

CHAPTER 8

Depth Perception and the Perception of Events

DENNIS R. PROFFITT AND CORRADO CAUDEK

DEPTH PERCEPTION 214

Depth Cues: What Is the Information? What Information Is Actually Used? 214

Integration of Depth Cues: How Is the Effective Information Combined? 219

Distance Perception 220

Size Perception 222

Geographical Slant Perception 223

EVENT PERCEPTION 224

Perceptual Organization 224

Perceiving Dynamics 226

Perceiving Our Own Motion 227

Visual Control of Posture 228

Perceiving Approaching Objects 228

CONCLUSION 229

REFERENCES 231

Our understanding of the perception of depth and events entails a paradox. On the one hand, it seems that there is simply not enough information to make the achievement possible, yet on the other hand the amount of information seems to be overly abundant. This paradox is a consequence of evaluating information in isolation versus evaluating it in context. Berkeley (1709) noted that a point in the environment projects as a point on the retina in a manner that does not vary in any way with distance. For a point viewed in isolation, this is true. From this fact, Berkeley concluded that the visual perception of distance, from sight alone, was impossible. Visual information, he concluded, must be augmented by nonvisual information. For example, fixating a point with two eyes requires that the eyes converge in a manner that does vary with distance; thus, proprioceptive information about eye positions could augment vision to yield an awareness of depth. If our visual world consisted of a single point viewed in a void, then depth perception from vision alone would, indeed, be tough. Fortunately, this is not the natural condition for visual experience.

As the visual environment increases in complexity, the amount of information specifying its layout increases. By adding a second point to the visual scene, additional information is created. If the two points are at different depths, then they will project to different relative locations in the two eyes, thereby providing information about their relative distances to each other. If the observer fixates on one point and

moves his or her head sideways, then motion parallax will be created that gives information about relative depth. Expanding the points into extended forms, placing these forms on a ground plane, having the forms move, or allowing the observer to move are some of the possible complications that create information about the depth relationships in the scene.

Complex natural environments provide lots of different kinds of information, and the perceptual system must combine all of it into the singular set of relationships that is our experience of the visual world. It is not enough to register the available information; the information must also be combined.

From this brief introduction, two fundamental questions emerge: What is the information provided for perceiving spatial relationships and how is this information combined by the perceptual system? We begin our chapter by reviewing the literature that addresses these questions. (Additional topics in visual perception are discussed in the chapter by Kubovy, Gepshtein, and Epstein in this volume.)

There is a third question that we also address: Do people perceive spatial layout accurately? The answer to this question depends upon the criteria used to define accuracy. Certainly, people act in the environment as if they represent its spatial relationships accurately; however, effective action can often be achieved without this accurate representation. This issue is developed and discussed throughout the chapter.

DEPTH PERCEPTION

Depth Cues: What Is the Information? What Information Is Actually Used?

The rules for projecting a three-dimensional object or three-dimensional layout onto a surface (for example, the retina) are unambiguously defined, whereas the inverse operation (from the image to the three-dimensional projected object or scene) is not. This is the so-called *inverse-projection problem*. Any two-dimensional projection is inherently ambiguous, and a central problem of visual science is to determine how the perceptual system is able to recover three-dimensional information from a retinal projection. This problem is usually attacked from two sides: first, by analyzing those properties of the image (hereafter called *sources of depth information*, or *cues*) that, in principle, allow for the recovery of some of the three-dimensional properties of the projected objects; second, by investigating the effectiveness of these sources of depth information for the human visual system. In this section, we discuss the problem of depth perception by clarifying what kinds of three-dimensional information can be recovered from each source of depth information in isolation, and by presenting psychophysical evidence suggesting whether and to what degree the visual system is actually able to use them. We start with the ocular motor sources of information, followed by binocular disparity, pictorial depth cues, and motion.

Ocular Motor

There are two potentially useful extraretinal sources of information for specifying egocentric distance: the vergence angle of the eyes and the state of accommodation. The vergence angle is approximately equal to the angle between the lines from the optical centers of the eyes to the fixation point and the parallel rays that would define gaze direction if the eyes were fixated at infinity. If the vergence angle and the interocular distance are known, then it is clear that the radial distance to the fixation point could in principle be recovered. This potential cue to depth, however, is limited to a restricted range of distances, because the eyes become effectively parallel (optical infinity) for fixation distances larger than 6 m. Moreover, the information content of vergence drops off rapidly with increasing distance: The variation in the vergence angle is very limited for fixation distances larger than about 0.5 m. Psychophysical evidence suggests that vergence information is not a very effective source of information about distance. Erkelens and Collewijn (1985), for example, showed that observers could make large tracking vergence eye movements without seeing any motion in depth

when the expansion or contraction of the retinal projection is controlled. In such a cue-conflict situation (with ocular convergence conflicting with the absence of expansion or contraction of the retinal image), extraretinal information fails to affect perceived distance.

Accommodation is a second source of extraretinal information about distance and refers to the change in shape of the lens that the eye performs to keep in focus objects at different distances. Changes in accommodation occur between the nearest and the farthest points that can be placed in focus by the thickening and thinning of the lens. Although in principle, accommodation could be a source of depth information, psychophysical investigations suggest that the contribution of accommodation to perceived depth is minimal and that there are large individual differences (Fisher & Ciuffreda, 1988). For single point-light targets in the dark, Mon-Williams and Tresilian (1999) reported that observers were unable to provide reliable absolute-depth judgments on the basis of accommodation alone, even within a stretched-arm distance. Observers, however, were able to recover ordinal-depth information for sequentially presented targets from accommodation alone, even though the depth-order judgments were only 80% correct.

In conclusion, the ocular-motor cues are not reliable cues for perceiving absolute depth, even though they may play a more important role in the recovery of ordinal-depth information. The effectiveness of the ocular-motor cues is limited to a small range of distances, and they are easily overcome when other sources of depth information are available.

Binocular Disparity

Since Euclid, we have known that the same three-dimensional object or surface-layout projects two different images in the left and right eye. It was Wheatstone (1938), however, who provided the first empirical demonstration that disparate line drawings presented separately to the two eyes could elicit an impression of depth. Since then, binocular disparity has been considered one of the most powerful sources of optical information for depth perception, as is easily realized by anyone who has ever seen a stereogram. It is likewise easy to realize that binocular disparity—although by itself sufficient to specify relative distance—may not be sufficient to specify absolute-distance information. The problem of scaling disparity information is one of the central themes in the literature on stereopsis.

Using the small angle approximation, the geometrical relation between disparity and depth is

$$\eta \approx \frac{l\delta}{D(D - \delta)}$$

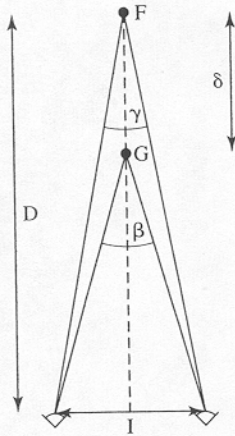


Figure 8.1 F = fixation point, I = interocular separation, δ = depth difference between F and G , γ = binocular parallax of F , and β = binocular parallax of G . The relative disparity, η , between F and G is $\beta - \gamma$.

where η is the angular binocular disparity, D is the viewing distance, δ is the depth, and I is the interocular distance. The angles and distances of stereo geometry are specified in Figure 8.1. From disparity, therefore, the depth magnitudes can be recovered as

$$\delta \approx \frac{\eta D^2}{I + \eta D}$$

The previous equation reveals that there is a nonlinear relationship between horizontal disparity (η) and depth (δ) that varies with the interocular separation (I) and the viewing distance (D). This means that disparity information by itself is not sufficient for specifying depth magnitudes, because different combinations of interocular separation, depth, and distance can generate the same disparity. In order to provide an estimate about depth, disparity must be scaled, so to speak, by some other source of information specifying the interocular separation and the viewing distance. This is the traditional *stereoscopic depth-constancy problem* (Ono & Comerford, 1977).

One proposal is that this scaling of disparity is accomplished on the basis of extraretinal sources. According to this approach, failures of veridical depth perception from stereopsis have been attributed to the misperception of the viewing distance. Johnston (1991), for example, showed random-dot stereograms to observers who decided whether they were seeing simulated cylinders that were flattened or elongated along the depth axis with respect to a circular cylinder. Johnston found that depth was overestimated at small distances (with physically circular cylinders appearing as elongated in depth) and underestimated at larger distances (with physically circular cylinders appearing as flattened). These depth distortions have been attributed to the hypothesis that

observers scaled the horizontal disparities with an incorrect measure of physical distance, entailing an overestimation of close distances and an underestimation of far distances. A second proposal is that disparity is scaled on the basis of purely visual information. Mayhew and Longuet-Higgins (1982), for example, proposed that full metric depth constancy could be achieved by the global computation of vertical disparities. The psychophysical findings, however, do not support this hypothesis. It has been found, in fact, that human performance is very poor in tasks involving the estimation of metric structure from binocular disparities, especially if compared with the precision demonstrated by performance in stereo acuity tasks involving ordinal-depth discriminations, with thresholds as low as 2 s arc (Ogle, 1952).

Koenderink and van Doorn (1976) proposed a second purely visual model of stereo-depth. This model does not try to account for the recovery of absolute depth, but only for the recovery of the affine (i.e., ordinal) structure from a combination of the horizontal and vertical local disparity gradient. This model, however, is inconsistent with the psychophysical data. It predicts that the local manipulation of either horizontal or vertical disparity should have the same effects on perceived shape; however, it has been shown that the manipulation of the local horizontal disparities has reliable consequences on perceived depth, whereas the manipulation of the local vertical disparities does not (Cumming, Johnston, & Parker, 1991). Moreover, several studies have shown that vertical disparity processing is not local, but rather is performed by pooling over a large area (Rogers & Bradshaw, 1993).

In conclusion, binocular disparity in isolation gives rise to the most compelling impression of depth, and for relatively short distances provides a reliable source of relative-, but not of absolute-depth information. Although the geometric relationship between binocular disparity and depth is well understood, a plausible psychological model of stereopsis has yet to be provided.

Pictorial

Pictorial depth cues consist of those depth-relevant regularities that are manifested in pictures. There is a long list of these cues, and although we have attempted to describe the most important ones, our list is not exhaustive.

Aerial Perspective. *Aerial perspective* refers to the reduction of contrast that occurs when an object is viewed from great distances. Aerial perspective is the product of the scattering of light by particles in the atmosphere. The contrast reduction by aerial perspective is a function of both distance and the attenuation coefficient of the atmosphere. Under hazy

conditions, for example, the contrast of a black object against the sky at 2000 m is only 45% of the contrast produced by the same object when it is viewed at a distance of 1 m (Fry, Bridgeman, & Ellerbrock, 1949).

Aerial perspective is an ambiguous cue about absolute distance: The recovery of distance from aerial perspective requires knowledge of both the reflectance value of the object and the attenuation coefficient of the atmosphere. It should also be noticed that the attenuation coefficient is changed by the scattering and blocking of light by pollutants. As a consequence, aerial perspective cannot be taken as providing more than ordinal-depth information, and several psychophysical investigations have indicated its effectiveness as a depth-order cue (Ross, 1971).

Height in the Visual Field and the Horizon Ratio. If an observer and an object are situated on the same ground plane, then the observer's horizontal line of sight will intersect the object at the observer's eye height. Because this line of sight also coincides with the horizon, the horizon intersects the object at the observer's eye height. The reference to the (explicit or implicit) horizon line therefore can be used to recover absolute size information as multiples of the observer's eye height. The geometry of the horizon ratio was first presented by Sedgwick (1973):

$$h = \frac{\tan \beta + \tan \gamma}{\tan \beta}$$

where γ is the visual angle subtended by the object above the horizon, and β is the visual angle subtended by the object below the horizon (see Figure 8.2). Although size information (h) is independent of the distance of the object from the observer, the object size (scaled in terms of eye height) and the visual angle are known; thus, distance itself can also be recovered. The recovery of absolute-size information from the horizon ratio requires two assumptions: (a) that both observer and target object lie on the same ground plane, and

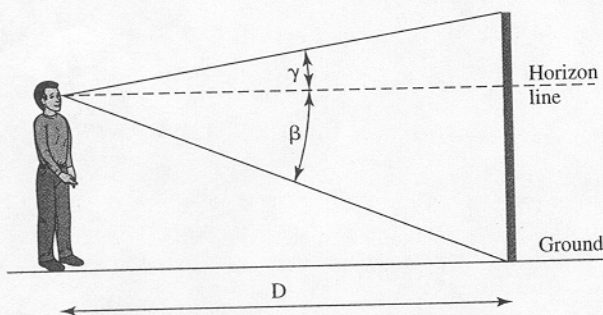


Figure 8.2 D = distance from the observer to the object, β = visual angle between the point where the object meets the ground and the horizon, and γ = visual angle between the top of the object and the horizon.

(b) that the observer's eye is at a known distance from the ground. If the second assumption is not met, then the horizon ratio still provides relative-size information about distant objects.

Evidence has been provided showing that the horizon ratio is an effective source of relative-depth information in pictures (Rogers, 1996). Wraga (1999) and Bertamini, Yang, and Proffitt (1998) reported that eye-height information is used differently across different postures. For example, Bertamini et al. investigated the use of the implicit horizon in relative-size judgments. They found that size discrimination was best when object heights were at the observers' eye height regardless of whether they were seated or standing.

Occlusion. Occlusion occurs when an object partially hides another object from view, thus providing ordinal information: The occluding object is perceived as closer and the occluded object as farther. Investigations of occlusion have focused on understanding how occlusion relationships are identified: that is, how the perceptual system decides whether a surface "owns" an image boundary or whether the boundary belongs to a second occluding surface. It is easily understood that the so-called border ownership problem is critical to correctly segmenting the spatial layout of the visual scene into surface regions at different depths. Boundaries that belong to an object are *intrinsic* to its form, whereas those that belong to an occluding object are *extrinsic* (Shimojo, Silverman, & Nakayama, 1989). Shimojo et al. (1989) created a powerful demonstration of the perceptual effect derived from changing border ownership by using the barber-pole effect. The barber-pole effect has been attributed to the propagation of motion signals generated by the contour terminators along the long sides of the aperture (Hildreth, 1984). By stereoscopically placing the contours behind aperture boundaries, Shimojo et al. caused the terminators to be classified as extrinsic to the contours (because they were generated by a near occluding surface), and the terminators to be subtracted from the integration process. As a consequence, Shimojo et al. found that the bias of the barber-pole effect was effectively eliminated.

Relative and Familiar Size. The relative-size cue to depth arises from differences in the projected angular sizes of two objects that have identical sizes and are located at different distances. If the assumption that the two objects have identical physical sizes is met, then from the ratio of their angular sizes, it is possible to determine the inverse ratio of their distances to the observer. In this way, metrically scaled relative-depth information can be specified. If the observer also knows the size of the objects, then in principle, absolute-distance

information becomes available. Several reports indicate that the familiar-size cue to depth does indeed affect perceived distance in cue-reduction conditions (Sedgwick, 1986), but it does not affect perceived distance under naturalistic viewing conditions (Predebon, 1991).

Texture Gradients. Textural variations provide information about both the location of objects and the shape of their surfaces. Cutting and Millard (1984) distinguished among three textural cues to shape: *perspective*, *compression* (or *foreshortening*), and *density*.

Perspective: Due to perspective, the size of the individual texture elements (referred to here as *texels*) is inversely scaled with distance from the viewer. Perspective (or scaling) gradients are produced by perspective projections, in the cases of both planar and curved surfaces. To derive surface orientation from perspective gradients, it is necessary to know the size of the individual texels.

Compression: The ratio of the width to the length of the individual texels is traditionally referred to as compression. If the shape of the individual texels is known a priori, compression can in principle provide a *local* cue to surface orientation. Let us assume, for example, that the individual texels are ellipses. In such a case, if the visual system assumes that an ellipse is the projection of a circle lying on a slanted plane, then the orientation of the plane could be locally determined without the need of measuring texture gradients. In general, the effectiveness of compression requires the assumption of *isotropy* (A. Blake & Marinos, 1990). If the texture on the scene surface is indeed isotropic, then for both planar and curved surfaces, compression is informative about surface orientation under both orthogonal and perspective projections.

Density: Density refers to the spatial distribution of the texels' centers in the image. In order to recover surface orientation from density gradients, it is necessary to make assumptions about the distribution of the texels over the object surface. Homogeneity is the default assumption; that is, the texture is assumed to be uniformly distributed over the surface. Variation in texture density can therefore be used to determine the orientation of the surface. Under the homogeneity assumption, the density gradient is in principle informative about surface orientation for both planar and curved surfaces under perspective projections, and only for curved surfaces under orthographic projections.

Even if the mathematical relationship between the previous texture cues and surface orientation is well understood for both planar (Stevens, 1981) and curved surfaces (Gårding, 1992), the psychological mechanism underlying the perception of shape from texture is still debated. Investigators are trying to determine which texture gradients observers use to

judge shape from texture (Cutting & Millard, 1984), and to establish whether perceptual performance is compatible with the isotropy and homogeneity assumptions (Rosenholtz & Malik, 1997).

Linear Perspective. Linear perspective is a very effective cue to depth (Kubovy, 1986), but it can be considered to be a combination of other previously discussed depth cues (e.g., occlusion, compression, density, size). Linear perspective is distinct from natural perspective by the abundant use of receding parallel lines.

Shading. Shading information refers to the smooth variation in image luminance determined by a combination of three variables: the illuminant direction, the surface's orientation, and the surface's reflective properties. Given that different combinations of these variables can generate the same pattern of shading, it follows that shading information is inherently ambiguous (for a discussion, see Todd & Reichel, 1989). Mathematical analyses have shown, however, that the inherent ambiguity of shading can be overcome if the illuminant direction is known, and computer vision algorithms relying on the estimate of the illuminant direction have been devised for reconstructing surface structure from image shading (Pentland, 1984).

Psychophysical investigations have shown that shading information evokes a compelling impression of three-dimensional shape, even though perceived shape from shading is far from being accurate. The perceptual interpretation of shading information is strongly affected by the pictorial information provided by the image boundaries (Ramachandran, 1988). Moreover, systematic distortions in perceived three-dimensional shape occur when the direction of illumination is changed in both static (Todd, Koenderink, van Doorn, & Kappers, 1996) and dynamic patterns of image shading (Caudek, Domini, & Di Luca, in press).

Thus, perceiving shape from shading presents a paradox. Shading information can, in principle, specify shape if illumination direction is known. Moreover, in some circumstances, observers recover this direction with good accuracy (Todd & Mingolla, 1983). Yet, perceived shape from shading is often inaccurate, as revealed by the studies manipulating the image boundaries and the illuminant direction.

Motion. The importance of motion information for the perception of surface layout and the three-dimensional form of objects has been known for many years (Gibson, 1950; Wallach & O'Connell, 1953). When an object or an observer moves, the dynamic transformations of retinal projections

become informative about depth relationships. When one object is seen to move in front of or behind another, dynamic occlusion occurs. This information specifies depth order. When an observer moves, motion parallax occurs between objects at different depths, and when an object rotates, it produces regularities in its changing image. These events provide information about the three-dimensional structure of objects and their spatial layout.

Dynamic Occlusion. Dynamic occlusion provides effective information for determining the depth order of textured surfaces (Andersen & Braunstein, 1983). In one of the earliest studies, Kaplan (1969) showed two random-dot patterns moving horizontally at different speeds and merging at a vertical margin. Observers reported a vivid impression of depth at the margin, with the pattern exhibiting texture element deletion being perceived as the farthest surface.

It is interesting to compare dynamic occlusion and motion parallax, because in a natural setting, these two sources of depth information covary. Ono, Rogers, Ohmi, and Ono (1988) put motion parallax and dynamic occlusion in conflict and found that motion parallax determines the perceived depth order when the simulated depth separation is small (less than 25 min of equivalent disparity), and dynamic occlusion determines the perceived depth order when the simulated depth separation is large (more than 25 min of equivalent disparity). On the basis of these findings, Ono et al. proposed that motion parallax is most appropriate for specifying the depth order *within* objects (given that the depth separation among object features is usually small), whereas dynamic occlusion is more appropriate for specifying the depth order *between* objects at different distances.

Structure From Motion. The phenomenon of the perceived *structure from motion* (SFM) has been investigated at (at least) three different levels: (a) the theoretical understanding of the depth information that, in principle, can be derived from a moving projection; (b) the psychophysical investigation of the effective ability of observers to solve the SFM problem; and (c) the modeling of human performance. These different facets of the SFM literature are briefly examined in the following section.

Mathematical analyses: A way to characterize the dynamic properties of retinal projections is to describe them in terms of a pattern of moving features, often called *optic flow* (Gibson, 1979). Mathematical analyses of optic flow have shown that, if appropriate assumptions are introduced in the interpretation process, then *veridical* three-dimensional object shape can be derived from optic flow. If rigid motion is assumed, for example, then three orthographic projections

of four moving points are sufficient to derive their three-dimensional metric structure (Ullman, 1979). It is important to distinguish between the *first-order* temporal properties of optic flow (*velocities*), which are produced by two projections of a moving object, and the *second-order* temporal properties of the optic flow (*accelerations*), which require three projections of a moving object. Although the first-order temporal properties of optic flow are sufficient for the recovery of affine properties (Koenderink & van Doorn, 1991; Todd & Bressan, 1990), the second-order temporal properties of optic flow are necessary for a full recovery of the three-dimensional metric structure (D. D. Hoffman, 1982).

Psychophysical investigations: A large number of empirical investigations have tried to determine whether observers actually use the second-order properties of optic flow that are needed to reconstruct the veridical three-dimensional metric shape of projected objects. The majority of these studies have come to the conclusion that, in deriving three-dimensional shape from motion, observers seem to use only the first-order properties of optic flow (e.g., Todd & Bressan, 1990). This conclusion is warranted, in particular, by two findings: (a) the metric properties of SFM are often misperceived (e.g., Domini & Braunstein, 1998; Norman & Todd, 1992), and (b) human performance in SFM tasks does not improve as the number of views is increased from two to many (e.g., Todd & Bressan, 1990).

Modeling: An interesting result of the SFM literature is that observers typically perceive a unique metric interpretation when viewing an ambiguous two-view SFM sequence (with little inter- and intraobserver variability), even though the sequence could have been produced by the orthographic projection of an infinite number of different three-dimensional rigid structures (Domini, Caudek, & Proffitt, 1997). The desire to understand how people derive such perceptions has led researchers to study the relationships between the few parameters that characterize the first-order linear velocity field and the properties of the perceived three-dimensional shapes. Several studies have concluded that the best predictor of perceived three-dimensional shape from motion is one component of the local (linear) velocity field, called deformation (*def*; see Koenderink, 1986). Domini and Caudek (1999) have proposed a probabilistic model whereby, under certain assumptions, a unique surface orientation can be derived from an ambiguous first-order velocity field according to a maximum likelihood criterion. Results consistent with this model have been provided relative to the perception of surface slant (Domini & Caudek, 1999), the discrimination between rigid and nonrigid motion (Domini et al., 1997), the perceived orientation of the axis of rotation (Caudek & Domini, 1998), the discrimination between

constant and variable three-dimensional angular velocities (Domini, Caudek, Turner, & Favretto, 1998), and the perception of depth-order relations (Domini & Braunstein, 1998; Domini, Caudek, & Richman, 1998).

In summary, the research on perceived depth from motion reveals that the perceptual analysis of a moving projection is relatively insensitive to the second-order component of the velocity field (accelerations), which is necessary to uniquely derive the metric structure in the case of orthographic projections. Perceptual performance has been explained by two hypotheses. Some researchers maintain that the perceptual recovery of the metric structure from SFM displays is consistent with a heuristical analysis of optic flow (Braunstein, 1976, 1994; Domini & Caudek, 1999; Domini et al., 1997). Other researchers maintain that the perception of three-dimensional shape from motion involves a hierarchy of different perceptual representations, including the knowledge of the object's topological, ordinal, and affine properties, whereas the Euclidean metric properties may derive from processes that are more cognitive than perceptual (Norman & Todd, 1992).

Integration of Depth Cues: How Is the Effective Information Combined?

A pervasive finding is that the accuracy of depth and distance perception increases as more and more sources of depth information are present within a visual scene (Künnapas, 1968). It is also widely believed that the visual system functions normally, so to speak, only within a rich visual environment in which the three-dimensional shape of objects and spatial layout are specified by multiple informational sources (Gibson, 1979). Understanding how the visual system integrates the information provided by several depth cues represents, therefore, one of the fundamental issues of depth perception.

The most comprehensive model of depth-cue combination that has been proposed is the *modified weak fusion* (MWF) model (Landy, Maloney, Johnston, & Young, 1995). *Weak fusion* refers to the independent processing of each depth cue by a modular system that then linearly combines the depth estimates provided by each module (Clark & Yuille, 1990). *Strong fusion* refers to a nonmodular depth processing system in which the most probable three-dimensional interpretation is provided for a scene without the necessity of combining the outputs of different depth-processing modules (Nakayama & Shimojo, 1992). Between these two extremes, Landy et al. proposed a modular system made up of depth modules that interact solely to facilitate *cue promotion*. As seen previously, visual cues provide qualitatively different types of information. For example, motion parallax can in principle provide

absolute depth information, whereas stereopsis provides only relative-depth information, and occlusion specifies a greater depth on one side of the occlusion boundary than on the other, without allowing any quantification of this (relative) difference. The depth estimates provided by these three cues are incommensurate, and therefore cannot be combined. According to Landy et al., combining information from different cues necessitates that all cues be made to provide absolute depth estimates. To achieve this task, some depth cues must be supplied with one or more missing parameters. If motion parallax and stereoscopic disparity are available in the same location, for example, then the viewing distance specified by motion parallax could be used to specify this missing parameter in stereo disparity. After stereo disparity has been promoted so as to specify metric depth information, then the depth estimates of both cues can be combined. In conclusion, for the MWF model, interactions among depth cues are limited to what is required to place all of the cues in a common format required for integration.

In the MWF model, after the cues are promoted to the status of absolute depth cues, it becomes necessary to establish the *reliability* of each cue: "Side information which is not necessarily relevant to the actual estimation of depth, termed an *ancillary measure*, is used to estimate or constrain the reliability of a depth cue" (Landy et al., 1995, p. 398). For example, the presence of noise differentially degrading two cues present in the same location can be used to estimate their different reliability.

The final stage of cue combination is that of a weighted average of the depth estimates provided by the cues. The weights take into consideration both the reliability of the cues and the discrepancies between the depth estimates. If the cues provide consistent and reliable estimates, then their depth values are linearly combined. On the other hand, if the discrepancy between the individual depth estimates is greater than what is found in a natural scene, then complex interactions are expected.

Cutting and Vishton (1995) proposed an alternative approach. According to their proposal, the three-dimensional information specified by all visual cues is converted into an ordinal representation. The information provided by the different sources is combined at this level. After the ordinal representation has been generated, a metric scaling can then be created from the ordinal relations.

The issue of which cue-combination model best fits the psychophysical data has been much debated. Other models of cue combination, in fact, have been proposed, either linear (Bruno & Cutting, 1988) or multiplicative (Massaro, 1988), with no single model being able to fully account for the large number of empirical findings on cue integration.

A similar lack of agreement in the literature concerns two equally fundamental and related questions: How can we describe the mapping between the physical and the perceived space? What geometric properties comprise perceived space?

Several answers have been provided to these questions. According to Todd and Bressan (1990), physical and perceived spaces may be related by an affine transformation. Affine transformations preserve distance ratios in all directions, but alter the relative lengths and angles of line segments oriented in different directions. A consequence of such a position is that a depth map may not provide the common initial representational format for all sources of three-dimensional information, as was proposed by Landy et al. (1995).

The problem of how to describe the properties of perceived space has engendered many discussions and is far from being solved. According to some, the intrinsic structure of perceptual space may be Euclidean, whereas the mapping between physical and perceptual space may not be Euclidean (Domini et al., 1997). According to others, visual space may be hyperbolic (Lunenburg, 1947), or it may reflect a Lie algebra group (W. C. Hoffman, 1966). Some have proposed the coexistence of multiple representations of perceived three-dimensional shape, reflecting different ways of combining the different visual cues (Tittle & Perotti, 1997).

A final fundamental question about visual-information integration is whether the cue-combination strategies can be modified by learning or feedback. Some light has recently been shed on this issue by showing that observers can modify their cue-combination strategies through learning, and can apply each cue-combination strategy in the appropriate context (Ernst, Banks, & Bühlhoff, 2000).

In conclusion, an apt summarization of this literature was provided by Young, Landy, and Maloney (1993), who stated that a description of the depth cue-combination rules "seems likely to resemble a microcosm of cognitive processing: elements of memory, learning, reasoning and heuristic strategy may dominate" (p. 2695).

Distance Perception

Turning our attention from *how* spatial perception is achieved to *what* is perceived, we are struck by the consistent findings of distortions of both perceived distance and object shape, even under full-cue conditions. For example, Norman, Todd, Perotti, and Tittle (1996) asked observers to judge the three-dimensional lengths of real-world objects viewed in near space and found that perceived depth intervals become more and more compressed as viewing distance increased. Given that many reports have found visual space to be distorted, the

question arises as to why we do not walk into obstacles and misguide our reaching. Clearly, our everyday interactions with the environment are not especially error-prone. What, then, is the meaning of the repeated psychophysical findings of failures of distance perception?

We can try to provide an answer to this question by considering four aspects of distance perception. We examine (a) the segmentation of visual space, (b) the methodological issues in distance perception research, (c) the underlying mechanisms that are held responsible for distance perception processing, and (d) the role of calibration.

Four Aspects of Distance Perception

The Segmentation of Visual Space. Cutting and Vishton (1995) distinguished three circular regions surrounding the observer, and proposed that different sources of information are used within each of these regions. *Personal space* is defined as the zone within 2 m surrounding the observer's head. Within this space, distance perception is supported by occlusion, retinal disparity, relative size, convergence, and accommodation. Just beyond personal space is the *action space* of the individual. Within the action space, distance perception is supported by occlusion, height in the visual field, binocular disparity, motion perspective, and relative size. Action space extends to the limit of where disparity and motion can provide effective information about distance (at about 30 m from the observer). Beyond this range is *vista space*, which is supported only by the pictorial cues: occlusion, height in the visual field, relative size, and aerial perspective.

Cutting and Vishton (1995) proposed that different sources of information are used within each of these visual regions, and that a different ranking of importance of the sources of information may exist within personal, action, and vista space. If this is true, then the intrinsic geometric properties of these three regions of visual space may also differ.

Methodological Issues. The problem of distance perception has been studied by collecting a number of different response measures, including verbal judgments (Pagano & Bingham, 1998), visual matching (Norman et al., 1996), pointing (Foley, 1985), targeted walking with and without vision (Loomis, Da Silva, Fujita, & Fukusima, 1992), pointing triangulation (Loomis et al., 1992), and reaching (Bingham, Zaal, Robin, & Shull, 2000). Different results have been obtained by using different response measures. Investigations of targeted walking in the absence of vision, for example, have produced accurate distance estimates (Loomis et al., 1992), although verbal judgments have not (Pagano & Bingham, 1998). Reaching has been found to be accurate when dynamic

binocular information was available, but not when it is guided by monocular vision, even in the presence of feedback (Bingham & Pagano, 1998). Matching responses for distance perception have typically shown that simulated distances are underestimated (Norman et al., 1996).

The fact that such a variety of results have been obtained by using different response measures suggests at least two conclusions. First, different response measures may be the expression of different representations of visual space that need not be consistent. This would mean that visual space could not be conceptualized as a unitary, internally consistent construct. In particular, motor responses and visual judgments may be informed by different visual representations. Second, there are reports indicating that accurate distance perception can be obtained, but only if the appropriate response measure is used in conjunction with the appropriate visual input. By using a reaching task, for example, Bingham and Pagano (1998) have recently reported accurate perception of distance with dynamic binocular vision (at least within what Cutting and Vishton (1995) call "personal space"), but not with monocular vision.

Conscious Representation of Distance Versus Action.

Some evidence suggests that different mechanisms may mediate conscious distance perception and actions such as reaching, grasping, or ballistic targeting. This leads to the hypothesis that separate visual pathways exist for perception and action. Milner and Goodale (1995) proposed that two different pathways, each specialized for different visual functions, exist in the visual system. The projection to the temporal lobe (the ventral stream) "permit[s] the formation of perceptual and cognitive representations which embody the enduring characteristics of objects and their significance," whereas the parietal cortex (the dorsal stream) "capture[s] instead the instantaneous and egocentric features of objects, [and] mediate[s] the control of goal-directed actions" (Milner & Goodale, 1995, p. 66). In Milner and Goodale's proposal, the coding that mediates the required transformations for the visual control of skilled actions is assumed to be separate from that mediating experiential perception of the visual world. According to this proposal, the many dissociations that have been discovered between conscious distance perception and locomotion "highlight surprising instances where what we think we 'see' is not what guides our actions. In all cases, these apparent paradoxes provide direct evidence for the operation of visual processing systems of which we are unaware, but which can control our behavior" (p. 177).

The dissociations most relevant for the present discussion are examined in the later section on egocentric versus exocentric distance. In general, this research suggests that accuracy

measured in action performance is usually greater than that found in visual judgment measures. It must be noticed, however, that action measures do not always produce accurate or distortion-free results (Bingham & Pagano, 1998; Bingham et al., 2000).

The Role of Calibration. It has been suggested that some response measures (such as visual matching) produce poor performance in distance perception because the task is unnatural, infrequently performed, and thus poorly calibrated. Bingham and Pagano (1998) suggest that in these circumstances, only relative-depth measures can be obtained. Conversely, absolute-distance perception may be obtained (within some tolerance limits) by using feedback to calibrate ordinarily scaled distance estimates. Evidence that haptic feedback reduces the distortions of egocentric distance has indeed been provided (Bingham & Pagano, 1998; Bingham et al., 2000), although feedback does not always reduce spatial distortions, and never does so completely.

Egocentric Versus Exocentric Distance. From the point of view of the observer, the horizontal plane extends in all directions from his or her current position. The observer's body is the center of egocentric space. Objects placed on the rays that intersect this center have an egocentric distance relative to the observer, whereas objects on different rays have an exocentric distance relative to each other.

It has been repeatedly found that even in full-cue viewing conditions, distances are perceptually compressed in the egocentric direction relative to exocentric extents, with most comparisons being between egocentric distances and those in the fronto-parallel plane (Amorim, Loomis, & Fukusima, 1998; Loomis et al., 1992; Norman et al., 1996).

In a typical experiment, observers were instructed to match an egocentric extent to one in the fronto-parallel plane (Loomis et al., 1992). Observers made the extents in the egocentric direction greater in order to perceptually match them to those in the fronto-parallel direction. This inequality in perceived distance defines a basic anisotropy in the perception of space and objects. The perceptions of extents and object shapes are compressed in their depth-to-width ratio. Loomis and Philbeck (1999) showed that this anisotropy in perceived three-dimensional shape is invariant across size scales.

Paradoxically, the compression in exocentric distance does not seem to affect visually guided actions. One method to assess the accuracy of visually guided behavior is to have observers view a target in depth, and then after blindfolding the observer, have them walk to the target location. This technique is called *blindwalking*. Numerous studies have found that people are able to walk to targets without showing

any systematic error (overshoot or undershoot) when viewing occurs in full-cue conditions and distances fall within about 20m (Loomis et al., 1992; Philbeck & Loomis, 1997; Philbeck, Loomis, & Beall, 1997; Rieser, Ashmead, Talor, & Youngquist, 1990; Thomson, 1983).

Two alternative explanations could account for this discrepancy between the compression found in explicit reports of perceived distance and the accuracy observed in blindwalking to targets. By one account, the difference is attributable to two distinct streams of visual processing in the brain. From the visual cortex, there is a bifurcation in the primary cortical visual pathways, with one stream projecting to the temporal lobe and the other to the parietal. Milner and Goodale (1995) have characterized the functioning of the temporal stream as supporting conscious space perception, whereas the parietal stream is responsible for the visual guidance of action. Other researchers have characterized the functioning of these two streams somewhat differently; for a review, see Creem and Proffitt (2001). By a two-visual-systems account, the compression observed in explicit judgments of egocentric distance is a product of temporal stream processing, whereas accurate blindwalking is an achievement of the parietal stream. The second explanation posits that both behaviors are grounded in the same representation of perceived space; however, differences in the transformations that relate perceived space to distinct behaviors cause the difference in accuracy. Philbeck and Loomis (1997) showed that verbal distance judgments and blindwalking varied together under manipulations of full- and reduced-cue viewing conditions in a manner that is strongly suggestive of this second alternative.

The dissociation between verbal reports and visually guided actions depends upon whether distances are encoded in exocentric or egocentric frames of reference. Wraga, Creem, and Proffitt (2000) created large Müller-Lyer configurations that lay on a floor. Observers made verbal judgments and also blindwalked the perceived extent. It was found that the illusion influenced both verbal reports and blindwalking when the configurations were viewed from a short distance; however, the illusion only affected verbal judgments when observers stood at one end of the configurations' extent. As was suggested by Milner and Goodale (1995), accuracy in visually guided actions may depend upon the egocentric encoding of space.

Size Perception

The size of an object does not appear to change with varying viewing distance, despite the fact that retinal size depends on the distance between the object and the observer. Researchers

have proposed several possible explanations for this phenomenon: the size-distance invariance hypothesis, the familiar-size hypothesis, and the direct-perception approach.

Taking Distance Into Account

The *size-distance invariance hypothesis* postulates that retinal size is transformed into perceived size after taking apparent distance into account (see Epstein, 1977, for review). If accurate distance information is available, then size perception will also be accurate. However, if distance information were unavailable, then perceived size would be determined by visual angle alone. According to the size-distance invariance hypothesis (Kilpatrick & Ittelson, 1953), perceived size (S') and perceived distance (D') stand in a unique ratio determined by some function of the visual angle (θ):

$$\frac{S'}{D'} = f(\theta)$$

The size-distance invariance hypothesis has taken slightly different forms, depending on how the function f in the previous equation has been specified. If the size-distance invariance hypothesis is simply interpreted as a geometric relation, then $f(\theta) = \tan \theta$ (Baird & Wagner, 1991). For small visual angles, $\tan \theta$ approximates θ , and so $S'/D' = \theta$. According to a psychophysical interpretation, the ratio of perceived size to perceived distance varies as a power function of visual angle: $f(\theta) = k\theta^n$, where k and n are constants. Foley (1968) and Oyama (1974) reported that the exponent of this function is approximately 1.45. According to a third interpretation, visual angle θ is replaced by perceived visual angle θ' (McCready, 1985). If θ' is a linear function of θ , then $\tan \theta' = \tan(a + b\theta)$, where a and b are constants.

The adequacy of the size-distance invariance hypothesis has often been questioned, given that the empirically determined relation between S' and D' for a given visual angle is sometimes opposite from that predicted by this hypothesis (Sedgwick, 1986). For example, observers often report that the moon at the horizon appears to be both larger and closer to the viewer than the moon at the zenith appears (Hershenson, 1989). This discrepancy has been called the *size-distance paradox*.

Familiar Size

According to the familiar-size hypothesis, the problem of size perception is reduced to a matter of object recognition for those objects that have stable and known sizes (Hershenson & Samuels, 1999). Given that familiar size does not assume information about distance, it can itself be considered a source

of that information: Familiar size may determine perceived size, which, in agreement with the size-distance invariance hypothesis, may determine perceived distance (Ittelson, 1960).

The hypothesis that familiar size may determine apparent size (or apparent distance) has been denied by the *theory of off-sized perceptions* (Gogel & Da Silva, 1987). According to this theory, the distance of a familiar object under impoverished viewing conditions is determined by the egocentric reference distance, which in turn determines the size of the object according to the size-distance invariance hypothesis.

Direct Perception

Epstein (1982) summarizes the main ideas of Gibson's (1979) account of size perception as follows: "(a) there is no perceptual representation of size correlated with the retinal size of the object; (b) perceived size and perceived distance are independent direct functions of information in stimulation; (c) perceived size and perceived distance are not causally linked, nor is the perception of size mediated by operations combining information about retinal size and perceived distance. The correlation between perceived size and perceived distance is attributed to the correlation between the specific variables of stimulation which govern these percepts in the particular situation" (p. 78).

Surprisingly little research on size perception has been conducted within the direct-perception perspective. However, the available empirical evidence, consistent with this perspective, suggests that perceived size might be affected by the ground texture gradients (Bingham, 1993; Sinai, Ooi, & He, 1998) and the horizon ratio (Carello, Groszofsky, Reichel, Solomon, & Turvey, 1989).

Size in Pictures Versus Size in the World

It is sometimes assumed that illusions, especially geometrical ones, are artifacts of picture perception. This assumption is false, at least with respect to the anisotropy between vertical and horizontal extents. The perceptual bias to see vertical extents as greater than equivalent horizontal extents is even greater when viewing large objects in the world than when viewing pictures (Chapanis & Mankin, 1967; Higashiyama, 1996; Yang, Dixon, & Proffitt, 1999).

Yang et al. (1999) compared the magnitude of the vertical-horizontal illusion for scenes viewed in pictures, in the real world, and in virtual environments viewed in a head-mounted display. They found that pictures evoke a bias averaging about 6%, whereas viewing the same scenes in real or virtual

worlds results in an overestimation of the vertical averaging about 12%.

Geographical Slant Perception

The perceptual bias to overestimate vertical extents pales in comparison to that seen in the perception of geographical slant. Perceived topography grossly exaggerates the geometrical properties of the world in which we live (Bhalla & Proffitt, 1999; Kammann, 1967; Ross, 1974; Proffitt, Bhalla, Gossweiler, & Midgett, 1995).

Proffitt et al. (1995; Bhalla & Proffitt, 1999) obtained two measures of explicit geographical slant perception from observers who stood at either the tops or bottoms of hills in either real or virtual worlds. Observers provided verbal judgments and also adjusted the size of a pie-shape segment of a disk so as to make it correspond to the cross-section of the hills they were observing. These two measures were nearly equivalent and revealed huge overestimations. Five-degree hills, for example, were judged to be about 20°, and 10° hills were judged to be about 30°. These huge overestimations seem odd for at least two reasons. First, people know what angles look like. Proffitt et al. (1995) asked participants to set cross-sections with the disk to a variety of angles and found that they were quite accurate in doing so. Second, when viewing a hill in cross-section, the disk could be accurately adjusted by lining up the two edges of the pie section to lines in the visual scene; however, the overestimations found for people viewing hills in cross-section do not differ from judgments taken when the hills are viewed head-on (Proffitt, Creem, & Zosh, 2001).

Proffitt and colleagues have argued that the conscious overestimation of slant is adaptive and reflective of psychophysical response compression (Bhalla & Proffitt, 1999; Proffitt et al., 2001; Proffitt et al., 1995). Psychophysical response compression means that participants' response sensitivities decline with increases in the magnitude of the stimulus. Expressed as a power function, the exponent is less than 1. Response compression promotes sensitivity to small changes in slant within the range of small slants that are of behavioral relevance for people. Overestimation necessarily results from a response compression function that is anchored at 0° and 90°. People are accurate at 0°—they can tell whether the ground is going up or down—and for similar reasons of discontinuity, they are also accurate at 90°.

All of Proffitt's and his colleagues' studies utilized a third dependent measure, which was a visually guided action in which participants adjusted the incline of a tilt board without looking at their hands. These action-based estimates of slant were far more accurate than were the explicit judgments.

It was proposed that the dissociation seen between explicit perceptions and visually guided actions were a symptom of the two visual streams of processing. The basis for this argument was a dissociation found in the influence of physiological state on the explicit versus motoric dependent measures.

Proffitt et al. (1995) found that, as assessed by verbal reports and the visual matching task employing the pie-shaped disk, hills appear steeper when people are fatigued. Bhalla and Proffitt (1999) replicated this finding; in addition, he found that hills appear steeper when people are encumbered by a heavy backpack, have low physical fitness, are elderly, or have failing health. None of these factors influenced the visually guided action of adjusting the tilt board. Thus, in the case of geographical slant, explicit judgments were influenced by the manipulation of physiological state, although the visually guided action was unaffected. This finding differs from Philbeck and Loomis's (1997) results showing that verbal reports of distance and blindwalking changed together with manipulations of reduced-cue environments.

EVENT PERCEPTION

Events are occurrences that unfold over time. We have already seen that the visual perception of space and three-dimensional form is derived from motion-carried information. The structure perceived in these cases is, of course, recovered from events. The literature in the field of visual perception, however, has partitioned the varieties of motion-carried information into distinct fields of inquiry. For example, perceiving spatial layout and three-dimensional form from motion parallax and object rotations falls under the topic of structure-from-motion. The field of event perception has historically dealt with two issues, perceptual organization (see chapter by Palmer in this volume) and perceiving dynamics.

Perceptual Organization

The law of common fate is among the Gestalt principles of perceptual organization that were proposed by Wertheimer (1923/1937). In essence, this law states that elements that move together tend to be perceived as belonging to the same group. This notion of grouping by common fate is at the heart of the event perception literature that developed as an answer to this question: *What are the perceptual rules that define what it means for elements to move together?*

In response to this question, three classes of events have received the most attention: (a) surface segregation from motion, (b) hierarchical motion organization, and (c) biological motion.

Surface Segregation From Motion

We have already discussed how dynamic occlusion specifies depth order within the section on motion-based depth cues. With respect to issues of perceptual organization, however, there are additional matters to relate to this event.

In a typical dynamic occlusion display, one randomly textured surface is placed on top of another (Gibson, Kaplan, Reynolds, & Wheeler, 1969). When the surfaces are stationary, there is no basis for seeing that two separate surfaces are, in fact, present. The instant that one of the surfaces moves, however, the distinct surfaces become apparent. The dynamic occlusion that occurs under these conditions provides an invariant related to depth order. The surface that exhibits an accretion and deletion of optical texture is behind the one that does not. Yonas, Craton, and Thompson (1987) used sparse point-light displays to show that perceptual surface segregation occurs even when optical texture elements do not actually overlap in the display. In their display, the point lights were turned on and off in a manner consistent with their being occluded by a virtual surface carrying other point lights. The surface segregation that is perceived in the presence of dynamic occlusion is different from the figure-ground segregation that is perceived in static pictures due to the lack of ambiguity of edge assignment and depth order. In dynamic occlusion displays, the edge is seen to belong to the surface that does not undergo occlusion; moreover, this surface is unambiguously perceived to be closer.

Hierarchical Motion Organization

Suppose that you observe someone bouncing a ball; the ball is seen to move up and down. Suppose next that the person who is bouncing the ball is standing on a moving wagon. In this latter case, you will likely still see the ball moving up and down and at the same time moving with the wagon. The wagon's motion has become a perceptual frame of reference for seeing the motion of the bouncing ball. This is an example of hierarchical motion organization in which the ball's motion has been perceptually decomposed into two components, that which it shares with the wagon and that which it does not.

Rubin (1927) and Duncker (1929/1937) provided early demonstrations of the perceptual system's proclivity to produce hierarchically organized motions; however, Johansson (1950) brought the field into maturity. Johansson (1950, 1973) provided a vector analysis description of the perceptual decomposition of complex motions. In the case of the ball's being bounced on a moving wagon, the motion of the ball relative to a stationary environmental reference frame is its *absolute motion*, whereas the motion shared by the ball and the

wagon is the *common motion* component, and the unique motion of the ball with respect to the wagon is the *relative motion* component. It is generally assumed that relative and common motion components sum to equal the perceptually registered absolute motion; however, this assumption has rarely been tested and, in at least one circumstance, has been shown not to hold (Vicario & Bressan, 1990).

Johansson frequently used point-light displays to demonstrate the nature of hierarchical motion organization. In a case analogous to the bouncing ball example, Johansson created a display in which a single point light moved along a diagonal path, the horizontal component of which was identical to that of a set of flanking point lights moving above and below it. Observers tend to see the single point light moving vertically relative to the flankers and the whole array of point lights moving with a common back and forth motion. Here, two sorts of perceptual organization are apparent. First, perceptual grouping is achieved on the basis of common fate; second, the common motion is serving as a moving frame of reference for the perception of the event's relative motion.

Another event that has received considerable attention is the perception of a few point lights moving as if attached to a rolling wheel. Duncker (1929/1937) showed that the motions perceived in this event depend upon where the configuration of point lights is placed on the wheel. If two point lights are placed on the rim of an unseen wheel, 180° apart, then the points will appear to move in a circle (relative motion) and at the same time translate horizontally (common motion). On the other hand, if the points are placed 90° apart, so that the center of the configuration does not coincide with the center of the wheel, then the points will appear to tumble. The reason for this difference is that the perceptual system derives relative motions that occur around the center of the configuration, with the observed common motion being of the configurational centroid (Borjesson & von Hofsten, 1975; Proffitt, Cutting, & Stier, 1979). When the configurational centroid coincides with the hub of the wheel, smooth horizontal translation is seen as the common motion. If it does not coincide, then the common motion will follow a more complex wavy path (a prolate cycloid).

Attempts have been made to provide a general processing model for perceiving hierarchical motion organization; however, little success has been achieved. Johansson's (1973) perceptual vector analysis model describes well what is seen in events, but does not derive these descriptions in a principled way. Restle (1979) applied Leeuwenberg's (1971, 1978) perceptual coding theory to many of Johansson's displays. By this account, the resulting perception is a consequence of a minimization of the number of parameters required to describe the event. The analysis worked well; however, it

evaluates descriptions after some other process has derived them—thus it fails to account for the process that produces the set of initial descriptions. Borjesson and von Hofsten (1975) and Cutting and Proffitt (1982) proposed models in which the perceptual system minimized relative motions. These accounts work well for wheel-generated motions, but they are not sufficiently general to account for the varieties of hierarchical motion organizations.

Not all motions are equally able to serve as perceptual frames of reference. In the case of the ball bouncing on a wagon, the common motion is a translation. Bertamini and Proffitt (2000) compared events in which the common motion was a translation, a rotation, or a divergence-convergence (radial expansion or contraction). They found that common translations and divergence-convergence evoked hierarchical motion organizations, but that rotations typically did not. In the cases in which rotations did serve as perceptual reference frames, it was found that they did so because there were also structural invariants present in the displays. A structural invariant is a spatial property that, in these cases, was revealed in motion. For example, if one point orbited around another at a constant distance, then this spatial invariant would promote a perceptual grouping. Given sufficiently strong spatial groupings, a rotational common motion can be seen as a hierarchical reference frame for the extraction of relative motions. One case in which this occurs is the class of events called biological motions.

Biological Motion

Few displays have captured the imagination of perceptual scientists as strongly as the point-light walker displays first introduced to the field by Johansson (1973; Mass, Johansson, Janson, & Runeson, 1971). These displays consist of spots of light attached to the joints of an actor, who is then filmed in the dark. When the actor is stationary, the lights appear as a disorganized array. As soon as the actor moves, however, observers recognize the human form as well as the actor's actions. In the Mass et al. movie, the actors walked, climbed steps, did push-ups, and danced to a Swedish folk song. The displays are delightful, as they demonstrate the amazing organizing power of the perceptual system under conditions of seemingly minimal information.

Following Johansson's lead, others showed that observers could recognize their friends in such point-light displays (Cutting & Kozlowski, 1977), and moreover that gender was reliably evident (Barclay, Cutting, & Kozlowski, 1978). Runeson and Frykholm (1981) filmed point-light actors lifting different weights in a box and found that observers could reliably judge the amount of weight being lifted. These are amazing perceptual feats. Upon what perceptual processes are they based?

When a point-light walker is observed, hierarchical nestings of pendular motions are observed within the figure, along with a translational common motion. Models have been proposed for extracting connectivity in these events; however, they are not sufficiently general to succeed in recovering structure in cases in which the actor rotates, as in the Mass et al. dancing sequence (D. D. Hoffman & Flinchbaugh, 1982; Webb & Aggarwal, 1982). These models look for rigid relationships in the motions between points that are not specific to biological motions. It may be that the perceptual system is attuned to other factors specific to biomechanics, such as necessary constraints on the phase relationships among the limbs (Bertenthal & Pinto, 1993; Bingham, Schmidt, & Zaal, 1999). It may also be the case that point-light walker displays are organized, at least in part, by making contact with representations in long-term memory that have been established through experiences of watching people walk. Consistent with this notion is the finding that when presented upside down, these displays are rarely identified as people (Sumi, 1984). Cats have also been found to discriminate point-light cats from foils containing the same motions in scrambled locations when the displays were viewed upright, but not when displays were upside down (R. Blake, 1993). Additional support for the proposal that experience plays a role is seen in the findings on infant sensitivities to biological motions. It has been found that infants as young as 3 months of age can extract some structure from point-light walker displays (Bertenthal, Proffitt, & Cutting, 1984; Fox & McDaniels, 1982). However, it is not until they are 7 to 9 months old that they show evidence of identifying the human form. At earlier ages, they seem to be sensitive to local structural invariants.

Perceiving Dynamics

Suppose that you are taking a walk and notice a brick lying on the ground. Now, suppose that as you approach the brick, a small gust of wind blows it into the air. You will, of course, be surprised and this surprise is a symptom of your violated expectation that the brick was much too heavy to be moved by the wind. Certainly, people form dynamical intuitions all of the time about quantities such as mass. A question that has stimulated considerable research is whether the formation of dynamical intuitions is achieved by thought or by perception.

Hume (1739/1978) argued that perception could not supply sufficient and necessary information to specify the underlying causal necessity of events. Hume wrote that we see forms and motions interacting in time, not the dynamic laws that dictate the regularities that are observed in their motions. Challenging Hume's thesis, Michotte (1963) demonstrated

that people do, in fact, form spontaneous impressions about causality when viewing events that have a collision-like structure. Michotte's studies created considerable interest in the question of how much dynamic information could be perceived in events.

In this spirit, Bingham, Schmidt, and Rosenblum (1995) showed participants patch-light displays of simple events such as a rolling ball, stirring water, and a falling leaf. Participants were found to classify these events based upon similarities in their dynamics.

Runeson (1977/1983) showed that there were regularities in events that could support the visual perception of dynamical quantities. So, for example, when two balls of different mass collide, a ratio can be formed between the differences in their pre- and postcollision velocities that specifies their relative masses. Moreover, it has been shown that, when people view collision events, they can make relative mass judgments (Gilden & Proffitt, 1989; Runeson, Juslin, & Olsson, 2000; Todd & Warren, 1982). Contention exists, however, on the accuracy of these judgments and on what underlying processes are responsible for this ability. Following Todd and Warren (1982), Gilden and Proffitt (1989) provided evidence that judgments were based upon heuristics, such as *the ball that ricochets is lighter* or *the ball with greatest postcollision speed is lighter*. Like all heuristics, these judgments reflect accuracy in some contexts but not in others. On the other hand, Runeson et al. (2000) showed that people could make accurate judgments, provided that these individuals are given considerable practice with feedback.

People also perceive mass when viewing displays of point-light actors lifting weights (Bingham, 1987; Runeson & Frykholm, 1983). It has yet to be determined how accurate people are in their judgments and upon what perceptual or cognitive processes these judgments are based.

In addition to noticing dynamical quantities when viewing events, people also seem more attuned to what is dynamically appropriate when they view moving (as opposed to static) displays. For example, many studies have used static depictions to look at people's apparent inability to accurately predict trajectories in simple dynamical systems (Kaiser, Jonides, & Alexander, 1986; McCloskey, 1983; McCloskey, Caramazza, & Green, 1980). For example, when asked to predict the path of an object dropped by a moving carrier—participants were shown a picture of an airplane carrying a bomb—many people predict that it will fall straight down (McCloskey, Washburn, & Felch, 1983). However, when shown a computer animation of the erroneous trajectory that they would predict, these people viewed these paths as being anomalous and choose as correct the dynamically correct event (Kaiser, Proffitt, Whelan, & Hecht, 1992).

The ability to discriminate between dynamically correct and anomalous events is not without limit. People's ability to perceptually penetrate rotational events has been found to be severely limited. Kaiser et al. (1992) showed people computer animations of a satellite spinning in space. The satellite would open or close its solar panels; as it did so, its spinning rate would increase, decrease, or reverse direction. Rotation rate should increase as the panels contracted (and visa versa), as does a twirling ice-skater who extends or contracts his or her arms. People showed virtually no perceptual appreciation for the dynamical appropriateness of the satellite simulations. Other than when the satellite actually changed its rotational direction, the dynamically anomalous and canonical events were all judged to be equally possible. Another event that does not improve with animation is performance on the water-level task. In paper-and-pencil tests, it has been found that about 40% of adults do not draw horizontal lines when asked to indicate the surface orientation of water in a stationary tilted container (McAfee & Proffitt, 1991). Animating this event does not improve performance (Howard, 1978).

Clearly, a theory of dynamical event perception needs to specify not only what people can do, but also what they cannot. Attempts to account for the limits of our dynamical perceptions have focused on perceptual biases and on event complexity. With respect to the former, perceptual frame-of-reference biases have been used to explain certain biases in people's dynamical judgments (Hecht & Proffitt, 1995; Kaiser et al., 1992; McAfee & Proffitt, 1991; McCloskey et al., 1983). As a first approximation toward defining dynamical event complexity, Proffitt and Gilden (1989) made a distinction between particle (easy) and extended body (hard) motions. *Particle motions* are those that can be described adequately by treating the object as if it were a point particle located at the object's center of mass. Free-falling is a particle motion if air resistance is ignored. *Extended body motions* make relevant other object properties such as shape and rotations. A spinning top is an example of an extended body motion. The apparent gravity-defying behavior of a spinning top gives evidence to our inability to see the dynamical constraints that cause it to move as it does. Tops are enduring toys because their dynamics cannot be penetrated by perception.

Perceiving Our Own Motion

In this section, we consider the perception of our own motion by examining three problems: how we perceive our direction of motion (*heading*); the illusion of self-motion experienced by stationary individuals when viewing moving visual surrounds (*vection*); and the visual control of posture.

Heading

In studying how the direction of self-motion (or *heading*) is recovered, researchers have focused on the use of relevant visual information. However, vestibular information (Berthoz, Istraël, George-François, Grasso, & Tsuzuku, 1995) and feedback from eye movements (Royden, Banks, & Crowell, 1992) may also play a role in this task. Gibson (1950) proposed that the primary basis for the visual control of locomotion is *optic flow*.

In general, instantaneous optic flow can be conceptualized as the sum of a translational and a rotational component (for a detailed discussion, see Hildreth & Royden, 1998; Warren, 1998). The translational component alone generates a radial pattern of velocity vectors emanating from a singularity in the velocity field called the *focus of expansion* (FOE). A pure translational flow is generated, for example, when an observer moves in a stationary environment while looking in the direction of motion. In these circumstances, the FOE specifies the direction of self-motion. In general, however, optic flow contains a rotational component as well, such as when an observer experiences pursuit eye movement when fixating on a point not in line with the motion direction. For a pure rotational flow, equivalent to a rigid rotation of the world about the eye, both the direction and magnitude of the velocity vectors are independent of the distance between the observer and the projected features. The rotational component, therefore, is informative neither about the structure of the environment, nor about the motion of the observer. The presence of the rotational flow, however, does complicate retinal flow. When both translational and rotational components are present, a singularity still exists in the flow field, but in this case it specifies the fixation point rather than the heading direction. Thus, "if observers simply relied on the singularity in the field to determine heading, they would see themselves as heading toward the fixation point" (Warren, 1995, p. 273).

Many theoretical analyses have demonstrated how the direction of heading could be recovered from the optic flow (e.g., Regan & Beverly, 1982). However, no agreement exists on whether, in a biologically plausible model of heading, the rotational component must first be subtracted from retinal flow in order to recover the FOE from the translational flow, or whether heading can be recovered without decomposing retinal flow into its two components. Most of the theoretical analyses of the compound velocity field have been developed for computer vision applications and have followed the first of these two routes (for a review, see Hildreth & Royden, 1998). The second approach has received less attention and has been advocated primarily by Cutting and collaborators (Cutting, Springer, Braren, & Johnson, 1992). Although

followers of this second proposal used more naturalistic settings, most studies on the perception of heading have used random-dot displays simulating the optical motion that would be produced by an observer moving relative to a ground plane, a three-dimensional cloud of dots, or one or more fronto-parallel surfaces at different depths.

Overall, empirical investigations on heading show that the human visual system can indeed recover the heading direction from velocity fields like those generated by the normal range of locomotion speeds. The psychophysical studies in particular have revealed the following about human perception of heading: It is remarkably robust in noisy flow fields (van den Berg, 1992); it is capable of making use of sparse clouds of motion features (Cutting et al., 1992) and of extraretinal information about eye rotation (Royden, Banks, & Crowell, 1992); and it improves in its performance when other three-dimensional cues are present in the scene (van den Berg & Brenner, 1994). Some of the proposed computational models embody certain of these features, but so far no model has been capable of mimicking the whole range of capabilities revealed by human observers.

Vection

Observers sometimes experience an illusory perception of self-motion while sitting in a stationary train and watching an adjacent train pulling out of the station. This *train illusion* is the best-known example of vection (Fisher & Kornmüller, 1930). Vection can be induced not only by visual, but also by auditory (Lackner, 1977), somatosensory (Lackner & DiZio, 1984), and combined somatosensory and kinesthetic (Bles, 1981) information. The first studies on visually induced vection can be dated back to Mach (1875) and were performed using a vertically striped optokinetic drum or an endless belt (Mach, 1922). Two kinds of vection can be distinguished: circular and linear. Circular vection typically refers to yaw motion about the vertical axis, whereas linear vection refers to translatory motion through a vertical or horizontal axis. Vection is called *saturated* when the inducing stimulus appears to be stationary and only self-motion is perceived (Wertheim, 1994).

Linear vection is typically induced about 1–2 s after the onset of stimulation (Giannopulu & Lepecq, 1998), circular vection after about 2–3 s, and saturated vection after about 10 s (Brandt, Dichgans, & Koenig, 1973). A more compelling vection is induced by faster speeds of translation or rotation (Larish & Flach, 1990), by low temporal frequencies (Berthoz, Lacour, Soechting, & Vidal, 1979), by more or larger elements (Brandt, Wist, & Dichgans, 1975); this is also the case when larger retinal areas are stimulated (Brandt et al., 1975) and when the inducing stimulus belongs to the background relative to the foreground (Nakamura & Shimojo, 1999).

Visual Control of Posture

Postural stability, or stance, is affected by visual, vestibular, and somatosensory information. Visual and somatosensory information are more effective in the low-frequency range of postural sway, whereas vestibular information is more effective in the high-frequency range (Howard, 1986). A device known as the *moving room* has been used to demonstrate that visual information can be used to control posture. In their original study, Lee and Aronson (1974) required infants to stand within a room in which the walls were detached from the floor and could slide back and forth. They reported that when the walls moved, infants swayed or staggered in spite of the fact that the floor remained stationary, a finding later replicated by many other studies (Bertenthal & Bai, 1989; for adult, see Lee & Lishman, 1975).

Two sources of visual information are available for postural control: the radial and lamellar motions of front and side surfaces, respectively, and the motion parallax between objects at different depths that is generated by the translation of the observer's head (Warren, 1995). Evidence has shown that posture is regulated by compensatory movements that tend to minimize both of these patterns of optical motion (Lee & Lishman, 1975; Warren, Kay, & Yilmaz, 1996). Three hypotheses have been proposed concerning the locus of retinal stimulation: (a) the *peripheral dominance hypothesis*, which states that the retinal periphery dominates both the perception of self-motion and the control of stance (Dichgans & Brandt, 1978); (b) the *retinal invariance hypothesis*, which states that self-motion and object motion are perceived independently of the part of the retina being stimulated (Crowell & Banks, 1993); and (c) the *functional sensitivity hypothesis*, which states that "central vision accurately extracts radial . . . and lamellar flow, whereas peripheral vision extracts lamellar flow but it is less sensitive to radial . . . flow" (Warren & Kurtz, 1992, p. 451). Empirical findings have contradicted the peripheral dominance hypothesis (Stoffregen, 1985) and the functional sensitivity hypothesis (Bardy, Warren, & Kay, 1999). Instead, they support the retinal invariance hypothesis by emphasizing the importance of the optic-flow structure for postural control, regardless of the locus of retinal stimulation.

Perceiving Approaching Objects

Time to Contact and Time to Passage

Coordinating actions within a dynamic environment often requires temporal information about events. For example, when catching a ball, we need to be able to initiate the grasp before the ball hits our hand. One might suspect that in order to plan for the ball's time of arrival, the perceptual system

would have to compute the ball's speed and distance; however, this turns out not to be the case. In determining time to contact, there exists an optical invariant, *tau*, which does not require that object speed or distance be taken into account.

First derived by Hoyle (1957) in his science-fiction novel, *The Black Cloud*, and later introduced to the vision community by Lee (1974), *tau* relates the optical size of an object to its rate of expansion in a manner that specifies time to contact. *Tau* is defined as follows:

$$\tau \approx \theta / \delta\theta / \delta t$$

where θ is the angular extent of the object in radians, and $\delta\theta/\delta t$ is the rate of its expansion. *Tau* specifies time to contact under the assumption that the object is moving with a constant velocity.

In the case in which the object and observer are not on a collision course, a similar relationship specifies time to passage (Lee, 1974, 1980). Let Φ be the angular extent between the observer's heading direction and the object, then time-to-passage, in terms of *tau*, is defined by

$$\tau \approx \Phi / \delta\Phi / \Phi t$$

Tresilian (1991) has refined the definition of *tau* by distinguishing between local and global *tau*. Local *tau* can be computed solely on the basis of information about angular extents and their rate of change. Global *tau* is only available to a moving observer and requires that the direction of heading be computed as well.

Research on *tau* has taken two forms. First, researchers have investigated whether people and animals are sensitive to time to contact and time to passage. Second, they have studied whether performance is actually based upon a perceptual derivation of *tau*. In summary, the literature suggests that time to contact and time to passage are accurately perceived; however, whether this perceptual feat is due to an appreciation of *tau* is currently a point of contention.

Optical expansion, or looming, evokes defensive postures in adults, animals (Schiff, 1965), and human infants (Bower, Broughton, & Moore, 1970). The assumption is that these defensive actions are motivated by a perception that the expanding object is on an imminent collision course. Although it is tempting to think of the link between looming and impending collision as being innate, human infants do not clearly show behaviors that can be defined as defensive in these conditions until 9 months of age (Yonas et al., 1977).

Adults are quite accurate in making time-to-contact judgments (Schiff & Oldak, 1990; Todd, 1981). Todd's data show relative time-to-contact judgments to be sensitive to less than 100-ms time differences. Relative time-to-passage judgments

are less accurate, requiring differences of about 500 ms (Kaiser & Mowafy, 1993).

How people actually make time-to-contact judgments is currently a topic of debate. In a review of the literature, Wann (1996) found that empirical support for the *tau* proposal was weak and that other optical variables, such as indexes of relative distance, could account for research findings as well as *tau*. Recently, Tresilian (1999) provided a revised *tau* hypothesis in which it is acknowledged that the effective information in time-to-contact situations is task- and context-specific, and moreover that it involves the utilization of multiple cues from diverse sources.

Intercepting a Fly Ball

In order to catch a fly ball, players must achieve two goals: First, they must get themselves to the location where the ball will land; and second, they must catch it. It seems reasonable to assume that satisfying the first goal of interception would require a determination of the ball's landing location, but this is not necessarily so. If a player looks at the ball in flight and runs so that the ball's perceived trajectory follows a straight path, then the ball will intersect the player's path on its descent (McBeath, Shaffer, & Kaiser, 1995). If players followed this simple control heuristic, then they would run along curved paths to the location of the ball's landing. If, instead, they knew where the ball would land, then they would run to that place in a straight line. In fact, outfielders run on curved paths that closely follow the predictions of the control heuristic formulation (McBeath et al., 1995). (See the chapter by Heuer in this volume for a discussion of motor control.)

This final experimental finding clearly points to the difficulty of disentangling conscious visual perceptions from the visual control of actions. Ask baseball players what they do when they catch fly balls, and they will tell you that they see the ball moving through space and run to where they can catch it. Without doubt, they perceive the ball to be moving in depth. On the other hand, the control heuristic that guides their running does not entail a representation of three-dimensional space. The heuristic applies to a two-dimensional representation of the ball's trajectory in the virtual image plan defined by their line of sight to the ball.

CONCLUSION

In perceiving depth and events, the relevant information is both limited and abundant. Viewed in isolation, visual information is almost always found lacking in its ability to uniquely specify those aspects of the environment to which it relates. However, combining different informational sources leads to

a more satisfactory state of affairs. In general, the more complex the visual scene, the more well specified it becomes.

Movement- and goal-directed behaviors are complications that add considerably to the sufficiency of optical information for specifying environmental layout and events; however, their study has recently led to the following conundrum: Conscious visual perceptions and visually guided actions do not always reflect a common underlying representation. For example, geographical slant is grossly overestimated; however, a visually guided adjustment of perceived slant is accurate. When catching a baseball, players perceive themselves to be moving in a three-dimensional environment even though the visual guidance of their running path is controlled by heuristics applied to a two-dimensional representation of the scene. The disparity between awareness and action in these cases may reflect the functioning of multiple perceptual systems.

Looking to the future, we see at least three developments that should have a significant impact on research on how people perceive depth and events. These developments include (a) improvements in research technology, (b) increased breadth in the interdisciplinary nature of research, and (c) increased sophistication in the theoretical approach.

Perceptual research has benefited enormously from computer technology. For example, Johansson (1950) used computers to create moving point-light displays on an oscilloscope thereby establishing the field of event perception. Current computer systems allow researchers to create almost any imaginable scene. Over the last 10 years, immersive displays have become available. Immersive displays surround observers and allow them to move and interact within a virtual environment. Head-mounted displays present images with small screens in front of the eyes and utilize tracking systems to register the position and orientation of the head and other tracked parts of the body. Another immersive display system is the Cave Automatic Virtual Environment, CAVE, which is a room having rear-projected images. The observer's head is tracked and the projected images transform in a manner consistent with the observer's movement through a three-dimensional environment. Such immersive display systems allow researchers to control optical variables that heretofore could only be manipulated within the confines of a computer terminal. Given the increased availability of immersive display systems, we expect to see more investigations of perceptions in situations entailing the visual control of action.

Understanding the perception of space and events is of interest to a wide variety of disciplines. The current chapter has emphasized the psychophysical perspective, which relates relevant optical information to perceptual sensitivities. However, within such fields as computer science and cognitive neuroscience, there is also considerable research on this topic.

Computer scientists are often interested in automating perceptual feats, such as the recovery of three-dimensional structure from optical motion information, and comparisons of digital and biological algorithms have proven to be useful (Marr, 1982). Another area of computer science that is ripe for interdisciplinary collaboration is in the computer-graphics animation of events. Interestingly, many movies today employ methods of motion capture to create computer-animated actors. These methods entail placing sensors on the head and joints of real actors and recovering an animation of a stick figure that can be fleshed out in graphics. One cannot help but think of Johansson's point-light walker displays when viewing such a motion capture system in use. Currently, there is considerable work attempting to create synthetic actors directly with algorithms. Perceptual scientists should be able to learn a lot by studying what works and what does not in this attempt to create synthetic thespians. Just as the pictorial depth cues were first discovered by artists and then articulated by psychologists, the study of computer-simulated events should help us better understand what information is needed to evoke the perceptions of such natural motions as a person walking, and perhaps more generally, perceptions of animacy and purpose.

Research in cognitive neuroscience has had an increasing impact on perceptual theory, and this trend is likely to continue. Advances in clinical case studies, functional brain imaging, and animal research have greatly shaped our current conceptions of perceptual processing. For example, the anatomical significance of the dorsal and ventral cortical pathways is currently receiving a lot of attention (Creem & Proffitt, 2001). These two pathways are the dominant visual processing streams in the cortex; however, there are many others visual streams in the brain. We have much to learn from functional anatomy that will help us constrain and develop our theoretical conceptions.

Finally, there have been a number of recent advances in the sophistication of our theoretical approach. One of the most notable of these was made recently by Cutting and Vishton (1995). Every text on depth perception provides a list of depth cues, as does the current chapter. How these cues are combined is still much debated. Given the huge number of cues, however, an account of how depth is perceived in the context of all possible combinations of these variables is probably unattainable. On the other hand, Cutting and Vishton showed that there is much to be gained by investigating the range of efficacy of different cues. For example, binocular disparity is useful at near distances but not far ones, whereas occlusion is equally useful at all distances. Looking at the problem of depth perception from this perspective motivates a search for the conditions under which information is

useful. A related theoretical approach is seen in the search for perceptual heuristics. Heuristics are simple processing rules having a precision that is no better than what is needed to effectively guide behavior. The control heuristics engaged when catching baseballs are examples (McBeath et al., 1995). From a pragmatic perspective, optical information is useful to the degree that it helps inform the requirements defined by the demands of the task at hand.

REFERENCES

- Amorim, M.-A., Loomis, J. M., & Fukusima, S. S. (1998). Reproduction of object shape is more accurate without the continued availability of visual information. *Perception*, 27, 69–86.
- Andersen, G. J., & Braunstein, M. L. (1983). Dynamic occlusion in the perception of rotation in depth. *Perception & Psychophysics*, 34, 356–362.
- Baird, J. C., & Wagner, M. (1991). Transformation theory of size judgment. *Journal of Experimental Psychology: Human Perception and Performance*, 17, 852–864.
- Barclay, C. D., Cutting, J. E., & Kozlowski, L. T. (1978). Temporal and spatial factors in gait perception that influence gender recognition. *Perception & Psychophysics*, 23, 145–152.
- Bardy, B. G., Warren, W. H., & Kay, B. A. (1999). The role of central and peripheral vision in postural control during walking. *Perception & Psychophysics*, 61, 1356–1368.
- Berkeley, G. (1709). *An essay towards a new theory of vision*. London: J. M. Dent.
- Bertamini, M., & Proffitt, D. R. (2000). Hierarchical motion organization in random dot configurations. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1371–1386.
- Bertamini, M., Yang, T. L., & Proffitt, D. R. (1998). Relative size perception at a distance is best at eye level. *Perception & Psychophysics*, 60, 673–682.
- Bertenthal, B. I., & Bai, D. L. (1989). Infants' sensitivity to optical flow for controlling posture. *Developmental Psychology*, 25, 936–945.
- Bertenthal, B. I., & Pinto, J. (1993). Complementary processes in the perception and production of human movements. In E. Thelan & L. Smith (Eds.), *A dynamic systems approach to development: Applications* (pp. 209–239). Cambridge, MA: MIT Press.
- Bertenthal, B. I., Proffitt, D. R., & Cutting, J. E. (1984). Infant sensitivity to figural coherence in biomechanical motions. *Journal of Experimental Child Psychology*, 37, 171–178.
- Berthoz, A., Istraël, I., George-François, P., Grasso, R., & Tsuzuku, T. (1995). Spatial memory of body linear displacement: What is being stored? *Science*, 269, 95–98.
- Berthoz, A., Lacour, M., Soechting, J. F., & Vidal, P. P. (1979). The role of vision in the control of posture during linear motion. *Progress in Brain Research*, 50, 197–209.
- Bhalla, M., & Proffitt, D. R. (1999). Visual-motor recalibration in geographical slant perception. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1076–1096.
- Bingham, G. P. (1987). Kinematic form and scaling: Further investigations on the visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception and Performance*, 13, 155–177.
- Bingham, G. P. (1993). Perceiving the size of trees: Form as information about scale. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 1139–1161.
- Bingham, G. P., & Pagano, C. C. (1998). The necessity of a perception-action approach to definite distance perception: Monocular distance perception to guide reaching. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 145–168.
- Bingham, G. P., Schmidt, R. C., & Rosenblum, L. D. (1995). Hefting for a maximum distance throw: A smart perceptual mechanism. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 507–528.
- Bingham, G. P., Schmidt, R. C., & Zaal, F. T. J. M. (1999). Visual perception of the relative phasing of human limb movements. *Perception & Psychophysics*, 2, 246–258.
- Bingham, G. P., Zaal, F., Robin, D., & Shull, J. A. (2000). Distortions in definite distance and shape perception as measured by reaching without and with haptic feedback. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 1436–1460.
- Blake, A., & Marinos, C. (1990). Shape from texture: Estimation, isotropy and moments. *Journal of Artificial Intelligence*, 45, 323–380.
- Blake, R. (1993). Cats perceive biological motion. *Psychological Science*, 4, 54–57.
- Bles, W. (1981). Stepping around: Circularvection and Coriolis effects. In J. B. Long & A. D. Baddeley (Eds.), *Attention and performance* (Vol. 9, pp. 47–61). Hillsdale, NJ: Erlbaum.
- Borjesson, E., & von Hofsten, C. (1975). A vector model for perceived object rotation and translation in space. *Psychological Research*, 38, 209–230.
- Bower, T. G. R., Broughton, J. M., & Moore, M. K. (1970). The coordination of visual and tactile input in infants. *Perception & Psychophysics*, 8, 51–53.
- Brandt, T., Dichgans, J., & Koenig, E. (1973). Differential effects of central versus peripheral vision on egocentric and exocentric motion perception. *Experimental Brain Research*, 16, 476–491.
- Brandt, T., Wist, E. R., & Dichgans, J. (1975). Foreground and background in dynamic spatial orientation. *Perception & Psychophysics*, 17, 497–503.
- Braunstein, M. L. (1976). *Depth perception through motion*. New York: Academic Press.
- Braunstein, M. L. (1994). Decoding principles, heuristics and inference in visual perception. In G. Johansson, S. S. Bergstrom, &

- W. Epstein (Eds.), *Perceiving events and objects* (pp. 436–446). Hillsdale, NJ: Erlbaum.
- Bruno, N., & Cutting, J. E. (1988). Minimodularity and the perception of layout. *Journal of Experimental Psychology: General*, *117*, 161–170.
- Carello, C., Groszofsky, A., Reichel, F., Solomon, H. Y., & Turvey, M. T. (1989). Perceiving what is reachable. *Ecological Psychology*, *1*, 27–54.
- Caudek, C., & Domini, F. (1998). Perceived orientation of axis of rotation in structure-from-motion. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 609–621.
- Caudek, C., Domini, F., & Di Luca, M. (in press). Illusory 3D rotation induced by dynamic image shading. *Perception & Psychophysics*.
- Chapanis, A., & Mankin, D. A. (1967). The vertical-horizontal illusion in a visually rich environment. *Perception & Psychophysics*, *2*, 249–255.
- Clark, J., & Yuille, A. (1990). *Data fusion for sensory information processing systems*. Boston: Kluwer.
- Creem, S. H., & Proffitt, D. R. (2001). Defining the cortical visual systems: "What," "where," and "how." *Acta Psychologica*, *107*, 43–68.
- Crowell, J. A., & Banks, M. S. (1993). Perceiving heading with different retinal regions and types of optic flow. *Perception & Psychophysics*, *53*, 325–337.
- Cumming, B. G., Johnston, E. B., & Parker, A. J. (1991). Vertical disparities and perception of three-dimensional shape. *Nature*, *349*, 411–413.
- Cutting, J. E., & Kozlowski, L. T. (1977). Recognizing friends by their walk: Gait perception without familiarity cues. *Bulletin of the Psychonomic Society*, *9*, 353–356.
- Cutting, J. E., & Millard, R. T. (1984). Three gradients and the perception of flat and curved surfaces. *Journal of Experimental Psychology: General*, *113*, 198–216.
- Cutting, J. E., & Proffitt, D. R. (1982). The minimum principle and the perception of absolute, common, and relative motions. *Cognitive Psychology*, *14*, 211–246.
- Cutting, J. E., Springer, K., Braren, P. A., & Johnson, S. H. (1992). Wayfinding on foot from information in retinal, not optical, flow. *Journal of Experimental Psychology: General*, *102*, 41–72, 129.
- Cutting, J. E., & Vishton, P. M. (1995). Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth. In W. Epstein & S. Rogers (Eds.), *Perception of space and motion* (pp. 69–117). San Diego, CA: Academic Press.
- Dichgans, J., & Brandt, T. (1978). Visual-vestibular interaction: Effects on self-motion perception and postural control. In H. Leibowitz & H.-L. Teuber (Eds.), *Handbook of sensory physiology* (pp. 755–804). New York: Springer-Verlag.
- Domini, F., & Braunstein, M. L. (1998). Recovery of 3-D structure from motion is neither euclidean nor affine. *Journal of Experimental Psychology: Human Perception and Performance*, *24*, 1273–1295.
- Domini, F., & Caudek, C. (1999). Perceiving surface slant from deformation of optic flow. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 426–444.
- Domini, F., Caudek, C., & Proffitt, D. R. (1997). Misperceptions of angular velocities influence the perception of rigidity in the kinetic depth effect. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 1111–1129.
- Domini, F., Caudek, C., & Richman, S. (1998). Distortions of depth-order relations and parallelism in structure from motion. *Perception & Psychophysics*, *60*, 1164–1174.
- Domini, F., Caudek, C., Turner, J., & Favretto, A. (1998). Discriminating constant from variable angular velocities in structure from motion. *Perception & Psychophysics*, *60*, 747–760.
- Duncker, K. (1937). Induced motion. In W. D. Ellis (Ed. and Trans.), *A source-book in Gestalt psychology*. London: Routledge. (Original work published 1929)
- Epstein, W. (1977). *Stability and constancy in visual perception*. New York: Wiley.
- Epstein, W. (1982). Percept-percept couplings. *Perception*, *11*, 75–83.
- Erkelens, C. J., & Collewijn, H. (1985). Eye movements and stereopsis during dichoptic viewing of moving random-dot stereograms. *Vision Research*, *25*, 1689–1700.
- Ernst, M. O., Banks, M. S., & Bühlhoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, *3*, 69–73.
- Fisher, S. K., & Ciuffreda, K. J. (1988). Accommodation and apparent distance. *Perception*, *17*, 609–621.
- Fisher, M. H., & Kornmüller, A. E. (1930). Optokinetisch ausgelöste Bewegungswahrnehmungen und optokinetischer Nystagmus [The perception of movement and optokinetic nystagmus initiated by optokinesis]. *Journal für Psychologie und Neurologie*, *41*, 273–308.
- Foley, J. M. (1968). Depth, size and distance in stereoscopic vision. *Perception & Psychophysics*, *3*, 265–274.
- Foley, J. M. (1985). Binocular distance perception: Egocentric distance tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *11*, 133–149.
- Fox, R., & McDaniels, C. (1982). The perception of biological motion by human infants. *Science*, *218*, 486–487.
- Fry, G. A., Bridgeman, C. S., & Ellerbrock, V. J. (1949). The effect of atmospheric scattering on binocular depth perception. *American Journal of Optometry*, *26*, 8–15.
- Gårding, J. (1992). Shape from texture for smooth curved surfaces in perspective projection. *Journal of Mathematical Imaging and Vision*, *2*, 327–350.
- Giannopulu, I., & Lepecq, J. C. (1998). Linear-vection chronometry along spinal and sagittal axes in erect man. *Perception*, *27*, 363–372.
- Gibson, J. J. (1950). *The perception of the visual world*. Boston: Houghton Mifflin.

- Lee, D. N., & Lishman, J. R. (1975). Visual proprioceptive control of stance. *Journal of Human Movement Studies*, 1, 224–230.
- Leeuwenberg, E. L. J. (1971). A perceptual coding language for visual and auditory patterns. *American Journal of Psychology*, 84, 307–349.
- Leeuwenberg, E. L. J. (1978). Quantification of certain visual pattern similarities: Saliency, transparency, similarity. In E. L. J. Leeuwenberg & H. F. J. M. Buffart (Eds.), *Formal theories of visual perception* (pp. 277–298). New York: Wiley.
- Loomis, J. M., Da Silva, J. A., Fujita, N., & Fukusima, S. S. (1992). Visual space perception and visually directed action. *Journal of Experimental Psychology: Human Perception and Performance*, 18, 906–921.
- Loomis, J. M., & Philbeck, J. W. (1999). Is the anisotropy of perceived 3-D shape invariant across scale? *Perception & Psychophysics*, 61, 397–402.
- Lunenburg, R. K. (1947). *Mathematical analysis of binocular vision*. Princeton, NJ: Princeton University Press.
- Mach, E. (1875). *Grundlinien der Lehre von der Bewegungsempfindungen* [Basic principles for the study of motion perception]. Leipzig, Germany: Engelmann.
- Mach, E. (1922). *Die Analyse der Empfindungen* [The analysis of sensations]. Jena, Germany: Gustav Fischer.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. F. Freeman.
- Mass, J. B., Johansson, G., Jansson, G., & Runeson, S. (1971). *Motion perception I and II* [Film]. Boston: Houghton Mifflin.
- Massaro, D. W. (1988). Ambiguity in perception and experimentation. *Journal of Experimental Psychology: General*, 117, 417–421.
- Mayhew, J. E., & Longuet-Higgins, H. C. (1982). A computational model of binocular depth perception. *Nature*, 297, 376–378.
- McAfee, E. A., & Proffitt, D. R. (1991). Understanding the surface orientation of liquids. *Cognitive Psychology*, 23, 483–514.
- McBeath, M. K., Shaffer, D. M., & Kaiser, M. K. (1995). How baseball outfielders determine where to run to catch fly balls. *Science*, 268, 569–573.
- McCloskey, M. (1983). Intuitive physics. *Scientific American*, 248, 122–130.
- McCloskey, M., Caramazza, A., & Green, B. (1980). Curvilinear motion in the absence of external forces: Naive beliefs about the motion of objects. *Science*, 210, 1139–1141.
- McCloskey, M., Washburn, A., & Felch, L. (1983). Intuitive physics: The straight-down belief and its origin. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 9, 636–649.
- McCready, D. (1985). On size, distance, and visual angle perception. *Perception & Psychophysics*, 37, 323–334.
- Michotte, A. (1963). *The perception of causality* (T. R. Miles & E. Miles, Trans.). London: Methuen.
- Milner, A. D., & Goodale, M. A. (1995). *The visual brain in action*. Oxford, UK: Oxford University Press.
- Mon-Williams, M., & Tresilian, J. R. (1999). Some recent studies on the extraretinal contribution to distance perception. *Perception*, 28, 167–181.
- Nakamura, S., & Shimojo, S. (1999). Critical role of foreground stimuli in perceiving visually induced self-motion (vection). *Perception*, 28, 893–902.
- Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science*, 257, 1357–1363.
- Norman, J. F., & Todd, J. T. (1992). The visual perception of 3-dimensional form. In G. A. Carpenter & S. Grossberg (Eds.), *Neural networks for vision and image processing* (pp. 93–110). Cambridge, MA: MIT Press.
- Norman, J. F., Todd, J. T., Perotti, V. J., & Tittle, J. S. (1996). The visual perception of three-dimensional length. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 173–186.
- Ogle, K. O. (1952). On the limits of stereoscopic vision. *Journal of Experimental Psychology*, 48, 50–60.
- Ono, H., & Comerford, T. (1977). Stereoscopic depth constancy. In W. Epstein (Ed.), *Stability and constancy in visual perception: Mechanisms and processes* (pp. 91–128). New York: Wiley.
- Ono, H., Rogers, B. J., Ohmi, M., & Ono, M. E. (1988). Dynamic occlusion and motion parallax in depth perception. *Perception*, 17, 255–266.
- Oyama, T. (1974). Perceived size and perceived distance in stereoscopic vision and an analysis of their causal relations. *Perception & Psychophysics*, 16, 175–182.
- Pagano, C. C., & Bingham, G. P. (1998). Comparing measures of monocular distance perception: Verbal and reaching errors are not correlated. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 1037–1051.
- Pentland, A. P. (1984). Local shading analysis. *IEEE, Transactions on Analysis & Machine Intelligence*, 6, 170–187.
- Philbeck, J. W., & Loomis, J. M. (1997). Comparison of two indicators of perceived egocentric distance under full-cue and reduced-cue conditions. *Journal of Experimental Psychology: Human Perception and Performance*, 23, 72–85.
- Philbeck, J. W., Loomis, J. M., & Beall, A. C. (1997). Visually perceived location is an invariant in the control of action. *Perception & Psychophysics*, 59, 601–612.
- Predebon, J. (1991). Spatial judgments of exocentric extents in an open-field situation: Familiar versus unfamiliar size. *Perception & Psychophysics*, 50, 361–366.
- Proffitt, D. R., Bhalla, M., Gossweiler, R., & Midgett, J. (1995). Perceiving geographical slant. *Psychonomic Bulletin & Review*, 2, 409–428.
- Proffitt, D. R., Creem, S. H., & Zosh, W. (2001). Seeing mountains in mole hills: Geographical slant perception. *Psychological Science*, 12, 418–423.

- Proffitt, D. R., Cutting, J. E., & Stier, D. M. (1979). Perception of wheel-generated motions. *Journal of Experimental Psychology: Human Perception and Performance*, 5, 289-302.
- Proffitt, D. R., & Gilden, D. L. (1989). Understanding natural dynamics. *Journal of Experimental Psychology: Human Perception and Performance*, 15, 384-393.
- Ramachandran, V. S. (1988). Perceiving shape from shading. *Scientific American*, 259, 76-83.
- Regan, D., & Beverly, K. I. (1982). How do we avoid confounding the direction we are looking and the direction we are moving? *Science*, 215, 194-196.
- Restle, F. (1979). Coding theory of the perception of motion configurations. *Psychological Review*, 86, 1-24.
- Rieser, J. J., Ashmead, D. H., Talor, C. R., & Youngquist, G. A. (1990). Visual perception and the guidance of locomotion without vision to previously seen targets. *Perception*, 19, 675-689.
- Rogers, B. J., & Bradshaw, M. F. (1993). Vertical disparities, differential perspective and binocular stereopsis. *Nature*, 361, 253-255.
- Rogers, S. (1996). The horizon-ratio relation as information for relative size in pictures. *Perception & Psychophysics*, 58, 142-152.
- Rosenholtz, R., & Malik, J. (1997). Surface orientation from texture: Isotropy or homogeneity (or both)? *Vision Research*, 37, 2283-2293.
- Ross, H. E. (1971). Do judgements of distance and greyness follow the physics of aerial perspective? In A. Garriga-Trillo, P. R. Minon, C. Garcia-Gallego, P. Lubin, A. Merino, & A. Villarino (Eds.), *Fechner Day '93: Proceedings of the International Society for Psychonomics* (pp. 233-238). Palma de Mallorca, Spain: International Society for Psychophysics.
- Ross, H. E. (1974). *Behavior and perception in strange environments*. London: George Allen & Unwin.
- Royden, C. S., Banks, M. S., & Crowell, J. A. (1992). The perception of heading during eye movements. *Nature*, 360, 583-585.
- Rubin, E. (1927). Visuell whrgenommene wirkliche Bewegungen [Visually perceived actual motion]. *Zeitschrift fur Psychologie*, 103, 384-392.
- Runeson, S. (1983). On visual perception of dynamic events. *Acta Universitatis Upsaliensis: Studia Psychologica Upsaliensis (Series 9)*. (Original work published 1977)
- Runeson, S., & Frykholm, G. (1981). Visual perception of lifted weight. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 733-740.
- Runeson, S., & Frykholm, G. (1983). Kinematic specification of dynamics as an informational basis for person and action perception: Expectation, gender recognition, and deceptive intention. *Journal of Experimental Psychology: General*, 112, 585-615.
- Runeson, S., Juslin, P., & Olsson, H. (2000). Visual perception of dynamic properties: Cue heuristics versus direct-perceptual competence. *Psychological Review*, 107, 525-555.
- Schiff, W. (1965). Perception of an impending collision: A study of visually directed avoidant behavior. *Psychological Monograph: General and Applied*, 79, (11, Whole No. 604).
- Schiff, W., & Oldak, R. (1990). Accuracy of judging time-to-arrival: Effects of modality, trajectory, and gender. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 303-316.
- Sedgwick, H. (1973). The visible horizon: A potential source of visual information for the perception of size and distance. *Dissertation Abstracts International*, 34, 1301-1302B. (UMI No. 73-22, No. 03B 530)
- Sedgwick, H. (1986). Space perception. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Vol. 1, pp. 1-57). New York: Wiley.
- Shimojo, S., Silverman, G. H., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, 29, 619-626.
- Sinai, M. J., Ooi, T. L., & He, Z. J. (1998). Terrain influences the accurate judgement of distance. *Nature*, 395, 497-500.
- Stevens, K. A. (1981). The information content of texture gradients. *Biological Cybernetics*, 42, 95-105.
- Stoffregen, T. A. (1985). Flow structure versus retinal location in the optical control of stance. *Journal of Experimental Psychology: Human Perception and Performance*, 11, 554-565.
- Sumi, S. (1984). Upside down presentation of the Johansson moving light spot pattern. *Perception*, 13, 283-286.
- Thomson, J. A. (1983). Is continuous visual monitoring necessary in visually guided locomotion? *Journal of Experimental Psychology: Human Perception and Performance*, 9, 427-443.
- Tittle, J. S., & Perotti, V. J. (1997). The perception of shape and curvedness from binocular stereopsis and structure from motion. *Perception & Psychophysics*, 59, 1167-1179.
- Todd, J. T. (1981). Visual information about moving objects. *Journal of Experimental Psychology: Human Perception and Performance*, 7, 795-810.
- Todd, J. T., & Bressan, P. (1990). The perception of 3-dimensional affine structure from minimal apparent motion sequences. *Perception & Psychophysics*, 48, 419-430.
- Todd, J. T., Koenderink, J. J., van Doorn, A. J., & Kappers, A. M. (1996). Effects of changing viewing conditions on the perceived structure of smoothly curved surfaces. *Journal of Experimental Psychology: Human Perception and Performance*, 22, 695-706.
- Todd, J. T., & Mingolla, E. (1983). Perception of surface curvature and direction of illumination from patterns of shading. *Journal of Experimental Psychology: Human Perception and Performance*, 9, 583-595.

- Todd, J. T., Reichel, F. D. (1989). Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychological Review*, *96*, 643–657.
- Todd, J. T., & Warren, W. H., Jr. (1982). Visual perception of relative mass in dynamic events. *Perception*, *11*, 325–335.
- Tresilian, J. R. (1991). Approximate information sources and perceptual variables in interceptive timing. *Journal of Experimental Psychology: Human Perception and Performance*, *20*, 154–173.
- Tresilian, J. R. (1999). Analysis of recent empirical challenges to an account of interceptive timing. *Perception & Psychophysics*, *61*, 515–528.
- Ullman, S. (1979). *The interpretation of visual motion*. Cambridge, MA: MIT Press.
- van den Berg, A. V. (1992). Robustness of perception of heading from optic flow. *Vision Research*, *32*, 1285–1296.
- van den Berg, A. V., & Brenner, E. (1994). Humans combine the optic flow with static depth cues for robust perception of heading. *Vision Research*, *34*, 2153–2167.
- Vicario, G. B., & Bressan, P. (1990). Wheels: A new illusion in the perception of rolling objects. *Perception*, *19*, 57–61.
- Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology*, *45*, 205–217.
- Wann, J. P. (1996). Anticipating arrival: Is the tau-margin a specious theory? *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1031–1048.
- Warren, W. H. (1995). Self-motion: Visual perception and visual control. In W. Epstein & S. Rogers (Eds.), *Perception of space and motion* (pp. 263–325). San Diego, CA: Academic Press.
- Warren, W. H. (1998). The state of flow. In T. Watanabe (Ed.), *High-level motion processing: Computational, neurobiological, and psychophysical perspectives* (pp. 315–358). Cambridge, MA: MIT Press.
- Warren, W. H., Kay, B. A., & Yilmaz, E. H. (1996). Visual control of posture during walking: Functional specificity. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 818–838.
- Warren, W. H., & Kurtz, K. J. (1992). The role of central and peripheral vision in perceiving the direction of self-motion. *Perception & Psychophysics*, *51*, 443–454.
- Webb, J. A., & Aggarwal, J. K. (1982). Structure from motion of rigid and jointed objects. *Artificial Intelligence*, *19*, 107–130.
- Wertheim, A. H. (1994). Motion perception during self-motion: The direct versus inferential controversy revisited. *Behavioral & Brain Sciences*, *17*, 293–355.
- Wertheimer, M. (1937). Laws of organization in perceptual forms. In W. D. Ellis (Ed. & Trans.), *A source book of Gestalt psychology* (pp. 71–88). London: Routledge. (Original work published 1923)
- Wheatstone, C. (1938). Contributions to the physiology of vision: I. On some remarkable and hitherto unobserved phenomena of binocular vision. *Philosophical Transactions of the Royal Society of London*, *128*, 371–394.
- Wraga, M. (1999). The role of eye height in perceiving affordances and object dimensions. *Perception & Psychophysics*, *61*, 490–507.
- Wraga, M., Creem, S. H., & Proffitt, D. R. (2000). Perception-action dissociations of a walkable Müller-Lyer configuration. *Psychological Science*, *11*, 239–243.
- Yang, T. L., Dixon, M. W., & Proffitt, D. R. (1999). Seeing big things: Overestimation of heights is greater for real objects than for objects in pictures. *Perception*, *28*, 445–467.
- Yonas, A., Bechtold, A. G., Frankel, D., Gordon, F. R., McRoberts, G., Norcia, A., & Sternfels, S. (1977). Development of sensitivity to information for impending collisions. *Perception & Psychophysics*, *21*, 97–104.
- Yonas, A., Craton, L. G., & Thompson, W. B. (1987). Relative motion: Kinetic information for the order of depth at an edge. *Perception & Psychophysics*, *41*, 53–59.
- Young, M. J., Landy, M. S., & Maloney, L. T. (1993). A perturbation analysis of depth perception from combinations of texture and motion cues. *Vision Research*, *33*, 2685–2696.