CHAPTER 1

The Origins of Object Perception

Philip J. Kellman

I. INTRODUCTION

Two streams of light enter tiny apertures of a child's eyes. Instantly, she is aware, in rich detail, of the objects that furnish her environment. How is this possible?

To explain object perception, we must connect facts of many different kinds. We need facts about the physical world, such as how light is absorbed and reflected by objects, and geometrical facts about how objects' projections change as they move or as the observer moves. We must also know about information processing: what properties and relationships in reflected light carry information about objects? How is information extracted, represented and transformed? We also need to know how this information processing is carried out biologically: facts about the functions of retinal receptors, single cortical cells, cortical areas, and so on. For some purposes, accounts at one level or another may be most important. For building a computer model or robot, information processing is the focus, since once understood in humans it may be simulated on the computer. The details of our biology will not be shared by the computer, which has much different circuitry. For treating pathology of the human visual system, on the other

Perceptual and Cognitive Development

hand, physiology is crucial, whereas knowing an algorithm for recovering shape from motion is irrelevant.

Woven through these multiple levels of understanding is a dimension we have not yet mentioned: Development. Available information, processes that extract it, and their biological substrates are not static. Growth and learning, especially during the first year of life, profoundly change perception. Our purpose in this chapter is to examine these changes, focusing on the question: How does object perception develop? We will emphasize perceptual abilities at or near the beginning of human life, and what is known about their transformation as a person grows and learns. Both early capacities and patterns of change have implications for early cognitive and social development.

Our treatment will necessarily be confined in several ways. Although objects are perceived via several perceptual systems, we will emphasize vision, both because it is primary in giving us spatial information at a distance, and because it has been heavily researched. We will also concentrate on how perceivers get knowledge about properties of the environment; accordingly, we will draw sparingly from the large literature characterizing sensory thresholds, selectively noting those facts about sensory limitations that can be clearly linked to perceptual performance. (For a more detailed discussion of the development of visual mechanisms, the interested reader may refer to Banks & Salapatek, 1983; Banks & Kellman, in press). Finally, in focusing on object perception we will often note its relation to other topics, such as space and motion perception, but we will not discuss them in detail. (For a detailed treatment of the development of space and motion perception, see Kellman, 1995.)

A. What Is an Object?

Much of perception is object perception. Having said that, it might be useful to say what we mean by an "object." Here "object" will mean a coherent, bounded volume of matter. A stick, a hat, or a cupcake is an object; a pile of sand, a loud noise, or a noun following a verb is not. Our usage suits the study of perception of physical objects. Even this straightforward and limited definition conceals many complexities. One worth mentioning is what might be called the *relativity* of objects.

The Relativity of Objects

Take an object to be a coherent physical unit, held together by forces and separable, by an action such as lifting or pushing, from other objects. A chair fits this definition, but what about a hydrogen atom or a spiral nebula? These latter examples are not objects, for us at least. What counts as an object depends on both physics and ecology (Gibson, 1966, 1979). When something is very large relative to the human body (e.g., the earth), we tend to treat it as a surface rather than as an object. When it is very small, we can still think of it as an object, but it is no longer detected by ordinary perception or acted upon by ordinary manipulation. It is interesting to ponder how our scientific understandings of the very small and large may implicitly contain aspects of our perception and representation of objects. However, we will not do so here. Closer to our focus, we may conjecture that the relativity of size may change with growth. To an infant, a table may appear as a terrain feature, like a hill. To an adult who can move the table, it is more objectlike. The point about relativity also involves *time*. Something that coheres, but only for milliseconds, will not be an object of our experience. Likewise, an apple and pencil are fine examples of objects, but they will not likely remain coherent over centuries. Finally, consider *forces*. How strongly or weakly matter must cohere to be a unit or to allow separation is relative to the capacities of the organism.

We have hardly done justice to the complexities of defining objects and elaborating their ecological basis, but we have some basis from which to proceed. Physical coherence and boundedness at the levels of scale and across the transformations most relevant for human functioning are the roots of, and motivations for, object perception.

B. The Function of Object Perception in Early Development

Before we embark on our excursion into early object perception abilities, a word is in order about the special function of object perception in infancy. Ecologically, it is obvious that perceiving objects allows humans and animals to obtain nutrition, avoid obstacles and predators, recognize conspecifics, return tennis serves, and make cellular telephone calls. It is striking that, early in development, human infants do virtually none of these things. By 5 months, an infant may reach for an object; by 7 months, she may crawl, and by 12 months, walk. These milestones, however, do not equip an infant to feed or protect itself (or even make phone calls). Yet this same infant, from its earliest days, possesses sophisticated object perception abilities. These have blossomed by 3–4 months and are adultlike by one year.

It may be argued that the function of these abilities in infancy is different from their function in adulthood. The young infant is not so much doing things with objects as exploring them. Much of what infants do serves primarily the process of learning about the physical and social worlds (Piaget, 1952, 1954). This difference in the task of infant and adult perception may have implications for the priorities of perceiving (Kellman, 1993). The adult may need split-second reactions to sometimes tentative information. The infant is not capable of rapid response, but must acquire *accurate* information. Consequences of misperception may be more profound in infancy, and opportunities for error correction more limited. For this reason I have conjectured that infant perception might be *risk averse* in that it is limited initially to those information sources of highest ecological validity. Of the variety of sources available to adults, we might expect that young perceivers will initially use those with the greatest accuracy in indicating what the world is really like (Kellman, 1993). We will return to this conjecture as we survey early object perception abilities.

C. A Taxonomy of Object Perception Abilities

Seeing an object means knowing something about the physical world. Certain chunks of the physical environment cohere: They function as units through various events. By the same token, they are separable from other objects and surfaces. A toothbrush may rest against the inside of a cup, which in turn rests upon a surface. When the toothbrush is lifted, its handle and bristles all move together, but no part of the cup or underlying surface moves with the toothbrush. We do not have to perform the action of lifting the toothbrush to know this outcome; it is easily *seen* in advance. Predictability about how things will cohere, separate, and function is the remarkable achievement of object perception.[1] It is central to most of behavior and thought.

Central but not simple. Seeing objects seems effortless and immediate, but the phenomenology conceals many mysteries. First, the structure of the physical world is not obvious in the array of energy that reaches the eye. Physical linkages and three-dimensional (3-D) arrangements are not given by simple properties of reflected light. Consider a convenient representation, used in computer graphics, of an array of light projecting from a scene. In a digitized image, we note for each location (pixel) numbers indicating luminance and spectral values. Strikingly, the pixel map contains no explicit information whatsoever about objects. Moving from one pixel to another, there is no indication that we move from one object to another in the scene.

This is not to say that the pixel map or the projection to the eyes does not contain information. If the latter did not, we could not see; if the former did not, computer vision would be a hopeless dream. But there is much work to do to make that information explicit. A first step is *edge detection*. Since objects are often made of different materials, boundaries between objects will often, but not always, produce optical discontinuities in luminance, color, or texture. Discontinuities in depth and in motion also indicate boundaries, even in the absence of other information (Gibson, Kaplan, Reynolds, & Wheeler, 1969; Julesz, 1971). Not all luminance discontinuities correspond to the boundaries of objects. Some are textural markings on a continuous surface; others are shadows, and so on. Likewise, there may be depth discontinuities within a single object, if its visible parts are at different depths and it is partly self-occluding. Movements of a nonrigid object may produce internal motion discontinuities, yet it may still be a physical unit. Therefore, a necessary next step in perceiving objects is *edge classification*. We need to know which luminance, depth, and motion contours specify object edges as opposed to other phenomena. Close on the heels of edge classification arises the issue of *boundary assignment*. Most visible contours mark the boundary of one object against a background that continues behind (Koffka, 1935). An example is a picture hanging on a wall. The visible contour at the edge of the frame is a boundary of the frame, but not of the wall. Important to perceiving objects is the correct assignment of which way each boundary bounds (c.f., Kellman & Shipley, 1991).

Unit Formation

Objects are continuous in space and persistent over time, but their effects on our senses are not. Behind the observation that most boundaries bound in one direction lies perhaps the hardest problem of object perception, what might be called the *fragmentation problem*. This problem is both spatial and temporal. Spatially, when objects continue behind others, they often project to separate locations on the retina. Temporal fragmentation arises from the fact that as objects and/or observers move, parts of objects go out of and come into view. We look briefly at each of these aspects of the fragmentation problem.

a. Spatial Fragmentation: The Problem of Occlusion

If one could solve all the riddles of edge detection and classification, it would be a great achievement, but it would not suffice to explain object perception. Consider a scene projected to an observer's eyes. Its projection contains areas homogeneous in lightness, color, and/or texture. Between these areas are edges. Are the homogeneous patches, encompassed by edges, objects? No. In general, most objects are partly concealed behind others. To get from the regions delimited by edges to perceived edges requires several minor miracles. Figure 1 gives an example. How many objects are in the display shown in A? To an adult observer, the display is immediately seen to contain three large objects of definite shapes (shown in B, C, and D) and a number of thin objects resembling blades of grass. Each of the grasslike objects is also seen effortlessly as a definite, separable object, although some effort is required to count them in this cluttered scene. (There are 12.) A much more difficult task is to count the number of relatively homogeneous regions that are connected to form each object. The object shown in C is perceived in display A by combining 11 regions; object

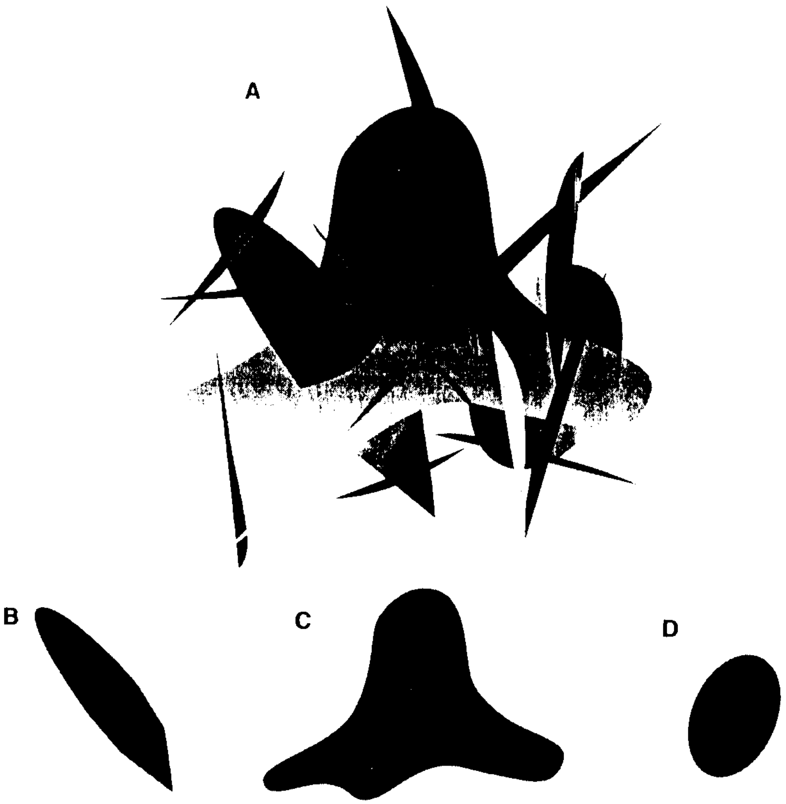


FIGURE 1 (A) Example of spatial fragmentation in unit formation. (B–D) The unoccluded objects shown in B, C, and D are readily perceived in A despite the numerous instances of partial occlusion and luminance variation. (See text.)

D is formed from 8 regions; and object B from 5 regions. Strikingly, when we look at the complex display, we take away little information about the shapes of the visible regions.

Figure 1A illustrates many of the perplexities that need to be overcome in unit formation. Technically, few of the regions are homogeneous in the simplest sense, because most contain luminance gradients (gradually changing shades of gray). Parts of single objects appear in separate visible areas. The complete objects are not seen because they are expected or familiar; for example, the shape in C was designed to be novel, yet it is seen effortlessly. Connections form between parts with different luminance, or different gradients of luminance, as is true to some degree for objects A, B, and C. Areas

of similar luminance are not always seen as connected. The figure contains several examples of cases in which objects are imaged against areas identical in luminance; here, the visual system generates *illusory contours*, such as occurs for the middle portion of the thin, black, vertical object at the top of the display. Thus, spatial gaps in the projections of objects are handled by the visual system by construction of occluded boundaries, behind other objects, and illusory contours, in front of other objects. The object shown in B has an even stranger existence in Figure 1A: Its visible parts are partly in front of and partly behind object C, and these parts connect by *passing through* object C!

Occlusion is pervasive in ordinary environments, and our example in Figure 1A is a simple case, compared to some. The ubiquity of partial occlusion derives from very basic facts: Light moves in straight lines; most objects are opaque, and environments usually contain objects at different distances from the observer. Luckily, human perceivers possess visual processes equal to the physical demands of occlusion: in Figure 1A, they turn the chaos of 45 projected regions into 3 objects and some stray foliage. Below we consider when and how these processes originate.

b. Temporal Fragmentation

Motion of the observer and of objects causes constant changes in patterns of occlusion. Also, because our acuity is relatively poor outside of a small region in the center of the visual field, we register even stationary environments by frequent changes of gaze. Clear views of a given object may be given to our retinas for one or two hundred milliseconds, or even several seconds in some cases, but seldom much longer.

Recovery of linkages and continuity of objects despite spatial and temporal fragmentation in the input to the eyes is called *unit formation*. Sometimes this process is described as *segmentation and grouping*. Either way, a fundamental issue in the perception of objects is how we determine coherent and persisting structures in the environment from intermittent and fragmented stimulation.

Having units in hand (or in sight), we arrive at more familiar object properties. When we do things with objects, when we recognize something as familiar or use them for a particular purpose, *three-dimensional form* and *size* are important properties. Finally, we have so far omitted one of the most important properties of all: perception of the *tangibility* or *substance* of objects.

To summarize, our taxonomy of object perception abilities includes:

1. Edge Detection
2. Edge Classification
3. Boundary Assignment

4. Unit Formation
5. Three-dimensional Form Perception
6. Size Perception
7. Perception of Substance

Our strategy in the remainder of the chapter will be to consider what is known about the developmental course of each of these abilities. For some of these abilities, it will be necessary to say a bit about how they work in adults and about the information available for perception. Most of the findings we will discuss have come in the last two decades. As we will see, more is known about some pieces of the puzzle than others, but on the whole, it is remarkable how complete a picture has been developed in a relatively short period of time.

II. EDGE DETECTION

Do infant perceivers detect edges from the very beginning? If so, how? Of the information used by adults for edge detection, such as luminance, texture, motion, or depth discontinuities, which are available initially and how do others come to function? There has been little explicit research on edge detection. Early performance on pattern and shape perception tasks, however, allows us to make some inferences.

Infants respond to differences among shapes and patterns from birth. For example, Fantz, Fagan, and Miranda (1975) presented patterns in pairs to newborn infants and found reliable fixation preferences. Preference for one of two patterns indicates detection of differences, perhaps implying detection of edges. This interpretation is consistent with another aspect of the data: Outermost contours of patterns were most important in evoking preferential looking. When different patterns were encompassed by a similar surround, such as an enclosing square, preferences were reduced.

Edge detection may not be proven by these results, however. Different patterns might evoke differential interest without their spatial boundaries being detected. Strange as it seems, there is some support for this hypothesis. Early stages of visual processing have been successfully modeled using linear systems analysis (DeValois & DeValois, 1988; Graham, 1989). Any two-dimensional image can be decomposed uniquely into a set of sinusoidally varying luminance components (often called Fourier components, after the mathematician), each having a particular orientation, spatial frequency, and phase. Initial encoding of patterns in the visual cortex involves cells sensitive to such components in local areas of the image. As in the example of the pixel map discussed earlier, this representation in human vision does not make explicit information about object boundaries. This is especially true if *phase* information (spatial relations among components) is left out.

If two patterns are encoded as sets of Fourier components, any detectable difference in their components or amplitudes (contrast) in the two patterns may be sufficient for telling them apart. It has been suggested that infants in the first several weeks of life actually respond on this basis (Banks & Ginsburg, 1983; Braddick, Atkinson, & Wattam-Bell, 1986). Banks and Ginsburg (1983) attempted to account for reported pattern preferences using measures gleaned from the Fourier amplitude spectra of patterns. They obtained good fits of their predictive measures to previously published data on infants' pattern preferences. Their predictive measure included no phase information. Patterns made up of the same spatial frequencies, but with different phase relations, appear radically different to adults, especially in terms of perceived edges. Braddick et al. (1986) carried out explicit tests of infant pattern discrimination for patterns having identical spatial frequency components but differing in phase. They found no evidence of phase sensitivity in infant pattern perception before 2 months of age. Without any phase information, edge detection and classification might be problematic.[2]

Other behavioral evidence, however, casts doubt on the idea that newborns lack edge detection abilities. In an important series of studies, Slater and colleagues (Slater, Mattock, & Brown, 1990; Slater & Morison, 1985) have provided strong evidence for size and shape constancy in newborn infants. Size constancy refers to the ability to detect physical size across variations in retinal size. Shape constancy in this case refers to infants' ability to detect the same planar (two-dimensional) shape across variations in slant toward or away from the observer. (See Section VIB and VII below.) It is hard to invent an explanation for the shape constancy results that does not require detection of boundary orientation in three-dimensional space. The argument from size constancy results is less direct. Detection of object depth is implied by size perception under the circumstances used. If other visible surfaces, as well as the target objects, are assigned depth appropriately, then the object's boundary cannot be a textural marking on a surface. Despite the indirectness of the size argument, these findings taken together suggest that edge detection and classification are possible even for neonates.

It is not clear how to resolve the inconsistency between behavioral evidence on pattern perception in the first 2 months of life and evidence suggesting the early absence of phase information. One possibility is that pattern discrimination in the relatively simple, meaningless, and periodic patterns used by Braddick et al. understates infants' capabilities for using phase information in detection of object boundaries. In particular, it may be the case that particular kinds of phase relations are important. An interesting property of an abrupt luminance edge is that at the location of the edge, many Fourier components of different frequencies will be in the same phase. Marr and Hildreth (1980) proposed that edge detection occurs at several different levels of scale, and that registration of an edge at multiple levels is particularly good information for surface edges. Perhaps infant edge detection abilities are better engaged by edges of ordinary objects than by stimuli composed of two spatial frequency components. Nevertheless, the data are clear in pointing to a conspicuous improvement in phase sensitivity over the first 6–8 weeks.

Other factors limiting the precision of edge detection are orientation selectivity, acuity, and contrast sensitivity. Some research indicates that orientation selectivity is not well developed until about 1 month of age (Atkinson, Hood, Wattam-Bell, Anker, and Tricklebank, et al., 1988; Held, 1993), although it does appear to function to some degree in newborns (Atkinson et al., 1988). Much of the fine detail and many of the low-contrast edges perceptible by adults are not detectable by infants. On the other hand, adult sensitivity far exceeds the minimum required for normal perception of objects and events. As noted by Hofsten (1983), newborn human visual acuity approximates that of an adult cat. Infant sensitivity improves quickly from birth to about 6 months (Banks & Dannemiller, 1987). We can conclude that edges of nearby and large objects may be perceivable quite early, but, in general, luminance edges are registered far less precisely in the first few months than later on.

Another source of information for edges derives from motion relationships. When objects move relative to each other, or when an observer moves while viewing a stationary array, visible texture on more distant objects is progressively *accreted and deleted* as it is revealed or occluded by nearer objects. This form of information supports perception of object boundaries and form by adults, even in the absence of other information, such as in displays comprised of random dots (Gibson et al., 1969; Kaplan, 1969). Kaufmann-Hayoz, Kaufmann, and Stucki (1986) found evidence that 3-month-old infants detected and discriminated shapes based solely on this kind of information. Studies with younger infants would be useful to determine when this ability first arises.

III. EDGE CLASSIFICATION

Evidence for early shape perception implies some capacity to *detect* edges. Implications for edge *classification* are less clear-cut. Detection of a planar object tilted in three-dimensional space may indicate that the shape is seen as an object separate from the background (Slater & Morison, 1985). If so, the edges are not only detected but classified as object boundaries. In this situation, depth discontinuities may provide crucial information. What about perception of figure and ground from luminance or color boundaries alone? Infants detect and discriminate shape from such information. Does this imply that they classify contours as object boundaries? Maybe. But it is also possible that a shape may be seen as a textural marking on a surface rather than as a separate object.

There is some evidence in favor of the latter possibility. Early in life, luminance, color, and texture discontinuities may not indicate object boundaries, a possibility raised by Piaget (1954). Observing his own child, Laurent, at 6 months, 22 days of age, he noted:

> Laurent tries to grasp a box of matches. When he is at the point of reaching it I place it on a book; he immediately withdraws his hand, then grasps the book itself. He remains puzzled until the box slides and thanks to this accident he dissociates it from its support.

When one object rests upon another, differences in luminance, color, and texture at the boundary and possibly the symmetry of each object readily lead adults to perceive object boundaries. Until a relatively late age (Piaget estimated 10 months), infants do not utilize this information, treating adjacent, stationary objects as unitary. In contrast, relative motion, such as the matchbox moving relative to the book, provides information about object boundaries much earlier.

Experimental research has solidified these interpretations. Spelke, Breinlinger, Jacobson, and Phillips (1993) presented displays in which the relationship of two object parts was varied. In homogeneous displays, the two adjacent parts were identical in luminance, color, and texture, and shared a smooth boundary. In heterogeneous displays, the two parts differed in luminance and color, and there were discontinuities (T junctions) where the two parts met. After familiarization with a display, infants viewed an event in which either both parts together or only the top part of the array was lifted. Greater looking time, relative to a baseline condition, to the event in which only the top part moved was interpreted as evidence that the original array was perceived as a single, connected object. Consistent with earlier results, infants at 3 months of age perceived the homogeneous and heterogeneous displays as connected, indicating that neither discontinuities along the outer boundary nor discontinuities of luminance, color, or texture are used for parsing arrays into objects. Results at 5 and 9 months were somewhat ambiguous, but they were consistent with some use of luminance and/or boundary discontinuities as information for object boundaries.

Hofsten and Spelke (1985) carried out experiments using infants' reaching behavior. Five-month-old infants were presented with displays consisting of a small, near object, a larger, further object, and an extended background surface. In some conditions the larger object moved, either rigidly with the background surface or with the smaller object. Infants' reaches were recorded under conditions with different spatial and kinematic relations between the objects. It was assumed that reaches would be directed to perceived boundaries of graspable objects. Much as Piaget (1954) had observed, when the array was stationary and the objects were adjacent, infants reached more to the edges of the larger, further object. This result suggests that infants perceived the two objects as a unit and distinguished the unit from a large extended background surface (as Piaget had also noticed). When the two objects were separated in depth, infants reached more to the nearer, smaller object, suggesting that depth discontinuities provided sufficient information for object segregation. Results of other experiments validated the role of motion in object segregation: when the larger object moved differently from the smaller object, more reaches were directed to the smaller object. Because this effect occurred when the small object was stationary, the findings disconfirmed the idea that infants merely reach for visible moving surfaces. Instead, these results support the idea that motion segregates objects and infants reach for perceived objects (usually the nearest in an array; Yonas & Granrud, 1984).

The informativeness of motion-carried information in infant object segregation was also demonstrated by Granrud et al. (1985). Accretion–deletion of texture, given by observer or object movement, can specify to adults not only contours but object boundaries (Gibson et al., 1969; Kaplan, 1969). Granrud et al. found that infants reached preferentially for the object specified as nearer by the accretion/deletion information.

We can make functional sense of these results. In terms of ecological validity—how well an information source specifies some aspect of the environment—some sources of object boundary information are better than others. Luminance and color changes are ambiguous: They characterize both textural variation on continuous surfaces and boundaries of objects. Depth and motion discontinuities, on the other hand, are unlikely to occur within a unified object. They are thus highly valid indicators of object boundaries. Early edge classification appears to fit the hypothesis that initial perceptual abilities rely on the most valid sources of information (Kellman, 1993). If there is an early stage in development in which luminance/color edges do not mark object boundaries, it remains an important question for future research when adjacent objects begin to be parsed based on their surface qualities, and what brings about this developmental advance.

IV. BOUNDARY ASSIGNMENT

Boundary assignment must occur whenever an object is seen as in front of another object or background. Evidence that infants distinguish shapes, or figures from grounds, might indicate that boundary assignment is occurring. It is not easy, however, to prove that a shape perceived is an object shape, as opposed to a marking on a surface or the shape of a hole.

There is little research that directly addresses this finer distinction relevant to development of boundary assignment. Again, the Slater and Morison (1985) shape constancy result indirectly supports both edge classification and boundary assignment, probably from discontinuities in depth at object edges. Accretion/deletion of texture also seems to determine boundary assignment (Granrud et al., 1985), since infants reached for the surface specified to be nearer by accretion/deletion, presumably indicating it "owned" its boundary. A limitation on reaching evidence is that infants do not achieve consistent, directed reaching until around 5 months of age. By this time infants have had considerable visual experience.

A way to infer younger infants' abilities involves another functionally interpretable behavior: infants' defensive responses to approaching objects. Head withdrawal, eyeblinks, and hand raising have all been reported as defensive behavior to approaching objects, even in the earliest weeks of life (Schiff, 1965; Yonas, 1981). Although there have been interpretive controversies, evidence supports the notion that some observed responses, especially blinking, indicate defensive behavior (Yonas, Arterberry, & Granrud, 1987). Defensive responding to a looming object may imply boundary assignment, but it does not specify much about the relevant information. Carroll and Gibson (1981) attempted to pinpoint effects of accretion/deletion of texture. They presented 3-month-old infants with arrays in which all surfaces were covered with random dot texture. In one condition a textured object approached the infant, whereas in the other condition, an aperture (opening in the surface) approached. This difference in physical events was specified by different accretion/deletion effects. Infants showed defensive behavior more frequently to the approaching objects than to apertures, suggesting that accretion/deletion of texture indicates object boundary ownership even at this early age.

V. UNIT FORMATION

Organizing the world into units is really the defining problem of object perception. Detecting many aspects of objects, for example, shapes, sizes, and familiarity, presupposes the successful operation of unit formation, that is, processes that determine connected regions and their boundaries. Unit formation is thus a foundation of many achievements in cognitive and social development. Some examples are social attachment (Regolin & Vallortigara, 1995), categorization of objects, mathematical skills such as counting, linguistic skills such as naming objects, and many more.

For adults, unit formation occurs routinely and effortlessly. This simplicity masks the fact that it depends on multiple sources of information, which may have different developmental origins. The Gestalt psychologists described a number of influences on unit formation (Michotte, Thines, & Crabbe, 1964; Wertheimer, 1923). The role of motion was expressed as a principle of *common fate*. Things that move together tend to be grouped together. Other principles applied to stationary arrays. By the principle of *good continuation*, things appear connected if their contours continue smoothly instead of changing abruptly. *Good form* refers to a tendency to organize arrays so that simple, symmetrical objects are perceived. The principle of *similarity* suggests a tendency to unify similar parts or areas, and *proximity*, a tendency to group closer things together.

The Gestalt descriptive principles do not permit precise or quantitative predictions. Part of the problem comes from terms that resist clear definition. These include the "goodness" in good continuation and good form, "simplicity," "common fate," and the notorious "similarity." Nevertheless, the principles contain important insights, most of which can be readily illustrated.

A. Two Processes in Unit Formation

Some contemporary work has made progress in giving more precise form to the Gestalt principles. Kellman and Shipley (1991) proposed dividing information for unity into two categories: the *rich* or *edge-sensitive* (ES) process and the *primitive* or *edge-insensitive* (EI) process.

The EI process is an elaboration of what Wertheimer (1923) called *common fate*. The process is edge-insensitive because the positions and orientations of the edges of visible parts play no role in determining their completion behind the occluding object. Certain motion relationships alone indicate connectedness. This information does not specify the exact form of the hidden parts under occlusion. For this reason Kellman and Shipley (1991) labeled it the "primitive process" (c.f., Hebb, 1949).

The ES process depends on edge positions and orientations, both in stationary and moving displays. Many of its formal properties, and some of its neural mechanisms, have been elucidated in recent years (Field, Hayes, & Hess, 1992; Kellman & Shipley, 1991; Polat & Sagi, 1993; Shapley & Gordon, 1987; von der Heydt, Peterhans, & Baumgartner, 1984). The input-output relations in this process may be thought of as a mathematical formalization of the Gestalt principle of good continuation, that is, that segmentation and connection of parts depend on straight lines and smooth curves.[3] Detailed models of the ES process may be found elsewhere (Grossberg, 1994; Kellman & Shipley, 1991, 1992). For our purposes, two points are most important. First, for adults, certain edge relationships support object completion (such edges are termed *relatable* edges), whereas others do not. Figure 2 gives some examples of relatable edges and nonrelatable edges. Second, the boundary interpolation process at work in occlusion cases is the same as in illusory figures (Kellman, Yin, & Shipley, 1995; Shapley & Ringach, 1994; Shipley & Kellman, 1992a). This identity allows us to obtain converging evidence on the developmental origins of the ES process.

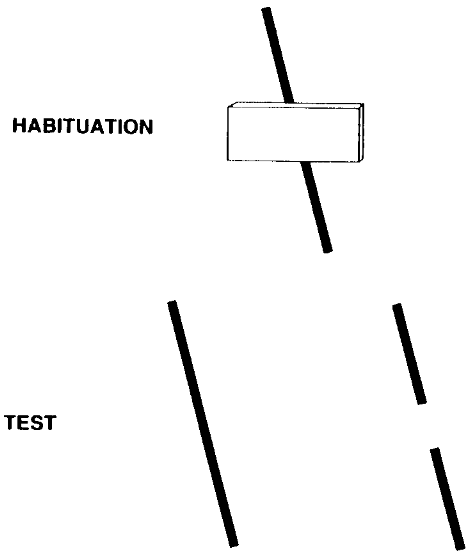


FIGURE 2 Relatable and nonrelatable edges. The ES process (see text) connects the visible black parts in each display when the edges are relatable (*top*), but not when the relatability criterion is not met (*bottom*).

Perception of object unity and boundaries was studied programmatically by Kellman and Spelke (1983) using habituation/dishabituation of visual attention. If infants perceive a partly occluded object as complete, then after habituation to such a display, they should generalize habituation more to an unoccluded complete object than to an unoccluded display containing separate pieces that correspond to the previously visible parts of the object. Figure 3 illustrates the paradigm. After habituation to a center-occluded display, broken and complete test displays are presented in alternation.

B. Common Motion and Relatable Edges in Combination

To test whether the EI process (common motion) or the ES process (completing relatable edges) functions in early infancy, an occlusion display combining these was constructed. It consisted of two visible, collinear parts of a rod, sharing a common lateral motion. Infants were habituated to this display and tested afterward with alternating presentations of two unoccluded test displays: a moving complete rod and moving broken display, consisting

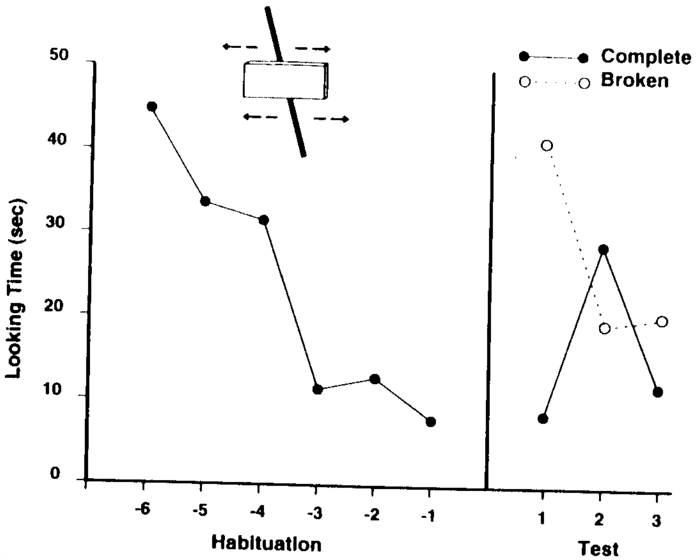


FIGURE 3 Design for habituation studies of occlusion. Infants are habituated to a partly occluded display and tested afterward for generalization of habituation or dishabituation to an unoccluded complete display and an unoccluded "broken" display containing two pieces separated by a gap where the occluder had been. (After Kellman & Spelke, 1983.)

of two rod pieces separated by a gap. Results are shown in Figure 4. After habituation, infants generalized habituation to the complete display and dishabituated to the broken display. This pattern suggests that the occlusion display was perceived as containing a unified object rather than two separate rod pieces. These data alone do not specify whether the EI (common motion), ES process, or both were responsible for unit formation.

C. Common Motion Alone: Revealing the EI Process

To assess the effect of common motion alone, Kellman and Spelke (1983) tested an occlusion display in which edge relationships would not be expected to support boundary completion. One visible piece was a black rod; the other was a red blob with black textural markings constructed randomly within certain constraints (see Figure 5). These two visible parts were not relatable nor similar in surface qualities. They did share a common lateral translation. Results indicated that here, too, infants perceive the visible parts as connected behind the occluder. The broken test stimulus, which included only the parts previously visible in the occlusion display, produced strong recovery of visual attention after habituation. The complete test display, constructed by continuing the rod halfway down and the random blob halfway up, induced little recovery.

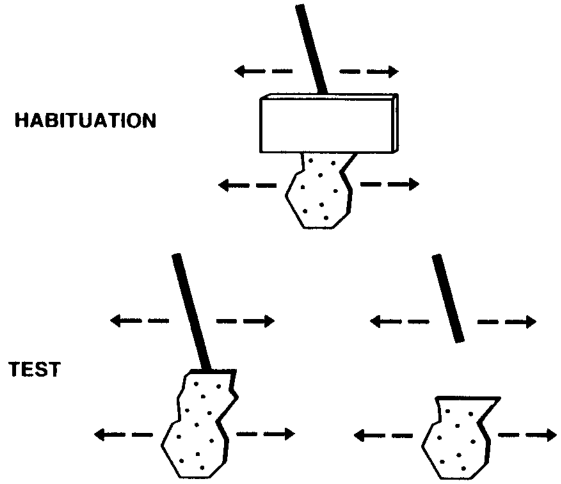
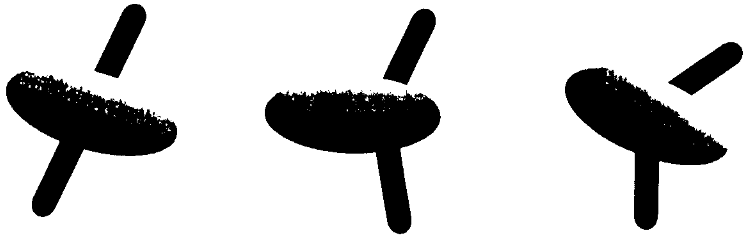


FIGURE 4 Results of experiment testing unity perception from common motion and relatable edges. Infants were habituated to two aligned, visible parts sharing a common lateral translation (*top*). Looking times are shown for the last six habituation trials (with the final one labeled −1 in Figure 3) and the test trials. Test trials consisted of successive presentations of unoccluded complete and broken displays, with half of the subjects seeing the complete display first. (From Kellman & Spelke, 1983.)

Relatable Edges



Non-Relatable Edges

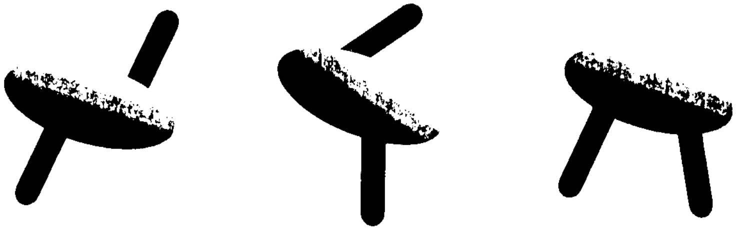


FIGURE 5 Displays for study of EI process alone. (See text.)

The outcome indicates that common motion alone can specify the unity of a partly occluded object to infants. These results occurred despite the fact that the specific form connecting the two visible parts was not specified. Visual attention in this situation seems to be controlled by unity, not specific form, a result confirmed by subsequent research (Craton & Yonas, 1990).

1. Motion Relationships in the EI Process

To say that common motion leads to perceived unity leaves much unsaid. When is motion "common" and what counts as "motion"? Later research addressed both of these questions.

Common motion must include at least the identical lateral translations shared by the object parts in the studies described so far. The class of unifying motions, however, might be larger. Rigid objects can undergo any combination of translations and rotations. One hypothesis, then, is that the two parts will appear connected if related by a rigid motion in three-dimensional space. A different way of thinking about the earlier displays is that each visible part underwent the same visible *event*. Kellman, Spelke, and Short (1986) contrasted these possibilities by testing infants' perception of a rod vertically translating behind an occluder. In this display, the two pieces shared a common translatory motion, but the visible events differed. As one visible part came more into view, the other progressively disappeared. Results supported the rigid motion hypothesis: Infants responded as if a single complete object were perceived in the occlusion display.

Another test of the rigid motion hypothesis involved rigid translation toward and away from the observer in depth. In this case, the stimulus for motion perception is not displacement of the object's projection across the retina, but optical expansion/contraction of the projection (or oculomotor changes in accommodation or convergence as the object moves). If unit formation in the EI process depends on the particular stimulus of optical displacement, then a depth-translating object should not engage the process. In contrast, on the rigid motion hypothesis, such a display should produce unit formation. Kellman et al. (1986) found clear evidence confirming the latter prediction.

A limitation on the class of rigid motions effective in unit formation was reported by Kellman and Short (1987a). Rotation of an object in the frontal plane is a rigid motion, but the two visible parts of the center-occluded object move in opposite directions. This rotation display, as well as others combining rotation and translation, yielded no evidence of unit formation at 16 weeks. A display in which the rotation and translation were phased so that the top and bottom generally always moved in the same direction (looking much like a windshield wiper) did support unit formation. It appears that the class of motions engaging the EI process in the early months is not the full class of rigid motions. In particular, common direction of visible parts seems to be required.

2. What Is Motion?

Motion ordinarily refers to the changing positions of objects in space. In studies of the EI process described so far, moving objects have been involved. Perceptually, motion of an object is detected by means of certain optical events; for example, lateral translation produces optical displacement.

Helmholtz (1885/1910) noted the similarities between optical events produced by motion of an observer and the motion of an object seen by a stationary observer. This similarity leads to the question: Does the EI process depend only on perceived motion of objects or on certain optical events? Could it, for example, be engaged by optical changes given when a moving observer views stationary objects? The question has many implications for perceptual development, only some of which we can consider here. A key question is whether infants can tell the difference between object motion and optical effects generated by observer motion. It has often been asserted (Helmholtz, 1885/1910; James, 1890) that young perceivers cannot distinguish these cases. If so, our question about the basis of the EI process is answered, or moot, at least until infants can tell the difference. J. Gibson (1966, 1979) suggested a more optimistic view of early motion perception. He noted that there is optical information distinguishing the case of observer movement from object motion. When an object moves, there are changes relative to the background that are in general different from what happens when the observer moves while viewing a stationary array. (Discussion of these differences and their role in infant perception may be found in Kellman, 1995.)

Whether unit formation depends on perceived motion or on optical change, and whether infants can tell the difference, were investigated by Kellman, Gleitman, and Spelke (1987). They tested moving infants in two conditions. In a *conjoint motion* condition, the infant's chair and a partly occluded object were rigidly connected beneath the display table, so that they rotated around a vertical axis in between. The conjoint motion condition contained real motion of the object in space, but no physical displacement relative to the subject. If real object motion underlies unit formation, and if it is detectable by a moving observer, this condition was expected to lead to perceived unity. In the other condition, the *observer movement* condition, the observer moved in an arc while viewing a stationary occluded rod. If optical displacement alone can specify unity, infants were expected to perceive a complete object in this condition. Unit formation was assessed, as in earlier studies, by dishabituation patterns following habituation to the occluded displays. Motion characteristics of the object parts were always the same in the test trials as in habituation.

Results were clear. Infants in the conjoint motion condition dishabituated robustly to the broken display, indicating that the occlusion display had been perceived as complete. (Fifteen of sixteen infants in this condition looked at least twice as long at the broken than at the complete test displays on the first test trial.) In the observer movement condition, there was no evidence of unit formation; infants dishabituated equally to the two test objects. Separate analyses, based on looking time differences to moving and stationary displays, indicated that conjoint motion infants perceived object motion during their own motion, whereas observer movement infants accurately perceived their occlusion display as stationary (Kellman et al., 1987).

These findings indicate that the EI process depends on perceived object motion. This outcome has significance for two more general theoretical issues in perceptual development. The first is the idea that early perception depends on sources of information of highest ecological validity. The nature of the EI process confirms this conjecture, for the following reason. The ecological validity of common motion of objects is much superior to that of common optical displacements. The latter occur any time an observer moves and views objects at similar observer-relative distances. It is sometimes the case, but not always, that object parts at similar distances are connected. On the other hand, real motion, such as common rigid translation of visible parts, almost never occurs for unconnected entities. Even when separate objects are subjected to the same force, as when leaves are blown by the wind, their motions virtually never maintain completely rigid relationships. (This situation can, however, be arranged in infant perception laboratories!) The EI process depends on a nearly foolproof principle. When an infant perceives two partly hidden things moving with a rigid relationship as physically connected, they are.

3. Is the EI (Common Motion) Process Innate?

Kellman and Spelke (1983) suggested that the common motion process is an unlearned foundation of object perception. Infants utilize this information at an early age, before other unit formation information operates. The data disconfirm proposals that object perception derives from action- or touch-based learning, because common motion leads to perceived object unity before the onset of skilled reaching or crawling.

Recent research, however, suggests that the common motion principle may not be present at birth. Slater, Morison, Somers, Mattock, Brown, and Taylor (1990) tested infants at 16 weeks and replicated the findings of Kellman and Spelke (1983). Their tests of newborns under identical conditions, however, led to a different result. After habituation to an occlusion display, newborns showed a preference for the complete object compared to a broken display. Newborns consistently showed this pattern after habituation to moving displays in several studies varying the depth separation of the occluder and partly occluded object (Slater, Johnson, Kellman, & Spelke, 1994). Slater et al. (1994) discuss two classes of explanation for this outcome. One possibility is that newborns lack basic visual sensitivity to detect the common motion of the parts. Direction selectivity of cells in visual cortex seems to mature between 1 and 2 months of age (Johnson, 1990; Wattam-Bell, 1991, 1992). It is surprising that before this time there is no behavioral or electrophysiological evidence that infants detect differences in motion direction (Wattam-Bell, 1991, 1992). Therefore, in occlusion situations, newborns may detect motion but not directional coherence of separate parts. If so, perception of separate, bounded fragments is not surprising. Newborns may *segregate* objects by motion at this stage but may not receive the information required for perceived unity.

The second possibility is that infants really do begin with an incorrect perceptual rule, assigning occlusion edges as object boundaries (Slater, Morison, Somers, Mattock, Brown, & Taylor, 1990). The "perceptual inference" (Slater, Morison, Somers, Mattock, Brown, & Taylor, 1990) for connecting visible parts based on common motion might be learned.

These two interpretations cannot be distinguished while infants lack directional sensitivity. The interpretations do make differing predictions about what should happen when directional sensitivity appears. If the EI process is an unlearned basis of unity perception, it should operate as soon as directional sensitivity becomes operative. On the learning account, directional sensitivity would be only the beginning of some learning process. Because the EI process is in place by 16 weeks, we know that if a learning process is involved, it is expeditious and does not require practice at skilled reaching or self-locomotion. These factors may favor a maturational explanation; however, further research would be useful.

D. The Edge-Sensitive Process

Relatability of edges, which underlies the ES process in adult perception (Kellman & Shipley, 1991), does not produce perceived unity for infants in the first half year of life. Typically, after habituation to a stationary, partly occluded rod display, infants show about equal looking times to the complete and broken test displays (Kellman & Spelke, 1983). This often-replicated finding is thought provoking. Greater dishabituation to a complete display might be expected if the visible rod pieces in the initial display were perceived as two separate objects. What might equal looking times, with some dishabituation to both test displays, mean? When infants do not perceive the unity of visible parts emerging from behind an occluder, what might they perceive? Equal dishabituation to both test stimuli suggests that the broken and complete displays are equally consistent with the initial display. A plausible interpretation is that infants at this age are "agnostic" about what happens behind the occluder. They do perceive the fact of occlusion, that the occluding object is nearer than the rod pieces, but their perceptual process renders no verdict on what happens behind the occluder (Kellman & Spelke, 1983). In the absence of the perceptual rules guiding the adult's perception in this situation, being agnostic makes sense. Consider the alternative. Suppose infants start out perceiving visible surfaces as ending where occluding objects intervene. If so, early perception would follow an incorrect perceptual rule: Every stationary, visible part of a partly occluded object would be incorrectly assigned a boundary at the point of occlusion. This chronic misperception would not only handicap early learning about specific objects, but it would impede learning of the correct rule. If rules about object interpolation arise later, whether by learning or maturation, the young perceiver might be better off seeing indeterminately rather than incorrectly.

Infants' inability to use the ES process extends at least through the first half year (Bertenthal, Campos, & Haith, 1980; Schmidt & Spelke, 1984; Spelke et al., 1993). Bertenthal et al. (1980) reported evidence of sensitivity to illusory contours at 7 months of age but not at 5 months. Similar results have been reported for kinetic illusory contours (which depend on relatable edges given sequentially in time) by Kaufmann-Hayoz, Kaufmann, and Walther (1988). Earlier perception of illusory contours was suggested by Ghim (1990), who used a familiarization and preference paradigm with 3- and 4-month-old infants. He predicted that if infants perceived subjective contours, novelty preferences would be greater between a subjective contour display and a display without subjective contours than between two displays with no subjective contours. Some comparisons were consistent with the hypothesis, but at least one predicted outcome failed to occur in each of five experiments. There were also plausible alternative explanations for the preferences that were observed in the studies. To generate control (nonsubjective contour) displays, an illusory square display made from four inducing elements was disrupted by changing the orientation of two or all four elements. This manipulation was elegant in allowing experimental and control display pairs to differ by the same local feature contrasts. Unfortunately, the illusory contour and control stimuli differed in several more global properties that could produce easier discrimination between illusory contour and control displays. For example, the outer perimeter of the illusory contour displays contained only smooth contours, whereas all control displays had either four or eight sharp corners and two or four deep concavities around their outer perimeters. Given these difficulties, these data do not disconfirm indications from other research that the ES process is absent in the first half year of life.

E. Origins of the ES Process

How does the ES process originate? Both neural maturation and learning mechanisms are possible explanations. Noting the onset of a number of pictorial depth cues around 5–7 months of age, Granrud and Yonas (1984) suggested they involve a perceptual module that matures at that time. Likewise, Gunderson, Yonas, Sargent, and Webster-Grant (1993) found that reaching behavior of 7- to 8-week-old macaque monkeys is influenced by pictorial depth cues, suggesting phylogenetic origins of these abilities. Edge-sensitive mechanisms for boundary interpolation might be closely related. In particular, the depth cue of interposition is closely connected to boundary interpolation under occlusion (Kellman and Shipley, 1991). Available evidence suggests that boundary interpolation may arise around this same time (Bertenthal et al., 1980).

Some accounts of the origins of the ES process invoke the traditional empiricist notion that perceiving unity and boundaries depends on experiences with objects (Nakayama & Shimojo, 1992; Spelke et al., 1993; Wallach & Slaughter, 1988). At present, no direct evidence from infant research supports this idea, but it remains a possibility.

There are numerous reasons to believe a contrary hypothesis: that the unit formation phenomena observed by the Gestalt psychologists, and embodied in recent computational models, depend on modular perceptual mechanisms, not recognition processes. Recent neurophysiological work suggests that boundary interpolation processes in occlusion and illusory contour perception are carried out at surprisingly early stages of visual processing, certainly as early as V2 and possibly V1 (von der Heydt et al., 1984). These findings, along with psychophysical results indicating boundary interpolation in cases where no familiar objects are present (Field et al., 1992; Kellman & Shipley, 1991), suggest that computing edge relatability is a basic visual function. It remains possible that effects in V2 or V1 result from some unknown feedback from higher levels, but there is no evidence for this proposition. Other indications that edge-sensitive unit formation is a modular perceptual capacity (c.f., Fodor, 1983) come from evidence indicating that illusory contours and occluded contours are processed by the same mechanisms (Kellman & Shipley, 1991; Kellman et al., 1995; Shapley & Ringach, 1994); local edge relationships override familiarity (Kanizsa, 1979); and the process obeys precise quantitative relationships (Shipley & Kellman, 1992b; Rubin, Shapley & Nakayama, 1995; Lesher & Mingolla, 1993). None of these findings would be expected if object completion depended on familiarity. For example, equivalent strength of boundary completion in illusory and occluded figure cases would not be expected, because occluded boundaries are far more common in ordinary visual experience than are illusory ones. (The latter require an exact visual match of luminance, color, and texture between an object and parts of its background, a situation that is not common.)

Whether or not these considerations are decisive, advancing our understanding of how edge-sensitive unit formation processes arise remains a high priority for future research.

F. Summary: Unit Formation

Unit formation in adult visual perception appears to be governed by two separate processes, what we have labeled EI and ES. The EI process utilizes motion, not edge, relationships, and begins to operate in the early weeks of life. The ES process is richer in specifying not only connectedness of objects but the forms of hidden boundaries, but it is long delayed in development relative to the EI process. The developmental sequence of the two unit formation processes parallels their differing ecological soundness. Coherence in motion is a deep, even defining, property of objects (Spelke, 1985), and if detected with precision, motion relationships are highly diagnostic of unity. Smoothness of object boundaries and connectedness of pieces that bear certain edge relations are common but not nearly universal characteristics of our physical environment. Accordingly, the ES process, sensitive to edge relations given simultaneously or over time, is a robust and useful perceptual process, but not of the highest ecological validity. The development of unit formation, then, fits our characterization of perceptual development as beginning with the most secure information sources and progressing toward more diverse but somewhat less trustworthy sources.

VI. THREE-DIMENSIONAL FORM PERCEPTION

Visual form perception is a great battleground of perceptual theory. Adults perceive three-dimensional (3-D) form from at least three different sources of information (Kellman, 1984), and each of these stands as the canonical example for a theory of form perception. We often perceive—or recognize—the whole form of a familiar object from a single, static view. This ability may depend on seeing a certain 3-D object from various viewpoints. Subsequently, any 2-D view calls up the associated views from memory. This associative account may be taken, as suggested by John Stuart Mill, to define what a 3-D object *is*: It is "the permanent possibilities of sensation." For the visual sense, this view implies that the products of vision are inherently two-dimensional; three-dimensionality can only be realized as the set of possible 2-D views from all vantage points.

The view of an object from a single, stationary point was also emphasized by Gestalt theorists, but for radically different reasons. Even unfamiliar objects could be perceived this way, because the 2-D stimulation sets in motion organizational forces in the nervous system, which in general lead to perception of simple, regular, 3-D forms. Although the Gestalt neurophysiological ideas seem implausible today, form perception might still depend on unlearned organizational tendencies. On the other hand, Brunswik (1956) suggested that laws of perceptual organization might be acquired by experience of object regularities.

Another class of information has become well understood only in the past few decades. Three-dimensional form may be specified by information in continuously changing optical projections, as an object rotates or as an observer walks around an object. It has been argued that this kind of information mathematically specifies object structure without utilizing any assumptions about object symmetry or regularities (Gibson, 1966, 1979; Johansson, 1970; Ullman, 1979).

Each of the several means by which adults perceive 3-D form suggests a developmental account (Kellman, 1984). If 3-D form is a product of accumulated 2-D views, then perceivers may initially have no notion of 3-D form at all. A specific object's form would develop from experience with different views, and perhaps from concurrent, active manipulation of the object (Piaget, 1954). Perceivers would have little competence with, and perhaps no notion of, 3-D form until they had undergone extensive learning. The possibility that perceived 3-D form is a direct response to certain optical transformations is usually linked with the hypothesis of evolved mechanisms sensitive to this kind of information (Fodor, 1983; Gibson, 1966; Shepard, 1984). On this account, perceivers might be sensitive to 3-D form from an early age. Finally, the use of general principles of object completion to derive 3-D form from particular views has been explained in two ways. On the Gestalt view, perceived form results from unlearned, organizational processes rooted in basic neurophysiology. Alternatively, in the position articulated by Brunswik (1949) and anticipated by Helmholtz (1885/1910): rules of perceptual organization might be abstractions from an individual's experience with many objects. These two accounts of how whole form may be gotten from a single view make disparate predictions about development. Gestalt organizational processes should operate as soon as the relevant brain mechanisms are mature, whereas Brunswikian learning has usually been hypothesized to be a protracted process.

A. Kinematic Information in Infant 3-D Form Perception

Which of these accounts of the origins of 3-D form perception is correct? Wallach (1985) raised the interesting possibility that in each perceptual domain (i.e., form, depth, motion), there is some information that is usable innately, whereas other cues are acquired later, perhaps through correlation with the innate process. In form perception, he hypothesized that motion-carried information was the innate foundation. In fact, the motivation for Wallach and O'Connell's classic (1953) studies of the "kinetic depth effect" was to shed light on the development of 3-D form perception. Knowledge of 3-D form seems to be available to congenitally monocular observers despite their having no access to stereoscopic information about 3-D form. Learning might allow 3-D form perception to occur even from pictorial information, but where might the *initial* notion of 3-D form come from? Wallach and O'Connell hypothesized that there must be an unlearned process of 3-D form perception, perhaps based on the optical changes given by motion.

Others have also given theoretical grounds for the primacy of motion-carried information about form. The speed and precision of adult processing of structure from motion (Braunstein, 1976; Johansson, 1975; Ullman, 1979) suggests dedicated neural machinery, especially given the complexity of the information itself. Another reason this information may be the earliest usable by infants is more rooted in developmental considerations. This source of form information has the highest ecological validity. Under reasonable constraints, it can be proven mathematically that perspective transformations contain sufficient information to specify uniquely an object's 3-D form (Ullman, 1979). When a stationary observer views a 3-D object from a single vantage point, its whole form may be predicted on the basis of simplicity, symmetry, or similarity to previously viewed objects. The accuracy of such predictions rests on probabilistic facts about the sorts of objects that exist and vantage points that occur. How often this information signals 3-D form accurately is hard to quantify, but it falls well short of the validity of kinematic information. Early cognitive and social development may be best served by perceivers getting only the most accurate information about 3-D form, even if this information is not obtainable under some circumstances.

These theoretical considerations have been put to empirical test. As predicted by Wallach and O'Connell (1953), evidence suggests that the earliest competence for perceiving overall form appears to be based on kinematic information (Kellman, 1984; Kellman & Short, 1987; Owsley, 1983; Yonas, Arterberry, & Granrud, 1987a).

To test infants' 3-D form perception, one must overcome an intrinsic problem. A viewed 3-D object is seen from a particular vantage point, or a changing sequence of such points. At these vantage points, particular 2-D projections of the object reach the eyes. Perception of 3-D form must somehow be disentangled from responses to these 2-D projections. For example, suppose infants are habituated to a stationary 3-D object from a particular vantage point. After habituation, suppose infants generalize habituation to this same display, but dishabituate to a novel 3-D object. This response pattern might indicate that infants detected the original 3-D form and discriminated it from the novel 3-D form. However, the observed responses could instead be based on differences in the 2-D projections of the original and novel object; 3-D form may not have been perceived at all.

A means of circumventing this problem is based on the geometry of form and motion. Information about a given 3-D form can be provided by rotation around various axes, provided there is some component of rotation in depth. If objects are chosen that are not too symmetric, one can test for detection of invariant 3-D form across rotation sequences that vary and have quite different 2-D appearances.

One experiment of this type (Kellman, 1984) tested 16-week-olds using the two objects depicted in Figure 6. In the kinematic condition, form was tested by habituating infants to videotaped displays of a single object rotating in depth. Two different axes of rotation in depth were used in habituation on alternate trials, so that the only constant from trial to trial was the 3-D form of the object. After habituation, subjects were tested on alternating trials with the same object, now moving around a third (new) axis of rotation, and a different object, rotating around the same new axis. The change to a new axis of rotation in the test period ensured that the particular 2-D views and transformations were novel for both the object shown previously and the new object. Generalization of habituation to the same object would thus reflect extraction of 3-D form, not a response to particular 2-D views.

Besides the kinematic condition, two groups viewed sequential stationary views (photographic slides) taken from the rotation sequences. The two

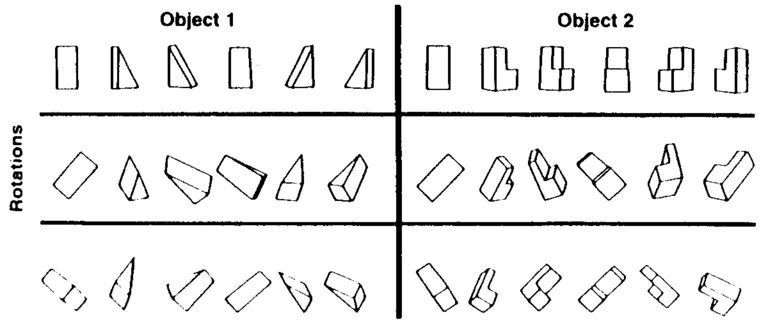


FIGURE 6 Schematic views of 3-D objects and axes of rotation used in a 3-D form perception experiment (Kellman, 1984). Successive views 60° apart are shown. All views in the same column are from the same 3-D object; views in the same row are from the same axis of rotation in depth.

groups differed in the number and spacing of the views. It was possible that infants could detect the 3-D forms of these objects from single views, or sequences of views. Adults can certainly do this; even the line drawings in Figure 6 allow 3-D form to be perceived from most views. If infants detected overall form from single views or sequences of static views, successful performance in the kinematic condition might not indicate use of optical change information; rather, it might indicate that transforming arrays are processed as sequences of static views.

Results indicated that infants perceived 3-D form from kinematic information. Those in the kinematic condition generalized habituation to the same object in a new rotation, but dishabituated to the new object, regardless of the object and particular axes of rotation used in the habituation and test trials. Infants did not appear to acquire 3-D form from static views: they showed no reliable differences in response to new views of an object versus views of a new object.

The finding that young infants have an early ability to perceive 3-D form, but only from continuous optical transformations, is supported by research from several laboratories using a variety of methods (Kellman & Short, 1987b; Owsley, 1983; Yonas et al., 1987a). Additional findings have led to a more precise understanding of this ability.

One interesting prediction was tested by Kellman and Short (1987b). Optical transformations that specify a particular form may be given in principle by either object or observer motion. The specificity of the motion patterns, insofar as 3-D form is concerned, is the same in both cases. Thus, an observer walking in an arc around a stationary object receives the same optical transformations, relevant to that object's form, as those obtained if the object rotates while the observer is stationary. If 3-D form depends on mechanisms sensitive to kinematic information, then moving infants who view stationary objects should detect 3-D form. Kellman and Short (1987b) found that 16-week-old infants did indeed perceive objects' 3-D forms from motion perspective.

1. Isolating Edge Transformations

It is natural to view information about 3-D form as carried by spatiotemporal changes in length and orientation of object edges caused by the object's rotation relative to the observer. However, the transforming optical projection of a rotating (solid) object also contains changes in brightness and texture gradients (Pentland, 1990). Shaw, Roder, and Bushnell (1986) argued that changes in brightness and texture are necessary for infants younger than 24 weeks of age to detect form. To disentangle the contributions of edge transformations from brightness changes during motion, Kellman and Short (1987b) used wire figures similar to those introduced by Wallach and O'Connell (1953). Such figures contain thin edges but no surfaces connecting them. In rotation, such objects provide the same geometric transformations of surface boundaries as do solid objects, but without transformations of surface brightness and texture. Lighting was arranged to eliminate visible shading changes even along the thin edges of the figures. The figures used by Kellman and Short (1987b) are shown in Figure 7.

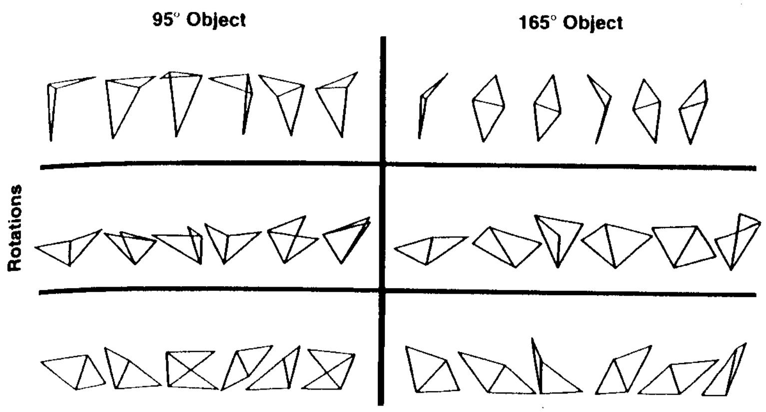


FIGURE 7 Objects and rotations used to test 3-D form perception from edge transformations alone. Successive views 60° apart are shown. (From Kellman & Short, 1987b. Copyright © 1987 by the American Psychological Association. Reprinted with permission.)

These wire figures allowed an additional way of ruling out contributions from 2-D similarities across axes of rotation. Not only were the two test objects designed to be very similar to each other; their structure virtually guaranteed that static, 2-D information could not give away 3-D form. A theorem of projective geometry states that all triangles are projectively equivalent; that is, any 2-D projection of one triangle could be the (polar) projection of any other triangle in some 3-D orientation and distance. By constructing each 3-D figure from two triangles, the overall structure of the object was minimized in this experiment. The effect of this manipulation was validated in an experiment with adults, whose sorting of static views of the two 3-D objects did not differ from chance (Kellman and Short, 1987b, Experiment 3b).

As shown in Figure 8, infants perceived 3-D forms of the wire figures from edge transformations alone. Two groups are shown, each habituated to one of the wire objects, in two different axes of rotation. Each group generalized habituation to the same 3-D object tested in a novel rotation and dishabituated to the new object. These findings do not rule out the possible informativeness of transformations of shading and texture, but show that the latter are not necessary for early 3-D form perception.

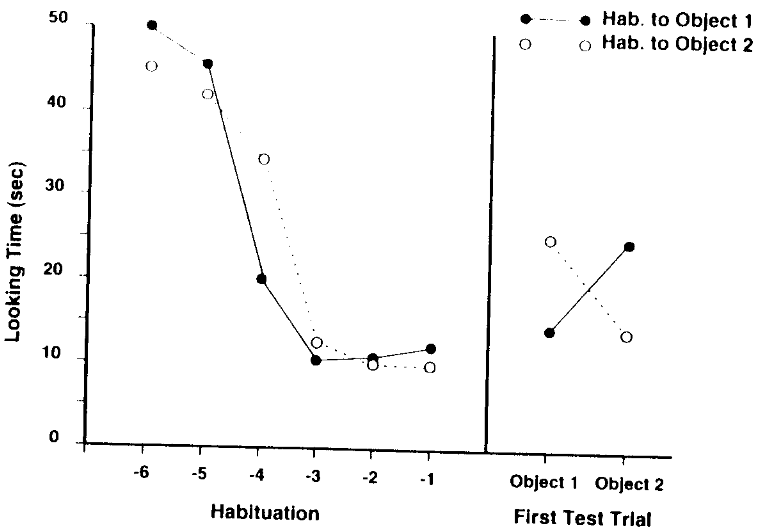


FIGURE 8 Results of experiment on 3-D form perception from edge transformations alone. Looking times are shown for the last six habituation trials (with the final one labeled −1) and the first test trial for each test object. Data are shown separately for infants habituated to each of the two test objects. (From Kellman & Short, 1987b. Copyright © 1987 by the American Psychological Association. Reprinted with permission.

B. Static 3-D Form Perception

In contrast to the infants' early ability to extract 3-D form from motion is their demonstrated inability to perceive whole form from single or multiple static views of objects (Kellman, 1984; Kellman & Short, 1987b; Ruff, 1978).

This failure of static information applies to 2-D stimuli, such as photographic slides, and to stationary views of 3-D objects, as well. Ruff (1978) found that 6-month-old infants failed to apprehend 3-D forms from stationary views of 3-D objects. The objects used were rather complex, however. Kellman and Short (1987b) found the same lack of 3-D form perception from multiple, stationary views of the simpler objects pictured in Figure 6 with infants aged 4 and 6 months, and other work suggests the problem still exists at 8 months (Kellman, 1993). Binocular, static views of objects may allow recognition of 3-D forms that have previously been perceived from kinematic information (Owsley, 1983; Yonas et al., 1987a), but there is no evidence that infants detect the overall form of an object initially from static, binocular views.

Given adult abilities to perceive 3-D form from single views, infants' limitations, persisting into the second half of the first year, are startling. What might be the problem in extracting 3-D form perception from static information? Certainly, some relevant abilities appear to be in place quite early. Earlier we noted that Slater and Morison (1985) tested newborns (mean age: 2 days, 8 hours) for perception of the constant shape of planar objects (e.g., flat rectangles and trapezoids) despite variations in their slant in depth. Results indicated that newborns detect invariant planar shape despite slant variations. Earlier work by Caron, Caron, and Carlson (1979) had found such an ability with 12-week-old infants. Although there is little discussion in these reports of the particular information underlying such performance, it appears that depth information must be combined with projective shape to determine actual planar shape. There are indications that binocular convergence provides the useful depth information in newborn shape and size constancy. (For discussion see Kellman, 1995.)

The results with planar shape perception by newborns make the failure of static 3-D form perception all the more mysterious at 6 months and beyond. By 6 months of age, virtually all infants have stereoscopic depth perception, which should provide extremely accurate information about surface slant. The problem seems to be developing a global 3-D representation. Infants do not appear to do this from successive views from different orientations (Kellman & Short, 1987b) nor from single views. The latter process is not fully understood with adults, but seems to involve symmetry or simplicity in representing unprojected object surfaces (Buffart, Leeuwenberg, & Restle, 1981). This process may be learned. In any case, it occurs relatively late in development. It is likely that the 3-D forms of stationary objects viewed from a stationary position are, to a young infant, indeterminate.

C. Nonrigid Unity and Form

Thus far we have been concerned with perception of objects whose forms are rigid. Some of the most important objects in the infant's world are nonrigid, such as people walking or a hand opening and closing. In this section we consider what is known about unity and form perception in nonrigid objects.

We can define nonrigid objects as those having points whose separations in 3-D space change over time. When we manipulate a glass, all of its points remain in a constant relationship (if it is not dropped), although the object may be rotated or translated in space. Now consider a human hand. When the hand changes from open to a closed fist, the point-to-point distance from a fingertip to the base of the wrist changes a great deal. Human movement can be considered *jointed* motion, because these changes are caused by operation of joints between relatively rigid segments. There are also elastic motions, such as bending, stretching, or squeezing of a rubbery substance (or a jellyfish) in which the nonrigid transformations are quite different. Analytically, it has been harder to describe processing constraints that allow recovery of nonrigid motions from optical information than is the case with rigid motion (Bertenthal, 1993; Cutting, 1981; Hoffman & Flinchbaugh, 1982; Johansson, 1975; Webb & Aggarwal, 1982). Perceptually, the notion of "form" perception must include objects whose forms change. Despite nonrigidity, a jellyfish does not have the same form as a walking person. What remains invariant in a nonrigid form? Some of the answer lies in connectivity: what is connected to what, where are joints located, and so on. These questions bear close relationship to issues of perceived unity in object perception, although they arise within a unit. Since Johansson's pioneering research (1950, 1975), most research on nonrigid motion has used displays comprised of separated points of light in a dark surround. This paradigm involves issues of unity as well as form; accordingly, we consider both below.

Johansson (1975) showed that the human form and its participation in events could be detected from motion information alone. He constructed films of people moving in which the only visible information came from small lights attached to the main joints of the body. Observers rapidly and effortlessly detect a person walking, a couple dancing, and various other events. Recognition of a human form does not occur when these displays are inverted (Sumi, 1984). The developmental origins of this ability have been investigated in programmatic research by Bertenthal, Proffitt, and their associates (Bertenthal, 1993; Bertenthal, Proffitt, & Cutting, 1984; Bertenthal, Proffitt, Kramer, & Spetner, 1987). An early study indicated that both 3- and 5-month-old infants discriminated motion sequences of the upright and inverted point-light walker displays, but did not discriminate static views. These results indicate some perceptual organization of the information that differs between upright and inverted, but they do not indicate whether infants actually perceived a person walking in either display. Some evidence suggests that infants do detect the familiar form of a person in point-light displays, beginning around 5 months of age (Bertenthal, 1993). Two manipulations that disrupt the appearance of a walking person for adult perceivers are phase shifting and inversion. Phase shifting refers to shifting the starting locations in the periodic motions of particular point lights. Both 3- and 5-month-olds discriminate normal from phase-shifted displays when they are upright. When inverted, however, only the 3-month-olds discriminate the two display types (Bertenthal & Davis, 1988). A group of 7-month-olds also showed poorer discrimination with the inverted displays than with upright displays. Although indirect, the results are consistent with the notion that 3-month-olds detect and discriminate certain properties of point-light displays, but do not see a walking person in upright displays. The change in response patterns between 3 and 5 months may reflect the onset of meaningful interpretation of the upright displays, in contrast with the inverted ones (which generally appear meaningless to adults).

D. Summary: 3-D Form Perception

Research on the development of 3-D form perception supports an ecological view (Gibson, 1966, 1979; Shepard, 1984). Recovery of object structure from optical transformations appears to depend on evolved perceptual mechanisms, present at birth or early maturing. Accounts attributing 3-D form perception to learning seem implausible given the developmental order in which kinematic and static information become useful. One might imagine that infants initially encode 2-D optical transformations and later learn their meaning in terms of 3-D form. Since 3-D form is not recoverable from static views, and infants perceive 3-D form from optical change before they attain skilled reaching and manipulation abilities, it is hard to see where initial information about 3-D form might come from (Kellman, 1984; Wallach & O'Connell, 1953).

VII. PERCEPTION OF SIZE

Detecting the physical size of objects requires relational information. The projective size (at the retina) of a given object varies as a function of its distance from the observer. To achieve size constancy (perception of true size despite changes in projective size) in many situations, projective size must be combined with information about viewing distance (Holway & Boring, 1941).

Day and McKenzie (1981) found evidence that 18-week-old infants are capable of perceiving size by taking distance into account. They habituated subjects to an approaching and receding object, and tested after habituation with the same object and another of different size whose retinal projections during its motion fell within the same range of visual angles (projective size) as the habituation object. Infants recovered habituation more to the object of novel size. Slater, Mattock, and Brown (1990) found that newborns also exhibit size constancy. In their design, subjects were familiarized with either a large or a small cube of a constant size, at varied distances across trials. After familiarization, the large and small cubes were presented successively. Distances for the two were selected so that they had identical projective size, and the familiarization object was presented at a novel distance. This arrangement made the (equal) projective sizes of the two objects novel, as well. All subjects looked longer at the object of novel real size.

Other research (Granrud, 1987; Slater & Morison, 1985) has yielded confirming results. Granrud (1987) measured rates of habituation to two kinds of sequences of objects containing identical variations in retinal sizes. Real size varied in one sequence, but did not change in the other. Infants showed slower rates of habituation to the sequence in which real size changed, suggesting that this sequence contained greater novelty.

Research to date thus supports the remarkable idea that size constancy is an innate visual capacity. This conclusion implies that at least one source of egocentric distance information is functional at birth. As yet, there is no direct evidence indicating what this source of distance information might be. A process of elimination along with some indirect evidence suggests that the information may be provided by binocular convergence (Kellman, 1995).

VIII. TANGIBILITY AND SUBSTANCE

It is odd to leave perception of substance for last, because it is so crucial to what an object is. We perceive objects in order to do things with them. We would not grasp, eat, throw, or step out of the way of objects if they did not have substance. We can be forgiven, in part, for our neglect because some considerations of substance are implicit in what we have said already. When a boundary is classified as belonging to one surface, not another, in effect we are determining where the tangible surface lies, that is, a certain shape is a thing, not a hole.

For other aspects of substance, perceiving what an object is made of, more must be said. Most of the research on this important problem has come from Eleanor Gibson and her colleagues. As in other domains, important carriers of information about substance are events. An object made of wood or steel or hard plastic will move rigidly, whereas one made of flesh or rubber will deform in certain characteristic ways as it moves. Gibson, Owsley, and Johnston (1978) tested 5-month-olds' sensitivity to substance from motion information. Infants were habituated to an object undergoing three different rigid motions and tested afterward with a fourth (new) rigid motion and a nonrigid (elastic deformation). Subjects generalized habituation to the novel rigid motion but dishabituated to the nonrigid deformation. Subsequent research (Gibson, Owsley, Walker, & Megaw-Nyce, 1979) showed a similar pattern of results at 3 months of age. A separate experiment (Walker, Owsley, Megaw-Nyce, Gibson, & Bahrick, 1980) produced a complementary result: When habituated to two deforming motions, infants generalized habituation to a new deformation but dishabituated to a rigid motion. These results are consistent with the interpretation that infants perceived a consistent object substance (rigid or nonrigid) in habituation in each case. An alternative interpretation is possible, however; infants might simply categorize the viewed events themselves as rigid motion or nonrigid motion, without attributing some consistent characteristic to the object.

Perception of rigid or nonrigid character of a surface was addressed in a different manner with older infants by Gibson et al. (1987). Crawling and walking infants were presented with narrow enclosed walkways, and their mothers beckoned to them from the far end. Infants' willingness to traverse a walkway varied as a function of visible surface qualities. Static qualities were varied between a homogeneous black velvet surface and a surface with many visible texture elements. Dynamic qualities were varied as well: In one condition, an experimenter pushed on the surface and caused undulations (a waterbed lay beneath the cloth). In another condition, hitting the surface did not lead to any deformation. Walking infants were reluctant to cross homogeneous surfaces and undulating ones. The results suggest that visible texture, as well as lack of deformation from contact events, specify a rigid surface, one that will offer support for locomotion (c.f., Gibson, 1979). Since walking infants have had considerable prior experience with surfaces, the role of learning in these results is not easily assessed.

IX. CONCLUSIONS

Our portrait of the development of object perception, although incomplete in some respects, has recognizable features. These include some understanding of the starting points of components of object perception and later developments.

A. How Object Perception Begins

Forty-three years ago, Piaget wrote about the perceiver's initial state with regard to the perception of objects. To the very young infant, physical objects produce "sensorial images" which:

> only constitute spots which appear, move and disappear without solidity or volume. They are, in short, neither objects, independent images, nor even images charged with extrinsic meaning. (1952, p. 65)

For Piaget, the adult's seemingly direct and immediate perceptual contact with physical objects results from a long process of associative learning:

> Perception of light exists from birth and consequently the reflexes which insure the adaptation of the perception (the pupillary and palpebral reflexes, both to light). All the rest (perception of forms, sizes, positions, distances, prominence, etc.) is acquired through the combination of reflex activity with higher activities. . . . (1952, p. 62)

Piaget's view was novel in his emphasis on action as the core of associative learning about objects. Classic, rather than novel, however, was his belief that perceivers begin with meaningless sensations and construct their meaning associatively (Berkeley, 1709/1910; Helmholtz, 1885/1910; Locke, 1690/1971; Mill, 1865/1965; Titchener, 1924).

This traditional empiricist view rested on logical considerations and anecdotal observations. Beginning in the late 1950s, systematic experimental research on perceptual development superseded these methods. Results of experimental research have in turn superseded the classical view. It is correct neither about the starting point of perception nor about developmental process.

During this same few decades, an alternative view of perception's origins was articulated, primarily by J. J. and Eleanor Gibson (Gibson & Gibson, 1955; J. Gibson, 1966, 1979; E. Gibson, 1969, 1984; Johansson, 1970). In this *ecological* view, perceptual systems evolved to provide meaningful contact with the environment, not to provide initially meaningless sensations. This view was accompanied by a new emphasis on sources of information that go beyond spatial relationships present in a single, static retinal image. Kinematic information, given by moving observers and objects, and stereoscopic information are less ambiguous indicators about objects and spatial layout than the classical cues, such as pictorial depth cues. Examples of functionally appropriate perceptual behavior by newborns of other species (e.g., Walk & Gibson, 1961; Sperry, 1943) illustrated the plausibility of the notion that perceptual systems are adapted to perceive objects in the same way that fingers and an opposable thumb are adapted to grasp objects.

The ecological premise that perception provides meaningful contact with the world from the start appears to be correct. In each perceptual domain (such as object, depth, and motion perception), there appear to be unlearned capacities to comprehend the environment. In other respects, however, neither the ecological perspective nor the traditional learning-oriented view captures the nuances of findings from the infant perception laboratory. We need a new perspective, one that combines an ecological starting point with later acquired, possibly learned, perceptual abilities.

Object perception begins very early, but the infant's abilities do not match the adult's. In every component of adult object perception, there are multiple sources of information, such as information carried by motion, information related to perceived depth, and information in static spatial (2-D) relationships. Infants in every case respond to a subset of these information sources. They show sensitivity to some information as early as tested but lack competence to use other sources until well into the second half of the first year. The pattern of competence and incompetence is not random. In most domains, kinematic information appears early, whereas information carried in 2-D spatial relationships appears later. Table 1 summarizes these developmental trends in object perception abilities.

Initial stages of visual processing in the human visual system appear to be specialized for edge extraction (DeValois & DeValois, 1988; Marr, 1982). Although maturation of certain mechanisms, such as those sensitive to spatial phase, may occur in the earliest weeks, there is little doubt that edge detection is an unlearned ability. Although the evidence is indirect, it appears that newborns detect edges from discontinuities in luminance and color, motion, and depth. At the same time, limitations are obvious. For

TABLE 1 Developmental Trends in Object Perception Abilities[a]

	Weeks of age						
	0	8	16	24	32	40	48
Edge detection	● Luminance, color discontinuities	● Accretion/deletion of texture					
Edge classification/Boundary assignment		● Optical expansion ● Accretion/deletion of texture Relative motion Depth discontinuities				● Luminance, color discontinuities	
Unit formation			● Common motion		● Edge relatability		
3-D form perception			● Structure-from-motion ● Biological motion			? Single or multiple static views	
Size perception	● Depth + projective size						
Substance perception		● Rigid versus nonrigid motions			● Texture		

[a] Components of object perception are given in the left column. Information sources are given by labels within the chart. Dots indicate that the labeled information can be used by the age shown in that column. Dots do not necessarily indicate the age of onset of an ability, but the earliest age at which competence has been observed. Question marks indicate evidence that the ability is lacking at least as late as the position marked and probably arises later. Some assignments are tentative, based on indirect or incomplete evidence.

Kellman, P. J., & Spelke, E. S. (1983). Perception of partly occluded objects in infancy. *Cognitive Psychology, 15*(4), 483–524.

Kellman, P. J., Spelke, E. S., & Short, K. R. (1986). Infant perception of object unity from translatory motion in depth and vertical translation. *Child Development*, 57(1), 72–86.

Kellman, P. J., Yin, C., & Shipley, T. F. (1995). A common mechanism for occluded and illusory contours: Evidence from hybrid displays. *Investigative Ophthalmology and Visual Science Supplements, 36*(4), S847.

Koffka, K. (1935). *Principles of Gestalt Psychology.* New York: Harcourt, Brace, & World.

Lesher, G. W., & Mingolla, E. (1993). The role of edges and lie-ends in illusory contour formation. *Vision Research, 33*(16), 2253–2270.

Locke, J. (1690/1971). *An essay concerning the human understanding.* New York: World Publishing Co.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Marr, D., & Hildreth, E. (1980). Theory of edge detection. *Proceedings of the Royal Society of London, B 207,* 187–217.

Marr, D., Poggio, T., & Hildreth, E. (1980). Smallest channel in early human vision. *Journal of the Optical Society of America, 70*(7), 868–870.

Michotte, A., Thines, G., & Crabbe, G. (1964). Les complements amodaux des structures perceptives. *Studia Psychologica.* Louvain: Publications Universitaires de Louvain.

Mill, J. S. (1865/1965). Examination of Sir William Hamilton's philosophy. In R. Herrnstein & E. G. Boring (Eds.), *A source book in the history of psychology.* Cambridge, MA: Harvard University Press.

Nakayama, K., & Shimojo, S. (1992). Experiencing and perceiving visual surfaces. *Science, 257*(5075), 1357–1363.

Owsley, C. (1983). The role of motion in infants' perception of solid shape. *Perception, 12*(6), 707–717.

Pentland, A. (1990). Photometric motion. *Investigative Ophthalmology and Visual Science Supplements, 31* 172.

Piaget, J. (1952). *The origins of intelligence in children.* New York: Norton.

Piaget, J. (1954). *The construction of reality in the child.* New York: Basic Books.

Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. *Vision Research, 33*(7), 993–999.

Regolin, L., & Vallortigara, G. (1995). Perception of partly occluded objects by young chicks. *Perception and Psychophysics, 57*(7), 971–976.

Rubin, N., Shapley, R., & Nakayama, K. (1995). *Rapid propagation speed of signals triggering illusory contours.* Paper presented at the Annual meeting of Investigative Ophthalmology and Visual Science (ARVO), Fort Lauderdale, FL.

Ruff, H. A. (1978). Infant recognition of the invariant form of objects. *Child Development, 49*(2), 293–306.

Schiff, W. (1965). Perception of impending collision: A study of visually directed avoidant behavior. *Psychological Monographs, 79*(Whole No. 604).

Schmidt, H., & Spelke, E. S. (1984). *Gestalt relations and object perception in infancy.* Paper presented at the International Conference on Infant Studies, New York.

Shapley, R., & Gordon, J. (1987). The existence of interpolated illusory contours depends on contrast and spatial separation. In S. Petry & G. E. Meyer (Eds.), (pp. 109–116). New York: Springer-Verlag.

Shapley, R., & Ringach, D. (1994). Similar mechanisms for illusory contours and amodal completion. *Investigative Ophthalmology and Visual Science Supplements, 35*(4), 1490.

Shaw, L., Roder, B., & Bushnell, E. W. (1986). Infants' identification of three-dimensional form from transformations of linear perspective. *Perception & Psychophysics, 40*(5), 301–310.

Shepard, R. N. (1984). Ecological constraints on internal representation: Resonant kinematics of perceiving, imagining, thinking, and dreaming. *Psychological Review, 91*(4), 417–447.

Shipley, T. F., & Kellman, P. J. (1992a). Perception of partly occluded objects and illusory figures: Evidence for an identity hypothesis. *Journal of Experimental Psychology: Human Perception & Performance, 18*(1), 106–120.

Shipley, T. F., & Kellman, P. J. (1992b). Strength of visual interpolation depends on the ratio of physically specified to total edge length. *Perception & Psychophysics, 52*(1), 97–106.

Slater, A., Johnson, S., Kellman, P. J., & Spelke, E. (1994). The role of three-dimensional depth cues in infants' perception of partly occluded objects. *Journal of Early Development and Parenting, 3*(3), 187–191.

Slater, A., Mattock, A., & Brown, E. (1990). Size constancy at birth: Newborn infants' responses to retinal and real size. *Journal of Experimental Child Psychology, 49*(2), 314–322.

Slater, A., & Morison, V. (1985). Shape constancy and slant perception at birth. *Perception, 14*(3), 337–344.

Slater, A., Morison, V., Somers, M., Mattock, A., Brown, E., & Taylor, D. (1990). Newborn and older infants' perception of partly occluded objects. *Infant Behavior & Development, 13*(1), 33–49.

Spelke, E. S. (1985). Perception of unity, persistence and identity: Thoughts on infants' conceptions of objects. In J. Mehler & R. Fox (Eds.), *Neonate cognition* (pp. 89–113). Hillsdale, NJ: Erlbaum.

Spelke, E. S., Breinlinger, K., Jacobson, K., & Phillips, A. (1993). Gestalt relations and object perception: A developmental study. *Perception, 22*(12), 1483–1501.

Sperry, R. W. (1943). Effect of 180 degree rotation of the retinal field on visuomotor coordination. *Journal of Experimental Zoology, 92,* 263–277.

Sumi, S. (1984). Upside-down presentation of the Johansson moving light-spot pattern. *Perception, 13*(3), 283–286.

Titchener, E. B. (1924). *A textbook of psychology.* New York: Macmillan.

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

von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224*(4654), 1260–1262.

von Hofsten, C. (1983). Foundations for perceptual development. *Advances in Infancy Research, 2,* 241–264.

Walk, R. D., & Gibson, E. J. (1961). A comparative and analytical study of visual depth perception. *Psychological Monographs, 75*(15).

Walker, A. S., Owsley, C. J., Megaw-Nyce, J., Gibson, E. J., & Bahrick, L. B. (1980). Detection of elasticity as an invariant property of objects by young infants. *Perception, 9*(6), 713–718.

Wallach, H. (1985). Learned stimulation in space and motion perception. Meeting of the American Psychological Association (1984, Toronto, Canada). *American Psychologist, 40*(4), 399–404.

Wallach, H., & O'Connell, D. N. (1953). The kinetic depth effect. *Journal of Experimental Psychology, 45,* 205–217.

Wallach, H., & Slaughter, V. (1988). The role of memory in perceiving subjective contours. *Perception & Psychophysics, 43*(2), 101–106.

Wattam-Bell, J. (1991). Development of motion-specific cortical responses in infancy. *Vision Research, 31*(2), 287–297.

Wattam-Bell, J. (1992). The development of maximum displacement limits for discrimination of motion direction in infancy. *Vision Research, 32*(4), 621–630.

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

Wertheimer, M. (1923/1958). Principles of perceptual organization. In D. C. Beardslee & M. Wertheimer (Eds.), *Readings in perception*. Princeton, NJ: Van Nostrand.

Yin, C., Kellman, P. J., & Shipley, T. (1995). A surface spreading process complements boundary interpolation under occlusion. *Investigative Ophthalmology and Visual Science Supplements, 36*(4), S1068.

Yonas, A. (1981). Infants' responses to optical information for collision. In R. N. Aslin, J. Alberts, & M. Petersen (Eds.), *Development of perception: Psychobiological perspectives: the visual system*, (Vol. 2, pp. 313–334). New York: Academic Press.

Yonas, A., Arterberry, M. E., & Granrud, C. E. (1987a). Four-month-old infants' sensitivity to binocular and kinetic information for three-dimensional-object shape. *Child Development, 58*(4), 910–917.

Yonas, A., Arterberry, M. E., & Granrud, C. E. (1987b). Space perception in infancy. *Annals of Child Development, 4*, 1–34.

Yonas, A., & Granrud, C. E. (1984). The development of sensitivity to kinetic, binocular and pictorial depth information in human infants. In D. Ingle, D. Lee, & M. Jeannerod (Eds.), *Brain mechanisms and spatial vision*. Amsterdam: Nijhoff.

Yonas, A., Pettersen, L., & Lockman, J. J. (1979). Young infants' sensitivity to optical information for collision. *Canadian Journal of Psychology, 33*(4), 268–276.

CHAPTER 2

Perceptual Classification and Expertise

Susan Carey

I. INTRODUCTION

Other chapters in this volume cover some key topics in perceptual development, including the maturation of mechanisms underlying face recognition (Johnson and Gilmore), depth and object recognition (Kellman; Spelke and Hermer), and the development of attentional mechanisms (Johnson and Gilmore). This chapter focuses on an aspect of perceptual development that is closely related to cognitive development—the development of the capacity for perceptual classification. Perceptual classification is a kind of categorization; we have the capacity to classify stimuli as members of such categories as cups, cars, dogs, and people. We also have the capacity to recognize *individuals* within such categories; I distinguish my dog Domino from other Labrador Retrievers, and we recognize our mother's face from among others.

There is reason to believe that the processes that allow us to recognize cups differ in important ways from those that allow us to recognize individual dogs or faces. Biederman (1987) has argued convincingly that our capacity to recognize artifacts such as cups and cars relies on representations built from an alphabet of primitive parts (called by him "geons"). For example, the handle of a cup and the handle of a basket are the same geon. Similarly, the body of a cup and the body of a basket (at least a round one) are the same

Perceptual and Cognitive Development