On the scaling of visual space from motion— in response to Pizlo and Salach-Golyska

JOSEPH S. LAPPIN

Vanderbilt University, Nashville, Tennessee

and

ULF B. AHLSTRÖM

Uppsala University, Uppsala, Sweden

Pizlo and Salach-Golyska (1994) have raised an important question about the validity of the interpretation of experiments reported by Lappin and Love (1992)—based on the fact that alternative cues in the image plane might have supported the shape discriminations in those experiments. The meaning of the hypothesis that visual space may be scaled by congruence under motion is clarified, pertinent evidence is reviewed, and new experimental evidence is reported. We conclude that visual space can be metrically scaled by congruence of moving shapes.

Pizlo and Salach-Golyska (1994) have pointed out some potentially important and previously overlooked properties of the stimulus displays used in experiments on the perception of metric spatial structure from planar motion (Lappin & Love, 1992). These properties are important, because they provide an alternative source of potential information that observers might have used to perform the shape-discrimination task in our experiments. This alternative information indicates an alternative hypothesis to account for the accurate discrimination performance obtained in our experiments—and this alternative is more conservative and more parsimonious in some ways than the interpretation suggested by Lappin and Love. The distinction between these two hypotheses involves general issues concerning the foundations of visual space and experimental evidence. Therefore, these alternative hypotheses merit both theoretical scrutiny and experimental tests.

The purpose of the present paper is to clarify and generalize the hypothesis offered by Lappin and Love (1992) and also assess current experimental evidence about its validity. Additional evidence, summarized briefly below, appears to support our original hypothesis: Visual space may be scaled by the congruence of moving objects.

The experiments reported herein were supported in part by NIH Grants EY05926 and P30EY08126. We are grateful to Warren Craft for his help in preparing the illustrations in Figure 1, to Karen Yu for help in collecting data, to Ted Payne for the computer software that controlled this experiment, and to Farley Norman and especially James Todd for discussions of both theoretical and experimental issues relevant to this research. Conversations with James Todd were partly responsible for the design of this experiment, and he also detected another potential 2-D artifact in our initial version of the experiment. Zygmunt Pizlo also provided helpful critical comments on a previous version of this paper. Correspondence should be addressed to J. S. Lappin, Department of Psychology, Vanderbilt University, Nashville, TN 37240 (e-mail: lappinjs@ctrvax.vanderbilt.edu).

Lappin and Love's (1992) Experiments

The aim of our experiments was to test the hypothesis that certain types of motion in depth provide sufficient information for perceiving metric structure in depth, even when stereoscopic information provides conflicting information about extension in depth. The observers' task was to discriminate between the shapes of nearly circular ellipses (described by 20 equally spaced dots on the circumference) that rotated ($\pm 45^\circ$ about a point below the center of the ellipse) in a plane slanted around the horizontal axis 40° – 60° from the frontal parallel. This discrimination task was intended to require perception of the metric structure of these rotating shapes because it involved discrimination of the relative lengths of the two principal axes of the ellipses in the plane of rotation slanted in depth.

The term *metric* is used in this context to refer to a scaling of spatial distances that preserves ratios of distances over changes in direction *on a surface*. Measures of contour length, angles between contours, and curvature on the surface are specified by this metric representation. Perception of metric structure intrinsic to the shape does *not* imply perception of the slant in depth, nor does it imply the perception of distances in the depth axis (orthogonal to the image plane) either between points within the shape or from the observer. The intrinsic metric structure of a shape on a surface involves fewer parameters than metric distances in empty three-dimensional (3-D) space—a maximum of three parameters in the former case and six in the latter (see Lappin, 1990; Lappin & Love, 1992, Appendix). Currently available evidence indicates that metric relations in depth per se cannot be visually perceived from images of moving objects (Norman & Todd, 1993; Todd & Bressan, 1990; Todd & Norman, 1991; Todd, Norman, Perotti, & Tittle, 1993).

One of the alternative ellipses in these experiments was compressed in one axis by 3%–6% relative to a circle and another was expanded relative to a circle by the same amount in the same axis. When slanted in depth around the horizontal axis, the stationary projected images of these shapes were very eccentric ellipses, compressed in the vertical axis by an amount much greater than the differences between the two principal axes within the shape. When the shapes rotated in the slanted plane, the projected images deformed with the changing orientations. Thus, discrimination of slight differences in shape of the two ellipses was intended to require perception of the shape in the plane slanted in 3-D rather than the projected shapes in the image plane.

Subjectively, when these shapes rotated in a plane in depth they did not appear to deform as they changed in position and orientation in the image plane; rather, they appeared to move rigidly in a plane slanted in depth. The shape appeared to remain constant and the alternative ellipses were easily discriminated. Surprisingly, large vari-

ations in the magnitude of binocular disparity were found to have no effect on discrimination accuracy and only occasional slight effects on the subjective appearance of the moving shapes. Theoretical understanding of these phenomena does not require examination of the stereoscopic disparities; for our present purposes, it suffices to consider the effects of motion on perceived spatial relations in the monocular images.

Now the effect on the images produced by orthographic projection from a plane slanted around the horizontal axis simply amounts to an affine coordinate transformation of the image plane, where the scale of the vertical axis is contracted by a factor equal to the cosine of the angle of slant from the frontal parallel. In contrast to the general case of objects moving in depth, in this case the images map onto an implicit surface that remains constant over time. That is, the projection from the slanted plane of rotation onto the image plane is described by an affine (linear) coordinate transformation that is constant both spatially and temporally; and the same is true for the inverse mapping from the image onto the plane of rotation in depth. Thus, perception of the constant metric structure of the moving shape depends only on the aspect ratio of the vertical and horizontal axes of the image plane. The aspect ratio for which the shape remains congruent under rotation in the plane specifies the image coordinate transformation that permits perception of the metric structure of the moving shape. Mathematically, this aspect ratio is determined by just two successive views, but visually, more views over a wider range of rotation seem to produce more accurate perception—on the basis of results from Lappin and Love (1992) and of informal pilot observations with displays like those described below.

Thus, moving objects in an image plane whose coordinates have been transformed by a smooth function produce the same optical and perceptual effects as do the images of objects moving over a smooth surface in depth—with no explicit computation of projected images of objects in 3-D space. In subsequent research, we have used this method of coordinate transformation followed by motion in the image plane, because this method corresponds conceptually to the geometry of the images. This arrangement is schematically illustrated in Figure 1.

A Hypothesis About the Scaling of Visual Space From Motion

The principal idea in our interpretation of these phenomena is that perceived visual space may be derived from congruence of the moving objects: Two spaces are congruent if and only if they can be covered by the same object; and two objects are congruent if and only if they can cover the same space. This relativistic conception of visual space as derived from congruence under motion contrasts with the common assumption that the spatial structure of the optical input to vision is defined by the 2-D metric structure of the retina or image plane. Thus, in Lappin and Love's (1992) experiments, the rotation of a simple planar shape in the transformed image plane determined a space—a plane slanted in depth—in which the shape re-

mained invariant under motion, and this derived space governed the perceived shape and its motion.

A more specific hypothesis applicable to these experiments is that this visual scaling of space from motion determines the *metric* structure of the moving shape when there is a temporally constant isometric mapping from the image plane onto the space—that is, when the derived visual space is either a plane or a surface curved in only one direction. When planar shapes are rotated in a plane and displayed by orthographic projection, as in Lappin and Love's (1992) experiments and in other experiments described below, the constant metric structure of the moving shape is determined simply by the aspect ratio of the two principal axes of the image plane.

Alternative Forms of Image Information

In an insightful analysis, Pizlo and Salach-Golyska (1994) have pointed out that the discrimination task in Lappin and Love's (1992) experiments might have been performed by discriminating other 2-D image properties that were correlated with the metric structure of the two or three alternative shapes, without actually perceiving the metric structure of the shapes. They described two such confounded properties of the shape of the image at either end of its range of rotation: (1) the temporal pattern of changes in projected width of the image, and (2) the 2-D orientation of the long axis of the ellipse-shaped image.

Other related forms of such image information may also be found—cues that have been correlated with specific alternative shapes in specific experiments but that are not in general correlated with the metric structure of moving shapes. For example, Farley Norman and James Todd (personal communication, 1993) have pointed out that if the translational components of the motion are removed, so that in effect the ellipse is rotated about its center, the boundary contours of the circular and two noncircular ellipses deform in qualitatively different patterns. This property of the images is closely related to the shape-discrimination task in Lappin and Love's (1992) experiments, but it would not constitute information about the metric structure of arbitrary shapes.

Still another possibility exists when such shapes are rotated through an angle of 90° or more: Since these shapes were bilaterally symmetric, they might have been discriminated by remembering and comparing the relative extensions of the two principal axes of the shape at the same orientation in the image. Similarly, the linear extension of any direction within the shape was maximized when it passed through the horizontal axis of the image plane in Lappin and Love's (1992) experiments, and this maximum extension could provide a standard for judging the shape.

Empirical Evidence Pertinent to These Alternative Hypotheses

The existence of such alternative forms of image information weakens the experimental support for the hypothesized scaling of visual space from motion, because

Figure 1. Schematic illustrations of the effect of several types of coordinate transformations on the images of a simple planar shape in two different positions in the image. The rectilinear coordinate grid has been added for illustrative purposes only; no such grid was visible in the displays in these experiments. (A) A standard cartesian coordinate system, with no transformation. The two patterns are circles, rotated from a central vertical position from a point in the lower third of the grid by 15° counter-clockwise and 60° clockwise. (B) The cartesian coordinate system in (A) has been changed by an affine transformation, $y' = 0.5y$, corresponding to a slant of (A) around the horizontal axis by 60° . (C) The same as (B) but with perspective projection from a point in depth from the image plane by a distance of twice the width of the cartesian grid. (D) The pattern in (A) has now been "wrapped" around a vertical cylinder with a diameter equal to two thirds the width of the cartesian grid in (A), and then projected orthographically. The transformations in (C) and (D) are nonaffine but isometric transformations of (A). That is, these transformations map the image plane isometrically onto a smooth surface with the same Gaussian curvature. Although these two transformations are not *globally* affine transformations of the image plane, they are described by *local* affine (linear) coordinate transformations of the image plane, where the parameters of the local coordinate transformation vary smoothly over the image.

observers might have discriminated these image properties rather than the metric shape per se. Whether or not observers do or can use any of these various forms of potentially useful image information is an empirically testable issue: The perceived properties can be characterized by the discriminable sets of image patterns and by convergence of discrimination performance over varying sets of shapes, motions, and behavioral tasks. Metric spatial structure is an abstract property generalizable over large sets of shapes, motions, and tasks, but in most experiments these alternative shapes, motions, and tasks are significantly restricted.

The *subjective appearance* of shapes moving in these transformed image planes under most of the conditions that we have studied does not support the alternative hypotheses corresponding to the cues identified by Pizlo and Salach-Golyska (1994), by Norman and Todd (personal communication, 1993), or those involving comparisons of features at common image positions. Subjectively, the moving shapes appeared to remain rigid and congruent over the changing positions in what appeared to be a plane slanted in depth; and the shapes of different ellipses appeared to differ in the curvature of their boundary contours and in the relative lengths of their principal axes—

consistent with the hypothesized scaling of space from motion. We were previously unaware of the image cues described by Pizlo and Salach-Golyska, and do not believe that observers in Lappin and Love's (1992) experiments were either aware of or used those properties. This is certainly not definitive evidence, however.

Consistency between subjective experience and hypothesized perceived properties may be regarded as necessary but not sufficient evidence for the hypothesized perception. Absence of confirming subjective experiences would usually raise problems for a given hypothesis about what is perceived, and the contrary subjective appearances raise problems for the hypotheses of Pizlo and Salach-Golyska (1994). Nevertheless, it is conceivable that observers actually discriminated the properties described by Pizlo and Salach-Golyska but attributed these discriminations to other subjective properties of the images.

On the other hand, subjective evidence is not sufficient. If the properties are indeed perceived, the perceptual sensitivity should be demonstrable in well-controlled tests of discriminability. Corroborating evidence of discriminability is especially important for hypotheses that vision is sensitive to the metric structure of perceived objects and spaces. Most observers seem to have subjectively compelling impressions of the metric shapes of perceived surfaces and often similarly compelling impressions of the spatial distances between neighboring objects. Despite these compelling impressions, Todd and his colleagues have convincingly demonstrated that in most such cases the discriminations are incorrect and unreliable, and where they are correct or reliable they may be based on discriminations of affine rather than metric structure (Norman & Todd, 1993; Todd & Bressan, 1990; Todd & Norman, 1991; Todd et al., 1993). Thus, Pizlo and Salach-Golyska (1994) have appropriately questioned the validity of apparent evidence that observers can discriminate metric structure.

The image information in Lappin and Love's (1992) experiments, however, differed from that available in most other experiments on perceiving structure from motion (on either the kinetic depth effect or motion parallax). In Lappin and Love's experiments, the differential structure of the image mapped onto a smooth surface (a 2-D manifold) that remained constant over time, was isometric with the image, and was specified by a globally constant affine coordinate transformation. In the general case of structure from motion, where the moving objects and features occupy a volume rather than a surface, none of these conditions holds.

As far as we know, the only other published experiments in which moving images were characterized by most of these geometric properties are those of Lappin and Fuqua (1983). Three collinear points were rotated in a plane slanted by varying amounts from the frontal parallel; and the plane of rotation was projected onto the image plane with exaggerated polar perspective, so that the relative positions of the points in 3-D space differed from the relative positions in the image plane. The observers'

task was to detect small differences in the position of the central point from bisection of the distance between the two outside points. Observers could detect deviations from bisection of only about 2% of the distance between the outside points.

Pizlo and Salach-Golyska (1994) point out, however, that these results do not imply perception of metric relations in 3-D space, because the bisection of a line segment is invariant under affine transformations. That is, the same discrimination performance could have been obtained even if distances in the depth axis relative to the image plane were incorrectly or unreliably perceived.

Nevertheless, Lappin and Fuqua's (1983) results indicate that relative distances within the plane of rotation were scaled by congruence—that is, *isometry*—under changing orientations: Because of the exaggerated polar perspective, the bisection property could not be obtained from any global affine transformation of the *image* plane. Indeed, observers were relatively inaccurate in discriminating deviations from bisection in the image plane itself. Subjectively, the imaginary line segment appeared rigid (isometric) under the apparent rotation in depth; but when the position of the central point bisected the image distance between the two outside points, it did not appear rigidly connected to the outside points and its relative image positions were not accurately discriminated. Thus, the perceived rigidity and discriminations of bisection were based on a spatial representation that derived from isometry under motion. (This does not imply that the scale of the depth axis per se was correctly perceived.) These perceptions and discriminations could not have been based on image properties like those described by Pizlo and Salach-Golyska (1994).

More recently, we have investigated the perceived spatial structure of other patterns moving in image planes altered by coordinate transformations different from those examined by Lappin and Love (1992). Specifically, we have investigated the discriminability of ellipse shapes like those used by Lappin and Love but randomly varied in size and rotating ($\pm 45^\circ$) in a transformed image plane that mapped onto a slanted plane with exaggerated polar perspective, with both perspective and slant varied between trials (Lappin & Craft, 1992). Generally speaking, the shape discriminations remained very accurate under these conditions, with only small decrements in accuracy from the variations in perspective, though some image deformations associated with the perspective were visible especially early in the practice in looking at such patterns.

In other recent but preliminary experiments, we have also examined the discriminability of differences in the relative lengths of two connected line segments (with angular separation varying from 110° to 130°) rotating ($\pm 45^\circ$) in an image plane in which both the horizontal and the vertical axes were multiplied by independent random values varying from 0.5 to 1.5 (Lappin, Craft, & Ahlström, 1993). Weber fractions for discriminable differences in relative length were about 5% for observers

with little practice, which is similar to performance for patterns in a nontransformed frontal parallel plane. These patterns contained neither of the cues mentioned by Pizlo and Salach-Golyska (1994), although they did contain other possible cues associated with rotation through 90° and with the visibility of maximum line lengths in the image.

We have also explored the perception of curved surfaces defined by transformed image planes that mapped isometrically onto cylindrical surfaces—where the coordinate transformation from the image plane onto the surface in depth varied nonlinearly but smoothly over the image plane (Lappin et al., 1993). Subjective estimates by naive observers of the perceived curvature and depth of the cylindrical surfaces were close to the correct values. This evidence is not definitive, however, because these patterns contained other theoretically detectable cues associated with relative velocity, and performance was also similar in a control condition theoretically lacking in potential metric information.

New Evidence

Because discrimination performance in the preceding experiments might conceivably have derived from perception of nonmetric spatial relations, we now report new psychophysical evidence that visual space can be metrically scaled by congruence of moving objects.

Method

Length discrimination task. The task was to identify the longer of two lines. The two lines were joined at a common vertex, at an angle of 70° or 110°. (These angles differed from 90° because we found in pilot experiments that a consistent right angle provided additional information about structure in 3-D that was visible even in stationary images.) This pair of lines was rotated around their common vertex $\pm 22.5^\circ$ (a total of 45°, in contrast to 90° in most of our previous experiments), so that in this case the two line segments never occupied the same position in the image, and only one of the two lines rotated through the image direction in which its length was maximized (minimized in the case of the other line).

Prior to applying coordinate transformation of the image plane, as described below, the length of one of the two lines was increased by a proportion that was constant throughout the block of trials. In the main coordinate transformation condition and in a frontal-parallel control condition, one line was 2% to 10% longer than the other; in another two-frame stationary control condition, which yielded poor discriminations, one line was 4% to 16% longer than the other.

In the coordinate transformation condition, the spatial coordinates of the image were transformed by randomly selecting a direction, with uniform probability over the 180° range, and then multiplying the cartesian coordinates in this direction by a random number uniformly distributed between 1.5 and 2.0. When the shapes were rotated in this transformed image plane, they appeared to rotate in a plane slanted in the direction perpendicular to the expanded axis, since the image was relatively contracted in this perpendicular direction. Of course these image transformations and implicit planes of rotation in depth were visible only by virtue of the image deformations of the rotating shapes and by the visual derivation of an isometric space from congruence under motion.

Another improvement over our previous experiments was that the line that was longer in the image was equally likely to be either

the longer or the shorter of the two lines in the implicit plane slanted in depth. This was accomplished by randomly placing the central position of the two-line shape and its 45° rotation range either -66.5° or $+66.5^\circ$ from the expanded axis of the image plane, so that either one of the lines was equally likely to move through this expanded image axis. Under this arrangement, the minimum ratio of the image lengths of two lines with equal lengths in 3-D was 1.27, a ratio much larger than the relative 3-D lengths of the lines tested in this discrimination experiment. This arrangement provides several controls on potential image cues, controls that were lacking in our previous experiments.

To provide additional image information about spatial structure, 30 dots were randomly positioned in the region between the two lines. When this whole pattern rotated in the image, the dots appeared to be rigidly connected to and in the same plane with the two lines. The positions of these dots were independently randomly generated in each display, and they provided no information about the relative lengths of the lines in the image or in 3-D. We have found in previous pilot experiments, however, that such additional spatial structure often produces more consistent perceptions of the structure and orientation of the plane containing the two lines and sometimes more accurate discriminations of line length. The apparent visual benefit from adding such extra spatial features may derive from information about the spatial connectedness and/or the second-order differential structure of the pattern (curvature and qualitative shape). Such second-order structure is theoretically irrelevant to the present patterns and task, but it may provide useful visual information (Koenderink & van Doorn, 1992).

Displays and equipment. These patterns were composed of white lines and dots on a black background displayed on a 19-in. gray-scale monitor (Silverview, with $1,152 \times 870$ pixels in an area of 39.4×29.5 cm) controlled by a Macintosh IIx computer. The monitor was viewed in a dimly lit room from a distance of about 57 cm. The luminance of the background was .03 cd/m², the luminance of the dots and one of the lines was 45.3 cd/m², and the luminance of the other line was 19.3 cd/m². The luminance of one line was reduced in order to distinguish the two alternatives for the discrimination task.

The lines were continuously connected between the end points, and the two shared a common vertex that was also the center of rotation, in the center of the display screen. The shorter of the two lines was 2.4 cm (2.4° arc) when perpendicular to the elongated axis, and lengthened to a maximum of 2.80–3.28 cm when it rotated, depending on the scale factor for the elongated axis between 1.5 and 2.0; when the same line was positioned near the elongated principal axis, its length ranged from a minimum of 3.30–4.25 cm to a maximum of 3.60–4.80 cm as it rotated, depending on the random scale factor in the range 1.5 to 2.0.¹ In the frontal-parallel control condition, the shorter line was always 3.0 cm, and the other line was longer by 2%–10%, depending on the increment discriminated in a given block of trials.

The 45° range of rotation was displayed in the coordinate transformation and frontal-parallel conditions by a series of 20 frames, each 45 msec, for a total of 900 msec for the 45° range of motion. In the two-view stationary condition, the pattern was displayed in the two positions at the ends of the range of rotation, with each frame displayed for 300 msec and a 150-msec interstimulus interval. On each trial, the pattern was presented for a maximum of eight such excursions, oscillating back and forth for a total of 7.2 sec or until the display was terminated by the observer's response. In the coordinate transformation and frontal-parallel conditions, the display was often terminated after about two excursions and seldom continued for the maximum eight excursions.

Experimental conditions. Three types of image conditions were examined: (1) the coordinate transformation condition, in which the image was altered by a random affine transformation that expanded the cartesian image coordinates in a randomly chosen direc-

tion by a random scale factor of 1.5–2.0; (2) a frontal-parallel control condition, in which the image coordinates were the normal x, y cartesian coordinates of the display screen; and (3) a two-frame stationary control condition, which displayed long-duration images of the patterns at the two positions at the ends of the rotation range with the same transformed images as in the coordinate transformation condition. In both the coordinate transformation and frontal-parallel conditions, discrimination accuracy was evaluated for five values of length difference, 2%–10% in steps of 2%. In the two-frame stationary condition, there were four values of length difference, 4%–16% in steps of 4%.

Procedures. Accuracies for this length-discrimination task in each of the 14 conditions were evaluated for 3 observers in two blocks of 48 experimental plus 4 practice trials, for a total of 288 data trials in each condition. Auditory feedback occurred after each correct response. For each of the three image conditions, the first block of trials was devoted to the largest length difference, and the difference was reduced in each succeeding block of trials until the length difference was the minimum for that condition. One block of trials was completed for all of the 14 conditions before beginning the second block for any condition. The full set of 14 conditions was usually completed in a single session of about 2 h, including breaks between the three types of spatiotemporal image conditions. The ordering of the three image conditions was varied between and within observers.

The observers were the two authors and a graduate student who was an experienced observer and knew the general purpose but not the specific manipulations of the experiment. All had practice in preliminary experiments with similar displays and tasks and additional practice with the present displays prior to collecting the data described below.

Results

The length-discrimination accuracies obtained for each observer in each condition are shown in Figure 2. Discrimination accuracy was measured by the choice theory measure, $-\ln \eta$ (essentially, the log odds of a correct response, corrected for response bias), which measures discriminability as a distance between two stimuli, similar to the signal detection measure, d' (Luce, 1963). As may be seen, length discriminations with randomly transformed image coordinates were similar to those in the frontal-parallel control condition. If performance is measured by a "threshold" value of relative length at $-\ln \eta = 1.0$ (where the percentage of correct responses equals 0.731), then, for the combined performance of all observers, the estimated Weber fraction in the coordinate transformation condition was 4.4%, and in the frontal-parallel control condition the estimated Weber fraction was 3.1%. (The threshold values were estimated by a minimum chi-square procedure that fit a straight line to the log odds of correct responses.)²

Subjectively, the spatial patterns in the coordinate transformation condition appeared to rotate rigidly in a plane slanted in 3-D space, and the relative line lengths usually seemed as clearly defined and discriminable as in the frontal-parallel condition. Indeed, even the frontal-parallel patterns sometimes appeared slanted in depth. Many errors in the frontal-parallel condition were associated with a horizontal-vertical illusion, where the more vertical line usually appeared longer. Todd et al. (1993) also recently found that when the relative orientations of two lines in

the image plane were randomized rather than matched, length discriminations declined, with a Weber fraction of about 5%. Evidently, vision does not provide especially precise information about metric relations in the frontal-parallel plane per se.

In contrast, in the two-frame stationary control condition the Weber fraction was 9.8%, and the rate of gain in accuracy with increasing length difference was much lower than in the other two conditions. Both the objective performance and the subjective appearance of patterns in this two-frame condition indicated that some information about 3-D structure was provided by this sequence of two images, though this information was imprecise and unreliable, associated in part with ambiguities about the angle between the lines in 3-D.

Conclusion

The discrimination performance obtained in the present experiment supports the interpretation suggested by Lappin and Love (1992): Visual space may be scaled by congruence of moving objects, and this scaling specifies metric structure in the special case in which there is a temporally constant affine transformation from the image onto a plane in which a rotating shape remains congruent (isometric) under motion. We think that successful discrimination performance in this task requires the perception of metric spatial relations, which specify relative distances invariant under changes in orientation and position within this implicit planar space. The patterns in the present experiments provide neither of the artificial image cues described by Pizlo and Salach-Golyska (1994) nor (except as described in note 1) any other information that we know about for performing this presumably metric discrimination task.

The capabilities for obtaining visual information about metric spatial structure may be more general than is indicated by the present experiment. Our working hypothesis is that such metric structure can be perceived whenever there is a temporally stable isometric mapping from the differential structure of a motion field in the image plane onto a surface, provided that the spatial complexity of the moving shapes and their range of motions in the image plane yield sufficient visual information for scaling the 2-D visual manifold by congruence of the moving shapes. Orthographic projection onto a plane—that is, by an affine transformation of the global image—is probably not necessary; an isometry under motions over the implicit surface, based on locally affine but smoothly varying transformations of the image plane, may be sufficient. That is, the perceived metric surface may be curved in a single direction and/or projected onto the image with polar perspective. This more general hypothesis is supported by observations in our laboratory, but it has not yet been confirmed by rigorous discrimination tests.

For cases in which there is a temporally stable mapping from the differential structure of the image onto that of a general quadric surface with curvature varying in two directions, where the Gaussian curvature varies over the

Figure 2. Discrimination accuracies in each of the 14 conditions for each of the 3 observers and for the combined data of all three observers.

surface, then of course shapes moving over this surface cannot remain isometric—but compelling perceptions of a curved surface can still be perceived. We have not yet determined the consistency and precision of such perceived surfaces, however.

In general, the hypothesis that visual space may be scaled by congruence of moving shapes seems to be well supported.

REFERENCES

- KOENDERINK, J. J., & VAN DOORN, A. J. (1992). Second-order optic flow. *Journal of the Optical Society of America A*, *9*, 530-538.
- 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., & CRAFT, W. D. (1992). 3-D structure from motion: A new method and new phenomena. *Bulletin of the Psychonomic Society*, *30*, 473. (Abstract)
- LAPPIN, J. S., CRAFT, W. D., & AHLSTRÖM, U. B. (1993). 3D structure from motion: A new method and new phenomena. *Investigative Ophthalmology & Visual Science*, *34* (Suppl. ARVO Abstracts), 1363.
- LAPPIN, J. S., & FUQUA, M. A. (1983). Accurate visual measurement of three-dimensional moving patterns. *Science*, *221*, 480-482.
- LAPPIN, J. S., & LOVE, S. R. (1992). Planar motion permits perception of metric structure in stereopsis. *Perception & Psychophysics*, *51*, 86-102.

- LUCE, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology* (Vol. 1, pp. 103-189). New York: Wiley.
- 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.
- PIZLO, Z., & SALACH-GOLYSKA, M. (1994). Is vision metric? Comment on Lappin and Love (1992). *Perception & Psychophysics*, **55**, 230-234.
- 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., NORMAN, J. F., PEROTTI, V. J., & TITTLE, J. T. (1993). The discrimination of 3D length from motion and stereopsis. *Investigative Ophthalmology & Visual Science*, **34** (Suppl. ARVO Abstracts), 1131.

NOTES

1. After reading the initial version of our report of this experiment, James Todd pointed out still another potential artifact that had escaped our notice: The minimum 2-D length of the shorter line in the image plane was always 2.4 cm when it corresponded to the shorter of the two lines. When this shorter line in the image corresponded to the longer line in 3-D, its minimum length in the image was always greater than 2.4 cm. Subsequent inspection of the displays with this artifact in mind supported our belief that the observers did not detect this information and indicated that they probably could not have used it even after direct practice: In our displays, the line that was shortest in the image plane was usually not subjectively obvious, and its minimum image length was not at all apparent. The lines usually appeared to maintain a constant length as they rotated in a slanted plane. Nevertheless, the objective was to remove all such extra 2-D image information, and we had not succeeded.

To check the possibility that this other potential source of information could be detected and used by our observers, we collected additional data under a better controlled version of the coordinate transformation condition in which this information was eliminated. Specifically, both principal axes of the display were expanded by randomly varying amounts. The shorter of these two axes was now expanded by a random factor of 1.0 to 1.2, with the magnitude of the scale factor selected randomly with uniform probability independently for each trial. The scale factor of the other axis was independently randomly varied between 1.5 and 2.0, as before. All other conditions were exactly the same as in the previous version of the coordinate transformation condition. Observer U.B.A., who was the most accurate of the observers in the main experiment, collected new discrimination data for both this modified version and the original version of the coordinate transformation condition. Data were collected for both a 4% and a 6% difference in lengths of the two lines. Discrimination data were collected for 192 trials, in four blocks of 48 trials, in each of these four conditions—4% and 6% differences in line lengths, for both the original and the modified versions of the coordinate transformation condition. Four brief sessions

were devoted to four replications of the experiment, with all four of the conditions occurring in each session in an order counterbalanced between sessions, and with blocks for the original and modified transformation conditions presented in immediate succession in counterbalanced order.

Discrimination accuracies were virtually identical for the modified and original versions of the coordinate transformation conditions: For a 6% difference in line lengths, judgments were correct on 164/192 trials (85.4%) in the original condition and on 163/192 trials (84.9%) in the modified condition. For the 4% difference in line lengths, the responses were correct on 137/192 trials (71.4%) in the original condition and on 136/192 trials (70.8%) in the modified condition. (Previously, this observer was 82.3% correct in the 6% condition and 76.0% correct in the 4% condition.) The observer reported that there was no visible difference between these two versions of the coordinate transformation conditions. Thus, we are confident that observers neither detected nor used information associated with the minimum image lengths of the two rotating lines.

2. The discriminability of differences in line length in this experiment was less than the discriminability of differences in shape in the study of Lappin and Love (1992). In Lappin and Love's Experiment 4, where two different ellipse shapes were discriminated, one expanded in one axis by 3% relative to a circle, and the other contracted in the same axis by 3%. The average accuracy in discriminating these two shapes was 87.5% correct. If one regards these shape discriminations as involving detections of 3% differences in relative spatial extent (rather than a 6% difference between the two alternatives), and if detectability (either $-\ln \eta$ or d') is assumed to be proportional to the difference in extent, then the "threshold" difference in line length (corresponding to $-\ln \eta = 1.0$, or 73.1% correct) would be just 1.54%. This value is almost one third that of the average threshold of 4.4% for a difference in lengths of the two lines in the present experiment. Obviously, different forms of spatial information were involved in these two experiments. In Lappin and Love's experiment, the two alternatives had qualitatively different shapes, defined by differences in curvature of the boundary contour of the ellipse as well as by differences in relative extent of the principal axes of the ellipses. In contrast, the differences in lengths of the two roughly perpendicular lines in the present experiment were less salient and less easily discriminated. Todd et al. (1993) also found that differences in line lengths in the 2-D image plane were significantly less discriminable when the relative orientations of the lines were made nonparallel, obtaining Weber fractions of about 5%. The motion of the patterns in our displays may have improved this discriminability slightly, as suggested by our Weber fractions of about 3.1% for the frontal-parallel condition. In addition to this difference in the discrimination task in Lappin and Love's experiment and in the present experiment, the range of motion was also less in the present experiment—45°, as opposed to 90° in most conditions of Lappin and Love's study. Lappin and Love found that discrimination declined rapidly with decreasing angular extent of rotation. Thus, the present discrimination task was more difficult for several reasons than that of Lappin and Love.

(Manuscript received July 22, 1993;
revision accepted for publication September 30, 1993.)