32

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

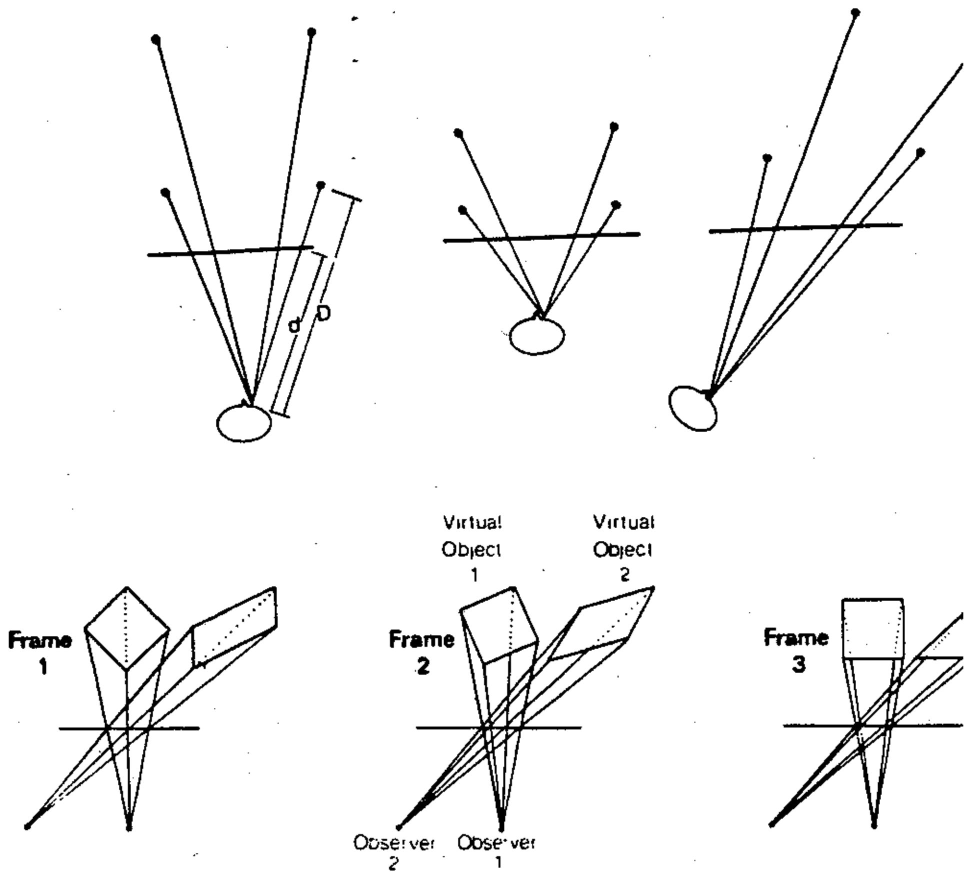

Figure 1. Reconstructive geometry and images. The upper panels show the struction of four pillars in depth. Consider the left-most panel a representation real depth relations projected onto the image plane. If that plane is now a photo the pillars are fixed in position on the image plane. Thus, when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

concerns viewing pictures either nearer or farther than the proper s point; the second and more dramatic concerns viewing pictures fro side. Both are shown in the top panels of Figure 1.

To consider either distortion one must reconstruct, as La Gourneri the geometry of pictured (or virtual space) behind the picture plane premise for doing so is that the image plane is unmoving, but invisibl that observers look through it into pictured space to make sense out of is depicted. Invisibility is, in many cases, obviously a very strong, false, assumption, but it yields interesting results. Possible changes in ing position are along the z axis, orthogonal to the picture plane, and 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 and 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

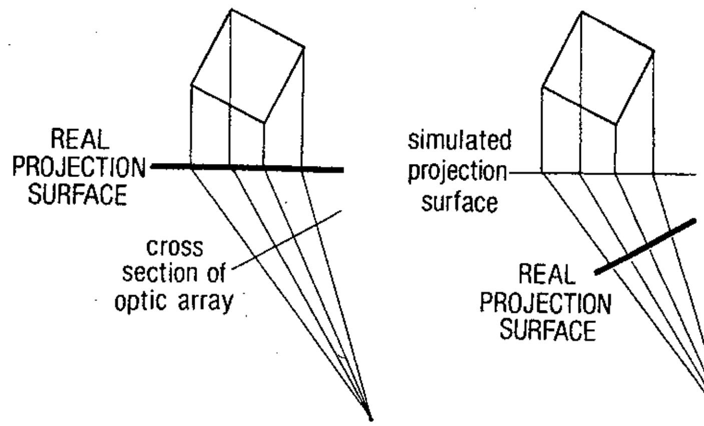

Figure 2. Arrangements of real and simulated projection surfaces that can re image information from objects projected onto slanted screens (from Cutting, 1.

section of that optic array from it. This would be an imaginary proje surface. Once considered this way, one can reverse the two, placing the surface in front of the imaginary, and this is what I did.

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

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

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

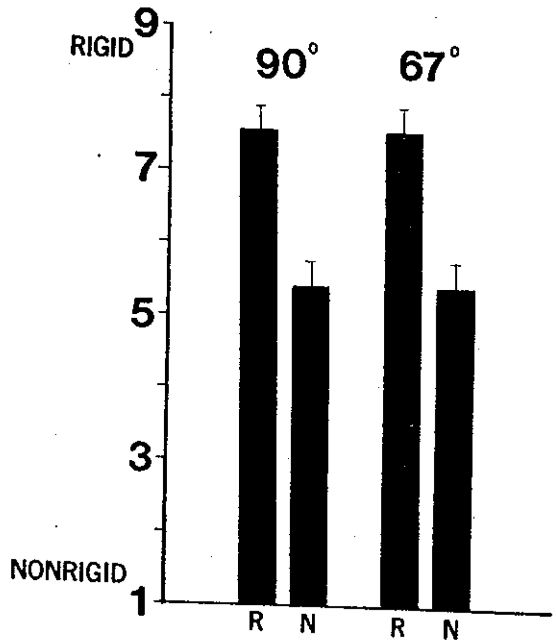

Figure 3. Selected results from Experiment 1. 90° and 67° are the two viewing conditions of interest, where 67° is the simulated screen slant as indicated in Figure 2. R = rigid stimuli, N = nonrigid stimuli.

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

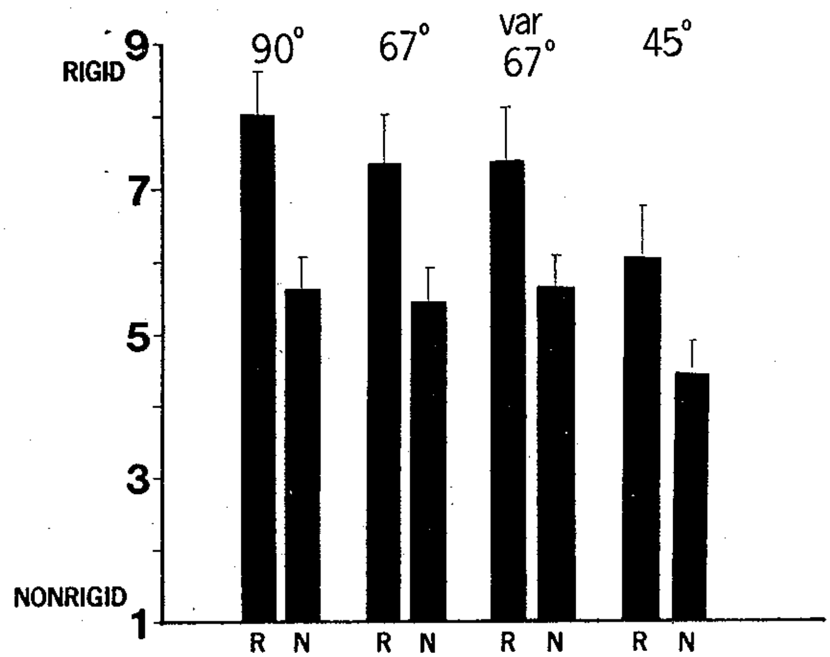

Figure 4. Selected results from Experiment 2. The added conditions are sin screen slants of 45°, and one of variable slant (between 80° and 55°), averagin R = rigid stimuli, N = nonrigid stimuli.

stimuli are shown in Figure 4, and show two striking effects. Firs variable 67° screen slant condition was not different from the nonv condition, and the lack of difference would seem to be embarrassing f structure-through-motion approach to the perception of these stimu includes screen slant as a variable to be solved for. Second, if sim screen slant is great enough, all stimuli begin to look nonrigid.

A more interesting result is an interaction concerning near-parall polar-projected stimuli, as shown in Figure 5, with the two 67° con collapsed, and all rigid and nonrigid trials collapsed. The near-paralle jected stimuli show no difference in perceived rigidity from any ang they are viewed; the more polar-projected stimuli, on the other hand, s sharp decrease in perceived nonrigidity as the angle of regard increase

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

In conclusion, let us be reminded that photographs and cinema are displays that are also powerful forms of art. Their efficacy, in part, from the fact that, although viewpoint is constrained when composing it is not nearly so constrained when viewing them. The reason that vie is relatively unconstrained, I claim, is not that viewers "take into ac the slant of the screen, but that the visual system does not seem to co

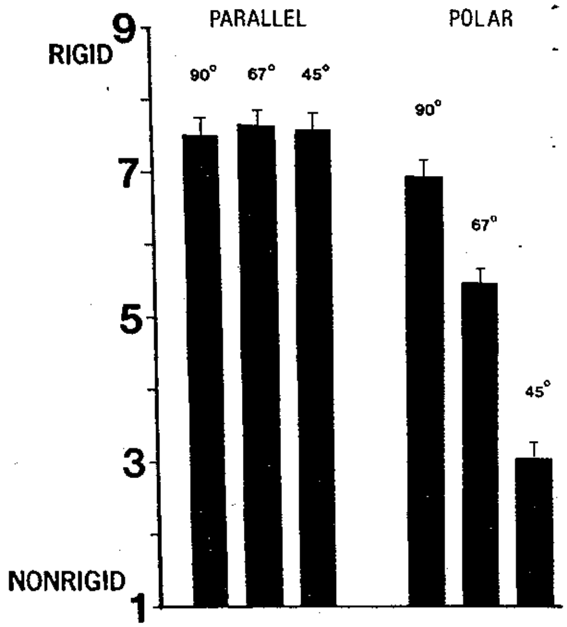

Figure 5. Another description of the results of Experiment 2, parsed according to projection.

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 the 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, J.E. (1986a). *Perception With an Eye for Motion*. Boston: MIT Press/Bradford Books.

Cutting, J.E. (1986b). The shape and psychophysics of cinematic space. *Beha Research Methods for Institutions and Computers*, **18**, 551–558.
Cutting, J.E. (1987). Rigidity in cinema seen from the front row, side aisle. *Joun Experimental Psychology: Human Perception and Performance*, **13**, 323–334.
Hagen, M.A. and Elliott, H.B. (1976). An investigation of the relationship bet viewing conditions and preference for true and modified perspective with ac *Journal of Experimental Psychology: Human Perception and Perfermance*, **4**, 479–
Hochberg, J. (1986). Representation of motion and space in video and cine displays. In Boff, K.R., Kaufman, L. and Thomas, J.P. (Eds), *Handbo Perception and Human Performance*. **1**, 22:1–63. New York: John Wiley.
Kubovy, M. (1986). *The Psychology of Linear Perspective and Renaissance Art*. C bridge, England: Cambridge University Press
Pirenne, M.H. (1970). *Optics, Painting, and Photography*. Cambridge, England: C bridge University Press.
Shepard, R.N. and Cooper, L.A. (1982). *Mental Images and Their Transform* Boston: MIT Press/Bradford Books.

Appendix

Mathematical proof of a formula for predicting perceived slant ang in a picture seen from the side

When one views a picture from the side, the virtual space behind the picture undergoes an affine transformation measured in any given xz plane (horizontal of the pictorial space), and a perspective transformation of any given yz (Cutting, 1986a). Since these two effects are additive, and since we are not conce with the latter, only the affine effects will be considered here. These concern predicted direction of perceived slant of an object in pictorial space as meas horizontally against the picture plane.

I. Consider how a picture is composed, as in Figure 6.

Let: O = center of projection of the picture (station point or compos point).
α = right angle, by definition (except in anamorphic art).
d = distance from the picture plane to the station point, O. Thus, the principal ray.
x = the point of intersection of the principal ray with the picture plane.
p = distance from the picture plane to the object that is slanted in picture (or to a line of slant extended from that object). Thus, the continuation of the principal ray to the slanted object (or its lin continuation; and $d + p$ is the reconstructed distance from the st point to the slanted object (or its continuation).
σ = true slant angle of the pictured object, as measured against the ture plane.

II. Consider a new viewpoint, as in Figure 6.

Let: O′ = new viewpoint to the picture.
α' = nonright angle between the picture plane and the new viewp at x.
d' = angled distance of new viewpoint to x.

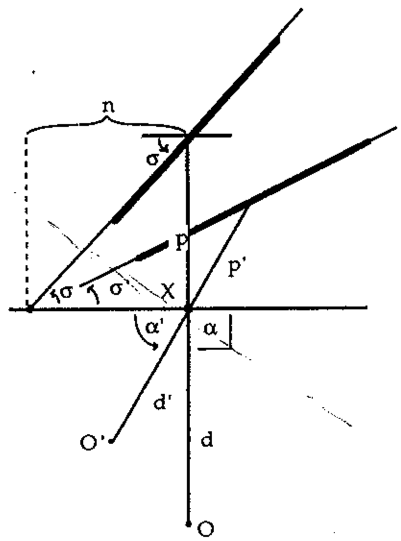

Figure 6. Geometry of a picture viewed along an oblique.

III. To derive:

σ' = predicted angle of perceived slant.

Step of Proof

(1) Since the pictorial space is affine, changes along the length of a line are proportional. Thus:

$$d/d' = p/p'$$

For further calculations we need p', thus

$$p' = d'p/d$$

(2) Since the pictorial space is affine, and since nothing in the picture plane moves, the line of the slanted object extended to the picture plane (or an extension of it) remains in place. Thus, we need to know, n, the distance from x to the point where the extension of the slanted object meets the picture plane.

$$n = p/\tan\sigma$$

(3) Since we now know n, p', and α', we can solve for σ'.

$$\sigma' = \operatorname{atan}\left[(p' \sin\alpha')/(n + p' \cos\alpha')\right]$$

or, going back to the terms defined originally:

$$\sigma' = \operatorname{atan}\left[((d'p/d)\sin\alpha')/((p/\tan\sigma) + (d'p/d)\cos\alpha')\right]$$

rearranging and simplifying the equation yields:

$$\sigma' = \text{atan}\,[(\sin\alpha' \times (d'/d))/((1/\tan\sigma) + (\cos\alpha'(d'/d))]$$

Thus, the predicted angle of perceived slant from affine geometry is a functi four variables (d', d, α', and σ) known about the original spatial relations i composition of the picture.

Pictorial communication in virtual and real environments

Second edition

Edited by
Stephen R. Ellis
NASA Ames Research Center and University of California, Berkeley, USA

Section Editor
Mary K. Kaiser
NASA Ames Research Center, USA

Assistant Section Editor
Arthur J. Grunwald
Technion, Israel

1991
1993 2nd ed

Taylor & Francis
London • Washington, DC