Chapter 11

Reconceiving Perceptual Space

James E. Cutting

11.1 INTRODUCTION

How do we perceive the space in pictures? In answering this question theorists typically consider standard photographs and other representational images, such as architectural drawings, engravings, and paintings in linear perspective. Adding motion augments this domain, including cinema and computer-graphic sequences. But cartoons, caricatures, and much of twentieth-century art and photography are seldom overtly considered. This, I think, is a mistake; without a broader perspective on pictures one is lulled into thinking too metrically, about pictures and about the real world as well.

Certain philosophers and psychologists have spilled a lot of ink discussing pictures and the space depicted within them; other philosophers and typically other psychologists have done the same about perceived space in the world around us. What I propose is that our perception of these spaces is done in pretty much the same way, that neither are guaranteed Euclidean but that they are built upon available information. In essence my chapter is on the promise of a constructive and cooperative measurement-theoretic approach to all perceived space. The more information, the more constraints, the more any perceived space will incrementally approach a Euclidean ideal.

A Précis of Measurement Theory

After his broadscale survey of how scientists measure their subject matters, S. S. Stevens (1951) reported that we measure the world in four ways—using nominal, ordinal, interval, and ratio (or metric) scales (see also Luce and Krumhansl 1988). Nominal scales name, or simply categorize, differences. Such categorization is always the beginning of science,

and taxonomies (which are nested nominal scales) remains critically important in biology today. Ordinal scales categorize *and* order, but they say nothing about distances between what is ordered. The division of U.S. college students into freshmen, sophomores, juniors, and seniors is an example, for this only gives a rough idea of their "distance" from graduation requirements. Interval scales categorize, order, *and* provide true distances, but there is no true zero on such a scale. Thus, one can say that the distance between 5°C and 10°C is the same as that between 10°C and 15°C, but one cannot say that 10°C is twice as warm as 5°C. Finally, ratio scales categorize, order, provide true distances, and have a true zero. Thus, one can say that 1 m is half as long as 2 m. In doing psychophysics, for example, one often manipulates a physical variable along a ratio scale, and records some psychological variable that is almost surely only ordinal.

Stevens's (1951) classification system of the four scales is itself ordinal. That is, one cannot really know the "distance" between nominal and ordinal scales, for example, as compared to interval and ratio. However true this is logically, one can nonetheless state psychologically that through various considerations one may find that ordinal information can be used to approximate a metric scale (Shepard 1980). And this is the crux of what I have to say. I will claim that all perceived spaces are really ordinal. However, sometimes these spaces can be said to converge on a metric space.

Framing the Problem(s)

What I have to say is framed by the title of the conference "Reconceiving Pictorial Space," on which this volume is based, and on three questions that seem to underlie its motivation. Each query comes in sequence as a slight refinement of the previous.

11.2 IS PERCEIVING A PICTURE LIKE PERCEIVING THE WORLD?

Yes—and for some pictures to a large extent. This is one reason nonprofessional, candid photographs work so well; the cinema can act as such a culturally important surrogate for the everyday world; and precious little experience, if any, is needed to appreciate the content of pictures (Hochberg and Brooks 1962) or film (Messaris 1994). Nevertheless, almost all theorists who bother to address this question—and few actually do—answer largely in the negative, choosing to focus instead on dif-

ferences between pictures and the world, regardless of how they define pictures or they might define "the world." Consider views from the humanities and then from psychology.

Among artists and art historians, statements about the similarity of pictures and the world are not prevalent. To be sure, few would deny the impressiveness of certain trompe l'oeil (e.g., Cadiou and Gilou 1989; Kubovy 1986) as having the power to be mistaken for a certain type of reality, but equally few would claim this is other than a relatively small genre, and it does not legitimately extend even to photographs. Moreover, the effectiveness of trompe l'oeil is predicated, in part, on a tightly constrained range of depicted distances from the observer, or simply depths.

Rather than concentrating on similarity or verisimilitude of images and their naturalistic counterparts, many in the humanities have focused on viewer response. Responses to images are often indistinguishable from responses to real objects, and this was an important problem in the development of the Protestant Church in the sixteenth century and in the Catholic Counter Reformation. Worship *with* images could not always be separated from the worship *of* images, a violation of the Old Testament's First Commandment (Michalski 1993). This flirtation probably contributed to the prohibition of images in Judaic and Islamic worship as well. Over a broader cultural sweep, Freedberg offered powerful analyses of how pictorial objects evoke the responses in people as real objects, but he never claimed that pictures are mistaken for reality. To be sure, "people are sexually aroused by pictures ...; they break pictures ...; they kiss them, cry before them, and go on journeys to them" (1989, p. 1), but they don't actually mistake them for the real objects they represent. Instead, images (and sculptures) stand in reference to the objects they represent, attaining an equal status with them, to be loved, scorned, appreciated, or decried in full value.

More generally, there are simply the difficulties of generating mimicry. Ernst Gombrich, for example, suggested that "the demand that the painter should stick to appearances to the extent of trying to forget what he merely 'knew' proved to be in flagrant conflict with actual practice.... The phenomenal world eluded the painter's grasp and he turned to other pursuits" (1974, p. 163). Indeed, with the invention of photography there was even a strong sentiment that, as Rodin suggested, "it is the artist who is truthful and it is the photograph which lies.... [H]is [the artist's] work is certainly much less conventional than the scientific image" (Scharf 1968, p. 226).

Within psychology, James J. Gibson used the picture versus real world difference as a fulcrum to make a distinction between direct and indirect (or mediated) perception. He is often quoted: "Direct perception is what one gets from seeing Niagara Falls, say, as distinguished from seeing a picture of it. The latter kind of perception is *mediated*" (1979, p. 147). Although the nature and the wider ramifications of this distinction have been much debated (e.g., see Cutting 1986a, 1998, for overviews), it is clear throughout all discussions that picture perception and real-world perception are conceived as different.[1] Although few other psychological theorists use pictures to discuss a direct/indirect distinction (but see Sedgwick 2001), Alan Costall (1990), Margaret Hagen (1986), Julian Hochberg (1962, 1978), John Kennedy (1974), Michael Kubovy (1986), Sheena Rogers (1995), and John Willats (1997) all emphasize differences between the world and pictures of it. Indeed, for William Ittelson (1996) there are few, if any, similarities between the perception of pictures and the everyday perception of reality.

In most psychological discussions of pictures versus reality, the essential element centers on the truism that pictures are two-dimensional surfaces and that the world around us is arrayed in three dimensions. At their photographic best, pictures are frozen cross sections of optical arrays whose elements do not change their adjacent positions when the viewer moves. In particular, what is left of a given object seen in a picture from a given position is always left of that object; what is right, always right; and so forth.[2] This is not always true in the natural environment. As one moves forward, to the side, or up or down, objects in the world cross over one another, changing their relative positions. In the world the projective arrangement of objects is not frozen.

To be sure, there are other differences between pictures and the world than the 2-D versus 3-D difference and the lack of motion in pictures. At a comfortable viewing distance, the sizes of objects as projected to the eye are generally smaller in pictures than in real life; pictures typically have a compressed range of luminance values (and often of color) compared to the real world; and there are lens effects that compress or dilate space. I will focus on lens effects later, but let's first consider a second question.

11.3 IS PERCEIVING PHOTOGRAPHIC SPACE LIKE PERCEIVING ENVIRONMENTAL SPACE?

Again, yes—more or less—and, I claim, certainly more so than less so. However, to answer this question, one must begin with an understanding

of the perceived space in the world around us. Variously, I will call this environmental space, physical space, and even reality. My intent in this multiplicity (other than to avoid semantic satiation) is to emphasize the assumption that I consider all of these identical.

Two interrelated facts about the perception of physical space must be considered. First, perceived space is anisotropic (Luneburg 1947; Indow 1991). In particular, perceived distances are somewhat foreshortened as compared to physical space, particularly as physical distances increase. This fact has been noted in various ways by many researchers (e.g., Gogel 1993; Loomis and Philbeck 1999; and Wagner 1985; see Sedgwick 1986, for a review), although some methods of judging distance yield quite different results from others (Da Silva 1985; Loomis et al. 1996).

Second, this compression is likely due to the decrease in information available. Such decreases in available information have been shown experimentally by Teodor Künnapas (1968) in the near range (<4 m) of physical environments. They have also been demonstrated for near space in computer-generated ones (Bruno and Cutting 1988). I contend that the dearth of information about depth is responsible for compressions throughout the visual range, particularly at extreme distances.

The Shape of Perceived Environmental Space

We don't notice that our perception of space is not veridical. Where it is compressed most, it is sufficiently far away that it has no consequence for our everyday behavior; any "errors" in distance judgment would not usually matter.[3] The amount of compression throughout perceived space is often said to be well captured psychophysically by an exponent. That is, perceived distance (ψ) is a function of physical distance (ϕ) with an exponent less than 1.0, $\psi = f(\phi^{<1})$. To my knowledge, all previous investigations have assumed that the exponent is constant throughout perceived space, and I think this is incorrect. Instead, it seems that exponents near the observer are near 1.0, then generally decrease with the distance range investigated. Shown in the top panel of figure 11.1 are plots for eleven studies done in naturalistic environments and including distances beyond 20 m (see Da Silva 1985, for a more complete review). The horizontal lines indicate the range of egocentric distances investigated (in meters); their height in the figure corresponds to the value of the exponent best fitting the data. Notice that exponents are generally from near 1.0 to .65 for studies investigating ranges out to about 300 meters but nearer to .40 for the only naturalistic study I know done beyond 1 km (Flückiger 1991).[4]

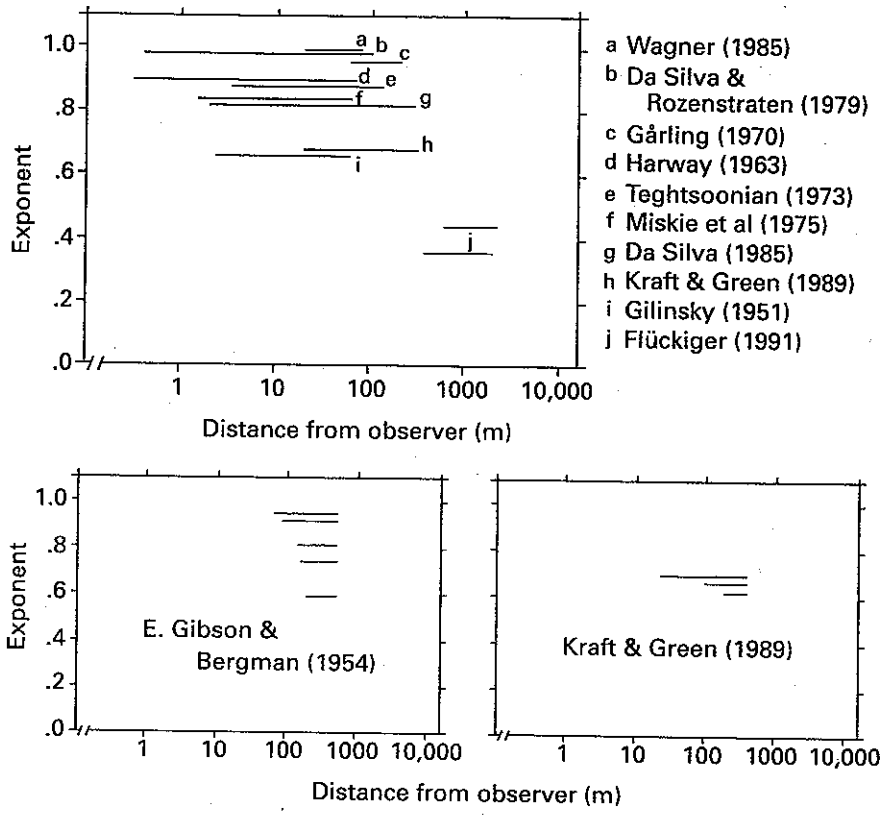

Figure 11.1

The top panel shows the exponents of perceived distance functions plotted by the range of egocentric distances investigated in eleven studies. Notice the general decline with increasing distance. This suggests an accelerated foreshortening of space, from near to far. All studies chosen were done in naturalistic environments. Those included, listed by order of decreasing exponents are Da Silva and Rozenstraten (1979, reported in Da Silva, 1985), method of fractionation; Gårling (1970), magnitude estimation; Harway (1963), method of partitioning; Teghtsoonian (1973), magnitude estimation; Miskie et al. (1975), magnitude estimation; Da Silva (1985), magnitude estimation; Kraft and Green (1989), verbal reports from photographs; Gilinsky (1951), method of fractionation; Flückiger (1991, two studies), method of reproduction. The bottom panels select the data of naive viewers from Gibson and Bergman (1954) and the data of Kraft and Green for special analysis. That is, data are shown first with exponents and full egocentric depth range but then with ranges incrementally truncated in the near range. This truncation systematically lowers the exponents. This means that the shape of perceived space beyond about 50 to 100 m compresses faster than is captured by exponential analysis. Exponents for the studies of Flückiger (1991), Gibson and Bergman (1954), and Kraft and Green (1989) were calculated from the published data for this analysis; others are available in Da Silva (1985).

With the addition of the data of Michelangelo Flückiger, the pattern in the top panel of figure 11.1 looks as if there is a trend: as the range moves outward from the observer, the exponent decreases. But is this just an anomaly of an outlier study? A cleaner analysis of this effect can be seen in the lower panels. Here, the data sets of Eleanor Gibson and Richard Bergman (1954, for their naive subjects) and of Robert Kraft and Jeffrey Green (1989) are considered. First, the complete published data sets were fit with a series of exponents and a best-fitting exponent found. Next, the data points nearest to the observer were omitted from the data sets and the procedure for finding best exponents repeated. Truncation and refitting procedures were repeated again until the data became unstable or too sparse for further analysis.[5] Notice that in both cases the truncation of near data caused the best-fitting exponent to diminish. This can only happen if perceptual space compresses faster than is captured by a single exponent. Consider next the information available in the real world as the beginnings of an explanation for such compression.

Information in Environmental Space

Peter Vishton and I (Cutting and Vishton 1995), among many others (Landy et al. 1995; Luo and Kay 1992; Massaro and Cohen 1993; Terzopoulos 1986), suggested that many sources of information about the layout of the world around us are combined in complex ways. Cutting and Vishton (1995) provided a list of nine different sources of information used—occlusion, relative size, relative density, height in the visual field, aerial perspective, binocular disparities, motion parallax, convergence, and accommodation—although similar lists are stated or implied in works from Euclid (Burton 1945) to Hermann von Helmholtz (1911/1925) and continuously to the present day.[6]

These sources of information often measure the world in different ways. For example, occlusion offers only ordinal information (what is in front of what but not by how much); relative size has the potential of offering ratio-scaled information (if two objects are identical in physical size and one subtends half the visual extent of the other, then it is twice the distance from the observer); and stereopsis and motion perspective may yield absolute distance information (e.g., Landy, Maloney, and Young 1991). Because it is not possible to compare the efficacy of measurement on different scales, we employed *scale convergence* (Birnbaum 1983). That is, we reduced the measures of information for each source to their weakest shared scale—ordinal depth. We next located threshold

measures for ordinal depth in the literature and then used them to make suprathreshold comparisons (Cutting and Vishton 1995). The results are shown in the top panel of figure 11.2.

In the panel, threshold functions for pairwise ordinal distance judgments are shown for nine sources of information known to contribute to the perception of layout and depth. They are based on various assumptions and data and applied to a pedestrian. Inspired by and elaborated from Shojiro Nagata (1991), the data are plotted as a function of the mean distance of two objects from the observer (log transformed) and of their *depth contrast*. Depth contrast is defined as the metric difference in the distance of two objects from the observer divided by the mean of their two depths. This measure is similar to that of Michelson contrast in the domain of spatial frequency analysis.

Notice that, plotted in this manner, some sources are equally efficacious everywhere. In particular, I claim that potency of occlusion (the most powerful source of information), relative size, and relative density do not attenuate with the log of distance. Depth contrasts of 0.1%, 3%, and 10%, respectively, between two objects at any mean distance are sufficient for observers to judge which of the two is in front. However, other sources are differentially efficacious. Most—accommodation, convergence, binocular disparities, motion perspective, and height in the visual field—decline with the log of distance. One, aerial perspective, actually increases with the log of distance (it is constant with linear distance), but this source of information, I claim, is used generally to support luminance contrast information for occlusion. Notice further that, integrated across all sources, the "amount" and even the "quality" of information generally declines with the log of distance. This seems the likely cause of compressed perceived distances.

The general shapes of the functions in figure 11.2, plus a few practical considerations, encouraged us to parse egocentric space into three regions (see also Grüsser and Landis 1991). *Personal space,* that which extends a little beyond arm's reach, is supported by many sources of information and is perceived almost metrically (Loomis et al. 1996), that is, with an exponent of 1.0. This means that observers can generally match a lateral distance of, say, 40 cm with a 40 cm distance extended in depth. After 2 m or so, height in the visual field becomes an important source (a normal adult pedestrian cannot use this information until objects are at least 1.5 m distant), and motion perspective for a pedestrian is no longer a blur; but accommodation and convergence cease to be effective.

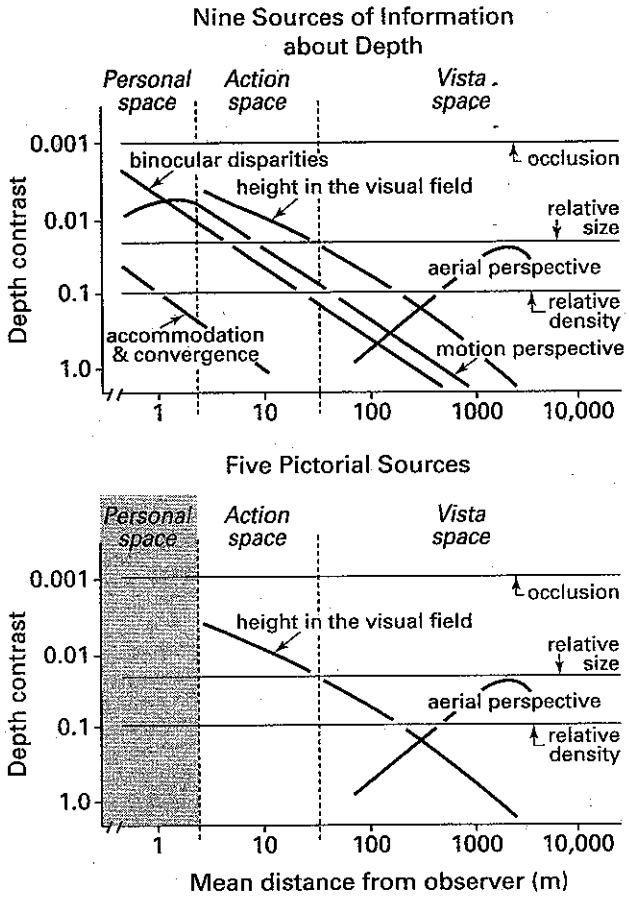

Figure 11.2
The upper panel shows the ordinal depth thresholds for nine sources of information (panel modified from Cutting and Vishton 1995). Egocentric depth (the abscissa) is logarithmically represented. Depth contrast (the ordinate) is measured in a manner similar to Michelson contrast in the spatial frequency domain. That is, the difference in the metric depths of two objects under consideration ($d_1 - d_2$) is divided by the mean egocentric distance of those objects from an observer [$(d_1 + d_2)/2$]. Egocentric space is also segmented into three general regions, which grade into one another. Personal space, out to about 2 m, is perceived as Euclidean (exponent of 1.0); action space, from about 2 to about 30 m, demonstrates some foreshortening (exponents near 1.0); and vista space, beyond about 30 m, often demonstrates considerable foreshortening (exponents declining to .40 or so, as suggested in figure 11.1). Different sources of information seem to work differentially in the different regions. An assumption made is that threshold measurements, shown in the panel, are good indicators of suprathreshold potency. The lower panel shows those thresholds isolated for pictorial sources of information, with personal space excluded because few pictures (before the twentieth century) depict space within 2 m of the viewer.

We suggested that at this distance (at about 1.5–2 m) there is a boundary, and it marks the beginning of *action space*. This new space extends out to about 30 m or so. Using 10% depth contrast as a general threshold of utility, both disparities and motion perspective cease to be useful at this distance. Most practically, an average individual can throw something reasonably accurately to a distance of 30 m. But perhaps most interestingly, 30 m is about the height of the Andrea Pozzo ceiling in the Chiesa di Sant'Ignazio in Rome (1694), where pillars and architectural ornaments are painted on a vaulted ceiling so that real pillars and painted pillars are indistinguishable when observed from a designated viewpoint (see Pirenne 1970). Action space is perceived near-metrically; its exponents are often only slightly less than 1.0.

Finally, there is *vista space,* everything beyond about 30 m. Here only the traditional pictorial sources of information are efficacious, and vista space becomes increasingly compressed with distance. Exponents in the near vista may be near 1.0, but those near the horizon become distinctively less.

Two caveats: First, although the boundaries between the spaces as we have drawn them are not arbitrary, neither are they well demarcated. Anywhere between 1 and 3 m could serve as the personal/action boundary and anywhere between 20 and 40 m could serve as the action/vista boundary. Training, stature, and a variety of other variables could also influence these boundaries. Most emphatically, one space should be expected to grade gracefully into the next.

Second, some functions can change (Cutting and Vishton 1995). If, for example, one places an observer in a British sports car driving through the countryside, then the function for height in the visual field would move a bit to the right. This is due to adjusting to the reduction in eye-height to about 1 m and because motion perspective would move considerably to the right (adjusting to an order of magnitude change in optical velocity). Some other functions change with the environment—height in the visual field as plotted is for vision standing on a plane (hills introduce changes in slant, altering the function, and occluded regions that create discontinuities in the function); aerial perspective is plotted for a clear day (humidity, mist, and fog would move the function considerably to the right). Still other functions may not prove useful in a given setting—there may be no objects on which to compare relative sizes. Finally, some functions change with characteristics of the observer: some people are stereoweak, others stereoblind, and sometimes this can occur after only a few hours of stereo

deprivation (Wallach, Moore, and Davidson 1963); many people over the age of forty can no longer accommodate; and, of course, adults are of different sizes and youngsters crawl and toddle, further changing the function for height in the visual field. All of these factors contribute, I claim, to the lability of perceived distances across people and environments.

Information in Pictorial Space

Most relevant to the purposes of this volume, I will claim that parts of this scheme can be applied directly to the perception of most photographs and to some art (particularly that using linear perspective). Again, such an application is suggested in the lower panel of figure 11.2, where pictorial sources of information—occlusion, relative size, relative density, height in the visual field, and aerial perspective—are shown. The first four comprise what was traditionally called artistic perspective (and with the addition of receding parallels, linear perspective; Kubovy 1986), with the fifth being added by Leonardo da Vinci (Richter 1883). In addition, at least with respect to art before the twentieth century, few paintings depict anything within personal space. Portraiture and still life—certainly the most proximal of artistic genres—typically portray their objects from just beyond arm's length, if not a greater distance. Moreover, this generally remained so even after the invention of photography. Thus, in the lower panel of figure 11.2, personal space is omitted.

Of course, there is additional information in pictures—information specifying that there is no depth (see discussion of flatness in Rogers, chapter 13 in this volume; Sedgwick, chapter 3 in this volume). When holding a photograph, the status of our accommodation and convergence, and the lack of binocular disparities and of motion perspective, all tell us that we are not really looking into the distance. Thus, pictures are endemic situations of information conflict. I think we fairly readily discount such conflict, although there may be residual influences. Reducing such potential conflict is one reason cinema is typically viewed from a distance, often 20 m or so, where accommodation and convergence are inactive and binocular disparities not salient.

A Case Study with Different Lenses

Representations of space often have manifold consequences; they even cause a perceptual elasticity. This changeableness can be demonstrated in at least two ways—in the effects of lenses (e.g., Farber and Rosinski 1978; Lumsden 1980; Pirenne 1970), and in the effects of viewing pictures from

noncanonical viewpoints of pictures, such as to the side (Cutting 1986b, 1987). Only the former will be discussed here.

Consider how camera lenses transform information for depth. Figure 11.3 shows two images from Charles Swedlund (1981) with the same content but taken with different lenses—a 35 mm lens (a somewhat short lens for 35 mm film) and a 200 mm lens (a telephoto). These images have been cropped, their sizes have been adjusted, and they were *not* taken from the same viewpoint. Instead, Swedlund normalized the relative sizes of the sign at the right of both images so that the sign takes up roughly the same proportion of the image space. To do this Swedlund was about four times farther away from the sign in the lower image, as noted in the plan view in the right panel of the figure. But look at the basketball backboards to the left in the background. These have grown enormously in size in the lower image, and their difference in relative size has diminished. The latter is due to the compressive effect of telephoto lenses. Ordinarily, a 200 mm lens blows up space 5.7 times compared to a 35 mm lens. With the 5.7-fold increase in image size, there should be a 5.7-fold decrease in apparent image depth. This is not quite true in the Swedlund (1981) images, the deviation due to cropping and sizing.

How exactly does a telephoto lens compress depth? Telephoto lenses enlarge the image for the observer. All enlargements compress depth—those created by longer lenses and those created by moving the observer forward toward the image from the original station point (the point of composition, or the point from which, when looking at the image, the optics are reconstructed most perfectly). How compression occurs is schematically suggested in figure 11.4. Because the points on the image do not change, and because the lines of sight from the observer to these points can be extended into "pictorial space," they find their end points at different pictorial depths. The closer the observer moves toward an image, the more compression in depth there should be; the farther he or she moves away from the image, the more dilation there should be. Sometimes these effects are demonstrable in experiments, but sometimes they are not. One point I would like to emphasize is that our perception of spaces is somewhat elastic; sometimes we pick up on information, sometimes not.

But what information is changed with a long lens? This is a bit more difficult to explain. Enlargements play with familiar size. A person 10 m away would ordinarily be in action space; a person 100 m away is ordinarily in vista space. Seen through a 500 mm lens, however, the latter

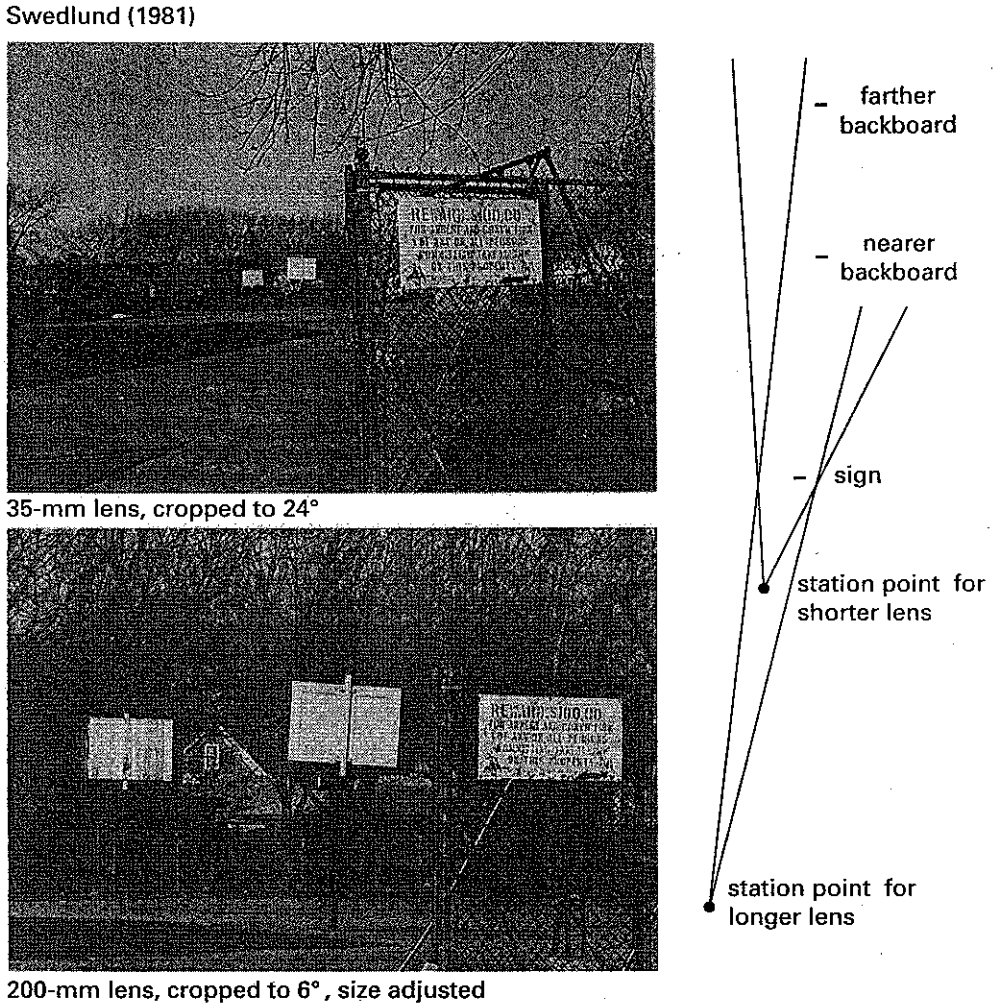

Figure 11.3
The two images on the left are from Swedlund (1981), reprinted with permission of Holt, Rinehart, and Winston. These images show the changes in pictorial space due to changes between a short and a long lens. The point from which the bottom image was photographed was about four times the distance of that for the top image, and both were cropped and sizes adjusted to make the appearance of the sign in the foreground the same. The diagram to the right is a plan view of the layout of the sign, the two backboards, and the two camera positions. It was reconstructed knowing that backboards are usually about 25 m apart and from the measurements of their heights (in pixels) in the digitized images. (Photographs from *Photography: A handbook of history, materials, and processes,* 2nd edition by Charles Swedlund and Elizabeth U. Swedlund, copyright © 1981 by Holt, Rinehart and Winston, reproduced by permission of the publisher.)

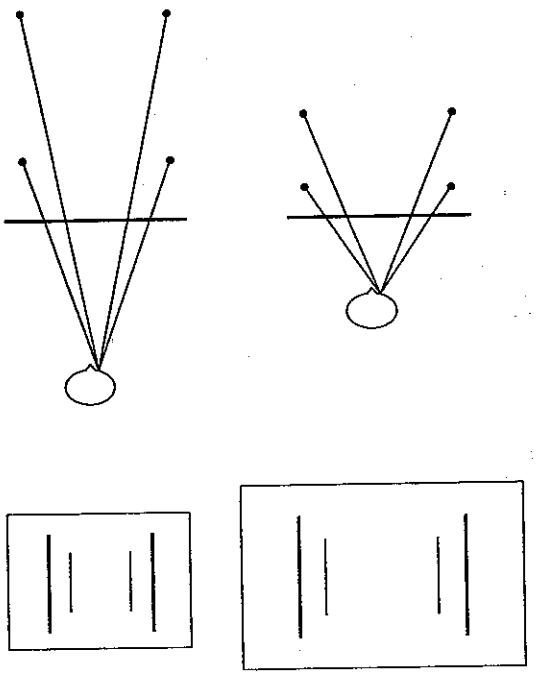

Figure 11.4
An affine reconstruction of pictorial space. The left panels show slices through untransformed pictorial space as seen from the composition point and the image seen by the observer; the other panels show transformations and views due to changes in observer position away from the composition point. Top panels show the affine transform in horizontal planes; the bottom-right panel shows a similarity transform (enlargement). After Cutting (1986b).

person has the same image size as the former seen through a 50 mm lens. Thus, familiarity with sizes of objects in the environment shifts, in this case, the whole of the environment shifts one log unit toward the observer; familiarity with what one can do with those objects move distant ones into what appears to be action space. (In telephoto images, as in regular images, there typically is no content in personal space.) Thus, what passes for action space now really extends out to perhaps 300 m.

If we carry with us a set of expectations about objects and the efficacy of information in real space to what we see in the telephoto image, perceptual anomalies will occur (Cutting 1997). Compared to the perception of environmental and normal pictorial spaces, relative size differences and relative density differences are reduced. Consider size. Basketball backboards are normally about 25 m apart. In the Swedlund images in figure 11.3 the ratio of sizes (linear extent) of the two backboards is 1.5:1 in the top image but only 1.24:1 in the bottom image. The perceived distance

difference in the top image could rather easily be 25 m; that in the lower image could not. In near space, we expect smaller relative size differences for similar objects closer together. In addition, height in the visual field also changes; the difference in height between the base of the poles for the two backboards is three times greater in the telephoto image. We expect larger height-in-the-visual-field differences for nearer objects.

The Shapes of Perceived Photographic Space

Consider next the data set of Kraft and Green (1989), already considered earlier as if it represented judgments about depth in the real world. Although gathered for different purposes,[7] these data are a definitive example of lens compression and dilation effects. Kraft and Green presented many photographs to observers. These were taken with a 35 mm camera and five different lenses: 17, 28, 48, 80, and 135 mm (where the first is called a *short* lens and the last a *telephoto* lens). With such lenses the horizontal field of view subtends about 105, 60, 45, 32, and 20°, respectively. Because a 48 mm lens is essentially a standard lens for a 35 mm camera (35 mm is the measure along the longer edge of the film frame; about 50 mm is the measure along its diagonal, and the functional diameter of the image circle; see, for example, Swedlund 1981), the layout of objects and their depth should appear relatively normal. Moreover, since the time of Leonardo, recommendations have been made to keep the larger, usually horizontal, subtense of a picture within 45°–60° to avoid distortions (Carlbom and Paciorek 1978). In two different outdoor environments, Kraft and Green (1989) planted poles at distances of 20, 40, 80, 160, and 320 m from a fixed camera. Viewing different arrangements of fifty slides, observers made judgments of the distance of each pole from the camera. Mean results are shown in the left panel of figure 11.5. This fan-shaped pattern provides grist for further analyses.

Notice first that perceived distances are generally quite a bit less than real distances, even for those stimuli shot with the standard lens. Ideally this should not occur.[8] To assess this general effects of foreshortening due to the experimental manipulation (the photographic lenses) and the experimental situation, I regressed the mean perceived distances in the left panel of figure 11.5 against a parameter cluster: $\phi \cdot (50/L)$. This is the physical distance of the posts from the camera multiplied by 50 (the standard lens length) over the lens length used. The second half of this composite variable is the extent to which perceived space ought to be compressed ($50/L < 1$) or dilated ($50/L > 1$; Cutting 1988; Lumsden 1980).

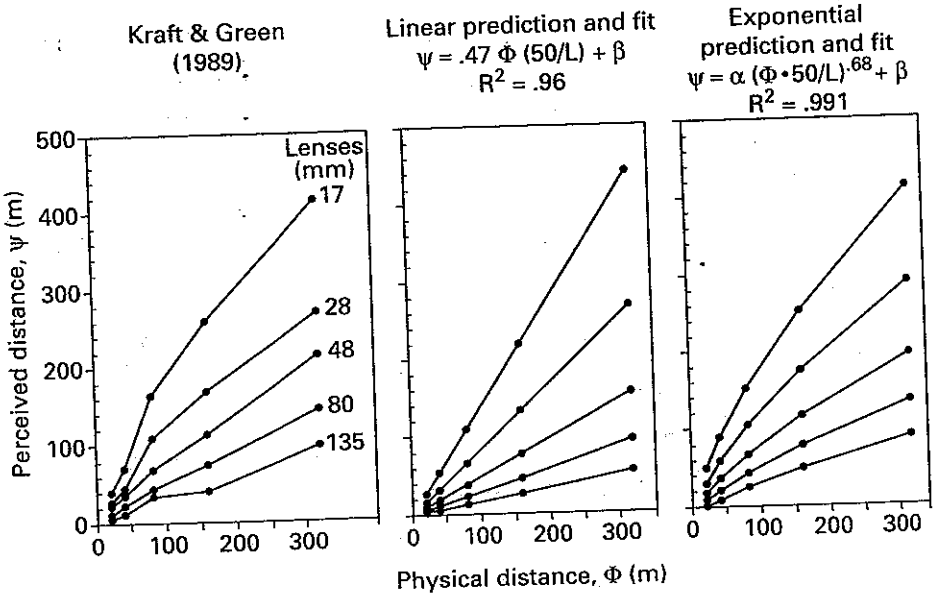

Figure 11.5
The data of Kraft and Green (1989) are presented in the left panel, showing perceived distances of posts as functions of their physical distance from a camera and its lens. The middle panel shows a model prediction based on the combination of physical distance and lens length; and the right panel shows a prediction based on this parameter and an exponent of .68. The fit shown in the left panel is statistically superior to that in the middle panel, revealing compression in perceived depth.

Thus, for example, a 135 mm lens should compress depth by a factor of about 2.8 (multiplying depth values by 0.37); a 17 mm lens should compress it by a 0.35 (actually dilating perceived space, multiplying values by 2.85).

This analysis accounted for 96% of the variance in the data ($R^2 = .96$, $F(1, 23) = 595$, $p < .0001$), fitting the pattern of data extremely well. The weight (a) given the parameter $\varphi \cdot 50/L$ in the equation was .47, about half what might have been expected. Perceived half-depth could be a compromise between full depth (carried by height in the visual field, relative size, relative density) and zero depth (enforced by no disparities and no motion perspective). Regardless, the schematized data, however, can serve as post hoc "predictions" for (i.e., model fits to) the real data. These are shown in the middle panel of figure 11.5. The match between them is quite good but not quite good enough.

Notice that there are some distinguishing characteristics in the shape of the predicted functions that are not generally wanted. Most salient is that

at large physical distances (φ) there is insufficient foreshortening in the perceived distances (ψ) when compared to the Kraft and Green data. This means that beyond general compression, there is general change in compression rates in the data, suggesting exponents less than 1.0. I reran the regression analysis with various exponents to determine the value of the best fit. A *single* exponent of .68 captured the most variance, increasing that accounted for to 99.1% ($F(1,23) = 2660, p < .0001$), a reliable improvement ($\chi^2 = 7.2$, $p < .01$) over the exponentless prediction. The model fits with the exponent are seen in the right panel of figure 11.5. (Remember, truncating the data at the near end decreases the exponent, as shown in the lower panels of figure 11.1.)

Of course, it may be simply a coincidence that judgments of distance in photographs (Kraft and Green 1989) and judgments in the real world (e.g., Da Silva 1985; Wagner 1985) are so similar, as shown in the top panel of figure 11.1. In may also be that the similarity in the decline in exponents found by truncating the near-distance data is similar in both photographic data (Kraft and Green 1989) and in real-world data (Gibson and Bergman 1954), shown in the lower panels of figure 11.1, is also a coincidence. But I think not. Most pleasantly, there are some elegant data gathered at ZiF (Center for Interdisciplinary Research), the site of the conference on which this volume is based. These show similar effects in the real world and in photographs of it (Hecht, van Doorn, and Koenderink 1999).[9] These data, if not the reanalyses that I have presented here, ought to be sufficiently convincing of the strong similarity—indeed functional identity—between photographic and real-world spaces.

11.4 IS PERCEIVING THE LAYOUT IN OTHER PICTURES LIKE PERCEIVING THAT IN ENVIRONMENTAL SPACE?

Yes—at least in terms of principles if more than the end result. Earlier I had suggested that psychologists and others interested in pictorial space have spent perhaps too much time with photographs and linear perspective paintings, drawings, and etchings. These are among the more recent of pictures that human beings have crafted. Manfredo Massironi and I (Cutting and Massironi 1998) suggested one might equally start at the other end of time, at least from the perspective of human culture. Cave paintings—some of which are at least 30,000 years old—as well as cartoons, caricatures, and doodles are made up of *lines*. These pictures are not copies of any optical array, and yet as images they depict objects well.

Many of these compositions can be read to represent animals in depth, often by occlusion of one by another.

How do line drawings work? What kind of space do they build? Cutting and Massironi (1998) suggested that there are four basic types of lines and that these ordinally parse the regions on either side of the line or parse themselves from those regions. The four types are shown in figure 11.6. *Edge lines* separate a figure from background, where one side belongs to the figure and is closer to the viewer and the other side belongs to the background. We notate this type of configuration [aAb] or [bAa], where letter order denotes ordinal position in depth, [a] in front of [b]; and uppercase [A] denotes the line, lowercase [a] one of the two regions on either side of the line. A second kind is an *object line,* where the line stands for an entire object in front of the background [bAb]. Its inverse is the *crack line* [aBa], where the line invites the viewer to imagine the interior space hidden from view. Finally, there are *texture lines,* which subdivide again. Texture lines can represent small edges, small objects, or small cracks; and they can also represent shading, color, or what the traditional pedagogical literature on drawing calls *mass* (Speed 1913).

Consider an example from the oldest known works of paleolithic art. Shown in figure 11.7 are several rhinoceroses from La Grotte Chauvet in the south of France (see also Bahn and Vertut 1997; Chauvet, Brunel Deschamps, and Hillaire 1995; Clottes 2001). As part of this figure one can see edge lines (outlining a rhinoceros), object lines (its horns), crack lines (its mouth), as well as texture lines (the coloring of the flank). The antiquity of this image, and the fact that it has the complete set of line primitives outlined by Cutting and Massironi (1998), suggests that line primitives are not culturally limited or even culturally determined. The fourfold use of lines seems deeply embedded in our genome.

I claim that each line serves to carve up a small bit of pictorial space; each contributes a local bit of ordinality. Nothing is needed in this case but a segregation of two different depth planes—one of the animal and one of the background against which it stands. As images get richer in information, depth can become richer as well. Consider another example from La Grotte Chauvet in figure 11.8.

Here a few well-crafted lines represent four horses. There is great controversy about whether or not these horses should be read as occupying the same space, each occluding another that is higher up in the pictorial place. Nonetheless, I think depth is inescapable. In the paleolithic corpus of images, one often finds the superimposition of lines from different

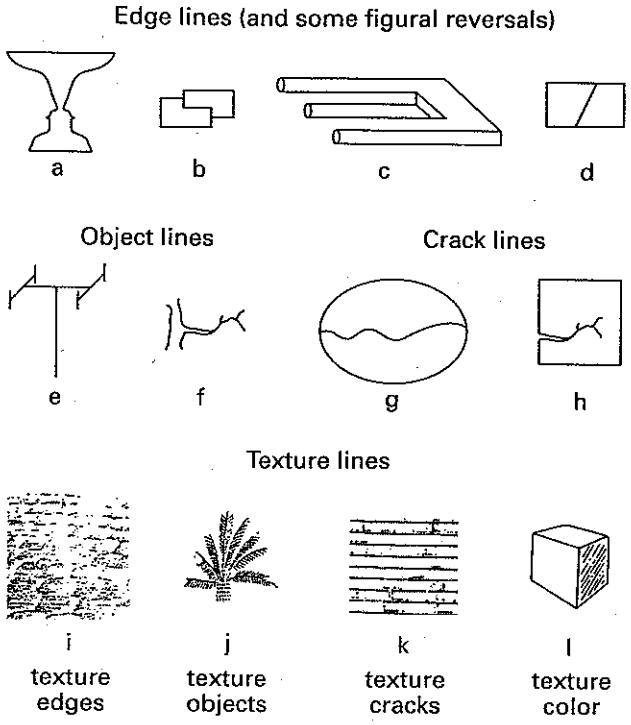

Figure 11.6
A taxonomy of lines. Edge lines separate the regions on either side and assign different ordinal depth. Four figures that play with this relationship are shown: (a) the faces/goblet illusion, after Rubin (1915); (b) an ambiguous occlusion figure, after Ratoosh (1949); (c) the devil's pitchfork, after Schuster (1964); and (d) a rectangle/window, after Koffka (1935). Next are shown some object lines: (e) is a mid-twentieth-century version of a television antenna; and (f) shows the twigs at the end of a tree branch. Third are figures with crack lines: (g) is the mouth of a clam, after Kennedy (1974); and (h) is a crack in a block. Note that (f) and (h) are exact reciprocals—switching from object line to crack line. Finally, four texture line types are shown: (i) texture edges of occlusion in cobblestones, after de Margerie (1994); (j) texture objects as palm fronds, after Steinberg (1966); (k) texture cracks as mortar between bricks, after Brodatz (1966); and (l) texture color, indicating shadow. (Adapted from Cutting and Massironi, 1998)

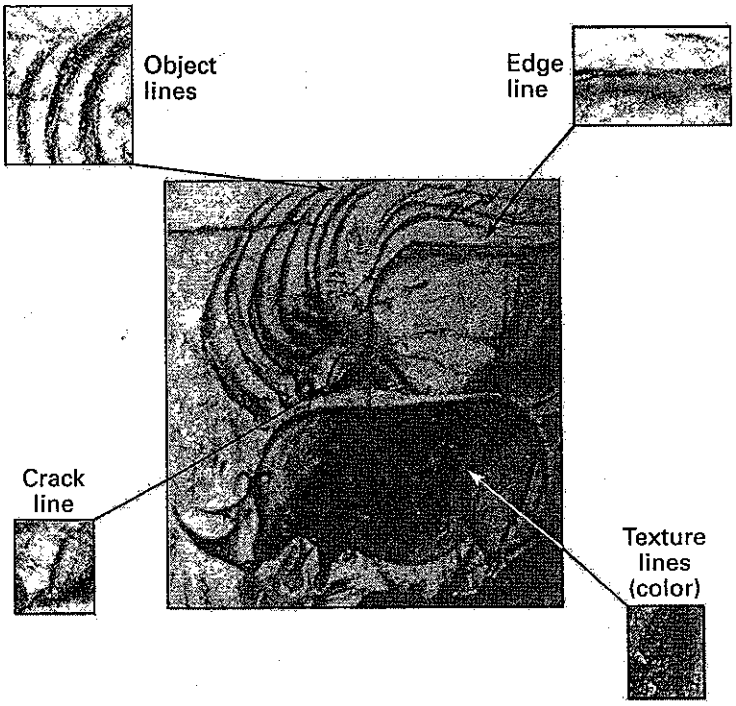

Figure 11.7
A thresholded detail from a painting in La Grotte Chauvet demonstrating the use of four types of lines—edges, objects, cracks, and textures. Because this image is at least 30,000 years old, it would appear that this typology of lines is not culturally relative but possibly biologically engrained. (Image reprinted from Clottes, 2001, with the kind permission of Jean Clottes, Ministère de la Culture)

animals in different orientations, lines crossing, suggesting that they were composed without respect for one other. Here, however, we find what appear to be true occlusions of horses each in the same orientation. Each edge line can be interpreted locally to generate depth, and local depth assignments can be written [bAa]. However, taken as an ensemble, several relations have to be rewritten, pushing some regions and lines farther into depth. Thus, the occluding edge of the second horse becomes [cBb], the third [dCc], and the fourth [eDd]. In this manner global coherence can be built up from local depth assignments, and multiple, ordered depth planes can be built up as well. I suggest this is done in a manner not much different than seen in cooperative stereo algorithms (Julesz 1971; Marr and Poggio 1976), although in pictorial cases it is much simpler. Depth is incrementally built up, in more difficult cases perhaps with many passes through the data. Moreover, there can be great play in this. Figure 11.6a–6d are examples from the psychological literature, but of course

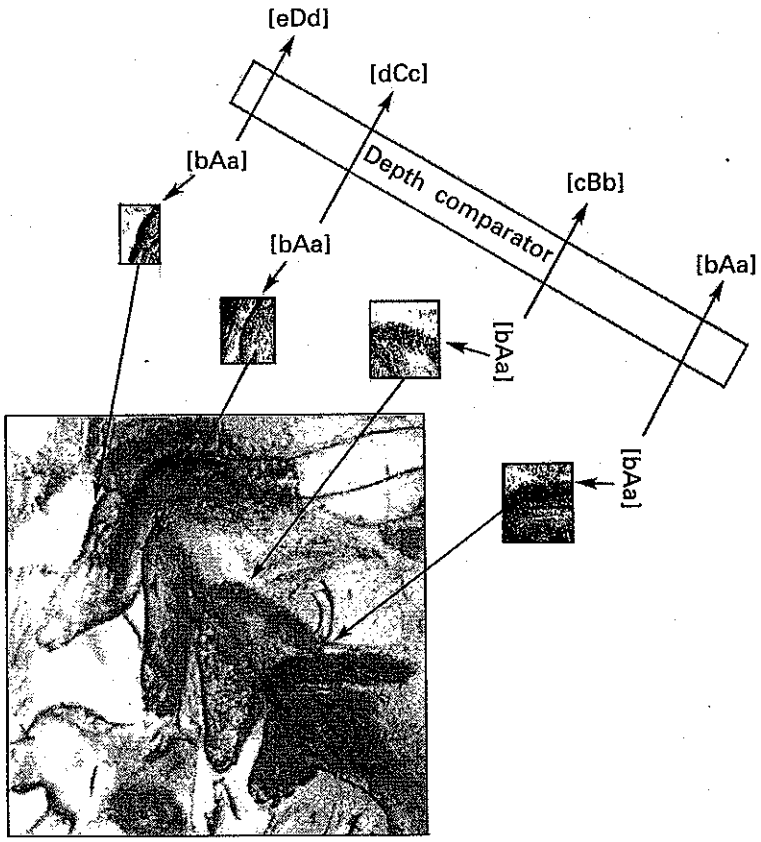

Figure 11.8
From pairwise ordinal depth relations to five ordinal depth planes. A thresholded detail from perhaps the most famous and controversial image in La Grotte Chauvet. Many claim it should not be read as four horses in depth, although I disagree. If it is read in depth, one can follow the assignments of occlusions, using edge lines [bAa]. Each of four pairs has this assignment. If this information is fed into a comparator that knows that objects can occlude, the end result is five ordinal pictorial depths, [a]–[e]. (Image reprinted from Clottes, 2001, with the kind permission of Jean Clottes, Ministère de la Culture)

much of the corpus of M. C. Escher (Bool et al. 1981) can be construed this way.

Assignments of edge lines [aAb] and [bAa] are often judgments of occlusion. Assignments of many lines (and all line types), however, could also be made on the basis of height in the visual field; assignments of similarly shaped objects (made up of lines) could also be made on the basis of relative size; and so forth. Thus, bringing knowledge to bear on a graphic situation, one can assign depth order to many elements, The five pictorial sources of information—occlusion, height in the visual field, relative size,

relative density, and aerial perspective—can constrain one another quite well so that a reasonably rich, affine space is established. Moreover, an approximation to a metric space can also be achieved, particularly with the application of linear perspective.

How might this be done? The analogy I wish to make is with respect to nonmetric multidimensional scaling (Kruskal 1964; Shepard 1980), and it is a deep lesson in measurement theory. By considering only pairwise ordinal relations among all (or even only most) members of a set of data, one can assemble a nearly metric space of that set. Metric distances, even ratios, are not needed. Yacov Hel-Or and Shimon Edelman (1994) demonstrated the logical efficacy of this approach to qualitative stereo, and Cutting and Vishton (1995) suggested it should work considering depth in general. Thus, one can think of perceived space—even at its articulated, near-Euclidean best—as built up incrementally from constraints of ordinality. These constraints, when sufficiently rich, converge on a near-Euclidean framework.

In summary, I claim that perceiving depth in pictures and perceiving depth in the real world are cut from the same informational cloth. Thus, if we misunderstand perceived pictorial space (as might be suggested by the title of the conference on which this volume was based), then it follows that we misunderstand perceived environmental space as well. I claim further that when ordinal depth information is sparse, perceived depth is also crude, confined to a few depth planes. When ordinal information is richer, perceived space becomes more articulated, allowing first for many depth planes (and an essentially affine representation). When that information is extremely rich, as with the addition of binocular disparities and motion perspective in cluttered environments and in daylight, ordinal constraints can become sufficiently tight to approach a metric representation. Thus, the concatenation of ideas from measurement theory and from the computational practice of multidimensional scaling allows promise of considering the perception of space in pictures and in the world as proceeding from the same principles.

ACKNOWLEDGMENT

I thank Claudia Lazzaro for many fruitful discussions and for her patience in listening to many ill-formed ideas, Robert Kraft for sharing methodological information about this study of pictorial depth, and Patrick Maynard for some insightful comments.

Notes

1. Earlier in his career, however, Gibson was less sure: "It is theoretically possible to construct a dense sheaf of light rays to a certain point in a gallery or a laboratory, one identical in all respects to another dense sheaf of light rays to a unique station point thousands of miles away" (1960, p. 223), although he denied it was obtainable in practice.

2. This is an axiom from Euclid's optics (Burton 1945). Euclid's axioms and his proofs deal with both stationary and moving observers.

3. This is not to say that people cannot be trained to perceive vista space more veridically; clearly they can, and the data of E. Gibson and Bergman (1954) and Galanter and Galanter (1973) show this. The focus of this chapter, however, is on what normal adults, without specific distance training, will perceive and judge.

4. Flückiger (1991) had observers judge the distance of boats floating on Lake Leman, looking from near Geneva toward Montreux. Only relative size and height in the visual field would provide firm information, perhaps with the addition of aerial perspective and relative densities of waves. Experimental results for judgments between 0.2 km and 2.0 km and between 0.75 km and 2.25 km were quite tidy; whereas those between 2.8 km and 5.6 km were not. Flückiger did not fit exponents to his data, but this is easily done. Mean exponents for the first complete data sets were .36 and .44, respectively. In addition, I have omitted the data of Galanter and Galanter (1973) here. Their observers were highly trained and, although they viewed objects up to 10 km into the distance, they viewed them from airplanes at altitudes of about 60 m. This raises the function of height in the visual field by a factor of 40, as discussed in the next section.

5. Exponents become unstable when a first pair exhibits overconstancy (more perceived distance between them than physically present). When this occurs, exponents tend toward zero. This also happened in the analysis of the data of Flückiger (1991).

6. Linear perspective, often cited as a single source of information, is really a concomitance of occlusion, relative size, relative density, and height in the visual field, using the technology of parallel straight lines and their recession.

7. Kraft and Green (1989) believed that the foreshortening of space seen in photographs due to the use of lenses of different lengths is caused by the truncation of the foreground. Although this is a possibility, the purist form of the argument is that the foreground is truncated but the remainder of space remains the same. Their own data (Kraft and Green, 1989, experiment 1) are not consistent with this idea; they show a fan effect rather than parallel lines.

8. The Kraft and Green article did not give full details of the presentation of the stimuli. Thus, I contacted Robert Kraft and he kindly provided additional experimental details (Kraft, personal communication, May 24, 2000). First, the length of the lens used to project the images was 100 mm. This means that the images are half as large than they would be with a 50 mm projector lens from the distance of

the projector. Ideally, when seen from the projector, such diminution should dilate depth by a factor of about 2.0 (multiplying distances by about 2.0; see Cutting, 1988, for a similar analysis). However, the distance of the projector to the screen was at 6 m, twice the mean distance of the observers (3 m). This latter effect enlarges the image by a factor of 2, compressing depth (multiplied by 0.5). These two effects are linear and should cancel. Theoretically one would expect perceived distances to be compressed by a factor of $0.5 \cdot 2.0$ (multiplied by a factor of 1.0). This also means that the images seen in the experimental situations subtended the same angle as the corresponding images in the real world.

9. The most convincing comparisons in the data of Hecht, van Doorn, and Koenderink (1999) are for angular judgments of the intersections of wall of buildings at ZiF, which has few right angles.

LOOKING INTO PICTURES

AN INTERDISCIPLINARY APPROACH TO PICTORIAL SPACE

edited by
Heiko Hecht
Robert Schwartz
Margaret Atherton

2003

A Bradford Book
The MIT Press
Cambridge, Massachusetts
London, England