

ON THE EFFICACY OF CINEMA, OR WHAT THE VISUAL SYSTEM DID NOT EVOLVE TO DO

James E. Cutting
Department of Psychology
Cornell University, Ithaca, New York

My topic concerns spatial displays, and a constraint that they do not place on the use of spatial instruments. Much of the work done in visual perception by psychologists and by computer scientists has concerned displays that show the motion of rigid objects. Typically, if one assumes that objects are rigid, one can then proceed to understand how the constant shape of the object can be perceived (or computed) as it moves through space. Many have assumed that a rigidity principle reigns in perception; that is, the visual system prefers to see things as rigid. There are now ample reasons to believe, however, that a rigidity principle is not always followed. Hochberg (1986), for example, has outlined some of the conditions under which a rigid object ought to be seen, but is not. Some of these concern elaborations of some of the demonstrations that Adelbert Ames provided us more than 35 years ago.

There is another condition of interest with respect to rigidity and motion perception. That is, not only must we know about those situations in which rigidity ought to be perceived, but is not, we also must know about those conditions in which rigidity ought not to be perceived, but is. Here I address one of these conditions, with respect to cinema. But before discussing cinema, I must first consider photography.

When we look at photographs or representational paintings, our eye position is not usually fixed. A puzzle arises from this fact: Linear perspective is mathematically correct for only one station point, or point of regard, yet almost any position generally in front of a picture will do for object identity and layout within the picture to appear relatively undisturbed. Preservation of phenomenal identity and shape of objects in slanted pictures is fortunate. Without them the utility of pictures would be vanishingly small. Yet the efficacy of slanted pictures is unpredicted by linear perspective theory.

This puzzle was first treated systematically by La Gournerie in 1859 (see Pirenne, 1970). I call it La Gournerie's paradox; Kubovy (1986) has called it the robustness of perspective. The paradox occurs in two forms: The first concerns viewing pictures either nearer or farther than the proper station point; the second and more dramatic concerns viewing pictures from the side. Both are shown in the top panels of figure 1.

To consider either distortion one must reconstruct, as La Gournerie did, the geometry of pictured (or virtual space) behind the picture plane. The premise for doing so is that the image plane is unmoving, but invisible, and that observers look through it into pictured space to make sense out of what is depicted. Invisibility is, in many cases, obviously a very strong, if not false, assumption, but it yields interesting results. Possible changes in viewing position are along the z axis, orthogonal to the picture plane, and along the x or y axes, parallel to it. Both generate affine transformations in depth in all xz planes of virtual space. Observer movement along x or y axes

also generates perspective transformations of the image, but these will not be considered here (Cutting 1986a, 1986b).

In the upper left panel, four points are projected onto the image plane as might be seen in a large photograph taken with a short lens. When the observer moves closer to the image, as in the upper middle panel, the projected points must stay in the same physical locations in the photo. Thus, the geometry of what lies behind must change. Notice that the distance between front and back pairs of points of this four-point object is compressed, a collapse of depth like that when looking through a telephoto lens. All changes in z axis location of the observer create compression or expansion of the object in virtual space. When an observer moves to the side, as seen in the upper right panel, points in virtual space must shift over, and do so by different amounts. Such shifts are due to affine shear. All viewpoints of a picture yield additive combinations of these two affine effects—compression (or expansion) and shear.

Such effects are compounded when viewing a motion sequence, as shown in the lower panels of figure 1. In particular, an otherwise rigid object should appear to hinge and become nonrigid over the course of several frames for a viewer seated to the side. Theoretically, the problem this poses for the cinematic viewer is enormous—every viewer in a cineauditorium has an eye position different than the projector and camera position, and thus, by the rules of perspective, no moving object should ever appear rigid. This is, I claim, the fundamental problem of the perception of film and television.

Most explanations for the perception of pictures at a slant are in sympathy with Helmholtz. Pirenne (1970, p. 99), for example, suggested that "an unconscious intuitive process of psychological compensation takes place, which restores the correct view when the picture is looked at from the wrong position." Pirenne's unconscious inference appears to unpack the deformations through some process akin to mental rotation (Shepard & Cooper, 1982). According to this view, the mind detransforms the distortions in pictured space so that things may be seen properly, and although Pirenne didn't discuss film, it might hold equally for film seen from the front row, side aisle. The force of my presentation is to show that this view is not necessary in the perception of slanted cinema. But first consider how this account might proceed.

Pirenne and others have suggested at least three sources of image surface information that might be used to "correct" slanted images—(1) the edges of the screen, which yield a trapezoidal frame of reference; (2) binocular disparities, which grade across the slanted surface; and (3) projection surface information such as texture and specularities. Since I am interested in none of these, I removed them from my displays through a double projection scheme, as shown in figure 2. If one considers the situation of viewing slanted cinema, one has the real, slanted surface and one can measure a cross section of that optic array from it. This would be an imaginary projection surface. Once considered this way, one can reverse the two, placing the real surface in front of the imaginary, and this is what I did.

In this manner, although the display frame was always rectangular for the observer, the shapes of rotating stimuli were like those seen from the side, with the right-edge elements in each frame longer than the left-edge ones and with the z axis compressed. This simulation yields a perspective transformation of the image screen, and a nonperspective transformation of the stimulus behind it in virtual space. I presented viewers with computer-generated, rotating, rectangular solids. Two factors are relevant to this discussion. (For a more complete analysis see Cutting, 1987.)

First, half the solids presented were rigid, half nonrigid. Nonrigid solids underwent two kinds of transformation during rotation—one affine, compressing and expanding the solid like an accordion along one of its axes orthogonal to the axis of rotation during rotation, and one non-affine, with a corner of the solid moving through the same excursion. Deformations were sinusoidal and were accomplished within one rotation of the stimulus. It was relatively easy to see the large excursions as making the solid nonrigid; it was more difficult in smaller excursions. This nonaffine deformation was much easier to see than the affine deformation, but there were no interactions involving types of nonrigidity, so here I will collapse across them (see Cutting, 1987, for their separate discussion).

Second, stimuli were presented with cinematic viewpoint varied; in Experiment 1, half were projected as if viewed from the correct station point, half as if seen from the side, with the angle between imaginary and real projections surfaces set at 23° . The latter condition allows investigation of La Gournerie's paradox, and compounds the nonrigid deformations of the stimulus in pictorial space with an additional perspective transformation of the image.

Viewers looked at many different tokens of all stimuli, and used a bipolar graded scale of rigidity and confidence, from 1 to 9—with 1 indicating high confidence in nonrigidity, 9 high confidence in rigidity, and 5 indicating no confidence either way.

Figure 3 shows the results of the first experiment for rigid and nonrigid stimuli, at both 90° and simulated 67° viewing angles. Two effects are clear. First, rigid stimuli were seen as equally rigid regardless of simulated viewpoint in front of the screen, and second, nonrigid stimuli were seen as equally nonrigid regardless of simulated viewpoint.

The lack of difference in the slanted and unslanted simulated viewing conditions is striking, but it could be due to the fact that the screen slant was relatively slight. Experiment 2, then, introduced a third viewing condition, a steeper angle— 45° . A fourth condition was also introduced. Its impetus came from structure-from-motion algorithms in machine vision research. Several people suggested to me that screen slant could be another parameter in rigidity-finding algorithms and that only a few more frames or points might be needed to specify slant. To test for this idea, I introduced a variable screen-slant condition, where the simulated slant of the screen oscillated between 80° and 55° , with a mean of 67° . It seemed highly unlikely that an algorithm could easily solve for both rigidity and a dynamically changing projection surface.

This time stimuli were generated in near-parallel and polar perspective. Again, stimuli could be rigid or nonrigid. Selected results for the nonrigid stimuli are shown in figure 4, and show two striking effects. First, the variable 67° screen slant condition was not different from the nonvarying condition, and the lack of difference would seem to be embarrassing for any structure-through-motion approach to the perception of these stimuli that includes screen slant as a variable to be solved for. Second, if simulated screen slant is great enough, all stimuli begin to look nonrigid.

A more interesting result is an interaction concerning near-parallel and polar projected stimuli, as shown in figure 5, with the two 67° conditions collapsed, and all rigid and nonrigid trials collapsed. The near-parallel projected stimuli show no difference in perceived rigidity from any angle that they are viewed; the more polar projected stimuli, on the other hand, show a sharp decrease in perceived nonrigidity as the angle or regard increases.

This latter effect adds substance to other results in the literature. For example, Hagen and Elliott (1976) found what they called a "zoom effect"—the general preference for static stimuli seen is more parallel than polar projection. Here, in cinematic displays, stimuli that are near-parallel-projected are seen as more rigid from more places in a cineauditorium.

In conclusion, let us be reminded that photographs and cinema are visual displays that are also powerful forms of art. Their efficacy, in part, stems from the fact that, although viewpoint is constrained when composing them, it is not nearly so constrained when viewing them. The reason that viewpoint is relatively unconstrained, I claim, is not that viewers "take into account" the slant of the screen, but that the visual system does not seem to compute the relatively small distortions in the projections, at least for certain stimuli that are projected in a near-parallel fashion.

It is obvious that our visual system did not evolve to watch movies or look at photographs. Thus, what photographs and movies present to us must be allowed in the rule-governed system under which vision evolved. Slanted photographs and cinema present an interesting case where the rules are systematically broken, but broken in a way that is largely inconsequential to vision. Machine-vision algorithms, to be applicable to human vision, should show the same types of tolerances.

But with regard to the use of camera lens in movies, it becomes quite clear why long lenses—those that are telephoto and nearly telephoto—are so popular and useful. First, and known for nearly a century, standard lenses tend to make people look like they have bulbous noses. Second, and corroborated by my results, long lenses provide a more nearly parallel projection of objects, and the distortions seen in these objects when a viewer looks at a slanted screen are significantly diminished. This enhances their efficacy considerably, despite the fact that it introduces the nonnatural situation of collapsing the apparent depth of a scene.

REFERENCES

- Cutting, James E. Perception With an Eye for Motion. MIT Press/Bradford Books, 1986(a).
- Cutting, James E. The Shape and Psychophysics of Cinematic Space. *Behav. Res. Meth. Inst. & Computers*, vol. 18, 1986(b), pp. 551-558.
- Cutting, James E. Rigidity in Cinema Seen From the Front Row, Side Aisle. *J. Exp. Psychol.: Human Perc. & Perf.*, vol. 13, 1987, pp. 323-334.
- Hagen, Margaret A.; and Elliott, Harry B. An Investigation of the Relationship Between Viewing Conditions and Preference for True and Modified Perspective With Adults. *J. Exp. Psychol.: Human Perc. & Perf.*, vol. 4, 1976, pp. 479-490.
- Hochberg, Julian. Representation of Motion and Space in Video and Cinematic Displays. In K.R. Boff, L. Kaufman, & J.P. Thomas (eds.), *Handbook of Perception and Human Performance*, Vol. 1, pp. 22:1-63. John Wiley, 1986.
- Kubovy, Michael. *The Psychology of Linear Perspective and Renaissance Art*. Cambridge University Press, 1986.
- Pirenne, M. H. *Optics, Painting, & Photography*. Cambridge University Press, 1970.
- Shepard, Roger N. & Cooper, Lynn A. *Mental Images and Their Transformation*. MIT Press/Bradford Books, 1982.

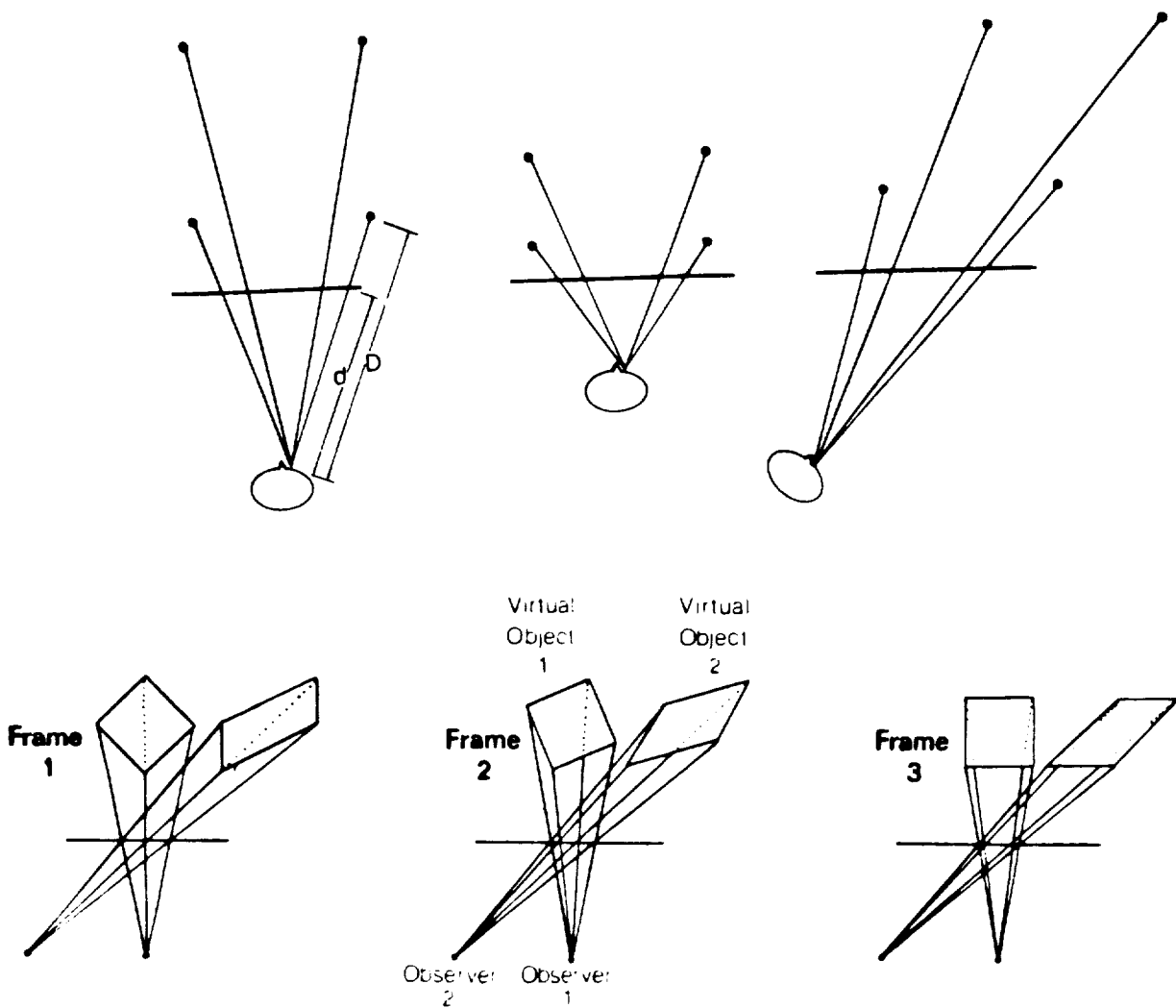


Figure 1. Reconstructive geometry and images. The upper panels show the reconstruction of four pillars in depth. Consider the left-most panel a representation of the real depth relations projected onto the image plane. If that plane is now a photograph, the pillars are fixed in position on the image plane. Thus, when an observer moves toward the plane, depth must be compressed, as in the upper middle panel. When the viewer moves to the side, all pillars slide over by differing amounts. The bottom panels show reconstructions of a moving square across three frames, from two viewpoints. Notice that the reconstruction for Observer 1 is rigid, but that that for Observer 2 is not (from Cutting, 1986a).

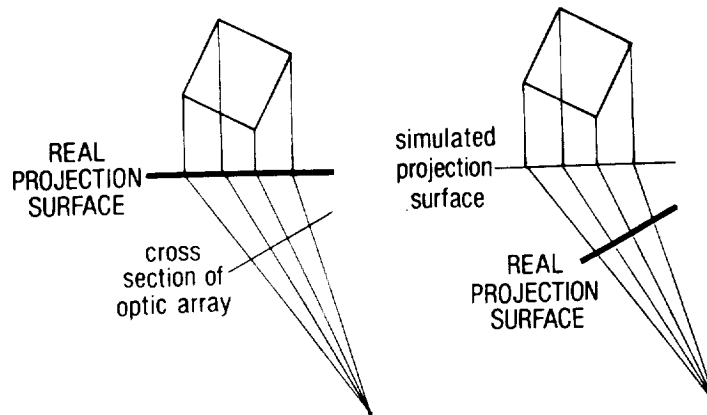


Figure 2. Arrangements of real and simulated projection surfaces that can remove image information from objects projected onto slanted screens (from Cutting, 1987).

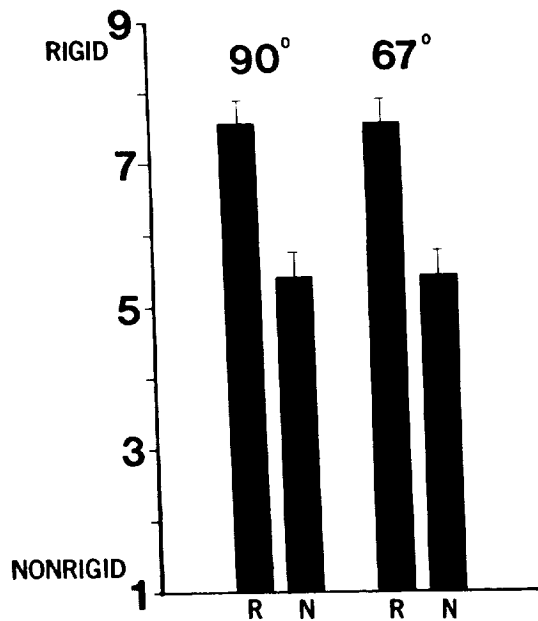


Figure 3. Selected results from Experiment 1. 90° and 67° are the two viewing conditions of interest, where 67° is the simulated screen slant as indicated in figure 2. R = rigid stimuli, N = nonrigid stimuli.

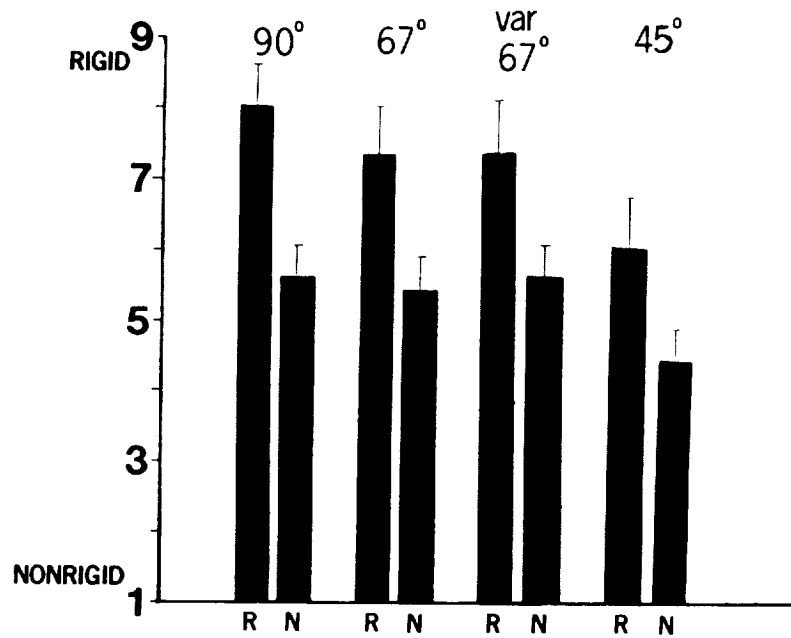


Figure 4. Selected results from Experiment 2. The added conditions are simulated screen slants of 45°, and one of variable slant (between 80° and 55°), averaging 67°. R = rigid stimuli, N = nonrigid stimuli.

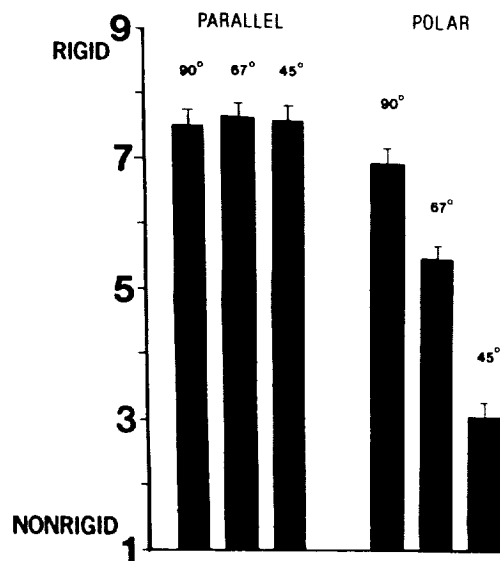


Figure 5. Another description of the results of Experiment 2, parsed according to projection.