

The shape and psychophysics of cinematic space

JAMES E. CUTTING
Cornell University, Ithaca, New York

For the film goer who sits to the front and side of a movie theater, the virtual space "behind" the screen undergoes affine and perspective transformations. These transformations should, one would think, make the rigidity of objects on the screen very difficult to discern. Despite the fact that it has long been known that viewers are not very sensitive to such distortions, a phenomenon I call La Gournerie's paradox, the effect is without a good theoretical account. Two possibilities are: (1) that viewers rectify the distortions of Euclidean space through the use of information about screen slant, and (2) that sufficient information is preserved under these transformations so that perception may be unperturbed. This paper presents preliminary arguments for the information-preservation view and introduces a new technique, that of using simulated projection surfaces, whose use in experimental situations suggests that Euclidean rectification is not necessary.

Psychophysics is now 125 years old. As a discipline, it has two roots, one in philosophy and the other in methodology, even pragmatics. The philosophical root, the one that gave psychophysics its name, was the mind-body problem, specifically the consideration of the mapping between mental and physical dimensions. The pragmatic and methodological root was the need for precision and stimulus control. The recognition of this need—no small feat in itself—and the manner in which psychophysics could deal with it won prompt and sustained attention. Given 19th century technology, stimulus control was usually achieved by using brass instruments that measured such physical units as mass, frequency, intensity, extent, and time. Typical experiments varied stimuli along only one of these dimensions.

Today, few researchers would deny the importance of the mind-body problem. Many, however, would suggest that psychophysics is no longer close to the core of this problem in psychology, much less philosophy. But this is not the major public-relations problem for psychophysics. Most researchers in the cognitive sciences have simply grown weary of results garnered from simple studies with simple stimuli. As a consequence, many researchers view psychophysics as an inherently dry discipline, essentially unrelated to the perception of complex, everyday phenomena.

This need not be the case and, in fact, it should not be. The reason for psychophysics' generally low esteem and visibility is a common view that confuses the pragmatics of the 19th century with those of the late 20th. The key idea in Fechner's (1860/1959) program was not that brass instruments must be used, but that the physical stimulus must be controlled and measurable (Haber, 1983). Since the advent of inexpensive fast computers and display

devices, one can control stimuli in domains previously unthinkable. In particular, and most important to me, such control can be achieved without compromising the naturalness of displays. Thus, the new visual psychophysics is capable of mapping a new set of physical and mental dimensions that are more relevant to human perceivers. As an example and as the focus of this paper, it is now possible to do a psychophysical study of cinema. First, it is important to discuss film in general to introduce the issues at stake.

EFFECTIVENESS OF CINEMA: WHAT'S "LOST"

Film may be art's most engrossing medium. We sit in a dark theater, look at a big bright screen, and lose ourselves in what is portrayed. This striking phenomenon is captured by the multiple-*entendred* title of Kael's (1965) collection of reviews, *I Lost It at the Movies*. One of the things we lose is the sense of ourselves; we become absorbed in the content of the film. Why film is so powerful in capturing viewers is not clear. Directors, cinematographers, script writers, sound engineers, composers, and actors certainly matter; they make the difference between good films and bad. Astonishingly, however, almost all films, good and bad, have the capacity to draw the viewer into them. This form of "losing it" is more than simply suspending disbelief; it is a suspension of nearly everything and a projection of oneself into a new space and time.

Photographic and cinematic standards are, in part, the cause of film's efficacy. The grain of 35-mm film is so dense that one image is equivalent to a raster-scan display of about 2,000 lines (Fox, 1983), about 4 times television quality along the raster axis and 16 times along both raster and scan axes. Even if such a display were as large as 40° × 40° (about twice the size of a movie screen as viewed from the middle front of a theater), it would still

This research was supported by NIMH Grant MH37467. Requests for reprints should be sent to James E. Cutting, Department of Psychology, Uris Hall, Cornell University, Ithaca, NY 14853-7601.

be near the nonvernier limit of foveal acuity; 70-mm and 140-mm format films obviously exceed such acuity. There are also considerable effects of screen size. Larger screens are more effective than small ones in giving the viewer an impression of "being there" (Hatada, Sakata, & Kusaka, 1980). It is as if filling the optic array also fills one's belief in what is seen. This belief may be enhanced by two other effects: Dimmed theater lights allow very little in the "real" world (other than exit signs) to be visible, and larger screens are much more likely to present moving stimuli to the retinal periphery, a region less devoted to acuity than to motion and change. In addition, sufficiently many separate frames strobed by an episotister (spinning shutter) guarantee that images of stroboscopic motion (often misnamed as apparent motion) appear like real motion. These separate images are sufficiently above flicker threshold for "flicks" (movies) to appear convincingly lifelike. Unlike resolution and flicker thresholds, however, the contrast ratios available in film are well below those of the eye in a given state of adaptation. Even so, most film is capable of ratios of about 500:1, a tenfold superiority over ratios used in television (Fox, 1983). The end result of grain, screen size, presentation rate, and contrast ratio considerations is that what is portrayed in cinema has many of the optical characteristics of what is portrayed in the real world.

Less easy to explain, however, are the many technical aspects of film and filmmaking that help engross the viewer. Hochberg and Brooks (1978) described the effects of pacing in cuts from one scene to the next. Fast pacing, for instance, can give a breath-taking quality to action sequences. Hochberg and Brooks also discussed the utility of pans and various other camera shots, which play on what a mobile eye might do in various environments. Why cuts and pans work as they do, however, is almost a complete mystery, since they do not mimic the experience and the scanning of everyday life.

EFFECTIVENESS OF CINEMA: WHAT'S NOT LOST

The most astonishing thing about film, however, is that it works at all as an optical projection. Every movie goes sits in a place different from the projector—and with obvious reason. This seating arrangement means that the image projected back from the screen to the viewer's eye is distorted and geometrically different from that of the scene as originally filmed. These distortions are substantial for most seat locations, and they create a bizarre virtual space "behind" the movie screen. One might think that the result would guarantee a loss of something in cinematic appreciation, particularly of our ability to discern certain aspects of the three-dimensionality of the virtual space. Obviously, however, these distortions matter little.

To sketch the shape of these distortions, let me start with a discussion of static pictures and their perception from different viewpoints. More detailed accounts of projections

in general have been given by Braunstein (1976), Cutting (1986), Pirenne (1970), Sedgwick (1983), and, in a particularly good account, Carlbohm and Paciorek (1978).

Pictures, Viewpoints, and La Gournerie's Paradox

Since the Renaissance era, most pictures in linear perspective have been composed from a particular point in space, often called a station point; photographs are shot from it. I call it the *point of composition*. In painting, the ideal (if not entirely practical) procedure for constructing such a picture comes from Alberti's *Della Pittura* in 1436: One holds up a flat pane of glass and traces on it what is seen with one eye at the point of composition. It is generally thought to be important to hold the pane so that the centric ray (the line of sight to the middle of the pane) is orthogonal to the image surface (Edgerton, 1975; White, 1957), although Pirenne (1970) has varied this experimentally in photography. If one is looking at a well-carpeted environment, tracings with linear perspective will result, with parallel lines of interior or exterior architectural features converging to a point or points on the horizon. But again, this array of geometric layout is correct only for the point of composition.

Perhaps the most fortuitous attribute of picture perception, then, is that one may view a picture from many locations other than the point of composition, and distortions of virtual space will interfere little with the perception of the picture's content. If this were not true, the utility and appreciation of pictures would be vanishingly small. Nonetheless, this fact is not predicted by linear perspective theory and is generally without good theoretical explanation. There is some debate about how much the distortions are noticed at various oblique viewing angles and at various distances from the image (Farber & Rosinski, 1978; Gibson, 1947, 1954; Kubovy, 1986; Lumsden, 1980; Pirenne, 1970; Purdy, 1960; Rosinski & Farber, 1980), but it is clear that identification and enjoyment of virtual objects remains high from many viewpoints (Hagen & Elliott, 1976; Hagen & Jones, 1978; Perkins, 1972, 1973, 1982).

At the end of the 15th century, Leonardo da Vinci (Richter, 1883/1970) discussed various oddities and problems of planar projections, but the first systematic treatment of pictures seen from noncomposition points was given by La Gournerie in the mid-19th century (Pirenne, 1970). Thus, I call the viewpoint nonspecificity of pictures *La Gournerie's paradox* (Cutting, 1986); Kubovy (1986) called it the *robustness of perspective*. Since La Gournerie's work, it has been useful to reconstruct the layout of virtual space (the space *in* the picture, behind the projection surface) to inspect these changes.

The idea behind virtual-space reconstruction is that perceptual properties, such as rigidity, can be investigated first by description and then by psychophysical manipulation. Justification for reconstruction is simply that it reverses the photographic process; if we make certain assumptions about the layout of objects in virtual space (e.g.,

the assumption that we are not looking into an Ames room), then relative sizes and distances of objects in space can be quite accurately determined.

Two sets of transformational distortions are important in this reconstruction. Consider first the changes in all x - z planes in (or horizontal slices through) virtual space, and then those in all x - y (or frontal) planes.

Horizontal-plane distortions. Distortions in horizontal planes of virtual space belong to a class called *affine transformations*; these preserve collinearities among points but not interpoint distances. The affine transforms of particular interest are shown in the top panels of Figure 1. The left-hand panel shows a view from the composition point and the reconstructed virtual layout of four poles. The first transform, shown in the central panels, is due to a change in viewpoint normal (perpendicular) to the image plane. Moving closer to the image creates a compression of the z -axis (or depth dimension) of virtual space analogous to looking through a telescope; moving away from the image creates expansion in depth analogous to looking through a microscope. The second transformation, and perhaps the more interesting, is caused by a change in viewpoint parallel to the plane, as in the top right-hand panel of the figure. This creates a shear of any given x - y (or frontal) plane against all its parallels in depth. Thus, the z -axis is no longer orthogonal to any frontal plane, but angles off to one side. All non-composition viewpoints at all angles and distances create additive combinations of these two types of affine distortions.

Frontal-plane distortions. Distortions in frontal planes, those parallel to the movie screen itself, are more familiar. Rectangular shapes become trapezoidal in their optical projection, as shown in the bottom-right panel of Figure 1. From a seat in the front of a movie theater on a side aisle, the vertical edge of the near side of the screen may be 40% to 50% larger than the far edge. Unlike horizontal planes, which have undergone affine transformations, frontal planes have undergone *perspective transformation*.

Hybrid geometry of virtual space. Although both types of distortion are reasonably well-known in the picture perception literature, the joint consequences of the two have not been fully appreciated. The combination of affine distortions in all horizontal planes and perspective distortions in all frontal planes creates a virtual space with a hybrid geometry. This hybrid geometry, and the objects that might fill its space, can be the subject of psychophysical analysis. Before considering this analysis, however, we must consider the general classes of theory that might be used to account for La Gournerie's paradox.

Classes of Explanations for La Gournerie's Paradox

Why are pictures useful and enjoyable when viewed from noncomposition points? How are affine and perspective transformations dealt with by the perceiver without affecting perception very much? Two classes of responses to these questions can be outlined, one of reconstructing

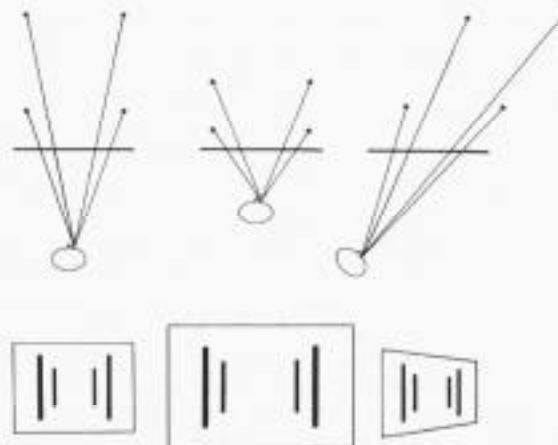


Figure 1. Left panels show slices through untransformed pictorial space as seen from the composition point and the image seen by the viewer; other panels show transformations and views due to changes in observer position away from the composition point. Top panels show the affine transforms in horizontal planes; the bottom-right panel shows a perspective transform of frontal planes. These two transformations together create a hybrid affine/perspective distortion of virtual pictorial space for an observer who is to the side of a photograph.

a Euclidean projection and the other of simply registering spatial information in the hybrid affine/perspective projection.

The first account, espoused by Pirenne (1970), Greene (1983), and Kubovy (1986), proposes that some mental transformation—perhaps something akin to mental rotation (Shepard & Cooper, 1982)—rectifies the physical transformation in virtual space. That is, on the basis of available information about the slant of the picture, typically in binocular disparities or the trapezoidal shape of the screen, perceivers either mentally rotate the picture or mentally relocate themselves in physical space so that they can detransform virtual space. I call this the *Euclidean-rectification view*, since the axes of the affine- and perspective-transformed virtual space of the picture as seen from the side are made Euclidean through some proposed psychological process, eventually matching coordinates under which the picture was composed.

Although this idea is plausible, it also has problems. One is that under many conditions the absolute slant of a regularly textured plane is not very well discerned by viewers (Braunstein & Payne, 1969; Epstein, Bontrager, & Park, 1962; Perrone, 1980), and considerable accuracy in slant estimation is needed to rectify virtual space in a picture or to recover the composition point. If such difficulties occur with well-textured surfaces, they must be compounded with relatively transparent surfaces, such as movie screens. A second problem is that some notion of "double" perception is entailed: the observer must first look at the picture and then derive the proper viewpoint to rectify the scene, as if he or she were then to look at the picture again. This idea seems inelegant; it provokes seeking another view.

The alternative account for viewing pictures from non-composition points is less well formulated at present. It is, simply, that the affine and perspective transformations of virtual space either are not well registered by the human visual system or are well tolerated. I call this the *information-preservation view*, which is based on the idea that object information is preserved in the affine and perspective transformations of virtual space.

The idea of information preservation in oblique views of pictures is attractive for four reasons. First, it makes rectification of affine and perspective transforms unnecessary, thereby making picture perception computationally easier and double perception unnecessary. Second, it promotes a psychophysical approach to the perception of pictured objects that allows for the quantification of detectable and allowable distortions. (A detectable transformation seems likely to prove different from an allowable one.) Third, the information-preservation view encourages exploration of the use of global geometric information as opposed to local information. The local information that has received most attention is the projected angles of intersection of a rectangular object's edges (what I call a right corner). This idea has been most thoroughly investigated by Perkins (1972, 1973) and by Shepard (1981), who have shown that right corners are perceived in line drawings when they can be perceived as such (that is, are allowable as projections). And fourth, this view allows the use and investigation of a powerful idea from mathematics: invariance under transformation.

What follows is a two-part argument for the information-preservation view. The first part is a demonstration of the possibility of information preservation, and the second is a demonstration of the nonnecessity of Euclidean rectification.

Film, Viewpoints, and the Cross Ratio

The phenomenon I wish to explore and explain occurs frequently when groups of people watch film or televi-

sion; it is a dynamic extension of La Gournerie's paradox, and it is a problem outlined by Hochberg and Brooks (1978). Even though the composition point is at the location of the projector for a film shot with a standard lens or is directly in front of the television set, a viewer can sit near the screen and well to the side, and yet a rotating object will not appear to deform. From such a viewpoint, however, a reconstruction of the shape of the virtual object, as in Figure 2, would show that it continually undergoes plastic deformations.

How is the perceiver able to see such a rotating object as rigid? One explanation concerns invariance and projective geometry. One projective invariant that might be useful to perception is the cross ratio of four collinear points (or four coplanar parallel lines), as suggested in Figure 3. I is the information (or invariant) projected to a point; it can be measured either on a planar projection surface, as in Equation 1, or as a relation among visual angles projected from four points, A, B, C, and D, as in Equation 2.

$$I = [AD \cdot BC] / [AC \cdot BD] = [A'D' \cdot B'C'] / [A'C' \cdot B'D'] \quad (1)$$

$$I = [\sin(\alpha + \beta + \gamma) \cdot \sin(\beta)] / [\sin(\alpha + \beta) \cdot \sin(\beta + \gamma)] \\ = \sin(\alpha' + \beta' + \gamma') \cdot \sin(\beta') / [\sin(\alpha' + \beta') \cdot \sin(\beta' + \gamma')] \quad (2)$$

where AD, A'D', and other analogous combinations of point names are segment lengths measured on the object and on the projection screen, respectively; and where α is the angle subtended between points A and B, β that between B and C, and γ that between C and D at Observer 1, and angles α' , β' , and γ' are those projected from A'B', B'C', and C'D', respectively, to Observer 2. What is important about these relations is that the value of I is constant regardless of the location of point of observation (the

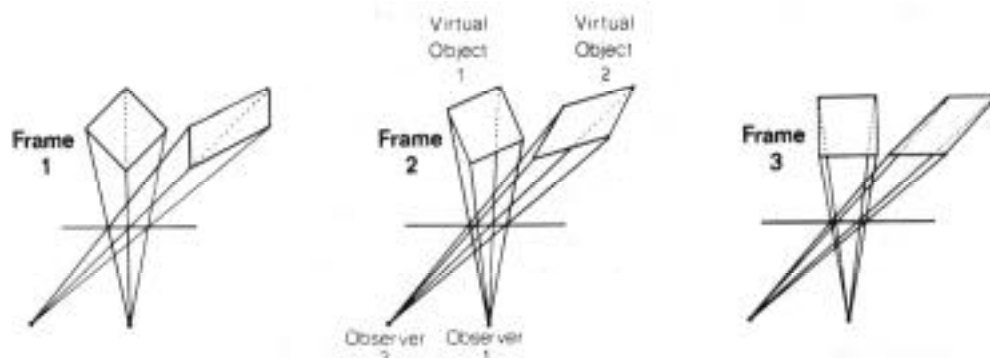


Figure 2. Two virtual objects reconstructed behind the image plane, one for an observer at the composition point and the other for an observer well off to the side. Reconstruction assumes an affine transformation in all $x-z$ slices through the virtual space, keeping distance between frontal-plane slices parallel to the picture plane but with considerable shear. Across the three panels, the virtual objects rotate 45° . Virtual Object 1 is rigid in Euclidean space; Virtual Object 2 undergoes plastic deformations. In this figure, only the affine deformations are shown; perspective deformations would be superimposed on all $x-y$ slices.

location of Observer 1), as long as that point is not collinear with A, B, C, and D. Thus, should the planar object rotate around point X or any other point, or should Observer 1 change his/her position in the theater, or should both occur, the cross ratio remains constant.

Proved by Pappus, the invariance of the cross ratio became a central theorem in projective geometry (Cutting, 1986; Gellert, Küstner, Hellwich, & Kästner, 1977). With increased understanding of such formal properties of projections at the end of the 19th century, Russell (1897) and Poincaré (1905/1952) suggested that perception and projective geometry ought to be intimately related. Gibson (1950) and Johansson, von Hofsten, and Jansson (1980) mentioned the cross ratio's potential utility for perception, but the researchers offered no tests. Ullman (1980), however, suggested that the cross ratio was unlikely to be useful, and Simpson (1983) demonstrated that it was not necessary for the perception of rigid planarity. But nongenerality and nonnecessity do not eliminate the possibility that cross ratios might be useful in some circumstances. In fact, I have demonstrated that the perception of rigidity in a rotating plane appears to be guided by changes in the cross ratio of four parallel lines and not by other, more local, factors (Cutting, 1986).

What is interesting about the cross ratio is its application to La Gournerie's paradox. If a rigid planar object with four parallel lines projects onto a planar surface, the cross ratio will remain invariant regardless of viewer position in the theater, as suggested in Figure 3. That is, if the planar object rotates, the cross ratio will remain constant for both Observer 1 and Observer 2. The proof of this, in brief, is that (1) the cross ratio projects invariantly throughout rotation of an object, as long as that object remains rigid; (2) the measure of the cross ratio on the object, on the projection screen, and at Observer 1 is constant (by Equations 1 and 2); and (3) change in viewer location (from the location of Observer 1 to that of Observer 2) within the theater is identical to rotation of the projection screen (Equation 2). Thus, what proves

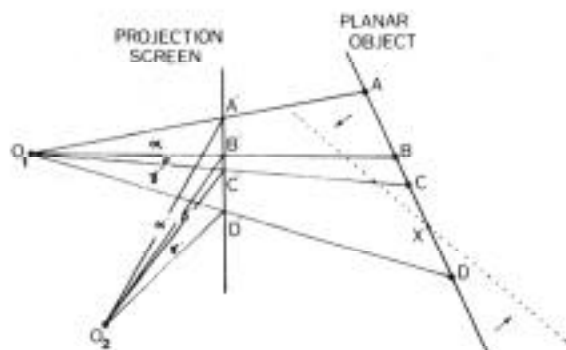


Figure 3. Measurement of the invariant cross ratio of four parallel lines (A, B, C, and D, seen from above) rotating around point X. The value of the cross ratio is the same regardless of whether it is measured on the virtual object or on the projection screen as a ratio of segment lengths (see Equation 1), or at Observer 1 or Observer 2 as a ratio of angles (see Equation 2).

invariant in one planar projection to its composition point (the film screen, which serves as a cross section of the optic array) is invariant to all noncomposition points in the theater. Moreover, observers appear to be able to use cross-ratio information just as readily from noncomposition points as from composition points (Cutting, 1986).

What is more interesting is that even with a slightly curved projection surface, such as a television screen, the cross ratio of a rotating planar object that fills about half of the screen is nearly invariant for almost all viewpoints. One must view the screen well to the side, from an angle of 79° to the surface normal at the center of the screen, for there to be slightly detectable changes in the cross ratio (Cutting, 1986). Thus, the cross ratio is an information source that survives affine and perspective transformations.

Nevertheless, as noted by Ullman (1980) and Simpson (1983), cross ratios are limited in scope. In particular, the measurement of cross ratios is confined to four elements in one dimension. Thus, although the invariance of the cross ratio exemplifies an information-preservation account of La Gournerie's paradox, it can hardly be considered general support. Because of the nongenerality of cross ratios, a second kind of investigation into the cinematic version of La Gournerie's paradox is needed. If the information-preservation account is thus far merely plausible, the Euclidean rectification account must be rendered unnecessary. To do this, one must explore a new methodology—that of using a simulated projection screen.

Simulated and Real Projection Surfaces

The heart of the Euclidean rectification view is that information about the slant of the projection surface is both usable and used. One way to rule out Euclidean rectification, then, is to remove the two main sources of information about its slant: (1) binocular disparities as they grade across the surface and (2) the trapezoidal shape of the screen. Instead of using one eye and a reduction screen, one can do this by hiding a second, or simulated projection, surface behind the actual surface of the screen. The manner in which this is done is indicated in Figure 4.

Such stimuli can be simulated as follows: The instantaneous (x_1, y_1, z_1) coordinates of the object are first polar projected to a station point through a projection surface orthogonal to the line of sight to the middle of the object. This is the simulated projection surface, and the particular station point is not shown in Figure 4. The instantaneous coordinates on the simulated screen are (x_2, y_2) . These coordinates are then polar projected to another station point, the one shown in Figure 4, through a second projection surface, this time the one that will serve as the scope face. Instantaneous coordinates for the object will now be (x_3, y_3) . As the object in (x_1, y_1, z_1) space rotates, the projections of the object are carried through two sets of coordinate changes, and the participant views the outcome.

Imagine a rotating, rectangular solid behind the two projection surfaces, one simulated and one real. Deform-

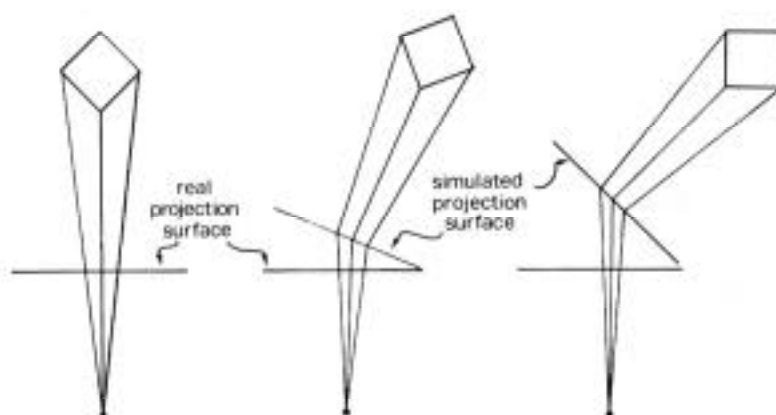


Figure 4. Simulated and real projection surfaces for the investigation of La Gournerie's paradox. The left panel shows normal viewing, the middle panel shows viewing with a 22.5° angle between the two projection surfaces, and the right panel shows viewing with 45° between them. This technique removes projection surface cues from the film-at-a-slant viewing situation.

mations due to noncomposition point viewing cannot be easily rectified from information external to the object. One experimental question is: How great an angle between simulated and real projection surfaces is tolerable before the rotating solid is noticeably and disturbingly distorted?

Meister (1966) drew iso-deformation contours that suggest intolerance partly depends on distance, but 45° is a good estimate for the angle beyond which deformations in moving images are disturbing. Unfortunately, Meister's contours are not based on empirical evidence, and Gibson (1947) presented performance data for viewing motion picture clips that showed no reduction in performance according to viewing angle. In judging the rigidity of rotating objects, I (Cutting, 1985) employed three conditions: normal viewing (with 0° between simulated and real projection surfaces) and viewing with 22.5° and 45° between the two projection surfaces. Results suggest that viewers do not see rotating solids projected normally as being more rigid than those projected with an angle of 22.5° between simulated and real projection surfaces. An angle of 45° difference, however, can yield noticeable distortions.

The difference between the first two conditions and the third is particularly strong for high-perspective polar projections, and less true for low-perspective near-parallel projections. This interaction lends generality to Hagen and Elliott's (1976) notion that parallel projection is more viewpoint independent than is polar projection: Not only are widely varying viewpoints tolerated along the surface normal (perpendicular) to the projection screen, but they are also better tolerated well off the surface normal as well.

It seems to me that the results of the first two conditions, those of 0° and 22.5° difference between projection surfaces, are quite problematic for the Euclidean-rectification view. Nonetheless, there is another way to save the concept of rectification. Since the angle between

the two projection surfaces is constant throughout stimulus presentation, it might be possible to derive that angle during stimulus rotation. Once that angle is derived, rectification could easily follow. Such an idea is consistent with structure-from-motion algorithms (Ullman, 1979, 1984); a few more points, or perhaps a few more views of those points, in space would probably suffice for determining the value of the additional parameter needed for simulated screen slant in structure-from-motion analysis.

To counter this possibility, I have investigated the perception of rotating solids in an additional condition in which the angle between simulated and real projection surfaces varies continuously as the stimulus rotates. Such a condition is somewhat complex, but imagine a condition similar to that shown in the middle and right panels of Figure 4, but with the angle between projection surfaces varying between 10° and 40° while the rectangular solid rotates. As it turns out, the perceived rigidity of such a stimulus is no different from that viewed normally.

Such a finding renders less plausible a structure-from-motion analysis that incorporates viewing angle (between projection surfaces) as a parameter in rigidity assessment. Instead, I suggest that the information for rigidity in cinema seen from a side view exists in the transformed stimulus in some form. What form is unclear. Local analyses on right corners, such as that proposed by Perkins (1972) and Shepard (1981), may prove to be explanatory; or perhaps some global tolerance for geometric distortions of the space may account best for perception. It is too soon to know.

CONCLUSION

What I conclude from this initial foray into the psychophysics of cinematic space is twofold: First, Euclid-

- SIMPSON, W. A. (1983). The cross-ratio and the perception of motion and structure. In *Motion: Representation and perception* (ACM SIGGRAPH/SIGART Interdisciplinary Workshop, pp. 125-129). New York: Association for Computing Machinery.
- ULLMAN, S. (1979). *The interpretation of visual motion*. Cambridge, MA: The MIT Press.
- ULLMAN, S. (1980). Against direct perception. *The Behavioral & Brain Sciences*, **3**, 373-415.
- ULLMAN, S. (1984). Maximizing rigidity: The incremental recovery of 3-D structure from rigid and nonrigid motion. *Perception*, **13**, 255-274.
- WHITE, J. (1957). *The birth and rebirth of pictorial space*. London: Faber & Faber.