PERCEIVING ENVIRONMENTAL PROPERTIES FROM MOTION INFORMATION: MINIMAL CONDITIONS

Dennis R. Proffitt
University of Virginia
Charlottesville, Virginia

Mary K. Kaiser
NASA Ames Research Center
Moffett Field, California

INTRODUCTION

Everyday perception occurs in a context of nested motions. Eyes move within heads, heads move on bodies, and bodies move in surroundings that are filled with objects, many of which can themselves move (Gibson, 1966). Motion is omnipresent in perception. Stabilize an image on the retina and it rapidly becomes imperceptible (Pritchard, 1961). Not only is motion a necessary condition for perception, but it is also a sufficient condition for the perception of a variety of environmental properties.

Until recently, spatial instruments had few degrees of freedom with respect to the sorts of motion-carried information that they could provide. With increasing opportunities to employ animation, spatial instruments can be crafted that are tied less to artificial conventions and more to the natural condition of everyday perceptual experience.

The implications of perception research for display design derive from the methods employed by visual scientists in their investigations of how people extract environmental properties from optical information. The approach taken in perception research involves a seeking of *minimal stimulus conditions* for perceiving these properties. Stimuli that typically evoke relevant perceptions are decomposed into minimal information sources, and these sources are evaluated separately. It is almost always found that we humans rely on a large variety of information sources in perceiving any particular aspect of the environment. Knowledge of minimal conditions for perceiving environmental properties can be utilized in the design of effective and technologically efficient spatial instruments.

Since motion information is a minimally sufficient condition for perceiving numerous environmental properties, its use in spatial instruments eliminates the need to employ most of the conventions typically found in static displays. Moreover, in some contexts animated displays can elicit more accurate perceptions than are possible for static displays.

In this chapter, we discuss the status of motion as a minimal information source for perceiving the environmental properties of surface segregation, three-dimensional (3-D) form, displacement, and dynamics. The selection of these particular properties was motivated by a desire to present research on perceiving properties that span the range of dimensional complexity.

SURFACE SEGREGATION

Surface segregation refers to the separation of distinct surfaces in depth. In order to represent surface segregation on a two-dimensional (2-D) display surface, the surfaces must be distinguished by some apparent optical differences. These distinctions can be achieved with either static images or animated displays; however, only with motion can surface segregation be specified by a single cue without introducing ambiguous depth-order relations. Moreover, the implicit viewer assumptions needed to interpret moving displays are derived from the laws of dynamics, and thus are more fundamental in nature than are those accessed in interpreting static displays.

Perceiving Surface Segregation in Static Images

In pictures, surfaces are typically distinguished by color contrasts produced by differences in intensity or wavelength. One surface thereby becomes separated from another at an edge. Figure 1 depicts the familiar faces-vase figure introduced by Rubin (1915). This figure exemplifies the inherent figure-ground ambiguity of all static displays. Here, depending upon which is taken as figure, the vase or the faces, depth-order relations reverse (depth order being a term that refers to what is in front of what).

In order to resolve this depth-order ambiguity, additional cues must be supplied. One effective cue is occlusion. As is shown in figure 2, having one surface appear to be partially covered by another is an effective convention for specifying depth order. It is important to realize, however, that the disambiguation of figure 2 is achieved only through the activation of implicit assumptions or biases on the part of the viewer. The viewer must assume that the apparent far surface does not, in fact, have a notch cut out of it. As the Ames demonstrations on the overlay show, if this assumption is violated, viewers will see erroneous depth-order relations (Ittelson, 1968).

Another static convention that helps to resolve depth-order ambiguity is the use of familiar surfaces. In figure 3, the "A" is typically seen in front of the background surface. As figure 1 showed, what is taken as figure—vases or face—is perceived as being in front of the apparent ground (Rubin, 1915). This perceptual bias can be exploited by representing the intended forward surface with a familiar figure. However, as with occlusion, this convention relies heavily on inherent viewer biases. The A is *assumed* to have been placed atop the surrounding surface, as opposed to having been cut out of it. This assumption may be in error.

The inclusion of additional cues, such as shading, perspective, or solid modeling, will further constrain depth-order interpretations. However, so long as the viewer cannot obtain multiple perspectives on the objects depicted, the display remains inherently ambiguous. Again, the Ames demonstrations serve to show that observers can always be made to have erroneous perceptions whenever they are constrained to view an object from a unique perspective.

Intermediate between static and animated displays are those that include flicker. Wong and Weisstein (1987) found that surface segregation is observed in displays consisting of randomly placed dots when a particular region is made to flicker. Moreover, the flickering region usually appears to be behind adjacent nonflickering regions. Spatial instruments have yet to exploit this perceptual influence of flicker.

Perceiving Surface Segregation in Motion Displays

The ability of motion information to specify surface segregation without depth-order ambiguity was demonstrated by Gibson et al. (1969). They produced movies of randomly textured surfaces. When the surfaces were superimposed and stationary, segregation could not be achieved. However, when one or both of the surfaces moved, they separated into distinct surfaces and their depth order became unequivocal.

It was thought that the ongoing occlusion of the far surface by the near one served as the essential source of information for the surface segregation demonstration of Gibson et al. Recently, however, Yonas, Craton, and Thompson (1987) showed that surface segregation could be achieved without ongoing occlusion occurring at surface edges. They created a computer-animated display in which surfaces were defined by randomly positioned points of light. As with the original Gibson et al. display, when the simulated surfaces were stationary, there was no information suggesting that more than one surface was present; however, when the surfaces moved, their segregation became apparent. In this case, segregation and depth order were specified by the relative motion of point-lights on different surfaces, and by the disappearance of the lights on the far surface when they passed beneath the subjective contour that defined the edge of the close surface.

There are, of course, implicit assumptions that must be made in interpreting moving displays; however, they are of a fundamentally different sort than those that were discussed for static presentations. For static displays, the assumptions are characterized by notions of likelihood and simplicity. It is highly unlikely that anyone would create a display such as figure 2 with the intent of depicting a square located behind a notched square. Moreover, by any criterion of simplicity, the obvious interpretation of figure 2 is the simpler of the two (or three) depth-order alternatives (see, for example, Leeuwenberg, 1982). For animated displays, the implicit assumptions reflect fundamental laws of dynamics. Surfaces are not destroyed or brought into being when they pass in front of, or go beyond, more distant surfaces. Unlike those accessed when viewing static displays, the assumptions engaged when perceiving animated displays are based upon dynamical laws.

THREE-DIMENSIONAL FORM

Any 2-D representation of a 3-D object is inherently ambiguous. This is true of both static and moving displays. The virtue of animated displays, however, is that time can substitute for the lost spatial dimension.

Implicit viewer assumptions are required to recover 3-D relations from either static or moving 2-D projections. As was found for perceiving surface segregation, those engaged when viewing animated displays are grounded in the laws of dynamics as opposed to the conventions of artifice.

Perceiving 3-D Form in Static Displays

Effective means for representing 3-D objects and scenes were discovered by pictorial artists and evolved over time (Gombrich, 1960). Following Berkeley (1709), these pictorial conventions have come to be called secondary or pictorial depth cues. Researchers are still attempting to discover the invented techniques by which artists produced their compelling spatial effects (Kubovy, 1986).

The list of secondary depth cues is a long one; however, all entries share a common origin in the motivation to overcome the ambiguity inherent in 2-D representations of a 3-D scene. The resolution of ambiguity through the implementation of such conventions as solid modeling, perspective, shading, occlusion, familiarity, and so forth is more apparent than real. Demonstrations, such as those of Ames (Ittleson, 1968), show that perception can always be in error when inferring 3-D structure from a single 2-D projection. The possibility of such errors reflect, in turn, on the processing assumptions made when interpreting static displays. As with surface segregation, assumptions grounded in likelihood and simplicity are prevalent. To these are added various assumptive geometric conventions (Kubovy, 1986).

Perceiving 3-D Form in Motion Displays

The use of geometry can show that the changing spatial pattern, produced when the image of a rotating rigid object is projected onto a 2-D surface, uniquely defines the 3-D configuration of the object. In addition, three projected images of four non-coplanar points undergoing rotation defines the minimal condition for the recovery of structure from motion (Ullman, 1979).

Wallach and O'Connell (1953) showed that people are able to recover 3-D form when viewing 2-D projections of rotating objects. They constructed wire forms and projected their shadows onto screens. Viewers of these shadows reported that they saw only 2-D configurations of lines when the wire forms were stationary; however, they accurately reported on the 3-D configurations when the forms were continuously rotated. Wallach and O'Connell called their demonstration the Kinetic Depth Effect, or KDE.

Interest in KDE has grown over the years. Braunstein (1962), Doner, Lappin, and Perfetto (1984), Todd (1982), and many others have investigated the psychophysics of the phenomenon. Recently, a good deal of research has been directed toward the rigidity assumption.

Recall that transforming a 2-D projection of a rotating form is unique to the form's 3-D configuration only so long as the form remains rigid. Psychologists are much in doubt as to whether the human perceptual system actually implements a rigidity assumption when extracting structure from motion in KDE (Hochberg, 1986).

When the veracity of interpretive assumptions is evaluated, the issue of whether people utilize a rigidity assumption is less important than that such a dynamical assumption is capable of serving as the sole basis for the recovery of structure from motion. Unlike the assumptions embodied in pictorial depth cues, the rigidity assumption is grounded in the following kinematic law: Objects do not distort when rotated. Our perceptual systems were formed in the context of natural constraints. The exploitation of these constraints does not require that they be embodied. The fundamental assumptive nature of the rigidity principle is not based upon whether or not it has been

internalized by the perceptual system, but rather upon this fact: Vision evolved in a context in which this rigidity assumption is inviolate.

It must be conceded that, in a few known circumstances, the assumptions of picture perception interact with those engaged by motion perception. Ames created a trapezoidal surface that looked like a rectangular window viewed at an angle. When observers viewed it monocularly as it underwent rotation, they typically reported seeing an oscillating rectangular window rather than a rotating trapezoid (Ittelson, 1968). It is important to note that this event's 2-D projection is, in fact, inconsistent with the rectangular percept; however, the strong influence of such pictorial assumptions as likelihood and simplicity outweigh, in this case, the motion-carried information defining the actual configuration.

Perceiving 3-D structure from motion information has also been shown to occur for jointed objects. Johansson (1973) placed point-lights on the joints of people and filmed them as they performed actions in the dark. When shown to observers, these movies were readily perceived as depicting people. It was later found that between 0.1 and 0.2 sec was a sufficient exposure duration for perceiving the human form in these films (Johansson, 1976).

Computational theorists have developed effective algorithms for extracting structure from these jointed events, given certain constraints on the motions of the walkers (Hoffman and Flinchbaugh, 1982; Webb and Aggarwal, 1982). These computational models implement assumptions about the local rigidity of moving limbs. In essence, the models assume that the act of rotating or translating a rod (bones in the case of point-light walkers) does not, itself, change the rod's length. This assumption is based upon a kinematic law of nature. The perceptual system may or may not have internalized this law (Proffitt and Bertenthal, 1988); however, it certainly evolved in a world that is governed by it.

DISPLACEMENT

The motion of an object relative to an observer is referred to as its displacement. Displacement information can be conveyed in static displays only through the use of very artificial conventions. In moving displays, displacement information is presented directly in the natural medium of time. In addition, the perceptual system effectively segregates those motions specifying form from those that define observer-relative displacement.

Perceiving Displacement in Static Displays

It is not difficult to represent in a static display the fact that an object is moving. What is difficult to represent is the future position that an object will achieve over time. Static representations of motion properties must rely on highly stylized conventions, the most prominent being vector depictions, such as those shown in figure 4. Interpreting such displays not only requires one to effectively read the intended meaning of the conventions, but he or she must also be able to mentally perform the transformation suggested in the representation. People are not very good at such tasks. In fact, when people attempt to extrapolate the future position of moving objects that become occluded behind barriers, they make sizable errors, particularly for complex motion functions (Jagacinski, Johnson, and Miller, 1983).

Perceiving Displacement in Motion Displays

It is rare in nature for an object to undergo a pure observer-relative translation such that every object point moves with exactly the same motion. In fact, only when objects move in horizontal circles around the observer do common linear motions project to the observer's point of observation; all nonorthogonal distal translations project a rotational component to the observer's viewpoint. The perceptual system deals effectively with complex motions by analyzing them into relative and common motion components (Johansson, 1950). To illustrate this analysis, consider the perception of a rolling wheel.

As is depicted in figure 4, except for the hub, every point on a rolling wheel follows a complex trajectory belonging to the family of cycloidal curves. These trajectories are referred to as the event's absolute motions. The perceptual system segregates these motions into two components, relative rotations and a common-observer relative displacement (Proffitt, Cutting, and Stier, 1979). This perceptual analysis selects the configural centroid as the center of relative rotations. Thus, for a rolling wheel, rotations are seen as occurring about the wheel's hub, and the common motion is seen as the hub's translation. However, if point-lights are attached to an unseen rolling wheel and the configural centroid of these lights does not correspond to the wheel's hub, then a different common motion is seen. Again, relative motions are seen as rotations about the configural centroid, but the common motion is, in this case, the prolate cycloidal path followed by this abstract centroid. This perceptual analysis has also been found to occur for configurations moving in depth (Proffitt and Cutting, 1979). It has been proposed that the selection of the configural centroid, as the center for perceived relative motions, reflects a perceptual preference to minimize relative motions; in centroid relative rotations, all instantaneous relative motions sum to zero (Cutting and Proffitt, 1982).

Research findings on the perceptual analysis of absolute motions into relative and common components have two implications for display design. First, object configuration interacts with displacement perception. Whenever an object undergoes a complex motion, its configural properties influence the common motions that are observed. Although the effects are somewhat different, robust configural influences have also been shown to occur in stroboscopically presented apparent motions (Proffitt et al., 1988). Second, relative and common motions have different perceptual significances (Proffitt and Cutting, 1980). As is depicted in figure 5, relative rotations are used to perceptually define 3-D form, whereas common motions are residual to form analysis, and define observer relative displacements.

DYNAMICS

The laws of dynamics place constraints on the sorts of motions that can occur in nature. Given these constraints, the patterns observed in natural motions reflect back upon underlying dynamical properties. The motions of colliding objects are a good example of this reciprocal specification of dynamic and kinematic properties.

When objects collide, the laws of linear momentum conservation state that post-collision motions must preserve the event's pre-collision momentum. (For the sake of simplicity, we

exclude considerations of friction and damping.) Given these laws, it can be shown that the ratio of masses for the objects involved in a collision are specified by ratios in their velocities (Runeson, 1977). It has been found that people are relatively good at judging mass ratios when observing collisions (Todd and Warren, 1982; Kaiser and Proffitt, 1984). In addition, people are able to accurately discriminate possible collisions from those that violate dynamical principles (Kaiser and Proffitt, 1987a).

These results do not necessarily imply that the human perceptual system has internalized physical conservation laws, and in fact, the results of recent studies strongly suggest that such laws are not inherent to perceptual processing (Gilden and Proffitt, 1989). However, as has been previously discussed for surface segregation and form perception, our sensory systems need not embody natural laws in order to take advantage of the fact that they evolved in an environment in which dynamical laws are always upheld. Motion information is fundamental because dynamical constraints shaped the natural environment in which vision evolved.

The interpretation of static displays require processing rules shaped in the context of pictorial conventions. The conceptual heritage of static information-processing rules is reflected in their subservience to cognitive beliefs. People hold inaccurate common-sense views about natural dynamics. These erroneous beliefs are reflected in their judgments of static, but not moving, displays.

Perceiving Dynamics in Static Displays

Recently, an intriguing literature has developed on people's naive beliefs about the laws of dynamics. Called "intuitive physics" by McCloskey (1983), these beliefs influence people's predictions about natural motions; moreover, they are often at odds with the laws of dynamics.

Figure 6 shows one of the problems used by McCloskey, Caramazza, and Green (1980). Depicted is a C-shaped tube that is lying flat on a horizontal surface. A ball is rolled through the tube, and upon exiting, the ball rolls across the surface. Subjects were asked to predict the path taken when the ball exited the tube. Approximately 45% of the undergraduate subjects who were asked this question incorrectly stated that the ball would continue to follow a curved path. McCloskey and his colleagues have conducted numerous similar experiments, all showing that judgments made about natural object motions often reflect erroneous beliefs.

All of these studies required people to make judgments while looking at pictures. The influence of intuitive physics beliefs is pervasive only in such static contexts. These beliefs have been found to have little or no effect on the perception of animated displays.

Perceiving Dynamics in Motion Displays

We replicated McCloskey et al.'s finding with the C-shaped tube problem, using a design in which observers were asked to judge which of a set of drawn trajectories appeared correct. Then, using the same design, we showed observers animated simulations of balls rolling through C-shaped tubes. Upon exiting the tubes, the balls followed a variety of paths. We found that people almost always chose as correct the natural trajectory when viewing these moving displays, and judged their erroneous predictions as being anomalous (Kaiser, Proffitt, and Anderson, 1985).

We have demonstrated this superiority of motion displays to evoke accurate dynamical judgments in other contexts (Kaiser and Proffitt, 1987b).

Static representations elicit intuitions that reflect cognitive beliefs. Obviously, people would have great difficulty getting about in the world if their perceptions were always tied to their knowledge of physical principles. A baseball outfielder, for example, would probably never succeed in catching a flyball if he was required to plan his pursuit using only his knowledge of physics.

Everyday perceptions necessarily occur in a context of naturally constrained motions. In such circumstances, our perceptual systems can function without recourse to memorial conceptions. Perception is good in motion context because motion is fundamental to the rules of perceptual processing.

CONCLUSIONS

Motion is an effective source of information for perceiving a variety of environmental properties. Because it is a minimally sufficient information source, it need not be simply added to the conventions employed in static displays. Rather, motion can replace many of these conventions, and in some contexts, motion can elicit more accurate perceptions than are possible for static displays.

Motion information is fundamental to everyday perception. The interpretive assumptions required to extract structure from motion are based upon the laws of nature—i.e., natural dynamics—whereas those evoked by static displays are based upon the artificial conventions of pictorial representations. The advantage that motion displays have over static ones derives from the heritages of the perceptual processes needed for their interpretation. The perceptual processes required to extract structure from motion information were formed in the context of dynamical constraints. The interpretation of static information relies more on perceptual processes that arise with conceptual development, and thus are grounded in such experientially based notions as simplicity, familiarity, and geometrical conventions.

REFERENCES

Berkeley, G. (1709). An essay towards a new theory of vision. London: Dent.

Braunstein, M. L. (1962). The perception of depth through motion. Psychol. Bull., 59, 422-433.

Cutting, J. E.; and Proffitt, D. R. (1982). The minimum principle and the perception of absolute, common, and relative motions. Cognitive Psychol., 14, 211-246.

Doner, J.; Lappin, J. S.; and Perfetto, G. (1984). Detection of three-dimensional structure in moving optical patterns. J. Exp. Psychol.: Human Percept. Perform., 10, 1-11.

Gibson, J. J. (1966). The senses considered as perceptual systems. Boston: Houghton Mifflin.

Gibson, J. J.; Kaplan, G. A.; Reynolds, H. N.; and Wheeler, K. (1969). The change from visible to invisible: A study of optical transitions. Perception & Psychophysics, 5, 113-116.

Gilden, D. L.; and Proffitt, D. R. (in press). Understanding collision dynamics. J. Exp. Psychol.: Human Percept. Perform.

Gombrich, E. H. (1960). Art and illusion: A study in the psychology of pictorial representation. Princeton, NJ: Princeton.

Hochberg, J. (1986). Representation of motion and space in video and cinematic displays. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), Handbook of perception and human performance. Vol. 1, New York: Wiley, 22:1-63.

Hoffman, D. D.; and Flinchbaugh, B. E. (1982). The interpretation of biological motion. Biol. Cybernetics, 42, 195-204.

Ittelson, W. H. (1968). The Ames demonstrations in perception. New York: Hafner.

Jagacinski, R. J.; Johnson, W. W.; and Miller, R. A. (1983). Quantifying the cognitive trajectories of extrapolated movements. J. Exp. Psychol.: Human Percept. Perform., 9, 43-57.

Johansson, G. (1950). Configuration in event perception, Uppsala, Sweden: Almqvist & Wiksell.

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. Perception & Psychophysics, 14, 201-211.

Johansson, G. (1976). Spatio-temporal differentiation and integration in visual motion perception. Psychological Res., 38, 379-396.

Kaiser, M. K.; and Proffitt, D. R. (1984). The development of sensitivity to causally-relevant dynamic information. Child Develop., 55, 1614-1624.

Kaiser, M. K.; and Proffitt, D. R. (1987a). Observers' sensitivity to dynamic anomalies in collisions. Perception & Psychophysics, 42, 275-280.

Kaiser, M. K.; and Proffitt, D. R. (1987b). Naive mechanics: Erroneous beliefs, but veridical perceptions. Manuscript submitted for publication.

Kaiser, M. K.; Proffitt, D. R.; and Anderson, K. (1985). Judgments of natural and anomalous trajectories in the presence and absence of motion. J. Exp. Psychol.: Learning, Memory, and Cognition, 11, 795-803.

Kubovy, M. (1986). The psychology of perspective and renaissance art. Cambridge, England: Cambridge.

Leeuwenberg, E. (1982). Metrical aspects of patterns and structural information theory. In J. Beck (Ed.), Organization and representation in perception, Hillsdale, NJ: Erlbaum, 57-71.

McCloskey, M. (1983). Intuitive physics. Scientific Amer., 248(4), 122-130.

McCloskey, M.; Caramazza, A.; and Green, B. (1980). Curvilinear motion in the absence of external forces: Naive beliefs about the motion of objects. Science, 210, 1139-1141.

Pritchard, R. M. (1961). Stabilized images on the retina. Scientific Amer., 204, 72-78.

Proffitt, D. R.; and Bertenthal, B. I. (1988). Recovering connectivity from moving point-light displays. In W. N. Martin and J. K. Aggarwal (Eds.), Motion understanding, Hingham MA: Kluwer, 298-328.

Proffitt, D. R.; and Cutting, J. E. (1979). Perceiving the centroid of configurations on a rolling wheel. Perception & Psychophysics, 25, 389-398.

Proffitt, D. R.; and Cutting, J. E. (1980). An invariant for wheel-generated motions and the logic of its determination. Perception, 9, 435-449.

Proffitt, D. R.; Cutting, J. E.; and Stier, D. M. (1979). Perception of wheel-generated motions. J. Exp. Psychol.: Human Percept. Perform., 5, 289-302.

Proffitt, D. R.; Gilden, D. L.; Kaiser, M. K.; and Whelan, S. M. (1988, in press). The effect of configural orientation on perceived trajectory in apparent motion. Perception & Psychophysics.

Rubin, E. (1915). Synsoplevede figurer. Copenhagen: Cyldendalske.

Runeson, S. (1977). On visual perception of dynamic events. Unpublished doctoral dissertation, Univ. Uppsala, Uppsala, Sweden.

Todd, J. T. (1982). Visual information about rigid and nonrigid motion: A geometric analysis. J. Exp. Psychol.: Human Percept. Perform., 8, 238-252.

Todd, J. T.; and Warren, W. H. (1982). Visual perception of relative mass in dynamic events. Perception, 11, 325-335.

Ullman, S. (1979). The interpretation of visual motion. Cambridge, MA: MIT Press.

Wallach, H.; and O'Connell, D. N. (1953). The kinetic depth effect. J. Exp. Psychol., 45, 205-217.

Webb, J. A.; and Aggarwal, J. K. (1982). Structure from motion of rigid and jointed objects. Artificial Intelligence, 19, 107-130.

Wong, E.; and Weisstein, N. (1987). The effects of flicker on the perception of figure and ground. Perception & Psychophysics, 41, 440-448.

Yonas, A.; Craton, L. G.; and Thompson, W. B. (1987). Relative motion: Kinetic information for the order of depth at an edge. Perception & Psychophysics, 41, 53-59.

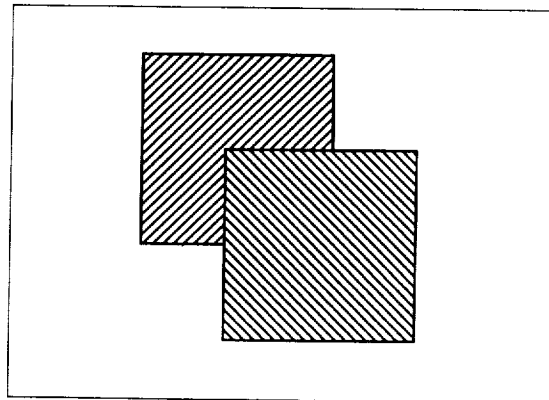Figure 1.– Rubin's (1915) faces-vase figure.

Figure 2.– Two surfaces are depicted. The one to the left appears to partially occlude the surface to the right.
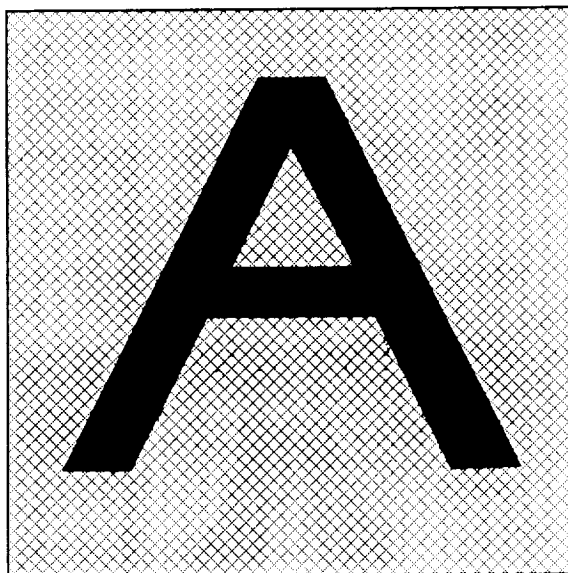
Figure 3.– The familiar figure, A, appears to be in front of the background surface.
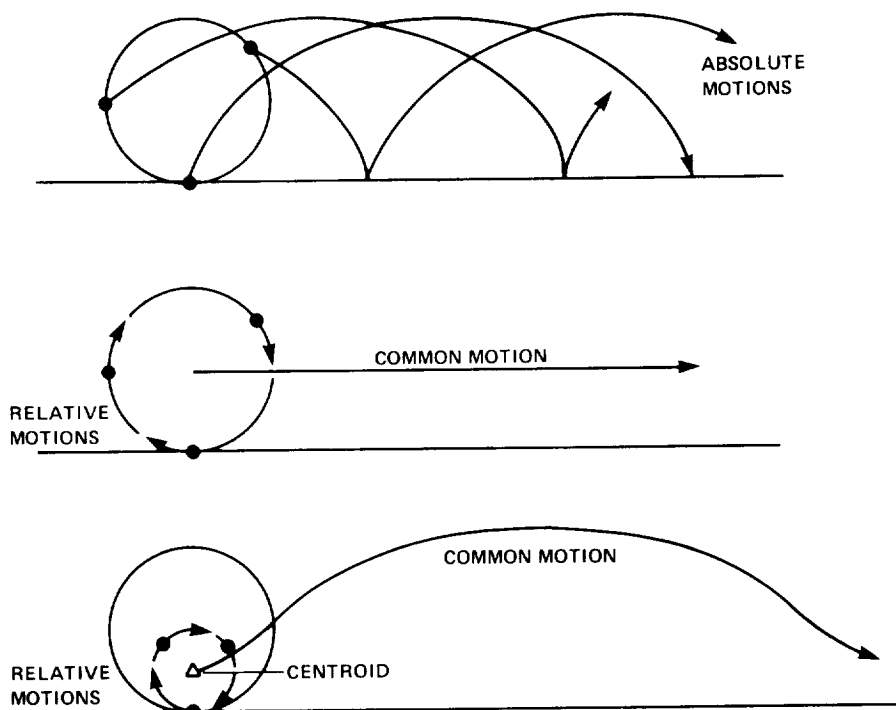

Figure 4.– The top panel depicts the absolute motions of three points on a rolling wheel. The middle panel shows the relative and common motions that are perceived in this event. The bottom panel depicts the perceived motions for three points on a rolling wheel in which the configural centroid of the points does not coincide with the wheel's hub.
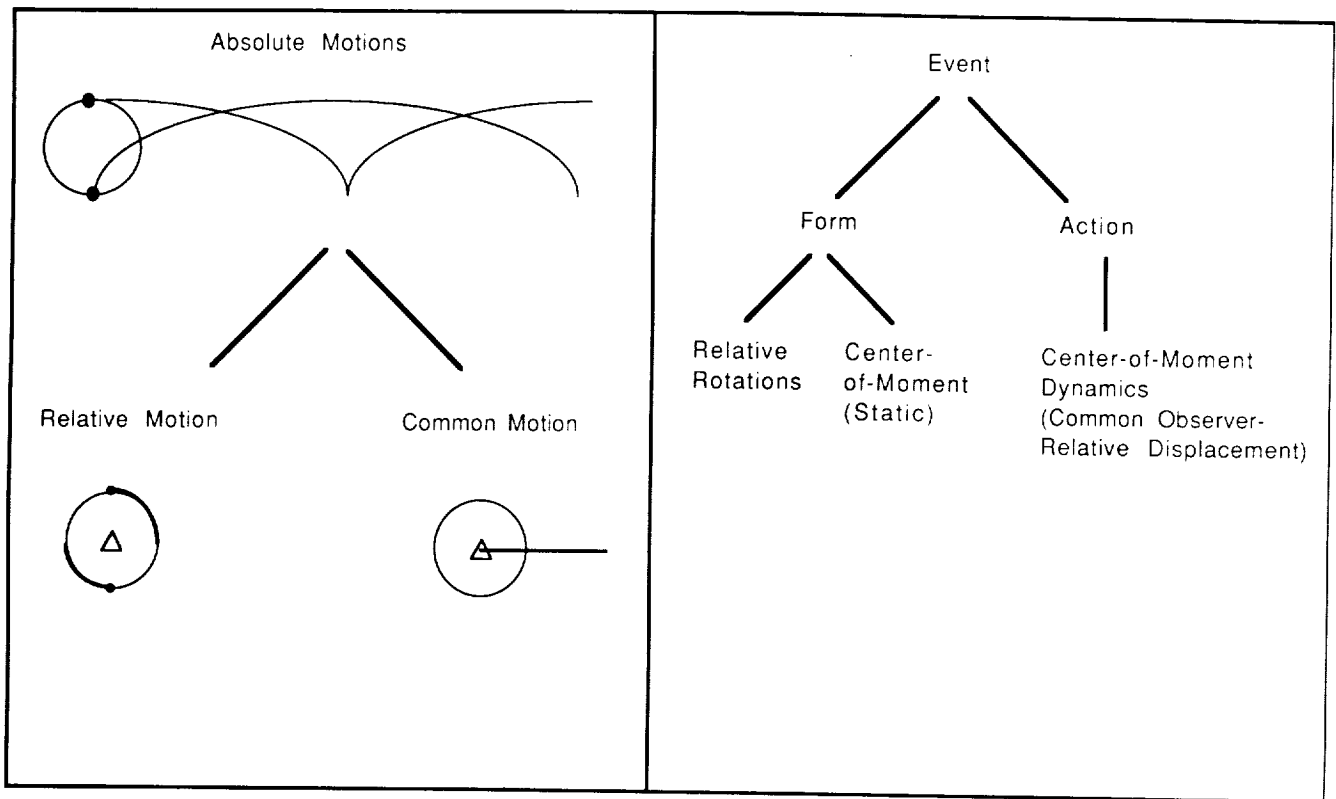
Figure 5.– The perceptual system divides absolute motions into relative and common components. The relative rotations are used in form analysis, whereas the form's common motion defines its observer-relative displacement.
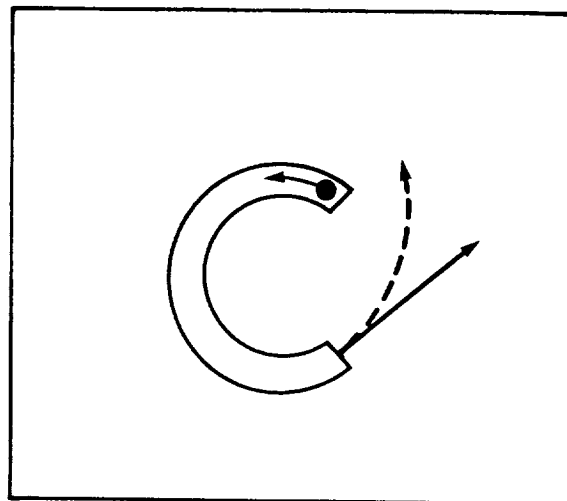

Figure 6.– Depicted is a horizontal C-shaped tube through which a ball is rolled. The two drawn trajectories represent the correct path that the ball takes upon exiting the tube, and a frequently drawn erroneous path.