1963b) and Wallach et al. (1963) found that after 10 min of passive observation of such rotating forms there was a reduction in the apparently expanded depth of a stationary form judged after the 10 min "training" period. Fisher and Ebenholtz (1986), however, concluded that this depth aftereffect did not involve adaptive recalibration of stereopsis by motion, since the aftereffects could be obtained after exposure to either stereopsis alone with no motion or to motion alone with no stereoscopic disparity. Moreover, artificial pupils eliminated these aftereffects, suggesting that they derived from changes in perceived distance associated with accommodation rather than from recalibration of either stereopsis or KDE. It should be noted that the differential structure of the images in these experiments was both complex and continually changing, because the sequential positions of the forms occupied a volume rather than a surface in  $E^3$ . Constant spatial structure in depth should have been difficult to perceive under these conditions, and the results of Fisher and Ebenholtz suggest that such constancy was not perceived.

The present study investigated real-time visual scaling of stereoscopic shapes rotating in a slanted plane. In Experiment 1, form discrimination was compared in conditions that varied in the presence or absence of (1) stereoscopic disparities, (2) distorted magnifications of the disparities, and (3) motion. The effects of the range of rotation and the complexity of the differential structure of the projected image motions were evaluated in Experiments 2 and 3. The opposite effects of motion on discriminations of form and on discriminations of the stereoscopic scale of depth were contrasted in Experiment 4.

## EXPERIMENT 1

The aim of Experiment 1 was to determine whether rotation of a form in depth could provide sufficient visual information about its metric structure to correct for spatial distortions produced by abnormal increases in binocular disparity. If stereoscopic perception of spatial structure requires extraretinal information about viewing distance, form perception should be seriously impaired by stereoscopic displays in which the disparities are magnified. Such stereoscopic distortion should produce an elongation of the form in depth relative to its extension in the frontal-parallel plane. On the other hand, if stereoscopic space can be scaled by congruence under planar rotations, such magnifications of the disparities may have little or no effect on the perceived shapes of the rotating forms.

The spatial forms employed in all the experiments were ellipses—approximately circular plane shapes differing only slightly in the length of one axis. In most experimental conditions, these forms were tilted in depth from the frontal-parallel plane by rotation around the horizontal axis. Of course, the projected shapes of these forms in the plane of the display screen were also ellipses contracted in the vertical axis of the screen, with the degree of contraction varying with the magnitude of slant and magnification of the stereoscopic disparities. The values of both the slant and stereoscopic disparities were randomly varied between trials, so that the projected shapes varied as a conjoint function of the slant, disparity, and rotational position in addition to the actual shape of the ellipse in  $E^3$ . The projected image shapes depended less on the shape of the ellipse in  $E^3$  than on the other three variables, and the image velocities also depended more on slant than on shape. Therefore, discriminations of these shapes required reliable information about metric structure in depth.

### Method

**Design.** Eight experimental conditions were obtained from the orthogonal combinations of three variables—moving versus stationary, stereoscopic versus nonstereoscopic (in which the patterns on the two monitors were identical), and a simulated viewing distance that was either correct or reduced and variable. (Reduction of the simulated viewing distance magnified the binocular disparities; the perspectivity of the monocular images was negligible.) In all these conditions, the stimulus forms were displayed as if slanted in depth by an amount that varied randomly between trials. In 2 additional conditions, the forms were presented only in the frontal-parallel

plane, either moving or stationary. The accuracy of form discrimination under each of these 10 conditions was evaluated in a separate block of trials.

**Apparatus and optical patterns.** The optical patterns were displayed on two cathode-ray tube (CRT) monitors (Tektronix 608, with P-31 phosphor) controlled by a computer (PDP-11/73) with a 12-bit digital-to-analog (D/A) interface (Data Translation DT2771). The two CRT displays were haploscopically combined by prisms and mirrors adjusted for each observer to match individual vergence and interocular separation. These displays were viewed from a distance of about 114.6 cm, with the observer's head position constrained by the two prisms in front of each eye. The binocular disparity of the two monocular patterns was scaled for an interocular separation of 6.3 cm, regardless of the observer's actual interocular separation. Each CRT screen was seen through a circular aperture 10 cm ( $5^\circ$  of visual arc) in diameter in a black baffle mounted immediately in front of the monitor.

The CRTs were viewed in a dimly lit room that provided ample visual cues about the distance of the displays. The background luminance on the CRT screens was about  $0.06 \text{ cd/m}^2$ , and the luminance of the individual points was about  $6 \text{ cd/m}^2$ , as measured by a Pritchard spot photometer from the observer's eye position. The diameter of an individual point was about  $0.25 \text{ mm}$  ( $45''$  of arc). The spatial resolution provided by the 12-bit D/A interface was  $.021 \text{ mm}$  ( $3.78''$  of arc). Corresponding points on the two CRTs were alternately refreshed with an interval of approximately  $16 \mu\text{sec}$  between successive pulses from the D/A interface. With 10 points on each CRT, the interval between successive refreshes of the same point was about  $320 \mu\text{sec}$  ( $= 16 \mu\text{sec} \times 2 \times 10$ ). When these patterns were moved, their positions were changed every 18 msec. The forms were rotated through an angle of  $90^\circ$ ,  $\pm 45^\circ$  from the vertical midline of the display, around a point midway between the center and the bottom edge of the ellipse. The forms rotated back and forth three times with an angular velocity of  $135^\circ/\text{sec}$  for 2 sec, occupying 38 separate positions over the  $90^\circ$  range of rotation.

The shapes to be discriminated were three ellipses defined by 10 dots equally spaced around the perimeter. When the forms were displayed in the frontal-parallel plane, the vertical axis was  $.94^\circ$ ,  $1.0^\circ$ , or  $1.06^\circ$  of arc, and the horizontal axis was always  $1.0^\circ$  of arc. Each of these three shapes occurred equally often in each block of trials.

**Conditions.** In 8 of the 10 experimental conditions, the ellipses were presented in a plane slanted in depth from the frontal-parallel plane by  $40^\circ$ ,  $50^\circ$ , or  $60^\circ$  around the horizontal axis. Each of these three alternative slants appeared equally often in a random sequence within each block of trials. In the other two conditions, the form was always in the frontal-parallel plane. Figure 2 shows the monocular projections of the three ellipses in the frontal-parallel plane and at the three values of slant.

In two conditions, the binocular disparities were displayed as if the observer's viewing distance was much closer than the actual 114.6 cm by an amount that varied randomly between trials—12.6, 18.9, or 25.2 cm. The width of the forms on the screen was held constant, independently of the simulated viewing distance. These varying viewing distances magnified the stereoscopic depth by an amount equal to the ratio of actual to simulated viewing distance. The amount of stereoscopic magnification also depended on the slant of the form, the predicted elongation being greatest for the  $60^\circ$  slant and 12.6-cm simulated viewing distance. This stereoscopic magnification would have lengthened the vertical axis of the circular shape from an undistorted value of 2.0 cm (measured in the slanted plane of the form) to a minimum value of 6.04 cm with a 25.2-cm simulated viewing distance and  $40^\circ$  slant, and a maximum value of 15.78 cm with the 12.6-cm viewing distance and  $60^\circ$  slant. A schematic illustration of this viewing situation is shown in Figure 3.

These two distorted-disparity conditions differed according to whether the forms were moving or stationary. In two correspond-

ing conditions with no stereoscopic distortion, the simulated viewing distance was constant at the correct value of 114.6 cm. In four corresponding nonstereoscopic conditions, the monocular patterns on each eye were identical—that is, they had no binocular disparity. In two of these nonstereoscopic conditions, the simulated viewing distances were also shortened and variable, just as above, although these variations in viewing distance had only a negligible effect on the projected monocular patterns.

Thus, the principal experimental comparisons involved the effects of stereoscopic distortion and motion on form discriminations. If visual information about spatial structure in depth can be obtained from the congruence of moving forms, independently of the specific disparities, then discriminations of the moving forms should be unaffected by the varying distortions of stereoscopic space. Discriminations among the stationary forms, however, should be seriously hindered by the stereoscopic distortions. Comparisons involving the other four nonstereoscopic conditions were included for control purposes. Two additional frontal-parallel control conditions provided evaluations of shape discriminations of 2-D forms that were either moving or stationary.

**Procedure.** The five observers included the 2 authors plus 3 paid volunteers who were undergraduate students in psychology at Vanderbilt. All 5 had normal or corrected-to-normal visual acuity; all had some understanding of the purpose of the experiment; all had participated in at least two practice sessions prior to the experimental data collection—until form discrimination had reached asymptotic accuracy under the condition in which the forms were moving, slanted in depth, and stereoscopic with correct and constant disparities.

Each of the observers participated in eight experimental sessions. Each experimental session consisted of six blocks of 27 trials. The first block in each session was a practice block in which the forms were moving, stereoscopic with correct and constant disparities, and slanted by  $40^\circ$ ,  $50^\circ$ , or  $60^\circ$  on each trial. In the next four blocks of each session, the forms were always either moving or stationary, with order of these two conditions counterbalanced over sessions and over observers. The conditions differed among these four trial blocks, involving the 4 conditions with and without binocular disparity and with simulated viewing distances either correct and constant or shortened and variable (12.6, 18.9, or 25.2 cm, randomly varying between trials). The ordering of these 4 conditions within sessions was also counterbalanced over sessions and observers. In the final block of trials in each session, the forms were always presented in the frontal-parallel plane. In these frontal-parallel control conditions, the forms were either moving or stationary, consistent with the preceding four blocks of trials. Thus, each of the 5 observers contributed data from 108 trials for each of the 10 experimental conditions.

The observer's task in all these conditions was the same—to identify which one of the three alternative shapes was presented on each trial. Each trial was initiated by the observer by depressing a switch on a keypad. The pattern was displayed for 2 sec. The observer responded by pressing one of three alternative keys. The response was followed by visual feedback indicating which of the three shapes had appeared on that trial. A fixation point reappeared on the screen when the next trial was ready.

## Results and Discussion

The principal results are shown in Figure 4, which gives the percentage of correct shape-discrimination responses for each condition and each observer as well as the combined accuracy totaled over all 5 observers. These results are easily summarized: Discriminations among moving shapes were very accurate, but discriminations were not reliably above chance when the forms were stationary and

slanted in depth. Variations in either the correctness or even the presence or absence of binocular disparities had no reliable effect on performance. Surprisingly, the moving forms were more accurately discriminated even in the frontal-parallel plane.

Subjectively, the shapes were more definite and distinct when they were moving. Since the orientations of the two axes of the ellipses rotated through an angle of 90°, the relative lengths of these two axes were visually scaled independently of the space in which they were located at any time—in  $E^3$ , on the display screen, or on the retina.

Such intrinsic optical information about the metric scale of the form and the space added to the visibility of the shape even when the forms were confined to the frontal-parallel plane. The observer's knowledge that the form was not slanted may not have fully resolved inherent visual ambiguity about its slant. A metric scale of the retinal coordinates may be less perceivable than the intrinsic geometry of the optical stimulation.

The failure of stereopsis to provide accurate shape discrimination even when the stationary slanted forms were displayed with the correct binocular disparities is interest-

SHAPES

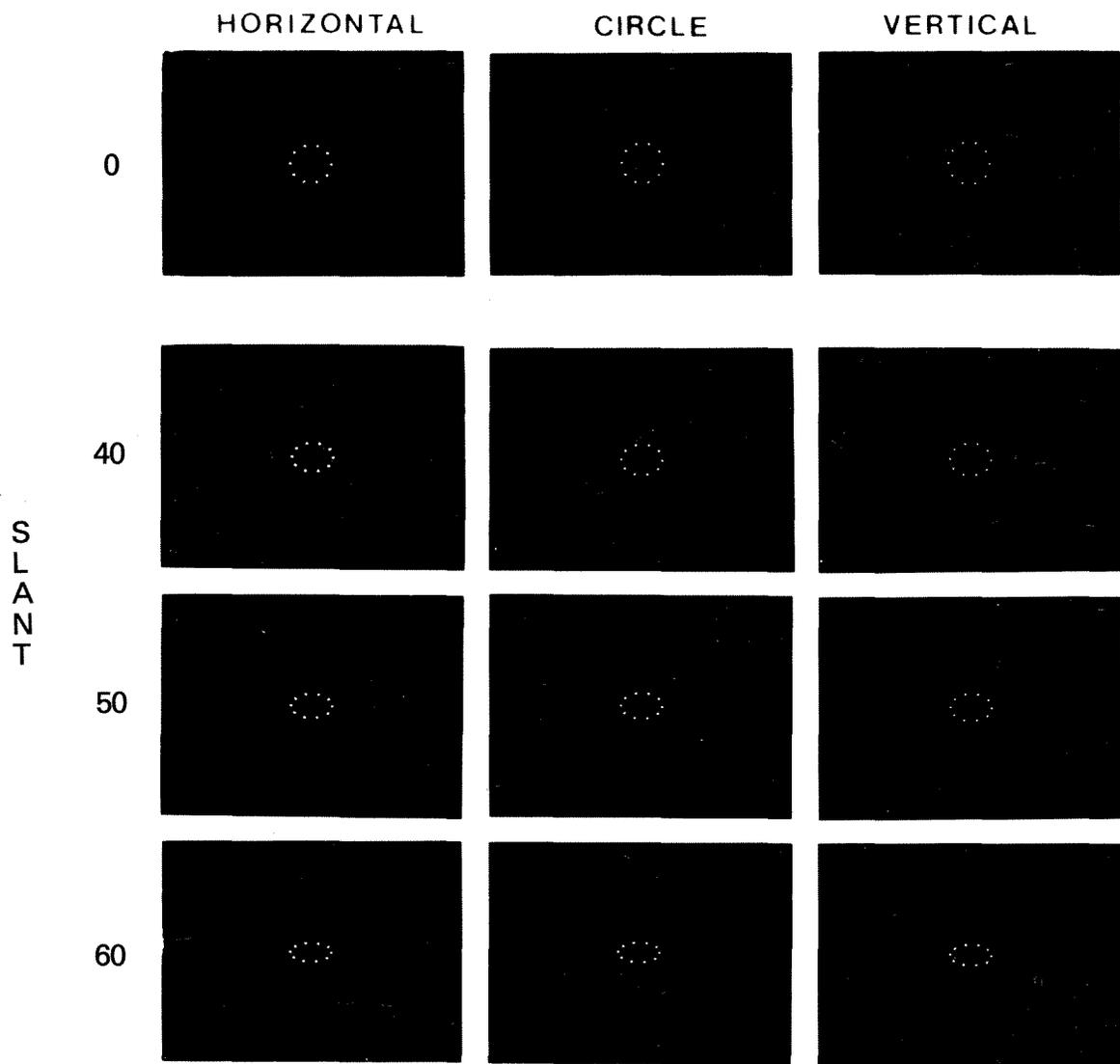


Figure 2. Photographic reproductions of the stationary images of the three shapes at each of four different magnitudes of slant. [From "The perception of geometrical structure from congruence," by J. S. Lappin & T. D. Wasson, 1991, in S. R. Ellis, M. K. Kaiser, & A. J. Grunwald (Eds.), *Pictorial communication in virtual and real environments* (pp. 425-448). London: Taylor & Francis. Copyright 1991 by Taylor & Francis Ltd. Reprinted by permission.]

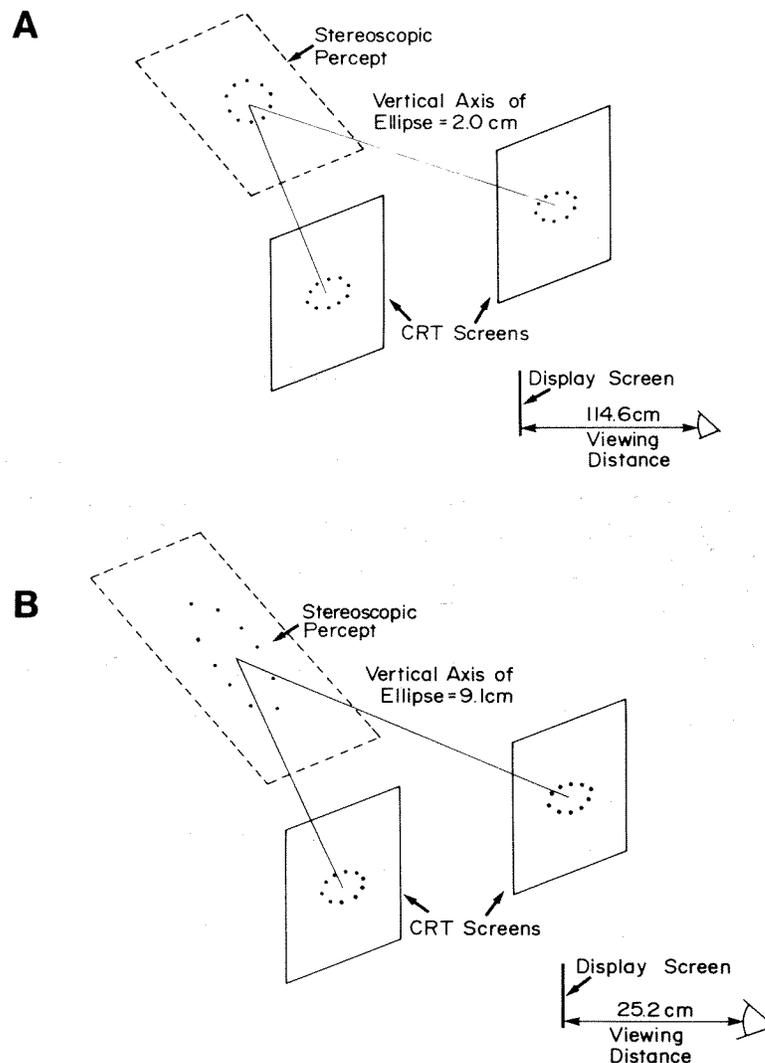


Figure 3. Schematic illustrations of the apparent shapes of a tilted circle as seen under two different conditions of stereoscopic projection. (A) The perceived shape of a circle slanted  $40^\circ$  as seen with the correct disparities at a distance of 114.6 cm. (B) The perceived shape of a circle slanted  $40^\circ$  as seen from a distance of 114.6 cm with the disparities appropriate for a viewing distance of 25.2 cm.

ing and was not anticipated. Unfortunately, the poor performance in this condition violates part of the rationale for this experiment: Since correct binocular disparities did not yield reliable discrimination of the stationary forms, accurate discrimination of the stereoscopically distorted moving forms cannot be attributed to stereopsis per se; the role of stereopsis in this shape-discrimination task is ambiguous.

Even though the obtained discrimination accuracies do not demonstrate a role of stereopsis in this task, the subjective impression was that both stationary and moving stereoscopic forms appeared to be enriched in depth. When the stationary forms were displayed with magnified disparities, they did in fact appear abnormally elongated in depth. One observer (J.S.L.) did discriminate the correctly disparate stationary slanted shapes with an ac-

curacy well above chance, but the other observers did not significantly exceed chance accuracy in this baseline condition. Even with the correct binocular disparities, the precise slant and shape in depth were surprisingly ambiguous. (In Experiment 4, a different method was used to demonstrate that variations in binocular disparities do affect perceived shape and depth in these patterns.)

The specific geometric properties that observers discriminated in this task were also described by a more detailed analysis of the stimulus-response correlations. The optical patterns in four experimental conditions consisted of the 27 combinations of three shapes, three slants, and three viewing distances. Although the observers were explicitly instructed to identify the shapes, their responses may have been guided by any combination of these three variables. To quantify the relative influences of these three

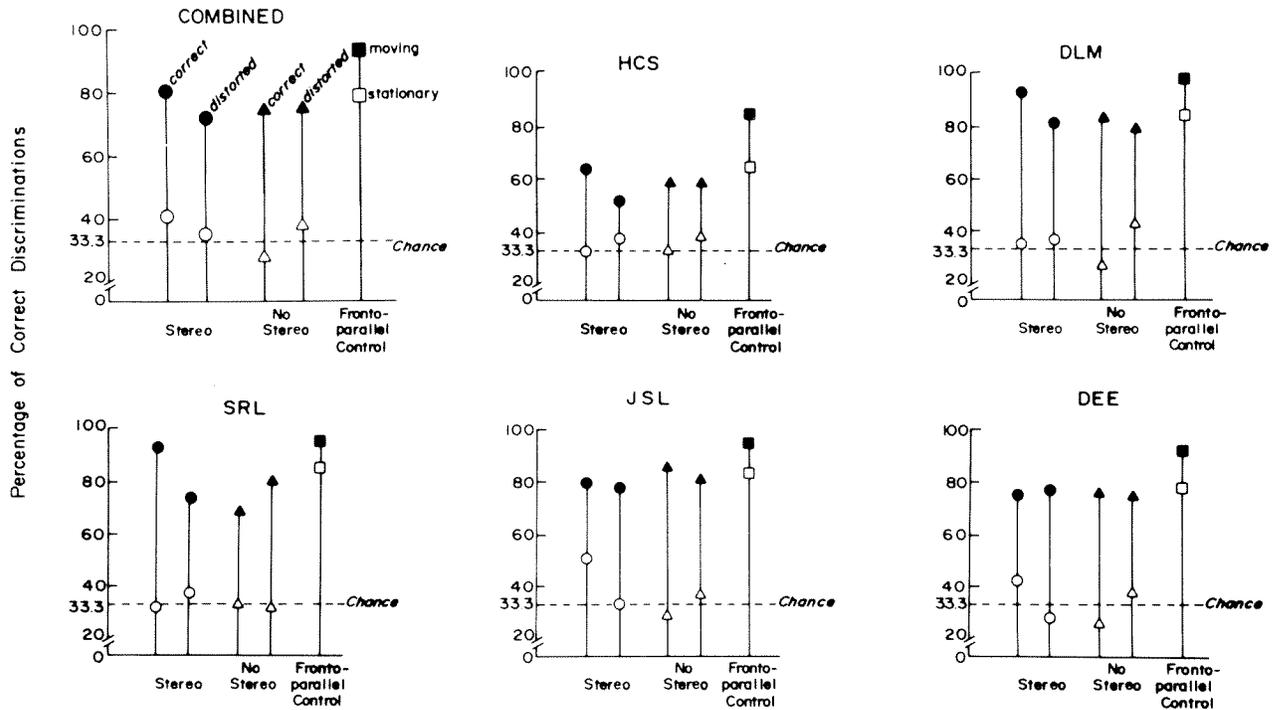


Figure 4. Accuracies of shape discrimination under each of 10 experimental conditions in Experiment 1 for each of 5 observers, and the corresponding average accuracies combined over the 5 observers.

variables, we calculated the information transmitted (in bits) about each of the variables by each observer's responses in each of the four relevant conditions. These stimulus-response correlations reflect the degree to which observers were able to distinguish variations in shape from variations in slant and disparity.

The results of this analysis of the combined responses of all 5 observers are shown in Figure 5. As can be seen, when the shapes were moving, the responses were correlated almost exclusively with shape. (These stimulus-response correlations were very similar among all the observers except H.C.S., who was less accurate than the other observers and whose discriminations of the moving shapes were influenced by slant about as much as by shape.) When the shapes were stationary, however, the observers' responses were influenced more by slant and by disparity than by shape—that is, observers were unable to discriminate variations in stationary shape from variations in slant and disparity. Thus, the optical information provided by motion was necessary for perceiving the intrinsic geometric shape independently of its slant and the scale of its stereoscopic disparities.

To summarize, the principal result was the clear and consistent superiority in discriminating moving as opposed to stationary shapes. Variations in binocular disparities had virtually no effect on discrimination of the moving shapes. Rotation in depth was necessary and sufficient for accurately perceiving metric shape in  $E^3$ .

## EXPERIMENT 2

Reductions in the range of rotation might be expected to produce concomitant reductions in visible information about the relative scaling of distances in the two perpendicular directions in the image plane. These effects were examined in Experiment 2.

### Method

With stimulus patterns and displays identical to those in the previous experiment, the range of angular rotation in this experiment was restricted to either 80°, 40°, 20°, or 10°. As before, these rotational motions occurred in a plane that was slanted in depth around the horizontal axis by 40°, 50°, or 60°, randomly varying between trials. The rotations were centered at the vertical midline of the display. The velocity of rotation was constant at 135°/sec, so that the number of cycles occurring during the 2-sec display duration increased as the range of rotation decreased. For each of these values of angular rotation, we also compared conditions in which stereoscopic forms were displayed with the correct or with incorrect and variable disparities, using the same values of simulated viewing distances as in the previous experiment. Thus, there were eight experimental conditions—four ranges of rotation and two types of stereoscopic disparity, orthogonally combined.

Three well-trained volunteers served as observers. All 3 had served in the previous experiment, and all were familiar with the purpose of the experiment. Each observer participated in four sessions, each of which comprised four blocks of 54 trials. Two different ranges of rotation were examined in each session, with two of the trial blocks devoted to the correct-disparity conditions and two to the distorted-disparity conditions.

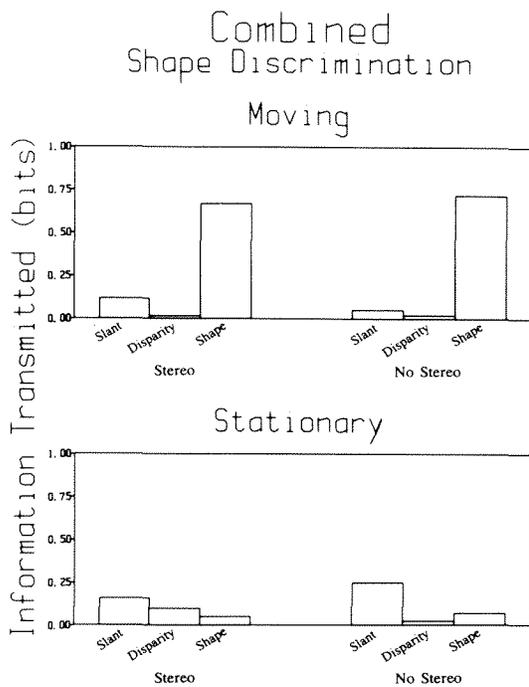


Figure 5. Amount of information transmitted by the shape-discrimination responses about the three independent stimulus variables—slant, disparity (as specified by the viewing distance parameter), and shape—for each of the four conditions in Experiment 1 in which these properties were varied. (Under the two non-stereoscopic conditions, the variations in viewing distance had a negligible effect on the projected patterns, and therefore the responses would not be expected to correlate with this variable.)

## Results

The results are shown in Figure 6, which gives the shape-discrimination accuracies for each observer and condition. As can be seen, discrimination accuracy declined systematically with reductions in the range of rotation. Performance was similar for both the correct-disparity and the distorted-disparity conditions at all ranges of rotation. Rotations of only  $10^\circ$  were sufficient to produce above-chance shape discriminations, even when the stereoscopic disparities were seriously distorted.

## EXPERIMENT 3

Theoretical analyses have indicated that the metric structure of a rotating form may be determined by two successive views only when both the form and motion lie within a single plane (Hoffman & Flinchbaugh, 1982; Appendix). Although Ullman (1979) and others have shown that full metric information about a rigidly rotating form and the  $E^3$  Euclidean space in which it is moving might in principle be determined from three or more views, the results of Todd and Bressan (1990) and Todd and Norman (1991) indicate that vision does not integrate such 3-D structure over three or more views. In general, the successive positions of a moving surface may occupy a volume rather than a surface in  $E^3$ —a 3-D manifold. Ac-

cordingly, the  $E^3$  embedding of the images of such moving forms must change over time to preserve the metric structure of the form. In contrast, the rotating forms in Experiments 1 and 2 occupied only a plane, and the embedding of their images into  $E^3$  remained constant over time. Clearly, recovery of the metric shape of a moving object is computationally much more difficult in the general case of arbitrary motions of solid objects. In Experiment 3, these ideas about the visual information about metric shape were tested by varying the departure of the trajectory of the rotating ellipse from a plane.

## Method

The independent variable in Experiment 3 was the angle between the ellipse and the axis of rotation. In Experiments 1 and 2, this angle was always  $90^\circ$ , so that the ellipse rotated in a single plane throughout its trajectory. In Experiment 3, this angle was reduced, yielding a curvilinear trajectory in which the successive positions of the ellipse were tangent to the surface of an imaginary cone (where the vertex coincided with the plane of the display screen at the center of rotation, the base of the cone was behind the display screen, and the axis of rotation, which was the central axis of the cone, was tilted around the horizontal axis of the display screen). As the angle between the ellipse and the axis of rotation decreased, the curvature of the trajectory increased and the variability of the tilt and slant of the ellipse also increased.

Four alternative values of this angle were used:  $90^\circ$  (replicating the preceding experiments),  $80^\circ$ ,  $70^\circ$ , and  $60^\circ$ . The value of this angle parameter was constant throughout each block of trials and was varied between the four blocks of 27 trials which composed each experimental session. Within each trial block, the slant of the axis of rotation was randomly varied among the three alternative values, which were  $40^\circ$ ,  $50^\circ$ , or  $60^\circ$ , as in Experiments 1 and 2. The same three alternative ellipses as those employed in Experiments 1 and 2 were viewed stereoscopically with the correct viewing distance parameter (114.6 cm). As in Experiment 1, the range of motion was constant at  $90^\circ$ . Other conditions and procedures were the same as in Experiments 1 and 2.

Three experienced observers who had served in Experiment 1 each served for four sessions in Experiment 3. (Experiment 3 was conducted before what is described as Experiment 2 in the present study.) There were four blocks of 27 trials in each session, for a total of 108 trials for each observer in each condition.

## Results

The results are displayed in Figure 7, which shows the effect on shape discrimination of the angle between the ellipse and axis of rotation. As can be seen, small amounts of curvature of the surface of rotation produced substantial decrements in accuracy. Merely changing the angle between the ellipse and axis of rotation from  $90^\circ$  to  $80^\circ$  more than tripled the percentage of errors (from about 10% to 36%) by the most accurate observer, D.L.M., and doubled the total errors of all 3 observers (from about 21% to 42%). Angles of  $70^\circ$  and  $60^\circ$  yielded accuracies only slightly above chance (44% and 48%, respectively, for the 3 observers combined). Subjectively, the increased curvature of the surface of rotation simply rendered differences in the three shapes indiscriminable; the shapes and motions did not appear ambiguous or confusing.

The obtained decrements in shape discrimination produced by relatively small deviations from planarity of the surface of rotation indicate that the resulting variability

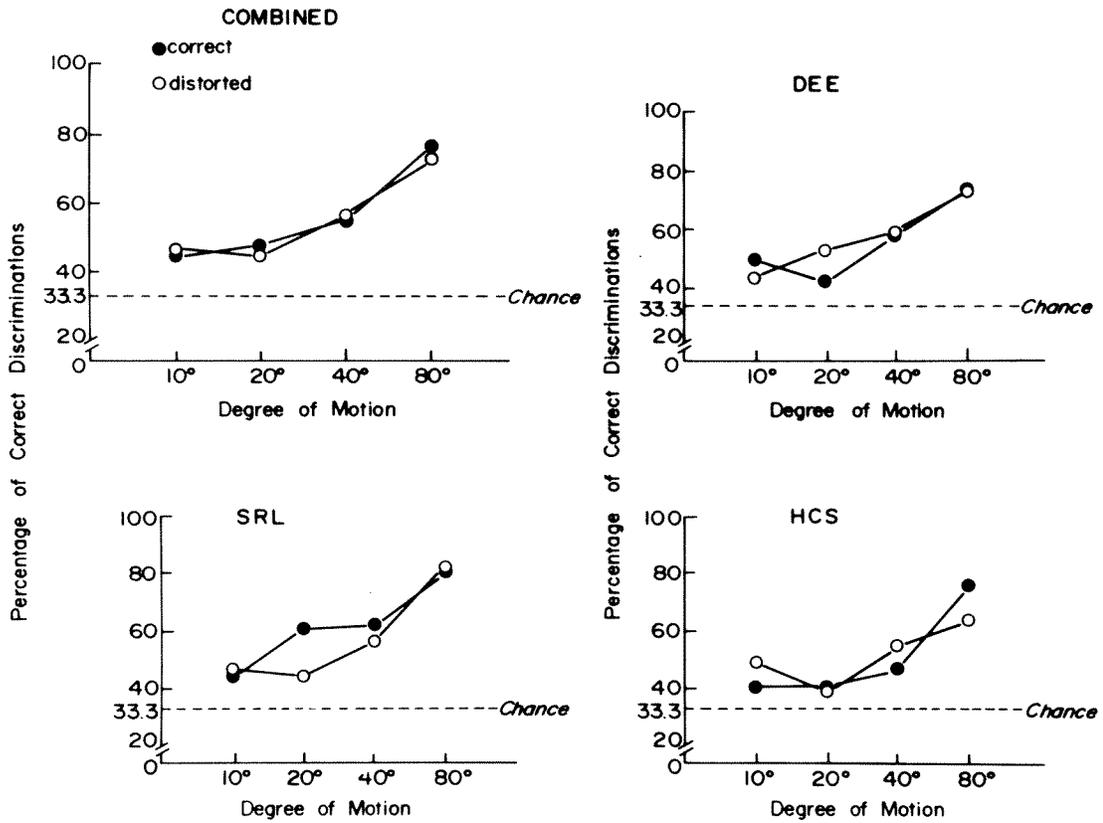


Figure 6. The accuracy of shape discrimination in Experiment 2 as a function of the range of motion and stereoscopic distortion for each of the 3 observers and averaged over observers.

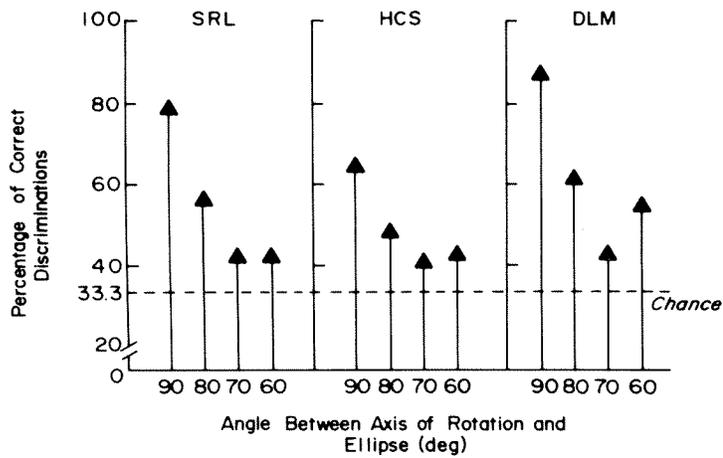


Figure 7. The accuracy of shape discrimination as a function of the angle between the axis of rotation and the planar forms for each of the three observers in Experiment 3. The deviation from planarity of the rotating form increased as the angle decreased from 90°.

in the metric tensors of the image motions significantly reduced the visual information about the intrinsic structure of the moving shapes. Obviously, all motions are not equal as sources of information about the invariant shape of the moving object.

#### EXPERIMENT 4

A surprising result of Experiments 1 and 2 was that the correctness, variability, or even presence of stereoscopic disparity had no reliable effect on discriminations of either moving or stationary shapes. When the shapes were stationary, discriminations were usually near chance even with the correct binocular disparities. One explanation for this poverty of shape from stereopsis may be an inherent ambiguity of stereoscopic information about the scale of depth, as many investigators have suggested. Even though stereoacuity for detecting a difference in depth exhibits exquisite sensitivity, the scale of extension in depth may be poorly resolved. In any case, the discrimination performance in Experiments 1-3 provided no direct evidence about the role of stereopsis in shape discrimination. Consequently, the role of motion in stereoscopic depth constancy remains ambiguous.

To demonstrate that binocular disparity did have a visible effect on perceived depths and shapes of these patterns, we evaluated discriminations of the viewing distance parameter for which the disparities were calculated. If these differences in disparity scale do alter the perceived shapes, they should be easily discriminated, at least when the forms are stationary. If the planar forms are rotated within the plane, however, their congruence under motion may calibrate the stereoscopic embedding into  $E^3$ , thereby destroying information about the disparity scale per se.

If these assumptions about the roles of stereopsis and motion in shape perception are correct, motion should have opposite effects on discriminations of shape and disparity. As found in the preceding experiments, moving shapes should be much more easily discriminated than stationary shapes, but variations in disparity may be discriminated better when the patterns are stationary. These four conditions—discriminations of shape or disparity, with moving or stationary patterns—constituted the main part of Experiment 4. As an additional control, the same four conditions were replicated under nonstereoscopic conditions. (The simulated viewing distance parameter could be manipulated independently of whether there was any disparity between the two monocular displays, and might conceivably be discriminated by very slight differences in perspective.)

#### Method

Several minor changes were made in the displays for Experiment 4. The number of alternative shapes, disparities, and slants was reduced to two from the three alternatives used in the preceding experiments. The purpose was to improve both the spatial discriminations and experimental efficiency.

The two alternative shapes were ellipses, whose vertical axes were either 3% greater than or 3% less than the horizontal axis. Two

alternative values of the stereoscopic disparity parameter corresponded to viewing distances of either one half or one fourth the actual 114.6-cm distance, thereby multiplying the disparities by a factor of two or four times the correct values. The two alternative slants were 50° and 60°. Thus, eight equally likely alternative patterns were composed from these three binary variables. A schematic illustration of the monocular projections of the two ellipses at each of the two slants is shown in Figure 8.

An additional minor change was to construct the ellipses from 20 points equally spaced around the perimeter, thereby describing the contours more accurately than in the preceding experiments. These forms rotated in the plane through an angle of 90°, as in Experiments 1 and 3. In other respects, the optical patterns were like those in the preceding experiments.

Four well-practiced observers each served for six experimental sessions. All had good stereoacuity as verified by other measures. All of the observers were knowledgeable about the purpose of the experiment. Because observer J.P.K. was less accurate in discriminating the two alternative shapes, the difference in the vertical axes of the two ellipses was increased to 8% ( $\pm 4\%$ ) rather than the 6% difference used for the other observers.

Each session was devoted to discriminations of either shape or disparity. Each session consisted of four blocks of 48 trials, with each block containing 40 experimental trials preceded by 8 practice trials. The eight alternative forms occurred equally often in a random sequence in each block. The four conditions defined by the moving versus stationary and stereoscopic versus nonstereoscopic conditions were varied between the four trial blocks in each session in counterbalanced order for the 4 observers. This provided a total of 120 trials for each observer in each of the eight conditions.

In other respects, the methods and procedures were the same as in the preceding experiments.

#### Results

The average accuracy of shape discriminations in each of the eight conditions and the corresponding results for

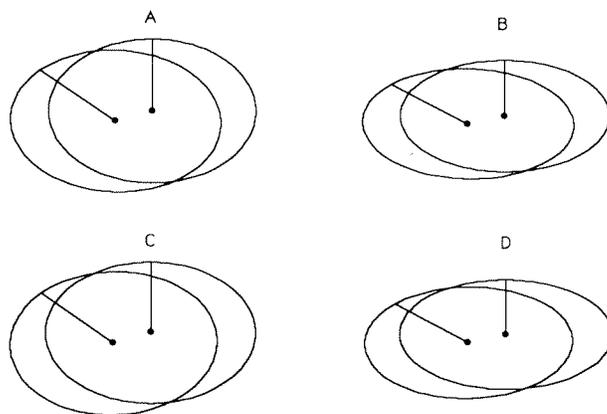


Figure 8. Schematic but accurately scaled illustrations of the projected optical images of the two ellipses at the two slants used in Experiment 4. Each section shows the same ellipse in two different positions on the same plane—in the central position, and rotated counterclockwise by 45°. (A) The larger ellipse (1.03 vertical/horizontal aspect ratio) at 50° slant. (B) The larger ellipse at 60° slant. (C) The smaller ellipse (0.97 vertical/horizontal aspect ratio) at 50° slant. (D) The smaller ellipse at 60° slant. The forms actually used in Experiment 4 were defined by 20 dots equally spaced around the perimeter; the central point and vertical line were added to these illustrations simply to indicate the positions of the two forms.

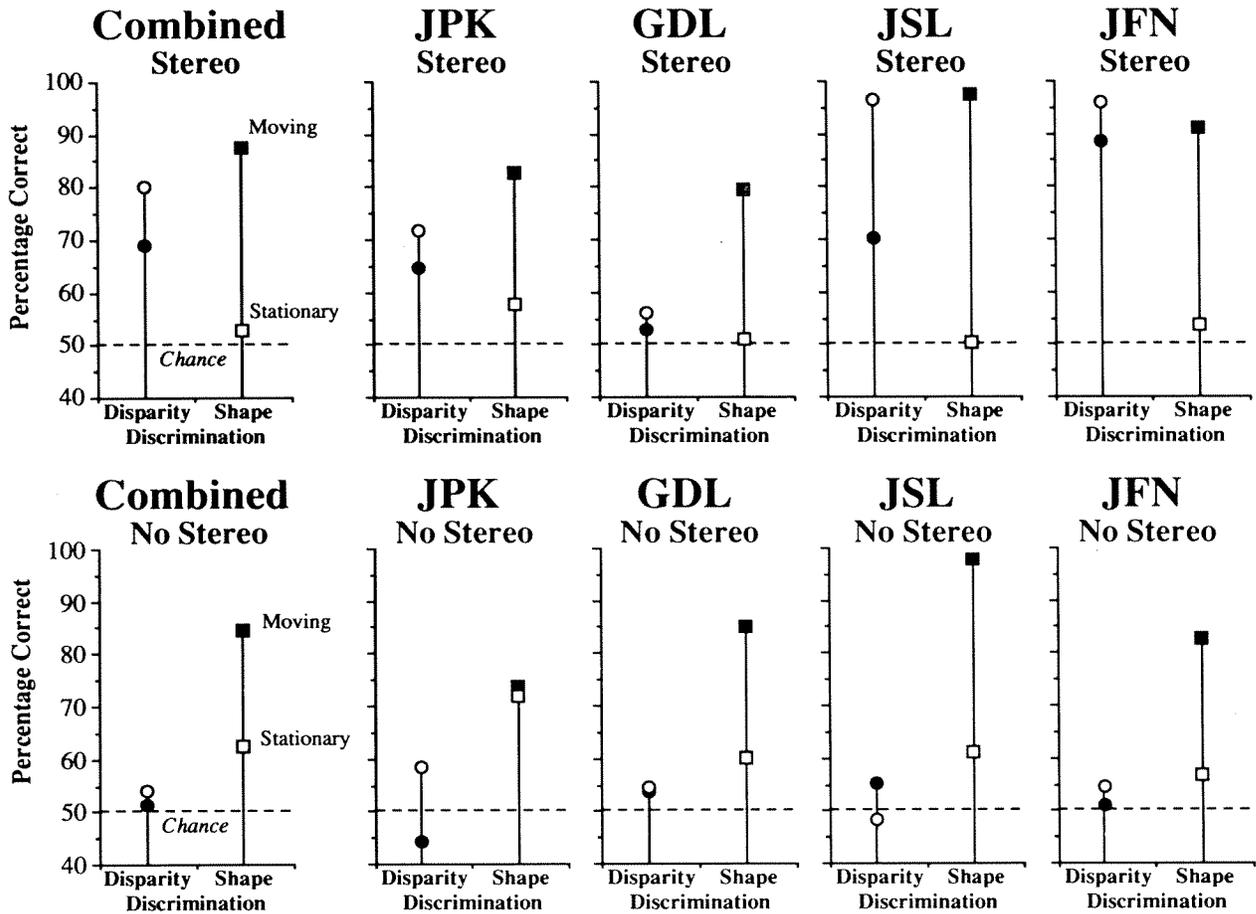


Figure 9. Discrimination accuracies for each observer in each condition of Experiment 4, and the average accuracies in each condition combined over observers.

each of the 4 observers are shown in Figure 9. The primary results are given in the upper half of Figure 9, showing the opposite effects of motion on discriminations of shape and disparity: As in the preceding experiments, moving shapes were accurately discriminated (87.5% correct) despite variations in slant and disparity, but they were essentially indiscriminable (52.9% correct) when stationary. In the disparity discrimination task, however, motion had the opposite effect, with the disparities of moving forms less accurately discriminated than those of stationary forms—68.8% versus 80.0%, a statistically reliable difference by chi-square test [ $\chi^2(1) = 15.35, p < .001$ ]. Disparity discriminations of the moving forms were well above chance, however.

Although disparities of the moving forms were discriminated less accurately by all 4 observers, there was considerable variability between the observers in both accuracy of disparity discrimination and in the detrimental effect of motion, as can be seen in Figure 9. The clearest effect was obtained for J.S.L., whose accuracy dropped from 96.7% in the stationary condition to 70.0% in the moving condition. The difference between these two conditions was statistically significant only for J.S.L., though

the chi-square value totaled over the 4 observers was statistically reliable [ $\chi^2(4) = 33.86, p < .001$ ]. Evidently, observers differ in their sensitivities to the magnitude of stereoscopic disparity.

Results for the nonstereoscopic conditions show that motion yielded shape-discrimination accuracies very similar to those for the stereoscopic conditions. Stationary shapes, however, were more accurately discriminated in the nonstereoscopic than in the stereoscopic conditions—62.3% versus 52.9%—a difference that was statistically reliable by chi-square test [ $\chi^2(1) = 8.26, p < .01$ ], based on the combined responses totaled over observers. Accuracy in discriminating the stationary stereoscopic shapes was not significantly above chance, however [ $\chi^2(1) = 1.52, p > .2$ ]. When the same tests were applied to the data of each individual observer, only J.P.K. was significantly more accurate in the nonstereoscopic than in the stereoscopic discriminations, and the chi-square value totaled over the 4 observers was not quite significant at  $\alpha = .05$  [ $\chi^2(4) = 8.93, p < .10$ ]. Three observers were significantly above chance in discriminating the nonstereoscopic stationary forms, however, and the chi-square totaled over the 4 observers was highly signifi-

cant [ $\chi^2(4) = 33.17, p < .001$ ]; but none of the observers was significantly above chance in discriminating the stereoscopic stationary shapes, and the total chi-square did not approach significance. Thus, the varying magnifications of the binocular disparities seem to have produced varying distortions of the perceived shapes. This result complements the obtained discriminations of disparities of both moving and stationary stereoscopic forms in showing that the binocular disparities affected the perceived shapes. Evidently, the ambiguities inherent in this stereoscopic information were bypassed by the intrinsic metric structure of the moving patterns.

### GENERAL DISCUSSION

The principal result of these experiments is that *planar motion provided an intrinsic metric scale for retinal images of planar forms in 3-D space*, independently of the form's slant in depth and binocular disparities—that is, independently of the extrinsic scales of the display screen or retinae. Evidently, the perceived metric structure was visually defined by the congruence of the spaces and forms under motion.

The precision of shape discrimination obtained in these experiments, with a Weber fraction in Experiment 4 of less than 3% (in the relative lengths of the two major axes of the ellipses), is similar to the precision obtained in other recent experiments on discriminations of relative positions in stationary line segments in the frontal-parallel plane (DeValois, Lakshminarayanan, Nygaard, Schluskel, & Sladky, 1990; Levi, Klein, & Yap, 1988). The precision of these discriminations of relative length is also similar to that obtained by Lappin and Fuqua (1983) for the relative positions of three points along a single line segment rotating in depth. The similarity of these relative length discriminations in 3-D and 2-D is remarkable. Evidently, vision is sensitive to spatial relations in the 3-D space of the environment rather than the 2-D space of the retina—spatial relations that remain invariant under motion.

The perception of metric structure was achieved in these experiments through the “planarity constraint” (Hoffman & Flinchbaugh, 1982). As shown by Hoffman and Flinchbaugh and in the present Appendix, metric structure is theoretically recoverable from two views of a KD pattern when a planar form rotates within the plane. The present experiments demonstrate that human observers are in fact sensitive to this optical information. In contrast, Todd and Bressan (1990), Todd and Norman (1991), and the present Experiment 3 demonstrate that metric structure is *not* perceived by human observers under a variety of conditions that violate the planarity constraint. Whether planarity is a strictly *necessary* condition for perceiving metric structure is uncertain. Ullman's (1979) “structure-from-motion theorem” shows that metric structure might in principle be recovered under much more general conditions involving three or more views (see also Bennett et al., 1989; Huang & Lee, 1989), but human observers have not yet been shown to be sensitive to this theoretically available information.

Planarity of the structure and motion imposes a strong and simple constraint on the optical patterns: Orthographic projection of the slanted plane merely alters the relative scales of two orthogonal axes of the projective plane. A slant of  $60^\circ$ , for example, simply reduces the scale perpendicular to the axis of rotation to one half that of the axis of rotation—an affine transformation. When a planar form is rotated within this rescaled plane, it is readily perceived as rotating rigidly in a slanted plane rather than as deforming in the image plane. This phenomenon is easily demonstrated in a computer-graphics display by first changing the relative scales of the horizontal and vertical axes of the display and then rotating a form in this rescaled plane.<sup>4</sup> Stereoscopic displays like those used in the present experiments are unnecessary.

Stereopsis proved a surprisingly unreliable source of information about the geometric shapes in these experiments. In Experiment 1, only 2 of the 5 observers were significantly better than chance at discriminating the correctly disparate stationary ellipses, and the average for these 2 observers was only 49% correct (33% was expected by chance). Compared with stereoacuity for detecting binocular disparities of  $10''$  of arc or less, this imprecision in measuring depth seems surprising.

Such imprecision of stereopsis in measuring depth has also been found in two other recent studies: Nawrot (1991) recently found that observers with excellent stereoacuity were very poor in discriminating which one of two KD patterns (rotating random-dot spheres) was displayed stereoscopically, even after training with feedback. They attribute this failure of stereoscopic detection to commonality of the neural mechanisms for detecting depth from either stereopsis or motion parallax, a hypothesis supported by other recent experiments on the visual integration of these two sources of depth information (Nawrot & Blake, 1989, 1991). McKee, Levi, and Bowne (1990) have found that Weber fractions for detecting increments to a standing disparity (of  $1' - 20'$  of arc) were about two to five times *greater* than monocular acuities for detecting increased horizontal separations in the same patterns. McKee et al. attribute this imprecision to an absence of binocular neural mechanisms sensitive to the separation between nonoverlapping distributions of neural excitation. But this hypothesis would not seem to explain the absence of stereoscopic sensitivity either in the experiment of Nawrot and Blake or in the present Experiment 1, in which the optical patterns were dense distributions of dots portraying a more nearly continuous gradient in depth than in the two-feature patterns of McKee et al. Perhaps much of the imprecision of stereoscopic depth results from the fact that two disparate views, from either stereopsis or motion, provide information about spatial structure only up to an affine transformation in depth, as shown by Koenderink and van Doorn (1991), Todd and Bressan (1990), and Todd and Norman (1991).

Compared to stereopsis, motion was a much more effective source of information about metric structure in depth. Three different characteristics of these moving patterns may have provided this advantage: (1) planarity of

the forms and motions, (2) a larger range of angular rotation, and (3) a larger number of views. All three of these characteristics were probably at least indirectly important to the perception of metric structure. First, as shown by Hoffman and Flinchbaugh (1982) and in the Appendix, two orthographic views of a planar form rotated within that plane differ simply by an affine transformation in the image plane—a two-parameter transformation (corresponding to the slant of the plane and the angle of rotation) determined by congruence of the form under rotation. Two stereoscopic half-images, however, differ by a perspective rather than an affine transformation, described by additional free parameters associated with the 3-D location of the fixation point. Thus, stereopsis admits a family of potential structures that differ from each other by nonaffine transformations of the images.

The planar motions of forms in these experiments also provided more optical information than did the stationary stereoscopic patterns, because both the range of rotation and the number of views were larger. At a viewing distance of 114.6 cm, the interocular separation of 6.3 cm corresponded to a rotation of only  $3.2^\circ$  around the center of the display screen. In contrast, the range of rotation was  $90^\circ$  for most of the moving patterns. The results of Experiment 2 indicate that a rotation of only  $3^\circ$  would have provided insufficient visual information about metric shape.

Evidently, vision integrated spatial information over substantially more than two successive views of these planar forms and motions. This result contrasts with those of Todd and Bressan (1990) and Todd and Norman (1991), who found that vision was insensitive to spatial relations over more than two successive views. In Todd and Bressan (1990) and Todd and Norman's (1991) KDE patterns, the depths at given image positions changed over time, but the image-to-depth embedding remained constant in the present experiments. The temporal integration of spatial structure in the present experiments probably resulted from the temporal consistency and spatial simplicity of the planarity constraint.

The present results might suggest that stereopsis was simply insensitive to moving patterns in which the disparities of given features changed over time. Indeed, experimental evidence demonstrates that vision is very slow in detecting changes in the binocular disparity of a given feature moving in depth only in the direction of view (e.g., Tyler, 1975; White & Odom, 1985). But for motion in other directions, where temporal variations in disparity are correlated with spatial directions in the cyclopean image, stereopsis is exquisitely sensitive to changing disparities (Lappin, Craft, & Payne, 1986; Lappin, Love, Cook, & Norman, 1992), with no apparent decrement in sensitivity produced by such motion in depth. There is no evidence and little reason to believe that stereopsis was silenced by the motion of these forms.

One of the questions raised by the results of this study is whether the perception of metric structure is restricted to image motions that are congruent under *affine* trans-

formations, or whether metric structure can be perceived as well from congruence under *perspective* transformations. The theoretical results of Hoffman and Flinchbaugh (1982) and in the present Appendix show that metric structure is obtainable from affine transformations associated with orthographic views of planar motion; and the present experimental results demonstrate that such motions in the monocular half-images are visually sufficient to perceive metric structure. But the stereoscopic relation between the two half-images involves a perspective transformation, geometrically and computationally more complex than affine, requiring estimation of additional parameters to obtain congruence. Insofar as the perceived spatial structure of these patterns may have involved the binocularly integrated stereoscopic images rather than the monocular half-images alone, stereopsis must have detected the perspective congruence of these two sets of half-images. The exact contribution of stereopsis in discriminating these moving shapes is ambiguous, but the obtained discriminations of shapes with magnified binocular disparities suggest that the congruent metric structure of these patterns was stereoscopically visible. In fact, the moving patterns with magnified disparities in Experiments 1 and 4 seldom appeared distorted. One of the observers mentioned that distortions were sometimes seen in patterns with the greatest disparity magnification, but typically the patterns appeared quite normal. Similarly, the results of Lappin and Fuqua (1983) and recent results of Norman (1990) also indicate that vision is adept at detecting congruent structure of moving patterns with exaggerated perspective projections. More experimental and theoretical work is needed on this issue, but vision appears to detect metric structure from congruence under perspective transformations. Theoretical analysis in the Appendix indicates that this is computationally tractable.

## REFERENCES

- BENNETT, B., HOFFMAN, D., NICOLA, J., & PRAKASH, C. (1989). Structure from two orthographic views of rigid motion. *Journal of the Optical Society of America A*, *6*, 1052-1069.
- BRAUNSTEIN, M. L., & ANDERSEN, G. J. (1981). Velocity gradients and relative depth perception. *Perception & Psychophysics*, *29*, 145-155.
- CORMACK, R. H. (1984). Stereoscopic depth perception at far viewing distances. *Perception & Psychophysics*, *35*, 423-428.
- CORMACK, R. [H.], & FOX, R. (1985). The computation of retinal disparity. *Perception & Psychophysics*, *37*, 176-178.
- DELONE, B. N. (1963). Analytic geometry. In A. D. Aleksandrov, A. N. Kolmogorov, & M. A. Lavrent'ev (Eds.), *Mathematics: Its content, methods, and meaning* (2nd ed.) (pp. 183-260). Cambridge: MIT Press.
- DEVALOIS, K. K., LAKSHMINARAYANAN, V., NYGAARD, R., SCHLUSSEL, S., & SLADKY, J. (1990). Discrimination of relative spatial position. *Vision Research*, *30*, 1649-1660.
- FISHER, S. K., & EBENHOLTZ, S. M. (1986). Does perceptual adaptation to telestereoscopically enhanced depth depend on the recalibration of binocular disparity? *Perception & Psychophysics*, *40*, 101-109.
- FOLEY, J. M. (1967). Disparity increase with convergence for constant perceptual criteria. *Perception & Psychophysics*, *2*, 605-608.
- FOLEY, J. M. (1980). Binocular distance perception. *Psychological Review*, *87*, 411-434.

- FOX, R., CORMACK, L., & NORMAN, J. F. (1987). The effect of vertical disparity on depth scaling. *Investigative Ophthalmology & Visual Science*, **28** (Suppl. No. 3: ARVO Abstracts), 293.
- GIBSON, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.
- GIBSON, J. J. (1966). *The senses considered as perceptual systems*. Boston: Houghton Mifflin.
- HOFFMAN, D. D., & FLINCHBAUGH, B. E. (1982). The interpretation of biological motion. *Biological Cybernetics*, **42**, 195-204.
- HUANG, T., & LEE, C. (1989). Motion and structure from orthographic projections. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **11**, 536-540.
- KILLING, W. (1892). Über die Grundlagen der Geometrie. *Journal für die reine und angewandte Mathematik*, **109**, 121-186.
- KOENDERINK, J. J. (1986). Optic flow. *Vision Research*, **26**, 161-179.
- KOENDERINK, J. J., & VAN DOORN, A. J. (1975). Invariant properties of the motion parallax field due to movement of rigid bodies relative to an observer. *Optica Acta*, **22**, 773-791.
- KOENDERINK, J. J., & VAN DOORN, A. J. (1976a). Geometry of binocular vision and a model for stereopsis. *Biological Cybernetics*, **21**, 29-35.
- KOENDERINK, J. J., & VAN DOORN, A. J. (1976b). Local structure of movement parallax of the plane. *Journal of the Optical Society of America*, **66**, 717-723.
- KOENDERINK, J. J., & VAN DOORN, A. J. (1976c). The singularities of the visual mapping. *Biological Cybernetics*, **24**, 51-59.
- KOENDERINK, J. J., & VAN DOORN, A. J. (1977). How an ambulant observer can construct a model of the environment from the geometrical structures of the visual inflow. In G. Hauske & E. Butenandt (Eds.), *Kybernetik* (pp. 224-247). Munich: Oldenbourg.
- KOENDERINK, J. J., & VAN DOORN, A. J. (1980). Photometric invariants related to solid shape. *Optica Acta*, **27**, 981-996.
- KOENDERINK, J. J., & VAN DOORN, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A*, **8**, 377-385.
- LAPPIN, J. S. (1990). Perceiving the metric structure of environmental objects from motion, self-motion, and stereopsis. In R. Warren & A. H. Wertheim (Eds.), *Perception and control of self-motion* (pp. 541-578). Hillsdale, NJ: Erlbaum.
- LAPPIN, J. S. (1991). Perceiving environmental structure from optical motion. In W. W. Johnson & M. K. Kaiser (Eds.), *Visually guided control of movement* (NASA Conference Publication 3118, pp. 39-61). Moffett Field, CA: NASA Ames Research Center.
- LAPPIN, J. S., CRAFT, W. D., & PAYNE, T. J. (1986). Stereoscopic acuity for bending motion in 3-D space. *Optics News*, **12**, 116.
- LAPPIN, J. S., & FUQUA, M. F. (1983). Accurate visual measurement of three-dimensional moving patterns. *Science*, **221**, 480-482.
- LAPPIN, J. S., LOVE, S. R., COOK, G. L., & NORMAN, J. F. (1992). *Stereoscopic acuity for patterns moving in three-dimensional space*. Manuscript in preparation.
- LAPPIN, J. S., & WASON, T. D. (1991). The perception of geometrical structure from congruence. In S. R. Ellis, M. K. Kaiser, & A. J. Grunwald (Eds.), *Pictorial communication in virtual and real environments* (pp. 425-448). London: Taylor & Francis.
- LEVI, D. M., KLEIN, S. A., & YAP, Y. L. (1988). "Weber's law" for position: Unconfounding the role of separation and eccentricity. *Vision Research*, **28**, 597-603.
- LONGUET-HIGGINS, H. C. (1986). Visual motion ambiguity. *Vision Research*, **26**, 181-183.
- MARR, D. (1982). *Vision*. San Francisco: Freeman.
- MAUK, D., CREWS, K., & FOX, R. (1987). Stereoscopic depth constancy in preschool children. *Investigative Ophthalmology & Visual Science*, **28** (Suppl. No. 3: ARVO Abstracts), 296.
- MAYHEW, J. E. W., & LONGUET-HIGGINS, H. C. (1982). A computational model of binocular depth perception. *Nature*, **297**, 376-378.
- McKEE, S. P., LEVI, D. M., & BOWNE, S. F. (1990). The imprecision of stereopsis. *Vision Research*, **30**, 1763-1779.
- NAWROT, M. (1991). *On the perceptual identity of kinetic depth and dynamic stereopsis*. Unpublished doctoral dissertation, Vanderbilt University.
- NAWROT, M., & BLAKE, R. (1989). Neural integration of information specifying structure from stereopsis and motion. *Science*, **244**, 716-718.
- NAWROT, M., & BLAKE, R. (1991). The interplay between stereopsis and structure from motion. *Perception & Psychophysics*, **49**, 230-244.
- NORMAN, J. F. (1990). *The perception of curved surfaces defined by optical motion*. Unpublished doctoral dissertation, Vanderbilt University.
- ONO, H., & COMERFORD, J. (1977). Stereoscopic depth constancy. In W. Epstein (Ed.), *Stability and constancy in visual perception* (pp. 91-128). New York: Wiley.
- POLLARD, S. B., PORRILL, J., MAYHEW, J. E. W., & FRISBY, J. P. (1985). Disparity gradient, Lipschitz continuity, and computing binocular correspondences. In *Proceedings of the Third International Symposium on Robotics Research* (pp. 19-26). Cambridge, MA: MIT Press.
- SPEERLING, G., LANDY, M. S., DOSHER, B. A., & PERKINS, M. E. (1989). Kinetic depth effect and identification of shape. *Journal of Experimental Psychology: Human Perception & Performance*, **15**, 826-840.
- SUPPES, P., KRANTZ, D. H., LUCE, R. D., & TVERSKY, A. (1989). *Foundations of measurement* (Vol. 2, pp. 13-78). New York: Academic Press.
- TODD, J. T. (1984). The perception of three-dimensional structure from rigid and nonrigid motion. *Perception & Psychophysics*, **36**, 97-103.
- TODD, J. T., & BRESSAN, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics*, **48**, 419-430.
- TODD, J. T., & NORMAN, J. F. (1991). The visual perception of smoothly curved surfaces from minimal apparent motion sequences. *Perception & Psychophysics*, **50**, 509-523.
- TODD, J. T., & REICHEL, F. D. (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychological Review*, **96**, 643-657.
- TYLER, C. W. (1975). Characteristics of stereomovement suppression. *Perception & Psychophysics*, **17**, 225-230.
- ULLMAN, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- WALLACH, H., & KARSH, E. (1963a). The modification of stereoscopic depth perception and the kinetic depth-effect. *American Journal of Psychology*, **76**, 429-435.
- WALLACH, H., & KARSH, E. (1963b). Why the modification of stereoscopic depth perception is so rapid. *American Journal of Psychology*, **76**, 413-420.
- WALLACH, H., MOORE, M., & DAVIDSON, L. (1963). Modification of stereoscopic depth perception. *American Journal of Psychology*, **76**, 191-204.
- WHITE, K. D., & ODOM, J. V. (1985). Temporal integration in global stereopsis. *Perception & Psychophysics*, **37**, 139-144.

## NOTES

1. The term *metric* is used in its conventional mathematical sense. A relation  $m(a,b)$  is said to be metric if it satisfies the following axioms for all points  $a, b, c$ : positivity,  $m(a,b) \geq 0$ ; symmetry,  $m(a,b) = m(b,a)$ ; reflexivity,  $m(a,a) = 0$ ; and triangle inequality,  $m(a,c) \leq m(a,b) + m(b,c)$ . The term *metric structure* is used in this paper to refer to an object or connected set of points within which the spatial relations among points satisfy the metric axioms and remain invariant (isometric, congruent) under arbitrary motions within the space. For spaces of two, three, or any odd number of dimensions, such isometry implies that the space is Euclidean, hyperbolic, or spherical (cf. Suppes, Krantz, Luce, & Tversky, 1989, chap. 12).

2. *Affine* spatial relations are those which remain invariant under arbitrary linear transformations of the Cartesian coordinates. An affine transformation  $t(x)$  of a vector  $x$  in an  $n$ -dimensional vector space has the form  $t(x) = A(x) + b$ , where  $A$  is a nonsingular  $n \times n$  matrix and  $b$  is an  $n$ -dimensional vector. Todd and Bressan (1990) and Todd and Norman (1991) have proposed that visually perceived structure from motion is invariant under a specific subgroup of affine transformations—under arbitrary scalar transformations of distances in depth in the direction of gaze. The term *affine structure* in the present paper also usually applies to this special case of invariance up to arbitrary scalar transformations of distances in depth relative to distances in the frontal-parallel plane.

3. *Congruence* is synonymous with isometry, referring to a mapping of one set of points onto another that preserves metric relations among all pairs of points. The group of congruencies is slightly more general than the group of rigid motions. In Euclidean space, congruent transformations include both rigid motions and bendings. Whereas "rigid motions" usually imply Euclidean space, congruent transformations are also definable in hyperbolic and spherical geometric spaces. Obviously, the projected *images* of objects rotating in  $E^3$  are not themselves congruent; the congruence is implicit, dependent on an embedding of the images into a space in which the congruence holds.

4. Symmetric forms such as the ellipses used in these experiments must be rotated around a point that is not the center of the form. Dots or other discrete position markers must also be used instead of smooth contours, to avoid ambiguities about the local direction of motion.

## APPENDIX

The purpose of this Appendix is twofold. First, we review briefly a theoretical framework and notational system for describing the optical information and perception of the moving spatial patterns in these experiments. This theoretical framework has been presented in more detail elsewhere (Lappin, 1990, 1991; Lappin & Wason, 1991). Second, we prove that the metric structure of planar forms rotating in a plane is theoretically determined by two views. This result and the required planarity of the form and motion were not given in the previously published presentations of this theory. This result is similar to that given earlier by Hoffman and Flinchbaugh (1982), although the present result is a bit more general—applying to arbitrary retinal coordinates as well as Cartesian coordinates, and applicable to perspective as well as orthographic projections—and the proof is different and in some respects simpler.

**PROPOSITION 1.** *Optical patterns on the retina may be regarded as images of smooth surfaces.*

The insight that the geometry of vision may be simplified by treating retinal patterns as images of environmental surfaces has been recognized and developed most clearly by Koenderink and van Doorn (e.g., 1975, 1976a, 1976b, 1976c, 1977; see also Koenderink, 1986). Related ideas were also discussed by Gibson (1950, 1966).

"Smooth" surfaces are differentiable almost everywhere—except at isolated local discontinuities corresponding to sharp peaks and corners and at self-occluding and boundary contours. Patterns of discrete features (e.g., dots) that can be simply connected by a smooth surface may also be regarded as constituting a smooth surface (cf. Pollard, Porrill, Mayhew, & Frisby, 1985).

Both environmental surfaces and their images constitute smooth 2-D manifolds. Moreover, there is a close correspondence between the differential structures of these two manifolds: Images defined by motion parallax, binocular disparity, and texture are *diffeomorphic* with the visible regions of the corresponding environmental surfaces. That is, there is a smoothly changing one-to-one mapping of the spatial derivatives of the visible regions of an environmental surface (its gradients, curvatures, singularities, and critical points) onto the spatial derivatives of the image of the surface; and the inverse mapping from the image onto the surface is also smooth, one-to-one, and onto. (This diffeomorphism does not hold for images defined by illuminance and shading, but a systematic correspondence between the differential structures of the two manifolds still exists; cf. Koenderink & van Doorn, 1980.)

It follows from Proposition 1 that the optical projection from a surface onto its image may be locally approximated by a linear coordinate transformation. Thus, let the  $2 \times 1$  column vector  $[dO]$

$= [do^1, do^2]^T$  represent an infinitesimal displacement within a local surface patch (sufficiently small that it contains negligible curvature) on the object surface; and let  $[dR] = [dr^1, dr^2]^T$  be the corresponding image of this vector in the retinal image of the object surface. Then the map of any such displacement within that surface patch onto its retinal image may be described by the following linear transformation:

$$[dR] = V[dO], \quad (A1)$$

where  $V$  is a  $2 \times 2$  Jacobian matrix of partial derivatives,

$$V = \begin{bmatrix} \partial r^1 / \partial o^1 & \partial r^1 / \partial o^2 \\ \partial r^2 / \partial o^1 & \partial r^2 / \partial o^2 \end{bmatrix}.$$

$V$  simply describes the retinal image of a surface patch; it represents the local structure of the image itself, and need not be computed or estimated. The four entries in this matrix vary as a smooth function of the position on the surface, changing with the orientation and distance of the object surface relative to the observer's retina. Of course, the inverse map from the image onto the surface is also described by a linear coordinate transformation:

$$[dO] = V^{-1}[dR].$$

An important consequence of this representation of the optical images is that the *metric* structure of a local surface patch is also given by a simple quadratic function of these local image data. The metric structure of the local surface patch is determined by just three unknown parameters that are coefficients of the quadratic terms. This equation for the metric structure of a surface patch is given by a basic formula in differential geometry known as the *first fundamental form* of the surface. In the present application, this formula may be expressed in terms of the retinal image data as follows. Let  $[dR]$  be the retinal image coordinates for an infinitesimal displacement on the object surface; and let  $[dX] = [dx^1, dx^2, dx^3]$  be the description of this same vector in the three orthogonal coordinate axes of  $E^3$ . Then

$$[dX] = P[dR], \quad (A2)$$

where  $P = [\partial x^k / \partial r^a]$  is a  $3 \times 2$  Jacobian matrix that specifies the embedding of the retinal image of a local surface patch into the three orthogonal coordinate axes of  $E^3$ . The metric structure of this image of the surface patch is then obtained from the Pythagorean formula for distance:

$$\begin{aligned} ds^2 &= [dX]^T [dX] \\ &= [dR]^T P^T P [dR] \\ &= [dR]^T P^* [dR] \\ &= [dO]^T V^T P^* V [dO], \end{aligned} \quad (A3)$$

where  $P^* = P^T P$  is a symmetric  $2 \times 2$  matrix with entries  $p_{ab} = \sum_k (\partial x^k / \partial r^a) (\partial x^k / \partial r^b)$ .  $P^*$  constitutes the *metric tensor* for the retinal coordinates of the local surface patch. Since  $p_{12} = p_{21}$ , this matrix contains three independent parameters.

It should be emphasized that the metric tensor is independent of the orientation of the surface patch; the metric tensor is specified by just three parameters,  $P^*$ , whereas six parameters, the entries of  $P$ , are required to describe the orientation and depth at each position on the surface. Theoretical analyses of the problem of perceiving 3-D structure have usually assumed that

surface structure must be derived from a depth map, but the problem is actually simpler. Accurate perception of surface shape does not require accurate perception of orientation or depth.

**PROPOSITION 2.** *The metric structure of visually perceived surfaces and spaces derives from congruence under motion.*

As given in the text, the logic of this idea was nicely stated by Killing (1892). The idea is related to the "rigidity constraint" employed by Ullman (1979) and others to infer the 3-D depth coordinates of a rotating object from the 2-D coordinate values from three or more 2-D images of at least four noncoplanar points, though the present approach does not assume that such coordinate systems with an implicit metric structure have an a priori definition in either the 3-D space of the environment or the 2-D space of the retina. The present development derives the metric structure of the objects, images, and spaces from their congruence under motion. A similar approach with generalized nonmetric coordinates was also discussed by Koenderink and van Doorn (1977).

The requirement that the metric structure of the image of a form remains constant over time may be formulated as an equality of the metric tensor of a surface patch under motion of the object in  $E^3$ . Thus, let  $V$  represent the retinal image of a surface patch in one temporal frame, and let  $U$  represent the retinal image of the same surface patch at a subsequent moment in time, after the object and its image have moved; let  $P^*$  be the metric tensor of the first image, and let  $Q^*$  be the corresponding metric tensor for the second image. The requirement that the object remain congruent under motion in  $E^3$  implies that

$$V^T P^* V = U^T Q^* U = G, \quad (A4)$$

where  $G$  is a symmetric  $2 \times 2$  matrix containing the three metric tensor coefficients of the first fundamental form of the surface patch. In general, this equation does not yield solutions for the unknown metric tensor parameters in the matrices  $P^*$  and  $Q^*$ , since there are six unknowns but only three independent equations. Under the planarity constraint, however, the situation is much simpler, because the projective mapping from the object's plane of motion in  $E^3$  onto the image plane remains constant over time. Hence, the image coordinates may be constructed in such a way that the metric embedding,  $P^*$ , of the image of a given surface patch remains constant as it moves from one position to another in  $E^3$  and in the image. For example, if the plane in which the form is moving is described with rectilinear Cartesian coordinates, the retinal image coordinates could be just the projection of this Cartesian coordinate system. In this case, a spatially invariant Pythagorean formula for distance can apply to any position in the object plane or to any position in the image plane. Equation A4 then simplifies to

$$V^T P^* V = U^T P^* U. \quad (A5)$$

Although the construction of such a retinal coordinate system with a uniform metric embedding may pose a significant computational problem in the general case, this turns out to be very easy for the case of orthographic projections of a plane. We will return below to the case of perspective projections of a plane.

Matrix Equation A5 may also be written in the form

$$g_{ij} = \sum_a \sum_b [p_{ab} (v_i^a v_j^b)] = \sum_a \sum_b [p_{ab} (u_i^a u_j^b)] \\ \sum_a \sum_b [p_{ab} (v_i^a v_j^b - u_i^a u_j^b)] = 0,$$

where  $v_i^a = \partial r^a / \partial o^i$  and  $u_i^a = \partial s^a / \partial o^i$ ,  $p_{ab}$  are the quadratic metric tensor coefficients given above for the retinal coordinates of the surface patch, and the terms  $r^a$  and  $s^a$  represent two different magnitudes of displacement on retinal coordinate axis  $a$ . Since  $g_{12} = g_{21}$ , there are three independent equations in three unknowns,  $p_{11}$ ,  $p_{22}$ , and  $p_{12} = p_{21}$ . For the general case in which  $p_{ab}$  are three independent coefficients, these equations have only the trivial solution  $p_{11} = p_{22} = p_{12} = 0$ , but for images of a plane  $p_{12}$  is a function of  $p_{11}$  and  $p_{22}$ . Without loss of generality, we may define  $r^1$  and  $r^2$  as two orthogonal coordinate axes, so that for orthographic projections of a plane  $p_{12} = 0$ . We then obtain the following three solutions for the aspect ratio of these parameters that embed the two retinal coordinate axes into  $E^3$ :

$$p_{11}/p_{22} = [(v_1^1)^2 - (u_1^1)^2] / [(u_1^1)^2 - (v_1^1)^2] \\ = [(v_2^2)^2 - (u_2^2)^2] / [(u_2^2)^2 - (v_2^2)^2] \\ = [v_1^1 v_2^2 - u_1^1 u_2^2] / [u_1^1 u_2^2 - v_1^1 v_2^2]. \quad (A6)$$

The terms on the right are ratios of differences in second-degree magnitudes of two successive images of a given local vector on the object surface, the numerators are differences in the projected magnitudes on retinal coordinate axis 2 ( $v_2^2 v_2^2 - u_2^2 u_2^2$ ), and the denominators are corresponding differences on coordinate axis 1 ( $u_1^1 u_1^1 - v_1^1 v_1^1$ ). Thus, the aspect ratio of the metric tensor coefficients for the two orthogonal retinal coordinates,  $p_{11}/p_{22}$ , is simply scaled to maintain invariant measures of object structure and motion over varying directions and positions on the retinal coordinates. Under the constraints of planarity and orthographic projection, both the image motion and this aspect ratio of the retinal coordinates constitute affine transformations, and the composition of these transformations is also merely an affine transformation.

The critical step in this proof is the assumption that the metric tensor for the image of a given object surface patch,  $P^*$ , remains invariant under motion. Although such image coordinates are easily constructed for orthographic projection of a plane, this requirement is less easily satisfied for perspective projections, in which the image motions involve nonaffine transformations. Additional parameters are required to specify the perspective projection of a plane, corresponding to the location of a vanishing point in  $E^3$ . Obviously, additional image points and/or views would be needed to determine these perspective parameters. In fact, four points, no three of which are collinear, are necessary and sufficient to specify the mapping from one perspective image of a planar form onto another, according to the "fundamental theorem of plane perspectivity" (cf. Delone, 1963). Although we cannot at present write down equations with which to compute the parameters for these perspective transformations, such perspective transformations are clearly computable from images of planar motions of planar forms. Moreover, the evidence of Lappin and Fuqua (1983) and Norman (1990) indicates that vision achieves such computations at least from patterns with rotational motions of about  $90^\circ$ .