The visual perception of 3-D shape from multiple cues: Are observers capable of perceiving metric structure?

JAMES T. TODD
Ohio State University, Columbus, Ohio

and

J. FARLEY NORMAN
Western Kentucky University, Bowling Green, Kentucky

Three experiments are reported in which observers judged the three-dimensional (3-D) structures of virtual or real objects defined by various combinations of texture, motion, and binocular disparity under a wide variety of conditions. The tasks employed in these studies involved adjusting the depth of an object to match its width, adjusting the planes of a dihedral angle so that they appeared orthogonal, and adjusting the shape of an object so that it appeared to match another at a different viewing distance. The results obtained on all of these tasks revealed large constant errors and large individual differences among observers. There were also systematic failures of constancy over changes in viewing distance, orientation, or response task. When considered in conjunction with other, similar reports in the literature, these findings provide strong evidence that human observers do not have accurate perceptions of 3-D metric structure.

One of the most important problems in the study of human perception is to identify the data structures by which objects in three-dimensional (3-D) space are perceptually represented. This is especially important for the development of computational models for determining 3-D structure from various sources of optical information, such as shading, texture, motion, or binocular disparity (Marr, 1982). After all, in order to devise an algorithm for the computation of 3-D shape, it is first necessary to define what *shape* is.

There are many possible attributes of an object's structure that could potentially be used for its perceptual representation, and it is sometimes convenient to categorize these attributes on the basis of their stability under change, as was first suggested over a century ago by the German mathematician Felix Klein. To better understand the basic insight of Klein's proposal, it is useful to consider a variety of possible perceptual judgments for the polyhedral object depicted in Figure 1 and how those judged properties might be affected by different types of object transformations.

Suppose, for example, that an observer is asked to determine whether the point labeled "a" is located on the same face as the point labeled "c" (Koenderink, van Doorn, Kappers, & Todd, 1997; Phillips, Todd, Koenderink, & Kappers, 1997) or to determine whether their respective faces are connected at an edge. These judgments both involve *topological* relations, which are the most stable attributes of all, because they remain invariant over all possible continuous transformations. Alternatively, observers could be asked to judge whether the points labeled "a," "b," and "c" are collinear (Koenderink, van Doorn, Kappers, & Todd, 2002; Koenderink, van Doorn, & Lappin, 2000). Within the Klein hierarchy of geometries, collinearity is categorized as a *projective* relation, because it is unaffected by projective transformations. Another possible judgment that has been used in psychophysical investigations involves *ordinal* relations, such as determining whether "a" is closer in depth than "b" (Koenderink, van Doorn, & Kappers, 1996; Norman & Todd, 1998; Todd & Reichel, 1989). Ordinal relations are much less stable than topological or projective structure, but they still remain invariant over a wide range of order-preserving transformations. Other possible judgments for this object could involve estimating whether the faces containing "a" and "c" are co-planar (Todd & Bressan, 1990), whether the interval "a–b" is parallel to the interval "c–d" (Domini, Caudek, & Richman, 1998), or which of two parallel intervals is longer (Todd, Oomes, Koenderink, & Kappers, 2001). These are all examples of *affine* relations that remain invariant over all possible linear transformations. One final
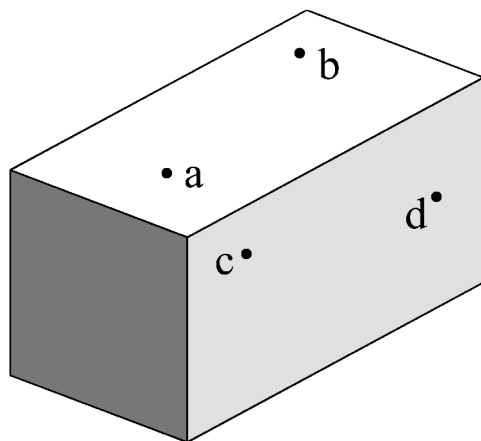
Figure 1. A polyhedral object with four labeled points. There are many relations among these points that could potentially be probed in a psychophysical experiment.

type of judgment to consider for this object involves *metric* relations among intervals in different directions, such as estimating the angle between "a–b" and "b–c" or deciding which of two nonparallel intervals is longer (Bradshaw, Parton, & Glennerster, 2000; Norman, Todd, Perotti, & Tittle, 1996; Tittle, Todd, Perotti, & Norman, 1995). Of all the possible aspects of geometric structure, these metric relations are the most unstable, since they are invariant only under a small set of rigid transformations.

In addition to categorizing invariance under change for different types of object transformations, the Klein hierarchy of geometries can also be used to categorize the mappings between objects in 3-D space and various aspects of optical structure, such as shading, texture, motion, or binocular disparity. For many of these potential sources of information, there are some aspects of 3-D shape that are uniquely specified, whereas others are inherently ambiguous. Theoretical analyses have shown, for example, that a shaded image of a Lambertian surface has an infinite number of possible 3-D interpretations that are all related by an affine transformation (Belhumeur, Kriegman, & Yuille, 1999). A similar ambiguity is also inherent in two-frame apparent motion sequences of objects observed under weak perspective (Koenderink & van Doorn, 1991). These sources of information are capable only of specifying the affine properties of an object, such as the parallelism of local surface patches or relative distance intervals in parallel directions, but they do not allow a unique determination of metric properties involving relative distance intervals in different directions.

It is especially interesting to note, in these cases, that observers will make judgments about 3-D metric structure from shaded images or two-frame apparent motion sequences even though it is not mathematically specified by the available information (see, e.g., Koenderink, van Doorn, Kappers, & Todd, 2001; Todd & Norman 1991). When asked to perform such judgments, they can apparently adopt unconscious strategies for filling in the unknown in-

formation. Recent evidence suggests, however, that these strategies can be strongly dependent on experimental context, so that the judged metric structure can vary significantly across different response tasks or among different observers (Koenderink et al., 2001).

There are other sources of visual information for which it is theoretically possible to uniquely determine the metric properties of objects. These include apparent motion sequences with three or more distinct views (Ullman, 1979), binocular stereograms with both horizontal and vertical disparities (Longuet-Higgins, 1981; Mayhew & Longuet-Higgins, 1982), or two-frame apparent motion sequences combined with horizontal binocular disparities (Richards, 1985). There is considerable controversy in the literature about the extent to which human observers are able to exploit this information. Although there have been some reports that judgments of metric structure are veridical when appropriate information is available (e.g., Brenner & Landy, 1999; Durgin, Proffitt, Olson, & Reinke, 1995; Johnston, Cumming, & Landy, 1994), a much larger number of studies have found that these judgments exhibit systematic patterns of distortion (e.g., Glennerster, Rogers, & Bradshaw, 1996; Hecht, van Doorn, & Koenderink, 1999; Tittle et al., 1995) and are often unreliable as well (e.g., Norman et al., 1996; Todd & Norman, 1995).

What could be the cause of these empirical discrepancies? Researchers on both sides of this issue have a standard argument for dismissing contradictory evidence. Those who believe that perception is veridical under most natural conditions will often focus on potential sources of information that are commonly excluded in psychophysical experiments. For example, when stimuli are presented on computer monitors, they do not produce the gradients of accommodative blur that are typically present when real objects are viewed at close distances. To the extent that these gradients are important for the perception of 3-D structure, experiments that employ computer displays could potentially produce misleading conclusions (Frisby, Buckley, & Horsman, 1995). However, it should also be noted in this context that distortions of perceived metric structure have also been obtained for judgments of real objects in fully illuminated natural environments (Baird & Biersdorf, 1967; Battro, Netto, & Rozestraten, 1976; Bradshaw et al., 2000; Cuijpers, Kappers, & Koenderink, 2000a, 2000b; Gilinsky, 1951; Harway, 1963; Hecht et al., 1999; Heine, 1900; Koenderink et al., 2002; Koenderink et al., 2000; Loomis, Da Silva, Fujita, & Fukusima, 1992; Loomis & Philbeck, 1999; Norman, Lappin, & Norman, 2000; Norman et al., 1996; Thouless, 1931; Toye, 1986; Wagner, 1985).

The standard argument for dismissing experiments in which observers produce veridical judgments of 3-D metric structure is that performance must have been based on some nongeneric cue that was inadvertently included in the experimental design. Consider, for example, a commonly used paradigm in which observers judge the apparent depth-to-width ratio of a rotating cylinder or a dihedral angle. If this type of surface is presented with

visible right-angle end cuts, its optical deformation allows a special case solution for the computation of 3-D structure from first-order motion measures (see the Appendix for details). This information is nongeneric, because it is available only for singly curved surfaces with right-angled end cuts whose orientation is perpendicular to the axis of rotation. Thus, to the extent that observers make use of such information, their performance would not generalize to other, more generic contexts. There have been several experiments reported in the literature in which observers judged the apparent depth-to-width ratios of rotating cylinders or dihedral angles (Braunstein, Liter, & Tittle, 1993; Hogervorst & Eagle, 1998; Johnston et al., 1994; Liter & Braunstein, 1998; Tittle et al., 1995). Johnston et al. is the only one of these for which the stimulus objects had visible right-angle end cuts. It is also the only one to obtain veridical performance.

Because of the potential use of nongeneric strategies, the only reasonable method of assessing observers' perceptual knowledge is to examine performance over a wide range of different contexts. If, as is argued by some, observers are sensitive to the relevant information from motion and/or binocular disparity for the perception of 3-D metric structure, the accuracy of their judgments should not be restricted to a limited set of stimulus configurations or response tasks. If, on the other hand, observers are insensitive to this information and they are forced to rely instead on nongeneric tricks for judgments of 3-D metric structure, there are likely to be significant variations in performance among different observers or different experimental paradigms.

The research described in the present article was designed to investigate the consistency of observers' judgments about 3-D metric structure over a variety of conditions that would be expected to be irrelevant if the observers were indeed sensitive to the potential information that is available within moving and/or stereoscopic displays. We measured the consistency among different observers, as well as the consistency of individual observers over different experimental sessions, response tasks, or stimulus configurations. Experiment 1 was similar in many respects to the earlier study by Johnston et al. (1994), in that observers made depth-to-width judgments for textured surfaces at multiple viewing distances with different combinations of motion and stereo. In order to assess consistency, however, we included an additional angle judgment task and presented the stimuli in multiple orientations. We also used surfaces with jagged edges in order to eliminate the end cut cue.

EXPERIMENT 1

Method

Apparatus. The stimuli were created and displayed on a Silicon Graphics Crimson VGXT workstation. The displays were viewed through liquid crystal display shuttered glasses that were synchronized with the monitor's 120-Hz refresh rate, so that the different views of a stereo pair would be visible only to the appropriate eye. The display monitor had a horizontal and vertical extent of 34.0 × 27.2 cm, and its spatial resolution was 1,280 × 1,024 pixels. The monitor was placed on a moving platform so that it could be positioned at different viewing distances. The observers viewed the displays while using a chinrest to restrict head movements.

Stimulus displays. A schematic diagram of the basic viewing geometry is shown in Figure 2, and stereograms of two representative displays are shown in Figure 3. The stimuli for this experiment were polar projections of concave dihedral angles formed by two differently oriented planes connected at an edge. Both planes were covered with a blue and red polka dot texture, and their flanking edges had irregular boundaries that varied randomly across trials. When
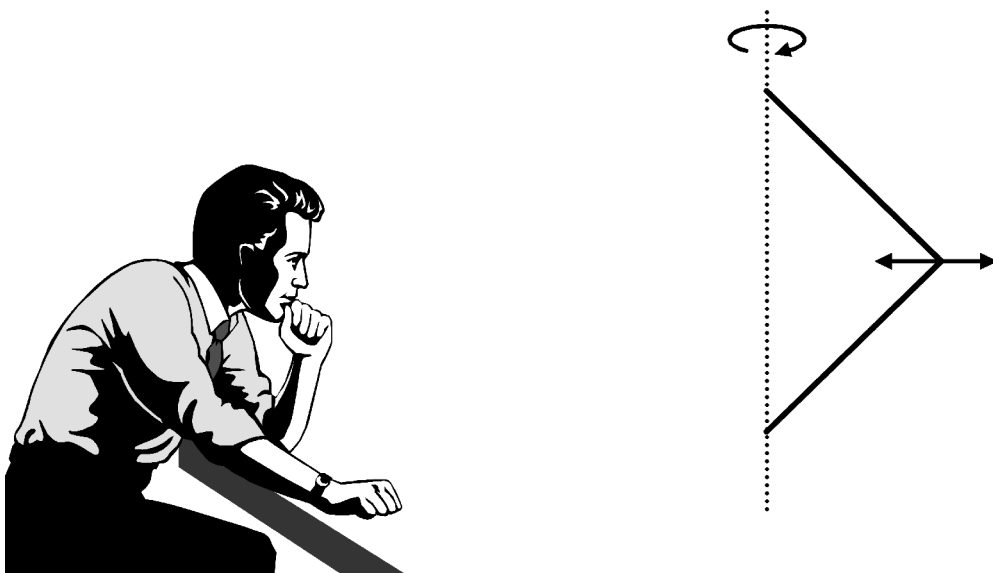

Figure 2. A schematic diagram of the viewing geometry of Experiment 1. Observers viewed a concave dihedral angle that could be stationary or rotating in depth about a vertical axis. The dihedral edge could be adjusted forward or backward in depth to achieve the desired setting.
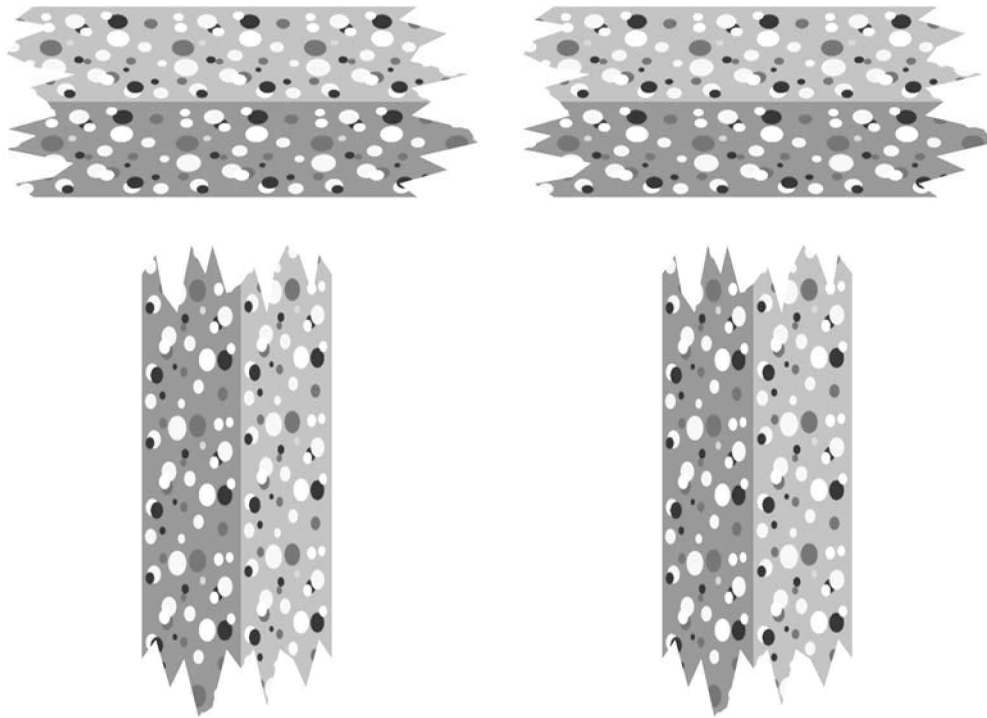
Figure 3. Some example stimulus configurations (designed for cross-fusion) that are similar to those employed in Experiment 1. The actual displays had a blue background with red polka dots.

presented in a fronto-parallel orientation, the two front edges of the dihedral angle were positioned in the image plane of the display monitor. The dimensions of its optical projection in that case were 20 cm along the length of the dihedral edge and 10 cm in width between the two front edges. The depth of the dihedral edge relative to the image plane could be adjusted by manipulating a hand-held mouse. The texture pattern was rescaled on every frame so that the shapes of the polka dots on the simulated surface would always be circular (i.e., the texture gradients were always consistent with the available information from motion and stereo).

Sources of information. There were three main viewing conditions in which the available sources of information about 3-D structure were systematically manipulated. In the static stereo condition, the dihedral angles were presented stereoscopically in a fronto-parallel orientation, as is depicted in Figure 3. In the monocular motion condition, the observer's nondominant eye was covered with an eye patch, and the surfaces were presented rotating back and forth in depth about a vertical axis through the center of the display screen. Each motion sequence consisted of 11 distinct frames in which the dihedral angle oscillated over a 20° range. The endpoints of this oscillation varied randomly across trials over a range of orientations between ±18° relative to fronto-parallel. These displays were presented with a frame-to-frame onset asynchrony of 67 msec, which was chosen to optimize the perception of rigid motion for this particular sequence length (see Todd & Bressan, 1990; Todd & Norman, 1991). We also included a stereo–motion combined condition, in which the moving dihedral angles were observed stereoscopically.

Spatial configurations. The depicted objects could be presented in two different orientations, so that the dihedral edge was either vertical or horizontal (see Figure 3 for an example of each). Note that the slants of the horizontal edges were affected by the object's rotation in depth but that the slants of the vertical edges remained invariant. The displays were also presented at two different viewing distances of 57 and 171 cm by moving the monitor platform. The parameters of the graphics simulation were adjusted accordingly for each distance.

Procedure. The observers were required to perform two separate tasks. In the depth-scaling task, they manipulated the depth of the dihedral edge until its distance from the plane of the two front edges appeared equal to the object width (i.e., the distance between the two front edges). In the angle-scaling task, in contrast, they adjusted the depth of the dihedral edge until the two planes appeared perpendicular. Note that these tasks were formally quite similar. For a correct setting of 90° on the angle-scaling task, the depth of the dihedral angle would be exactly one half its width.

To summarize the experimental design, there were 24 distinct conditions: 2 tasks × 2 viewing distances × 2 orientations × 3 combinations of motion and/or stereo. The experiment was performed over eight separate sessions, each consisting of three blocks. Within a given block, the two possible orientations were presented five times each in a random sequence, and all of the other parameters were held constant. All of the blocked conditions were repeated over two experimental sessions, so that each observer produced a total of 10 adjustments in each condition.

Observers. Six experienced psychophysical observers participated in the experiment, including the 2 authors and 4 other participants who were naive about the overall purpose of the study or how the displays had been generated. All the observers had normal or corrected-to-normal vision.

Results

During their debriefing sessions, all of the observers reported that the displays employed in this experiment produced perceptually vivid impressions of 3-D structure. However, they also all complained that the tasks they were

asked to perform were exceedingly difficult and that they did not have much confidence in the accuracy of their judgments. The observers reported that their adjustments produced clear variations in the apparent depth of a surface and the relative orientations of its two planar facets, but they experienced difficulties in trying to compare depth with width or in trying to decide when the dihedral angle was precisely 90°. It is important to keep in mind that it would have been theoretically possible to determine the correct 3-D metric structure in all of these conditions if the observers had been sensitive to the relevant information. This could have been achieved, for example, by analyzing the higher order relations among three or more views of an apparent motion sequence (Ullman, 1979) or by scaling binocular disparities, using first-order motion measures (Richards, 1985) or other sources of information about the distance of the monitor.

Although the observers were instructed in the angle-scaling task to adjust the planes of the dihedral angle so that they appeared to be orthogonal, these adjustments were actually recorded by the program as depth settings, which made it possible to analyze both tasks with the same measurement units. It is important to recognize that interpreting the raw settings on these tasks can be somewhat confusing. Suppose, for example, that a setting on the depth-scaling task is 50% of the actual stimulus width. Because the observers were instructed to make the depth and the width appear equal, a setting of .5 would indicate that the depth was perceptually overestimated by a factor of two. In order to avoid this confusion and to enhance the interpretability of the data, we measured the judged depth on each trial by computing the reciprocal of the observer's raw setting.

Figure 4 shows the average judged depth as a proportion of the correct setting for each observer on each response task. Note in particular that there was a wide variation of constant errors across the different observers that ranged from 5% to 75% of the correct setting. There were also large individual differences in relative performance on the two tasks. On average, the depths of these stimuli were systematically overestimated by 27% in the depth-scaling task and by 40% in the angle-scaling task.

An analysis of variance (ANOVA) for these data revealed several significant differences among the various conditions. For the depth-scaling task, there were significant main effects of orientation $[F(1,5) = 14.16, p < .01]$, viewing distance $[F(1,5) = 44.79, p < .01]$, and source of information $[F(2,10) = 7.962, p < .01]$, and a significant interaction between the latter two factors $[F(2,10) = 10.61, p < .01]$. The right panel of Figure 5 shows the average judged depth on the depth-scaling task as a function of orientation for each of the different sources of information. It is interesting to note in this figure that the vertically oriented dihedral angles produced much higher depth judgments than did those oriented in a horizontal direction (cf. Bradshaw & Rogers, 1999; Cagenello & Rogers, 1993; Cornilleau-Pérès & Droulez, 1989; Gillam, Chambers, & Russo, 1988; Gillam, Flagg, & Finlay, 1984; Nor-
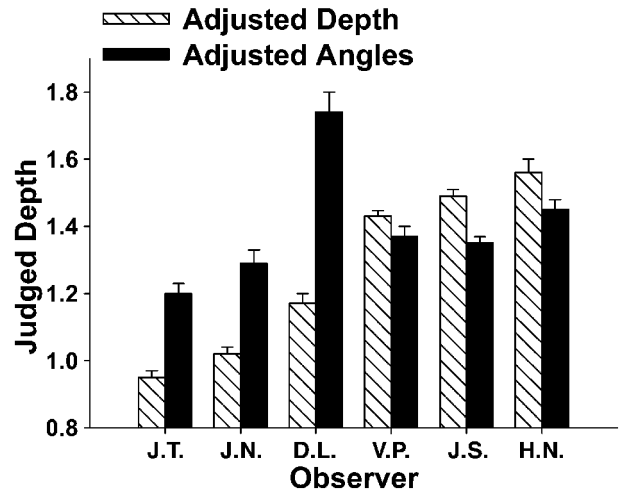

Figure 4. The average judged depth in Experiment 1 as a proportion of the correct setting for each combination of observer and response task.

man & Lappin, 1992; Rogers & Graham, 1983). With respect to the different sources of information, the judged depth was highest in the monocular motion condition and lowest in the static stereo condition. The ratings for the combined displays fell roughly midway between the other two. The left panel of Figure 5 shows the average judged depth for this task as a function of viewing distance. As has been reported previously by many other investigators, the magnitude of judged depth was significantly reduced with increasing viewing distance. A similar effect was also obtained for the monocular motion and combined conditions, although it was significantly attenuated relative to the static stereo displays (see also Tittle et al., 1995).

For the angle-scaling task, there were significant main effects of orientation $[F(1,5) = 43.55, p < .01]$ and source of information $[F(2,10) = 4.72, p < .05]$ and a significant interaction between those factors $[F(2,10) = 9.02, p < .01]$. The right panel of Figure 6 shows the average judged depth as a function of orientation for each of the different sources of information. The pattern of results was similar to that obtained for the depth adjustment task, except for the effect of viewing distance. When the observers were required to set the dihedral angle to 90°, as opposed to matching the depth and width, the manipulation of viewing distance had no effect whatsoever.

In order to assess the reliability of these effects, we computed the correlations between each pair of observers and the test–retest correlation for each individual observer across the different experimental sessions. The results of this analysis are shown in Table 1. Although the two authors (J.N. and J.T.) were able to perform these judgments with a reasonable degree of consistency, the 4 naive observers were much more variable. The correlations between observers were even lower, in some cases close to zero. This overall pattern is reminiscent of the results obtained by Koenderink et al. (2001), using displays for
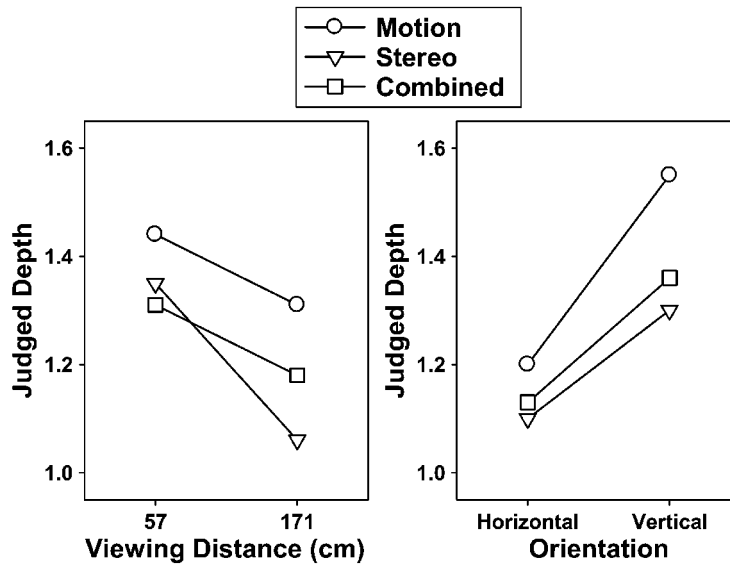
Figure 5. The average judged depth as a proportion of the correct setting for the depth-scaling task of Experiment 1. These judgments are plotted as a function of viewing distance and orientation for each combination of depth cues.

which a unique interpretation of 3-D metric structure was mathematically impossible.

There are two important aspects of these results that deserve to be highlighted. First, it should be noted that the variations in viewing distance had a significant effect on the judged depth-to-width ratios, but not on the judged angles, even though those properties were in perfect one-to-one correspondence on the adjusted surfaces. Glennerster

et al. (1996) found a significant effect of viewing distance for both of these tasks, using somewhat different stimuli (see also Tittle et al., 1995), but this effect disappeared when the observers compared the relative depths of two surfaces that were viewed simultaneously at different distances. If one believes that there is a single task-independent representation of the visual world, these variations over tasks should be considered quite troubling. Either the apparent
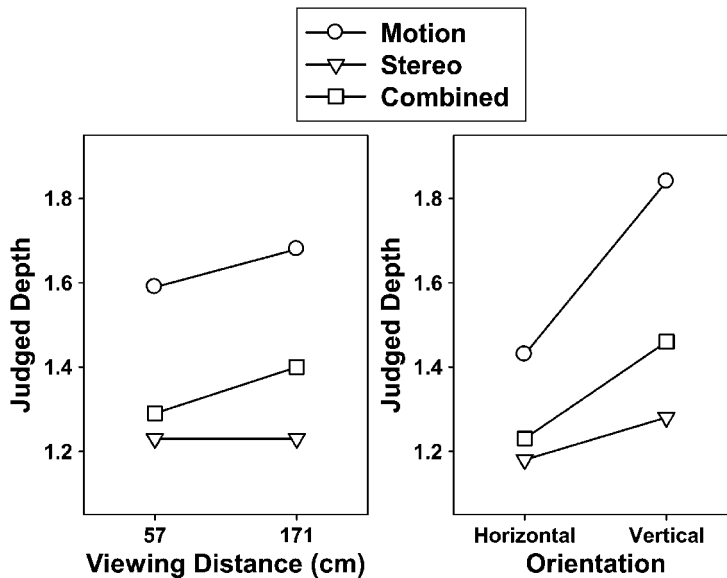

Figure 6. The average judged depth as a proportion of the correct setting for the angle-scaling task of Experiment 1. These judgments are plotted as a function of viewing distance and orientation for each combination of depth cues.

Table 1
The Test–Retest Correlation Squared (r^2) for Each Observer Across the Different Experimental Sessions of Experiment 1 and the Correlation Squared Between Each Pair of Observers

	D.L.	H.N.	J.N.	J.S.	J.T.	V.P.
D.L.	.65	.01	.58	.12	.60	.06
H.N.		.40	.18	.38	.28	.61
J.N.			.83	.01	.76	.19
J.S.				.21	.02	.21
J.T.					.80	.37
V.P.						.61

depth of an object changes with viewing distance, or it does not. The results should not depend on subtle details in how the apparent depth is measured psychophysically.

Another interesting aspect of this study involves the relative effects of stereo and motion on the magnitude of judged depth. During the debriefing sessions, we presented each observer with one of the displays from the stereo–motion condition, and we asked them to attend carefully to the apparent depth of the surface with both eyes open. We then asked them to close one eye and to describe whatever changes this produced in the object's appearance. All of the observers reported that the magnitude of perceived depth was reduced significantly when they switched from binocular to monocular viewing. Note, however, that these subjective reports are contradicted by the objective data from the same observers. All of them judged the stereo–motion combined displays to have at least 15% less depth than the monocular motion displays (see also Tittle et al., 1995).

This incompatibility of the objective data with the observers' phenomenal impressions provides strong evidence that there was a strategic component of their responses that was not based entirely on their conscious perceptions. The observers reported during their debriefings that they had compelling perceptions of depth from these displays but that they did not have a clear sense of how depth intervals are scaled relative to other physical dimensions. Thus, it is likely that these nonperceptual strategic components were elicited by their having to identify a specific metrical relationship between height, width, and depth. There is other evidence to suggest, moreover, that the strategies employed for metrical scaling are highly dependent on contextual factors, such as the nature of the experimental instructions (see Figures 5 and 6), and that there are large variations among different observers (see Figure 4).

EXPERIMENT 2

In an effort to further investigate how experimental context can influence observers' judgments of 3-D metric structure, Experiment 2 employed the same depth adjustment task and the same observers as those in Experiment 1, but with a different set of stimulus objects.
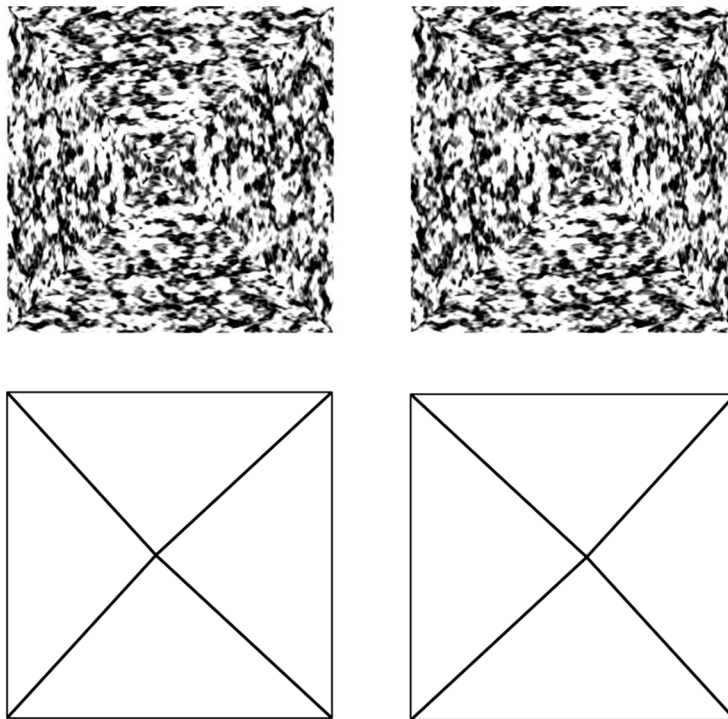

Figure 7. Some example stimulus configurations (designed for cross-fusion) that are similar to those employed in Experiment 2. In the actual displays, the texture had a reddish brown color.

Method

The apparatus and procedure were identical to those used in Experiment 1, except that only one task was employed. The displays in this case depicted stereoscopic and/or moving pyramids (see Figure 7) at viewing distances of 57 or 171 cm. The base of each pyramid had a height and width of 12 cm and was positioned in the plane of the display screen. The observers could adjust the position of the apex until the apparent depth of the pyramid appeared equal to its apparent width. The pyramids could be presented as wireframe figures, or they could be covered with a high-frequency texture resembling granite. As in Experiment 1, the texture pattern was rescaled on every frame as the object was adjusted so that its distribution on the surface would always be isotropic.

The displays were presented with the same three combinations of motion and stereo as in the previous experiment. In the moving conditions, the observer's nondominant eye was covered with an eye patch, and the surfaces were presented rotating back and forth in depth over a 20º range about a vertical axis through the center of the display screen. The endpoints of this oscillation varied randomly across trials over a range of orientations between ±18º relative to fronto-parallel, and the angular velocity was also varied randomly across trials over a continuous range between 1.5º and 2.5º per frame transition. All the displays had a fixed frame-to-frame onset asynchrony of 67 msec.

To summarize the experimental design, there were 12 distinct conditions: 2 viewing distances × 2 textures × 3 combinations of motion and/or stereo. The experiment was performed over four separate sessions, each consisting of three blocks. Within a given block, the textured and wireframe surfaces were presented five times each in a random sequence, and all of the other parameters were held constant. All of the blocked conditions were repeated over two experimental sessions, so that each observer produced a total of 10 adjustments in each condition. Judgments were obtained from the same 6 observers as those who had participated in Experiment 1.

Results

Figure 8 shows the average judged depth relative to the correct setting for the depth-scaling judgments of each observer in both Experiments 1 and 2. It is interesting to note that there was a wide variation of constant errors across the different observers and that the pattern of these variations was quite different in the two experiments. For 2 of the observers, the apparent depth-to-width settings were approximately equal for both pyramids and dihedral angles. Two observers produced significantly larger depth judgments for the pyramids, and the remaining 2 produced significantly larger depth judgments for the dihedral angles.

An ANOVA for these data revealed significant main effects of texture [$F(1,5) = 33.51, p < .01$], viewing distance [$F(1,5) = 7.20, p < .05$], and source of information [$F(2,10) = 29.18, p < .01$], and significant interactions of source of information with texture [$F(2,10) = 13.74, p < .01$] and viewing distance [$F(2,10) = 19.25, p < .01$]. The right panel of Figure 9 shows the average judged depths of the textured and wireframe objects for each of the different sources of information. It is interesting to note in this figure that the observers' depth judgments were significantly increased by the presence of texture (Peterson & Gibson, 1993), even though that texture provides virtually no information about 3-D structure when presented as a static pictorial depth cue (see Figure 7). The left panel of Figure 9 shows the average judged depth for this task as a
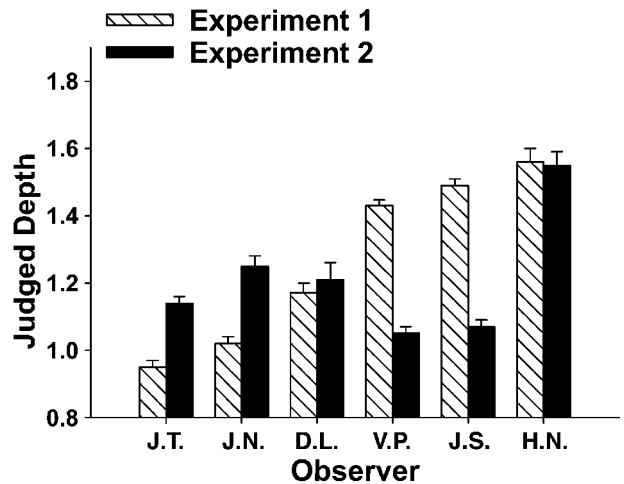

Figure 8. The average judged depth as a proportion of the correct setting for each observer on the depth-scaling tasks of Experiments 1 and 2.

function of viewing distance. As in Experiment 1, there was a significant decrease in judged depth with viewing distance, especially in the static stereo conditions. With respect to the different sources of information, the judged depth was highest in the monocular motion condition and lowest in the static stereo condition, although the magnitude of this effect varied significantly with viewing distance and the presence of texture.

The correlations between the patterns of judged depth for each pair of observers is presented in Table 2, as is the test–retest correlation for each individual observer across the different experimental sessions. With the exception of D.L., the observers were more reliable in this study than they were in Experiment 1, and the correlations between observers were also higher. This could be due to their increased experience in performing depth-to-width judgments or, perhaps, to some property of this particular stimulus configuration that made it easier to perform these judgments reliably.

These results provide further evidence that observers' judgments about the metrical relations between height, width, and depth are based on response strategies that can vary greatly as a function of experimental context. These judged metrical relations can be altered significantly by changing the nature of the experimental probe task (see Figures 5 and 6) or by manipulating the geometry of the depicted stimulus objects (see Figures 8 and 9). There are also large differences among individual observers (see Figures 4 and 8). When considered as a whole, the results of Experiments 1 and 2 are generally consistent with the earlier findings of Bradshaw et al. (2000) and Tittle et al. (1995). The evidence from these studies indicates that observers are unable to make accurate judgments about 3-D metric structure from motion or binocular disparity and that performance is not improved when both sources of information are presented in combination.
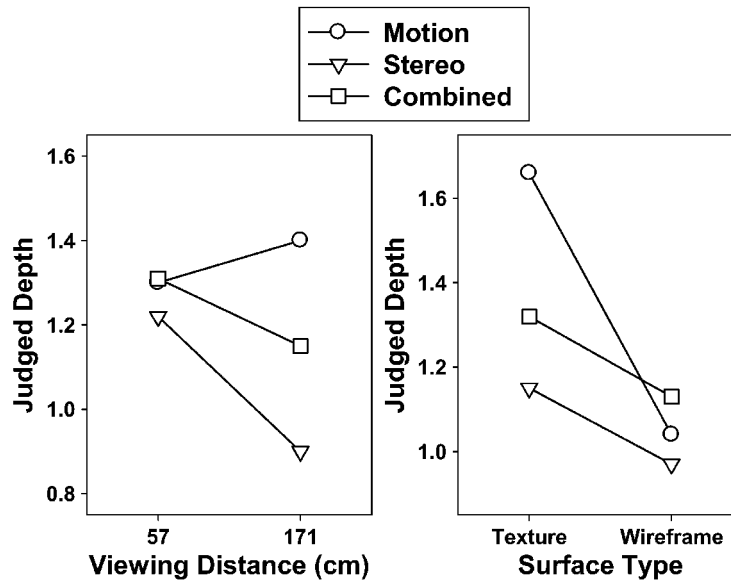
Figure 9. The average judged depth as a proportion of the correct setting for the pyramid adjustments of Experiment 2. These judgments are plotted as a function of viewing distance and surface type for each combination of depth cues.

EXPERIMENT 3

An interesting caveat to this conclusion is suggested by the recent findings of Glennerster et al. (1996) and Bradshaw et al. (2000). Using the same two tasks as those in the present study, they too found large systematic errors in observers' judgments of 3-D metric structure and a compression of perceived depth with viewing distance. They also employed an additional task, however, in which observers adjusted the depth of one surface until it appeared to match the depth of another that was positioned at a different viewing distance. Observers' judgments on this task were quite accurate, and they exhibited near-perfect constancy. It is important to recognize that accurate performance on this matching task did not require knowledge of metrical relations, because the intervals to be compared were parallel to one another. This might suggest, therefore, that the available information from stereoscopic vision provides veridical information about affine structure. An alternative possibility, identified by Glennerster et al., is that the use of identical monitors to present stereograms at different viewing distances may have allowed a nongeneric strategy for achieving accurate performance on this task. In an effort to further examine this issue, in Experiment 3, we used a similar matching task for judgments of real objects under full illumination.

Method

A schematic diagram of the experimental setup is presented in Figure 10. The stimuli consisted of three cardboard pyramids that were mounted on a cardboard box. Each pyramid had a height and width at its base of 24 cm. Their extensions in depth from the apex to the base were 12, 24, and 36 cm, respectively. The colors of the

cardboard were dark blue and white, which were alternated on adjacent faces. Each pyramid was identified by a number (1, 2, or 3) that was painted just above it on the box. The matching display depicted a stereogram of a pyramid that was identical in all respects to the textured stimuli in Experiment 2 (see Figure 7). That is to say, the simulated object had a height and width of 12 cm at the base and a depth that could be adjusted by manipulating a hand-held mouse. The monitor for the adjustment stimulus was always located at a fixed distance of 115 cm from the point of observation. The cardboard pyramids, in contrast, could be positioned at two possible distances of 75 or 235 cm. These objects were viewed, with a chinrest, under full illumination within a typically cluttered laboratory environment.

On each trial, a number was presented just above the stereogram in the monitor display plane to indicate which pyramid was to be matched. The observers terminated the trial with a mouse click when they were satisfied that the matching stimulus had the same shape as the target object. During each experimental session, the observers made five adjustments for each of the three target pyramids at a single viewing distance, and each distance was employed on two separate sessions. The displays were judged by the same 6 observers who had participated in Experiments 1 and 2 and by 4 new observers, who had no previous experience in psychophysical experiments.

Table 2
The Test–Retest Correlation Squared (r^2) for Each Observer Across the Different Experimental Sessions of Experiment 2 and the Correlation Squared Between Each Pair of Observers

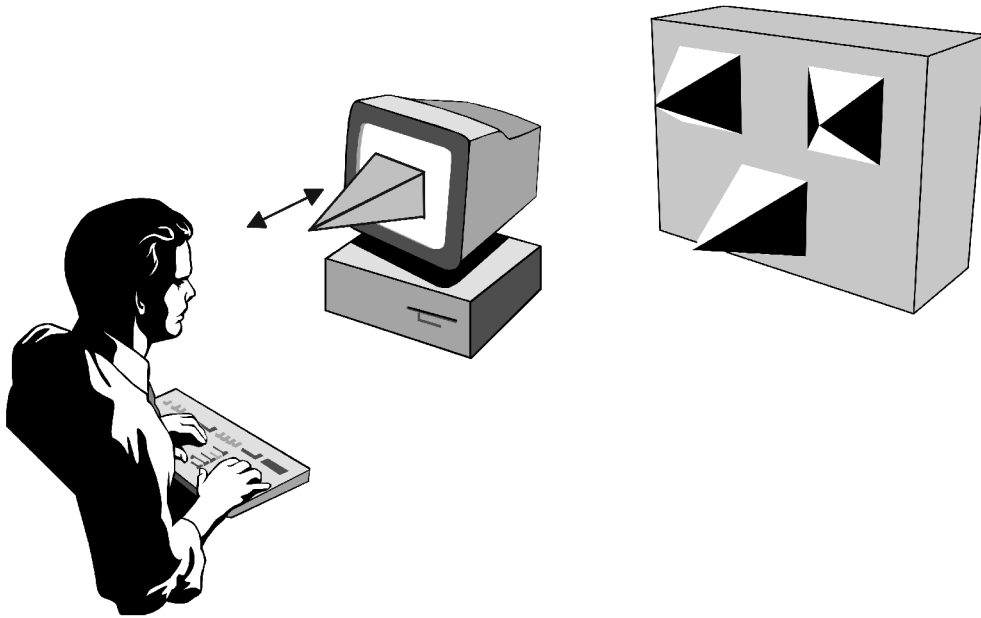|	J.T.	H.N.	D.L.	J.N.	J.S.	V.P.
J.T.	.93	.73	.43	.64	.55	.91
H.N.		.94	.68	.95	.80	.71
D.L.			.19	.77	.77	.46
J.N.				.82	.85	.61
J.S.					.82	.63
V.P.						.80

Figure 10. A schematic diagram of the viewing geometry used in Experiment 3. Observers adjusted the shape of a stereoscopic pyramid so that it appeared to match the shape of a real pyramid at a different viewing distance.

Results

Although these judgments could have been performed, in principle, by adjusting the matching stimulus to have the same apparent depth-to-width ratio (or apex angle) as the designated target object, that was not an absolute necessity in this case. An alternative strategy involving only affine relations would be to set the ratio between the depths of the target and matching stimuli so that it equals the ratio of their respective widths. The latter strategy could potentially be used to obtain the correct setting even if the relative depth-to-width ratio of each object was perceptually indeterminate.

During their debriefing sessions, all of the observers who had participated in Experiments 1 and 2 commented that this task was much easier than the depth-to-width or angle settings they had been required to perform in those earlier experiments, and the 4 new observers agreed that the task seemed quite natural. The observers' overall comfort with this task was also reflected in the reliability of their settings. When we compared the condition means of a given observer across different experimental sessions, the average r^2 for the test–retest correlations was .98. Similarly, the average r^2 for the correlations between observers was .95.

Despite this high level of reliability, however, the observers' judgments had relatively large constant errors, and they did not exhibit constancy over changes in viewing distance. Figure 11 shows the average judged depth as a proportion of the correct setting at the near and far viewing distances for each observer. Note in this figure that there was a wide variation of constant errors among the

different observers. Observer J.S., for example, consistently underestimated the depths of the target objects by 25%, whereas Observer V.P. consistently overestimated them by 20%. It should also be noted that when the viewing distance was increased from 75 to 235 cm, the judged depths were decreased by an average of 17%. This effect was significant for 9 of the 10 individual observers.

One potential criticism of this experiment (as well as of Experiments 1 and 2) is that the perceptual distortions may have been caused by the absence of some important source of information, such as gradients of accommodative blur (Frisby et al., 1995), that is not available when 3-D objects are simulated on computer monitors. To the extent that there is a gradient of accommodative blur on a 12-cm object at a viewing distance of 115 cm, this could, in principle, be a contributing factor for the constant errors in the observers' judgments. Note, however, that this cannot be a factor in the significant effects of viewing distance. Because the position of the computer monitor was never varied, these effects of viewing distance could be due only to the perceptual appearance of the real cardboard target objects.

It is clear from these results that the available sources of visual information in this experiment were insufficient for observers to make accurate judgments of either affine or metric structure. This would suggest, moreover, that the accurate depth-matching performance obtained by Glennerster et al. (1996) and by Bradshaw et al. (2000) may have a limited generality. This conclusion is reinforced by several other studies that have been reported previously in the literature. For example, Baird and Biersdorf (1967)
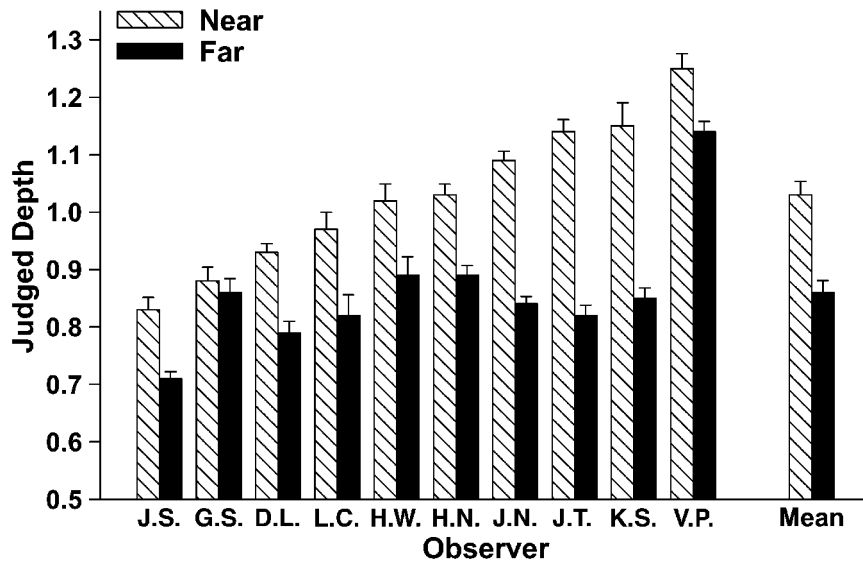
Figure 11. The average judged depth in Experiment 3 as a proportion of the correct setting for each observer at each viewing distance.

employed another type of depth-matching paradigm, in which observers were required to match the apparent lengths in depth of objects at different distances on a table top under normal illumination. As the viewing distance of the standard object was increased from 61 to 462 cm, its apparent depth was reduced by 24%. Other researchers have employed similar tasks over much larger viewing distances, in which observers indicated successive intervals in depth along the ground that appeared perceptually to be equal in length (Gilinsky, 1951; Harway, 1963). Although these judgments were performed in daylight under full cue conditions, the results revealed that perceived intervals in depth can be systematically compressed by over 50% at sufficiently long viewing distances.

GENERAL DISCUSSION

The research described in the present article was designed to investigate the abilities of human observers to judge the metrical relations between distance intervals in different directions. The results indicate that observers' judgments of 3-D metric structure can exhibit large constant errors and large individual differences and that they do not remain constant over changes in viewing distance or orientation. The pattern of performance can also be influenced by contextual factors, so that the relative scaling of height, width, and depth can vary significantly for different response tasks and different types of stimulus objects. A particularly interesting finding from Experiments 1 and 2 is that moving monocular displays produced significantly higher depth judgments than did moving stereoscopic displays (see also Tittle et al., 1995), yet the observers indicated in their debriefing sessions that the apparent magnitude of perceived depth was noticeably reduced when they switched from binocular to monocular viewing

of a moving stereogram. This finding suggests that the observers' response strategies may have involved both perceptual and nonperceptual components.

When the capabilities of human observers are evaluated, there are two methodological pitfalls for which researchers must always be vigilant. When one attempts to demonstrate that perception is incapable of some function, the primary danger is excluding some generic source of relevant information that would ordinarily be available in the natural environment. Conversely, when one attempts to demonstrate that a perceptual judgment can be performed accurately, there is a complementary problem of including nongeneric cues that would be useful only within a narrowly constrained experimental context. Because it is difficult to determine whether any given experiment has adequately controlled for all nongeneric sources of information and has included all relevant generic sources, the only viable procedure for evaluating the capabilities of human perception is to examine performance over a wide range of converging operations.

In order to assess these issues with respect to judgments of 3-D metric structure, it is important to consider all of the available evidence on this topic. Thus, in the discussion that follows, we will provide a general overview of prior research in this area for several different sources of information from which it is theoretically possible to compute metrical scaling relations among distance intervals in different directions.

Metric Structure From Motion

One of the most powerful sources of visual information about 3-D structure is provided by patterns of optical motion when objects are observed rotating in depth. Theoretical analyses of this information have shown that metric relations can be uniquely specified by motion sequences

containing three or more distinct views (Ullman, 1979) but that two-frame sequences are inherently ambiguous (Bennett, Hoffman, Nicola, & Prakash, 1989; Koenderink & van Doorn, 1991). There are, however, some non-generic special cases, in which metric relations can be specified from first-order motion measures. These include the motions of cylinders or dihedral angles with right-angle end cuts (see the Appendix) and the rotations of objects within a fixed slanted plane (Hoffman & Flinchbaugh, 1982).

There have been a few experiments reported in the literature in which human observers have made accurate judgments of 3-D metric structure from motion (Johnston et al., 1994; Lappin & Ahlström, 1994; Lappin & Love, 1992), but they have all involved one of the two special cases identified above. A more typical result in the vast majority of experiments on this topic is that judged metrical relations almost always deviate significantly from the physically specified structure and that they are often unreliable as well (Bocheva & Braunstein, 2000; Bradshaw et al., 2000; Braunstein & Andersen, 1984; Braunstein et al., 1993; Braunstein & Tittle, 1988; Caudek & Proffitt, 1993; Cornilleau-Pérès & Droulez, 1989; Domini & Braunstein, 1998; Domini & Caudek, 1999; Domini, Caudek, & Proffitt, 1997; Domini et al., 1998; Durgin et al., 1995; Eagle & Blake, 1995; Hogervorst & Eagle, 1998; Liter & Braunstein, 1998; Liter, Braunstein, & Hoffman, 1993; Loomis & Eby, 1988, 1989; Norman & Lappin, 1992; Norman & Todd, 1993; Norman et al., 1996; Norman, Todd, & Phillips, 1995; Perotti, Todd, Lappin, & Phillips, 1998; Perotti, Todd, & Norman, 1996; Tittle & Braunstein, 1993; Tittle et al., 1995; Todd, 1984, 1985; Todd & Bressan, 1990; Todd & Norman, 1991; Todd & Perotti, 1999; Turner & Braunstein, 1995; Werkhoven & van Veen, 1995). The relative scaling of height, width, and depth reported in these studies can vary dramatically depending on the details of each individual experiment. This overall pattern of results suggests that observers are insensitive to the higher order relations among three or more views that are required to compute 3-D metric structure from motion and that they must rely, instead, on heuristic strategies that are adapted for different experimental contexts (see, e.g., Braunstein et al., 1993; Caudek & Proffitt, 1993; Domini & Caudek, 1999; Todd & Perotti, 1999).

Metric Structure From Binocular Disparity

Another potential source of visual information for the perception of 3-D structure is provided by binocular disparity. Like two-frame motion sequences, the horizontal disparity between each eye's view is inherently ambiguous with respect to the metric structure of an observed scene. The ambiguity in this case arises from the fact that the disparity produced by a given depth interval varies with viewing distance and must, therefore, be scaled somehow in order to determine the correct metric relations between height, width, and depth (e.g., Longuet-Higgins, 1981; Mayhew & Longuet-Higgins, 1982).

There are two recent studies by Frisby, Buckley, and Duke (1996) and Durgin et al. (1995) that are frequently cited as evidence that human observers can accurately perceive 3-D metric structure from binocular disparity under natural viewing conditions. In Frisby et al.'s study, observers judged the relative 3-D lengths of gnarled sticks presented in different orientations and at different viewing distances. These judgments exhibited near-perfect constancy and low constant errors, although it is interesting to note that performance was nearly as good under monocular conditions in which no recognized depth cues were present. The latter finding provides a clear indication that there were nongeneric cues available in this experiment, which may have been necessary to achieve such high levels of accuracy. In Durgin et al.'s study, observers judged the depth-to-width ratios of wooden cones presented at different viewing distances. The effects of viewing distance were tested using a between-subjects design, however, and no data were presented for individual subjects. Given the large individual differences that were observed for similar judgments in our experiments (see, e.g., Figure 11), it is possible that their procedure may not have had sufficient statistical power to detect any failures of constancy that may have been present in the observers' perceptions.

Another important reason to be cautious in interpreting the results of these studies is that they are not representative of the overall literature on this topic. The vast majority of experiments on the perception of metric structure from binocular disparity have shown significant errors in the judged relations between height, width, and depth and systematic failures of constancy. Some of these experiments have involved judgments of virtual objects presented on computer graphics displays (Bradshaw, Glennerster, & Rogers, 1996; Brenner & Landy, 1999; Brenner & van Damme, 1999; Collett, Schwarz, & Sobel, 1991; Glennerster et al., 1996; Glennerster, Rogers, & Bradshaw, 1998; Johnston, 1991; Johnston et al., 1994; Norman & Todd, 1998; Tittle et al., 1995; Todd et al., 2001) and may, therefore, not contain all of the information that is available in natural vision. There are also many others, however, involving judgments of real objects in fully illuminated natural environments (Baird & Biersdorf, 1967; Battro et al., 1976; Bradshaw et al., 2000; Cuijpers et al., 2000a, 2000b; Gilinsky, 1951; Harway, 1963; Hecht et al., 1999; Heine, 1900; Koenderink et al., 2002; Koenderink et al., 2000; Loomis et al., 1992; Loomis & Philbeck, 1999; Norman et al., 2000; Norman et al., 1996; Thouless, 1931; Toye, 1986; Wagner, 1985).

There are two main patterns of perceptual distortion that are typically reported in these experiments: Physically straight lines in the environment can appear perceptually to be curved (see, e.g., Cuijpers et al., 2000a, 2000b), and apparent intervals in depth become systematically compressed with increased viewing distance (see, e.g., Bradshaw et al., 2000; Hecht et al., 1999). Although we are generally oblivious to these distortions of perceived space under ordinary conditions, it is possible to notice them

under certain circumstances if one pays careful attention. A particularly compelling example can be experienced by examining the apparent lengths of the dashed lines that separate lanes on a highway. If one looks off in the distance, these lines appear quite short, on the order of 1–2 m, but when viewed out of the side window, they appear several times larger.

Metric Structure From Multiple Cues

One possible method of resolving the ambiguities in motion and binocular disparity is to analyze them in combination (see, e.g., Landy, Maloney, Johnston, & Young, 1995). Because each of these potential sources of information allows a different family of possible interpretations, they could be used to mutually constrain one another when both are simultaneously available (Richards, 1985). There have been some reports in the literature that human observers may indeed be capable of combining information from motion and binocular disparity to achieve veridical judgments of 3-D metric structure (Brenner & Landy, 1999; Brenner & van Damme, 1999; Johnston et al., 1994), but we should again be cautious in evaluating the generality of these findings, because of the potential use of nongeneric cues. For example, in the procedure used by Brenner and van Damme and by Brenner and Landy, observers adjusted the shape of a rotating ellipsoid surface until it appeared to be spherical. Note that it is possible in this task to achieve the correct setting by nulling any visible distortions in the object's occlusion boundary, without having any knowledge of 3-D shape whatsoever. In an effort to address this criticism, Brenner and Landy showed that performance was reduced when distortions of the occlusion boundary were presented in isolation, but this does not preclude the possibility that they may have interacted with other sources of information to improve the accuracy of observers' judgments. The only way of demonstrating that the presence of this cue had no effect on performance would be to show that the same level of performance is maintained under conditions in which that cue is no longer available.

Another reason to be cautious before accepting these results is that they have not been replicated by other researchers. Several other studies in which the perception of 3-D metric structure from motion and binocular disparity (often in combination with other cues) has been investigated have found significant errors in the judged metrical relations between height, width, and depth and systematic failures of constancy (Bradshaw et al., 2000; Koenderink, Kappers, Todd, Norman, & Phillips, 1996; Norman et al., 1996; Norman et al., 1995; Tittle & Braunstein, 1993; Tittle et al., 1995; Todd, Koenderink, van Doorn, & Kappers, 1996). Other studies have shown, in addition, that observers have high discrimination thresholds for local metric properties on smoothly curved surfaces even when they are depicted in full cue displays with motion, stereo, shading, and texture (Norman & Todd, 1996; Norman et al., 1995; Phillips & Todd, 1996; Todd & Norman, 1995).

Conclusions

When one reviews the literature on perceived metric structure from various aspects of visual information, it is remarkable to note how little consistency there is among different experiments. Although there are a few common trends, such as the compression of perceived depth with viewing distance, the precise metrical relations between height, width, and depth can vary dramatically across different experimental conditions and procedures. This is an important empirical fact, we believe, that needs to be addressed by any adequate theory of the perceptual representation of 3-D structure.

One fundamental finding that is consistent among the vast majority of experiments in this area is that observers' judgments of 3-D metric structure can be systematically distorted and often exhibit large failures of constancy over changes in viewing distance and/or orientation. These results pose an interesting conundrum: If the perception of 3-D metric structure does not remain constant over changes in viewing distance or orientation, how can we have the phenomenal experience of constancy as we move through the environment? This apparent contradiction between conscious perception and the empirical data from psychophysical experiments could perhaps suggest that much of our perceptual awareness may be based on lower order geometric properties, such as ordinal or topological relations (Koenderink et al., 2002; Todd, Chen, & Norman, 1998; Todd & Reichel, 1989).

Of course, it is not necessarily the case that all aspects of perceptual knowledge are accessible to consciousness (e.g., Bridgeman, Kirch, & Spirling, 1981; Goodale & Milner, 1992), so we cannot rule out the possibility that an analysis of metric structure may be involved subliminally in other aspects of visually guided behavior. A particularly compelling phenomenon relating to this issue has been reported by Goodale, Milner, Jakobson, and Carey (1991). They studied a patient with visual form agnosia who was unable to subjectively distinguish between horizontal and vertical orientations of a visually presented rectangular object, but who was always able to orient her hand appropriately when asked to reach out and grab the object. A similar dissociation between conscious perception and action has been reported by Loomis et al. (1992). When observers in their study were asked to judge the relative depth-to-width ratios of lines on the ground, they produced systematic compressions of judged depth with viewing distance, as in many previous studies. However, when the same observers were shown a spot on the ground and then asked to walk to it while blindfolded, they were able to do so with little or no constant error (Rieser, Ashmead, Talor, & Youngquist, 1990; Rieser, Pick, Ashmead, & Garing, 1995; Steenuis & Goodale, 1988; Thomson, 1983).

One complication in evaluating these findings is that there have been few systematic analyses of what exactly an observer must "know" to successfully control complex visually guided behaviors, such as walking or reaching. Many tasks, such as tossing an object to a target, could be

performed, in principle, with just an ordinal representation of depth and a corresponding ordinal representation of limb force. On the basis of visual feedback after each toss, an observer could raise or lower the force associated with any given distance to achieve and maintain accurate performance.

In order to demonstrate that performance on a motor task involves accurate knowledge of affine structure, it would be necessary to show that distances can be subdivided. An affine variant of the blindfolded walking task, for example, would be to show observers a visible target and then ask them to walk some proportion of the distance after they are blindfolded.

In order to demonstrate that performance on a motor task involves accurate knowledge of metric structure, it would be necessary to show that observers can compare distances over multiple directions. We know of two different paradigms reported in the literature that satisfy this criterion, but they appear, at first blush, to be in conflict with one another. First, there is a considerable body of evidence that the relative perceived extents of arm movements in different directions are systematically distorted (Cheng, 1968; Daviddon & Chang, 1964; Day & Wong, 1971; Deregowski & Ellis, 1972; Hogan, Kay, Fasse, & Mussa-Ivaldi, 1990; Reid, 1954; Von Collani, 1979), which indicates that observers do not have accurate knowledge about the metric structure of their movement trajectories. There is another body of evidence, however, that blindfolded observers can successfully walk to a target via an indirect triangular path (Fukusima, Loomis, & Da Silva, 1997; Philbeck, Loomis, & Beall, 1997), which suggests that they can integrate distances over multiple directions for controlling locomotion.

One potential hypothesis for reconciling these findings is that that the perceived metric structure of visual space may be systematically distorted but that the structure of haptic-motor space has been adapted through visual feedback to have a compensating distortion that makes it possible to achieve accurate performance on visually guided metric tasks (see Loomis et al., 1992). This explanation is highly speculative, given that it is based on such a small number of studies. It does, however, suggest an intriguing hypothesis about why a metric representation of visual space might be shielded from consciousness. If our conscious representations of the visual world were metrically distorted, the environment would appear unstable, and we would be unable to experience perceptual constancy.

REFERENCES

BAIRD, J. C., & BIERSDORF, W. R. (1967). Quantitative functions for size and distance judgments. *Perception & Psychophysics*, **2**, 161-166.

BATTRO, A. M., NETTO, S. P., & ROZESTRATEN, R. J. A. (1976). Riemannian geometries of variable curvature in visual space: Visual alleys, horopters, and triangles in big open fields. *Perception*, **5**, 9-23.

BELHUMEUR, P. N., KRIEGMAN, D. J., & YUILLE, A. L. (1999). The basrelief ambiguity. *International Journal of Computer Vision*, **35**, 33-44.

BENNETT, B., HOFFMAN, D., NICOLA, J., & PRAKASH, C. (1989). Structure from two orthographic views of rigid motion. *Journal of the Optical Society of America*, **6**, 1052-1069.

BOCHEVA, N., & BRAUNSTEIN, M. L. (2000). The contributions of slant and tilt to the detection of local surface orientation in structure from motion. *Vision Research*, **40**, 3637-3649.

BRADSHAW, M. F., GLENNERSTER, A., & ROGERS, B. J. (1996). The effect of display size on disparity scaling from differential perspective and vergence cues. *Vision Research*, **36**, 1255-1264.

BRADSHAW, M. F., PARTON, A. D., & GLENNERSTER, A. (2000). The task-dependent use of binocular disparity and motion parallax information. *Vision Research*, **40**, 3725-3734.

BRADSHAW, M. F., & ROGERS, B. J. (1999). Sensitivity to horizontal and vertical corrugations defined by binocular disparity. *Vision Research,* **39**, 3049-3056.

BRAUNSTEIN, M. L., & ANDERSEN, G. J. (1984). Shape and depth perception from parallel projections of three-dimensional motion. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 749-760.

BRAUNSTEIN, M. L., LITER, J. C., & TITTLE, J. S. (1993). Recovering three-dimensional shape from perspective translations and orthographic rotations. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 598-614.

BRAUNSTEIN, M. L., & TITTLE, J. S. (1988). The observer-relative velocity field as the basis for effective motion parallax. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 582-590.

BRENNER, E., & LANDY, M. S. (1999). Interaction between the perceived shape of two objects. *Vision Research*, **39**, 3834-3848.

BRENNER, E., & VAN DAMME, W. J. (1999). Perceived distance, shape and size. *Vision Research*, **39**, 975-986.

BRIDGEMAN, B., KIRCH, M., & SPIRLING, A. (1981). Segregation of cognitive and motor aspects of visual function using induced motion. *Perception & Psychophysics*, **29**, 336-342.

CAGENELLO, R., & ROGERS, B. J. (1993). Anisotropies in the perception of stereoscopic surfaces: The role of orientation disparity. *Vision Research*, **33**, 2189-2201.

CAUDEK, C., & PROFFITT, D. R. (1993). Depth perception in motion parallax and stereokinesis. *Journal of Experimental Psychology: Human Perception & Performance*, **19**, 32-47.

CHENG, M. (1968). Tactile-kinesthetic perception of length. *American Journal of Psychology*, **81**, 74-82.

COLLETT, T. S., SCHWARZ, U., & SOBEL, E. C. (1991). The interaction of oculomotor cues and stimulus size in stereoscopic death constancy. *Perception*, **20**, 733-754.

CORNILLEAU-PÉRÈS, V., & DROULEZ, J. (1989). Visual perception of surface curvature: Psychophysics of curvature detection induced by motion parallax. *Perception & Psychophysics*, **46**, 351-364.

CUIJPERS, R. H., KAPPERS, A. M. L., & KOENDERINK, J. J. (2000a). Investigation of visual space using an exocentric pointing task. *Perception & Psychophysics*, **62**, 1556-1571.

CUIJPERS, R. H., KAPPERS, A. M. L., & KOENDERINK, J. J. (2000b). Large systematic deviations in visual parallelism. *Perception*, **29**, 1467-1482.

DAVIDDON, R. S., & CHANG, M. (1964). Apparent distance in the horizontal plane with tactile-kinesthetic stimuli. *Quarterly Journal of Experimental Psychology*, **16**, 277-281.

DAY, R. H., & WONG, T. S. (1971). Radial and tangential movement directions as determinants of the haptic illusion of an L figure. *Journal of Experimental Psychology*, **87**, 19-22.

DEREGOWSKI, J., & ELLIS, H. D. (1972). Effect of stimulus orientation upon haptic perception of the horizontal–vertical illusion. *Journal of Experimental Psychology*, **95**, 14-19.

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

DOMINI, F., & CAUDEK, C. (1999). Perceiving surface slant from deformation of optic flow. *Journal of Experimental Psychology: Human Perception & Performance*, **25**, 426-444.

DOMINI, F., CAUDEK, C., & PROFFITT, D. R. (1997). Misperceptions of angular velocities influence the perception of rigidity in the kinetic depth effect. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 1111-1129.

DOMINI, F., CAUDEK, C., & RICHMAN, S. (1998). Distortions of depth-order relations and parallelism in structure from motion. *Perception & Psychophysics*, **60**, 1164-1174.

Durgin, F. H., Proffitt, D. R., Olson, T. J., & Reinke, K. S. (1995). Comparing depth from motion with depth from binocular disparity. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 679-699.

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

Frisby, J. P., Buckley, D., & Duke, P. A. (1996). Evidence for good recovery of lengths of real objects seen with natural stereo viewing. *Perception*, **25**, 129-154.

Frisby, J. P., Buckley, D., & Horsman, J. M. (1995). Integration of stereo, texture, and outline cues during pinhole viewing of real ridge-shaped objects and stereograms of ridges. *Perception*, **24**, 181-198.

Fukusima, S. S., Loomis, J. M., & Da Silva, J. A. (1997). Visual perception of egocentric distance as assessed by triangulation. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 86-100.

Gilinsky, A. S. (1951). Perceived size and distance in visual space. *Psychological Review*, **58**, 460-482.

Gillam, B., Chambers, D., & Russo, T. (1988). Postfusional latency in stereoscopic slant perception and the primitives of stereopsis. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 163-175.

Gillam, B., Flagg, T., & Finlay, D. (1984). Evidence for disparity change as the primary stimulus for stereoscopic processing. *Perception & Psychophysics*, **36**, 559-564.

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1996). Stereoscopic depth constancy depends on the subject's task. *Vision Research*, **36**, 3441-3456.

Glennerster, A., Rogers, B. J., & Bradshaw, M. F. (1998). Cues to viewing distance for stereoscopic depth constancy. *Perception*, **27**, 1357-1365.

Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, **15**, 20-25.

Goodale, M. A., Milner, A. D., Jakobson, L. S., & Carey, D. P. (1991). A neurological dissociation between perceiving objects and grasping them. *Nature*, **349**, 154-156.

Harway, N. I. (1963). Judgment of distance in children and adults. *Journal of Experimental Psychology*, **65**, 385-390.

Hecht, H., van Doorn, A., & Koenderink, J. J. (1999). Compression of visual space in natural scenes and in their photographic counterparts. *Perception & Psychophysics*, **61**, 1269-1286.

Heine, L. (1900). Über Orthoskopie oder über die Abhängigkeit relativer Entfernungsschätzungen von der Vorstellung absoluter Entfernung [On "orthoscopy" or on the dependence of relative distance on the representation of absolute distance]. *Albrecht von Graefe's Archiv für Ophthalmologie*, **51**, 563-572.

Hoffman, D. D., & Flinchbaugh, B. E. (1982). The interpretation of biological motion. *Biological Cybernetics*, **42**, 195-204.

Hogan, N., Kay, B., Fasse, E. D., & Mussa-Ivaldi, F. A. (1990). Haptic illusions: Experiments on human manipulation and perception of "virtual objects." *Cold Spring Harbor Symposia on Quantitative Biology*, **55**, 925-931.

Hogervorst, M. A., & Eagle, R. A. (1998). Biases in three-dimensional structure-from-motion arise from noise in the early visual system. *Proceedings of the Royal Society of London: Series B*, **265**, 1587-1593.

Johnston, E. B. (1991). Systematic distortions of shape from stereosis. *Vision Research*, **31**, 1351-1360.

Johnston, E. B., Cumming, B. G., & Landy, M. S. (1994). Integration of stereopsis and motion shape cues. *Vision Research*, **34**, 2259-2275.

Koenderink, J. J., Kappers, A. M. L., Todd, J. T., Norman, J. F., & Phillips, F. (1996). Surface range and attitude probing in stereoscopically presented dynamic scenes. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 869-878.

Koenderink, J. J., & van Doorn, A. J. (1991). Affine structure from motion. *Journal of the Optical Society of America A*, **8**, 377-385.

Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. L. (1996). Pictorial surface attitude and local depth comparisons. *Perception & Psychophysics*, **58**, 163-173.

Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (1997). The visual contour in depth. *Perception & Psychophysics*, **59**, 828-838.

Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (2001). Ambiguity and the "mental eye" in pictorial relief. *Perception*, **30**, 431-448.

Koenderink, J. J., van Doorn, A. J., Kappers, A. M. L., & Todd, J. T. (2002). Pappus in optical space. *Perception & Psychophysics*, **64**, 380-391.

Koenderink, J. J., van Doorn, A. J., & Lappin, J. S. (2000). Direct measurement of the curvature of visual space. *Perception*, **29**, 69-79.

Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, **35**, 389-412.

Lappin, J. S., & Ahlström, U. B. (1994). On the scaling of visual space from motion—in response to Pizlo and Salach-Golyska. *Perception & Psychophysics*, **55**, 235-242.

Lappin, J. S., & Love, S. R. (1992). Planar motion permits perception of metric structure in stereopsis. *Perception & Psychophysics*, **51**, 86-102.

Liter, J. C., & Braunstein, M. L. (1998). The relationship of vertical and horizontal velocity gradients in the perception of shape, rotation and rigidity. *Journal of Experimental Psychology: Human Perception & Performance*, **24**, 1257-1272.

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

Longuet-Higgins, H. C. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature*, **293**, 133-135.

Loomis, J. M., Da Silva, J. A., Fujita, N., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 906-921.

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

Loomis, J. M., & Eby, D. W. (1989). Relative motion parallax and the perception of structure from motion. In *Proceedings: Analysis, Motion/Workshop on Visual Motion* (pp. 204-211). Washington, DC: IEEE Computer Society Press.

Loomis, J. M., & Philbeck, J. W. (1999). Is the anisotropy of perceived 3-D shape invariant across scale? *Perception & Psychophysics*, **61**, 397-402.

Marr, D. (1982). *Vision*. San Francisco: Freeman.

Mayhew, J. E., & Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. *Nature*, **297**, 376-378.

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

Norman, J. F., Lappin, J. S., & Norman, H. F. (2000). The perception of length on curved and flat surfaces. *Perception & Psychophysics*, **62**, 1133-1145.

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.

Norman, J. F., & Todd, J. T. (1996). The discriminability of local surface structure. *Perception*, **25**, 381-398.

Norman, J. F., & Todd, J. T. (1998). Stereoscopic discrimination of interval and ordinal depth relations on smooth surfaces and in empty space. *Perception*, **27**, 257-272.

Norman, J. F., Todd, J. T., Perotti, V. J., & Tittle, J. S. (1996). The visual perception of three-dimensional length. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 173-186.

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

Perotti, V. J., Todd, J. T., Lappin, J. S., & Phillips, F. (1998). The perception of surface curvature from optical motion. *Perception & Psychophysics*, **60**, 377-388.

Perotti, V. J., Todd, J. T., & Norman, J. F. (1996). The visual perception of rigid motion from constant flow fields. *Perception & Psychophysics*, **58**, 666-679.

Peterson, M. A., & Gibson, B. S. (1993). Shape recognition inputs to figure–ground organization in three-dimensional displays. *Cognitive Psychology*, **25**, 383-429.

Philbeck, J. W., Loomis, J. M., & Beall, A. C. (1997). Visually perceived location is an invariant in the control of action. *Perception & Psychophysics*, **59**, 601-612.

PHILLIPS, F., & TODD, J. T. (1996). The visual perception of local shape. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 930-944.

PHILLIPS, F., TODD, J. T., KOENDERINK, J. J., & KAPPERS, A. M. L. (1997). Perceptual localization of surface position. *Journal of Experimental Psychology: Human Perception & Performance*, **23**, 1481-1492.

REID, R. L. (1954). An illusion of movement complementary to the horizontal–vertical illusion. *Quarterly Journal of Experimental Psychology*, **6**, 107-111.

RICHARDS, W. A. (1985). Structure from stereo and motion. *Journal of the Optical Society of America*, **2**, 343-349.

RIESER, J. J., ASHMEAD, D. H., TALOR, C. R., & YOUNGQUIST, G. A. (1990). Visual perception and the guidance of locomotion without vision to previously seen targets. *Perception*, **19**, 675-689.

RIESER, J. J., PICK, H. L., JR., ASHMEAD, D. H., & GARING, A. E. (1995). Calibration of human locomotion and models of perceptual–motor organization. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 480-497.

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

STEENUIS, F. W., & GOODALE, M. A. (1988). The effects of time and distance on accuracy of target directed locomotion: Does an accurate short-term memory for spatial localization exist? *Journal of Motor Behavior*, **20**, 399-415.

THOMSON, J. A. (1983). Is continuous visual monitoring necessary in visually guided locomotion? *Journal of Experimental Psychology: Human Perception & Performance*, **9**, 427-443.

THOULESS, R. H. (1931). Phenomenal regression to the real object: 1. *British Journal of Psychology*, **21**, 339-359.

TITTLE, J. S., & BRAUNSTEIN, M. L. (1993). Recovery of 3-D shape from binocular disparity and structure from motion. *Perception & Psychophysics*, **54**, 157-169.

TITTLE, J. S., TODD, J. T., PEROTTI, V. J., & NORMAN, J. F. (1995). Systematic distortion of perceived three-dimensional structure from motion and binocular stereopsis. *Journal of Experimental Psychology: Human Perception & Performance*, **21**, 663-678.

TODD, J. T. (1984). The perception of three-dimensional structure from rigid and nonrigid motion. *Perception & Psychophysics*, **36**, 97-103.

TODD, J. T. (1985). The perception of structure from motion: Is projective correspondence of moving elements a necessary condition? *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 689-710.

TODD, J. T., & BRESSAN, P. (1990). The perception of three-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics*, **48**, 419-430.

TODD, J. T., CHEN, L., & NORMAN, J. F. (1998). On the relative salience of Euclidean, affine, and topological structure for 3-D form discrimination. *Perception*, **27**, 273-282.

TODD, J. T., KOENDERINK, J. J., VAN DOORN, A. J., & KAPPERS, A. M. L. (1996). Effects of changing viewing conditions on the perceived structure of smoothly curved surfaces. *Journal of Experimental Psychology: Human Perception & Performance*, **22**, 695-706.

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. (1995). The visual discrimination of relative surface orientation. *Perception*, **24**, 855-866.

TODD, J. T., OOMES, A. H., KOENDERINK, J. J., & KAPPERS, A. M. L. (2001). On the affine structure of perceptual space. *Psychological Science*, **12**, 191-196.

TODD, J. T., & PEROTTI, V. J. (1999). The visual perception of surface orientation from optical motion. *Perception & Psychophysics*, **61**, 1577-1589.

TODD, J. T., & REICHEL, F. D. (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychological Review*, **96**, 643-657.

TOYE, R. C. (1986). The effect of viewing position on the perceived layout of space. *Perception & Psychophysics*, **40**, 85-92.

TURNER, J., & BRAUNSTEIN, M. L. (1995). Size constancy in structure from motion. *Perception*, **24**, 1155-1164.

ULLMAN, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.

VON COLLANI, G. (1979). An analysis of illusion components with L- and T-figures in active touch. *Quarterly Journal of Experimental Psychology*, **31**, 241-248.

WAGNER, M. (1985). The metric of visual space. *Perception & Psychophysics*, **38**, 483-495.

WERKHOVEN, P., & VAN VEEN, A. H. C. (1995). Extraction of relief from visual motion. *Perception & Psychophysics*, **57**, 645-656.

APPENDIX

Consider a horizontal elliptical cylinder with a length of α, a height of β, and a depth of χ, whose optical projections under weak perspective are α', β', and χ', respectively. Suppose that this cylinder is rotating in depth about a vertical axis. At some time T_1 in the motion sequence, the cylinder will be oriented perpendicular to the line of sight, as shown in the upper panel of Figure A1. Its optical projection at that moment is defined by the following equations:

$$\alpha'_1 = \alpha,$$

$$\beta'_1 = \beta,$$

and

$$\chi'_1 = 0.$$

At some later time T_2, the cylinder will have rotated by an angle ω, as shown in the lower panel of Figure A1, and its optical projection will be transformed to

$$\alpha'_2 = \alpha \cos(\omega),$$

$$\beta'_2 = \beta,$$

and

$$\chi'_2 = \chi \sin(\omega).$$

After terms are rearranged and appropriate substitutions are made, the depth of the cylinder is uniquely specified from optical information by the following equation:

$$\chi = \chi'_2 \sqrt{1 - \frac{\alpha'^2_2}{\alpha'^2_1}}.$$

Note that this computation is based on just two views of a motion sequence, one of which must be a fronto-parallel view. It does not depend on the cross-sectional shape of the cylinder, so the same analysis can be applied to half cylinders (Johnston et al., 1994) or dihedral angles.

A similar source of information is also available at slanted orientations from measures of optical velocity to specify the ratio of depth to length:

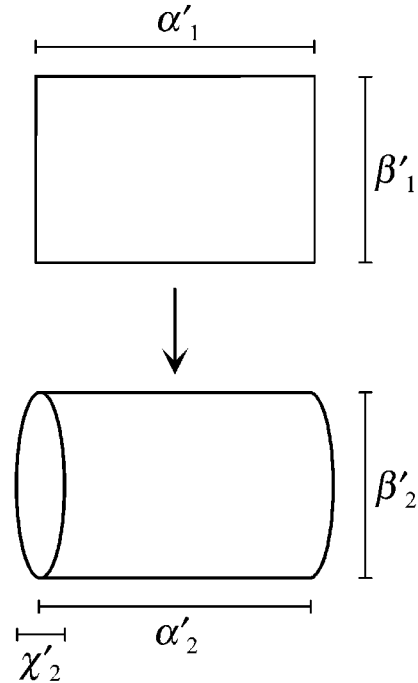$$\frac{\chi}{\alpha'} = \sqrt{-\frac{\chi' \frac{d\chi'}{dt}}{\alpha' \frac{d\alpha'}{dt}}}.$$

Figure A1. When a horizontal cylinder with right-angle end cuts rotates about a vertical axis, its optical deformation allows a special case solution for the computation of three-dimensional structure from first-order motion measures.

Both of these methods of computing metrical relations are highly nongeneric, because they are available only for singly curved surfaces with right-angled end cuts whose orientation is perpendicular to the axis of rotation. To prevent these strategies from being used in psychophysical investigations, the nongeneric cues can easily be eliminated by occluding the right-angle end cuts (Braunstein et al., 1993; Liter & Braunstein, 1998; Tittle et al., 1995), using end cuts that are not right angles, as in Experiment 1 of the present article, or by orienting the cylinders parallel to the axis of rotation (Hogervorst & Eagle, 1998).