Lawrence Erlbaum Associates, Inc., Publishers
10 Industrial Avenue
Mahwah, New Jersey 07430

Cover design by Kathryn Houghtaling Lacey

Library of Congress Cataloging-in-Publication Data

Perceptual organization in vision : behavioral and neural perspectives / edited by Ruth Kimchi,
Marlene Behrmann, Carl R. Olson.
p. cm.
Includes bibliographical references and index.
ISBN 0-8058-3872-4 (c. : alk. paper)—ISBN 0-8058-3873-2 (pbk. alk. paper)
1. Visual perception—Congresses. 2. Behavioral optometry—Congresses.
3. Cognition—Congresses. I. Kimchi, Ruth. II. Behrmann, Marlene. III. Olson, Carl R.
RE960 .P474 2003
152.14—dc21
2002033868

Books published by Lawrence Erlbaum Associates are printed
on acid-free paper, and their bindings are chosen for strength
and durability.

Printed in the United States of America
10 9 8 7 6 5 4 3 2 1

PERCEPTUAL ORGANIZATION IN VISION
BEHAVIORAL AND NEURAL PERSPECTIVES

Edited by
Ruth Kimchi
University of Haifa, Israel
Marlene Behrmann
Carnegie Mellon University
Carl R. Olson
Carnegie Mellon University
University of Pittsburgh

LEA LAWRENCE ERLBAUM ASSOCIATES, PUBLISHERS
2003 Mahwah, New Jersey London

5

Visual Perception of Objects and Boundaries: A Four-Dimensional Approach

Philip J. Kellman
University of California, Los Angeles

To see an object means to detect and represent a bounded volume of matter. To see objects accurately means that our perceptual representations correspond to facts about the physical world: what things cohere and where the world breaks apart. Carving the world into objects links fundamental aspects of physical reality to a primary format of cognitive reality. Our comprehension of scenes and situations is written in the vocabulary of objects, and to objects we assign basic properties important for thought and action, such as shape, size, and substance.

In ordinary situations, object perception is fast and effortless. Some lines of research suggest that we can obtain representations of object unity and shape in a third of a second or perhaps even much less. It is probable that we can do this with more than one object at a time. Perceiving on average several objects a second, in 16 waking hours of a normal day, object perception is something we do perhaps 10^5–10^6 times. We might suspect that something we do so often is important. Of course, the frequency of object perception is not the reason for its importance; rather, we do it so frequently because object perception organizes thought and action, is central to language, is salient in cognitive development, and anchors learning throughout life.

Mapping the world's objects into mental representations does not come easily, however. Despite its introspective ease, the underlying processes of object perception have proved complicated and have resisted satisfactory explanations.

In recent years, researchers' efforts have brought clear progress. The requirements for object perception, the information that makes it possible, and processes and brain mechanisms that carry it out are all beginning to be understood in greater detail. In this chapter, I suggest an overall framework for thinking about object perception. Within that framework, I focus on several important issues at the frontiers and some research efforts that may help to push them back.

The chapter is organized into three sections. In the first, I consider a schematic model of object perception, indicating the particular tasks that need to be accomplished and how they might fit together. In that context, I explore in some detail basic issues in segmentation and grouping, especially the idea of complementary contour and surface interpolation processes and the formal notion of contour relatability as an account of the information underlying contour interpolation.

In the second section, I take up issues that challenge existing models of object perception by falling outside of their scope. Components of object perception, such as edge finding, boundary assignment, and contour and surface interpolation have most often been studied in static, two-dimensional cases. I describe research showing that accounts of segmentation and grouping in object perception will require at least a four-dimensional account, incorporating all three spatial dimensions and time. These phenomena and findings will require major additions to object-perception models, especially existing neural-style models that build on two-dimensional spatial relations of oriented units in early cortical areas. At the same time, recent findings suggest continuity between existing geometric models of segmentation and grouping and object formation from three-dimensional and spatiotemporal (motion-based) information. Specifically, straightforward extensions of the geometry of contour relatability provide a unified formal account of contour interpolation in two- and three-dimensional and spatiotemporal object perception. This unified geometric account may help suggest more general models of neural mechanisms.

In a final section, I take up the question of the general character of object perception. Whereas my discussion emphasizes relatively local, autonomous perceptual mechanisms, other investigations have suggested a role for more global or top-down factors, or both. One line of current research may help to clarify the separate contributions of these different components in the visual processing of objects.

TASKS AND PROCESSES IN OBJECT PERCEPTION

Object perception involves several conceptually distinct information-processing tasks and representations. (These may or may not be realized as separate processes or mechanisms in the brain.) Fig. 5.1 combines current knowledge with some hypotheses about processes, representations, and how they interact. It is

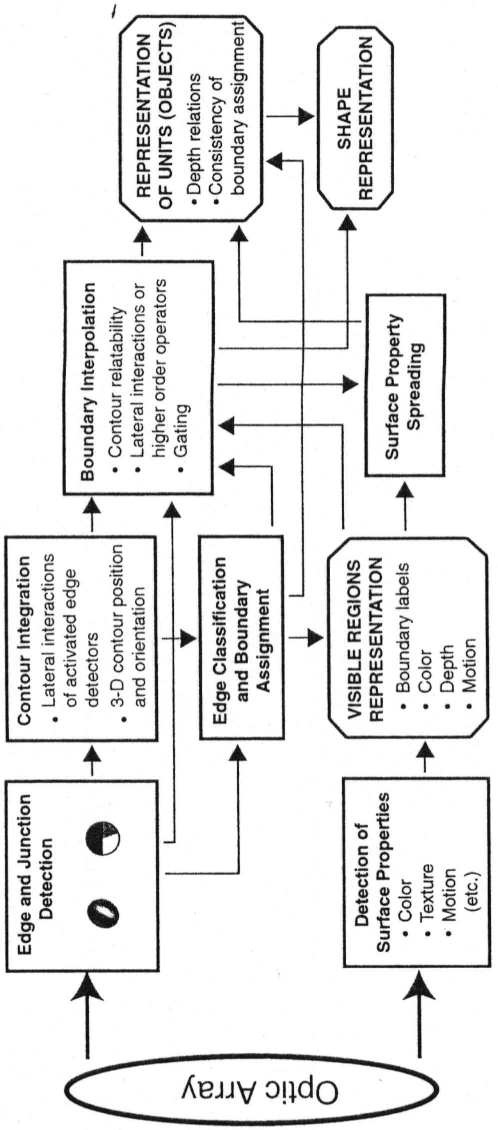


FIG. 5.1. A framework for object perception. Rectangles indicate functions or processes and octagons indicate representations. (See text for details.) From Kellman, Guttman, Wickens "Geometric and Neural Models of Object Perception." In *From Fragments to Objects: Segmentation and Grouping in Vision*, by T. F. Shipley and P. J. Kellman (Eds.), 2001, Amsterdam: Elsevier. Copyright 2001 by Elsevier Science. Reprinted with permission.

perhaps a framework more than a model, intended to provide a context for research discussed later. In the diagram, some components and their connections reflect established findings, whereas others represent newer conjectures. There are significant differences in how much is known about what goes on inside the specific boxes indicating particular representations or processes.

Rectangular boxes indicate functions or processes, and octagonal ones indicate representations. The model has few representations: output representations of shape and unity (specification of regions belonging to a single object) and one intermediate representation—*the visible regions representation.* I discuss the nature of these representations and the reasons for including them later.

General Considerations in the Model

The input to the model is the optic array. Although we will consider first information in momentary, static views of the environment, the true inputs include the time-varying optic array, sampled by two eyes. Later I focus on evidence supporting the idea that object formation is at least a four-dimensional process (i.e., involving three spatial dimensions and time).

Two processing streams handle complementary aspects of object formation. One deals with contours and junctions; the other deals with surface characteristics. The surface-processing stream extracts characteristics of luminance, color, texture, depth, and motion and represents these properties with relation to their surface locations; this information is used later to determine connections among spatially distinct visible regions. The contour stream locates discontinuities in luminance, color, texture, depth, and motion; these are used to locate edges and junctions, and, later, meaningful contours and object shape. This division into edge and contour processes was suggested previously (e.g., Grossberg & Mingolla, 1985), and it was foreshadowed by seminal work by Yarbus (1967). (See also Humphreys, Cinel, Wolfe, Olson, & Klempen, 2000, and Humphreys, this volume.)

The major processes and representations are briefly described later. For a more detailed discussion of the model, see Kellman, Guttman and Wickens (2001).

Edge and Junction Detection

An important early step on the way to object representations is locating significant edges and junctions in the scene. Not only is edge detection a point of entry into scene segmentation, but it is also crucial for defining shapes of perceived objects. For segmentation, the importance of edges is straightforward. For any visual processor to segment the world into objects, differences must exist between different objects or between objects and (projectively) adjacent visible surfaces. Abrupt changes in surface characteristics mark the locations of contours or edges. The changes are of two general types. Receiving most attention in models of edge detection have been discontinuities in luminance, chromatic, or textural properties in a scene. Their importance derives from the ecological fact that objects tend to be relatively homogeneous in composition (and thus relatively homogeneous in the properties of the light they reflect). At object boundaries, the likelihood is high that the luminance, chromatic, or textural information, or all three, will change.

Edge detection would be possible, however, even in a world of objects of homogeneous surfaces and even illumination, if the surfaces had visible texture. It might be argued, in fact, that the most robust sources of information for edges would still be available in such a world. These are discontinuities caused by the spatial arrangements of objects. At object boundaries, there are likely to be depth discontinuities that lead to gradient discontinuities in stereoscopic depth perception (e.g., Gillam, Chambers, & Russo, 1988) and also to accretion and deletion of texture when the scene is viewed by a moving observer. The ecological basis of these information sources derives from the objects being arrayed in three-dimensional space, and only rarely do adjacent objects' surfaces join to form a smooth gradient. In general, discontinuities of depth and motion are available in optical information whenever there are depth differences in the world across a visible contour. We say that these information sources are more robust because they are more highly correlated with object boundaries than are luminance or chromatic discontinuities. The latter arise often within continuous surfaces of a single object.

Computations leading to perception of contours and edges appear to begin in the early cortical areas, V1 and V2. Cells in these areas respond to oriented luminance contrast at particular spatial frequencies in particular retinal locations (Campbell, Cooper, & Enroth-Cugell, 1969; Hubel & Wiesel, 1968). By area V2, and perhaps earlier, many cells respond selectively to particular binocular disparities, providing the basis for stereoscopic depth perception (Fischer & Poggio, 1979). Some cells in the early cortical areas also respond preferentially to motion, although areas upstream, particularly area V5 (the human homologue to macaque area MT), appear to be specialized for motion processing. A number of specific proposals have been advanced regarding how these early cortical responses can be used to detect luminance edges in the optical projection (e.g., Marr & Hildreth, 1980; Morrone & Burr, 1988). The modeling of object perception based on other types of edge inputs, such as discontinuities in stereoscopic depth and motion, is less advanced (Gibson, Kaplan, Reynolds, & Wheeler, 1969; Julesz, 1971; Shipley & Kellman, 1994). Likewise, further research is needed to determine how the visual system integrates edge information arising from these various sources.

Edge Classification

Detected edges originate from several kinds of sources. Some are boundaries of objects or surfaces, whereas others are textural markings on surfaces or illumination edges (shadows). These different origins have markedly different consequences for perception. *Edge classification* refers to the labeling of edges of different kinds (or their differential use in further processing). One important distinction is the difference between *illumination* and *reflectance edges.* Illumination edges arise when surfaces having the same reflectance properties receive different amounts of illumination. Cast and attached shadows are examples of illumination edges. Note that an edge given by a cast shadow across some surface is not particularly helpful for scene segmentation (of that surface, at least) because there is no object boundary at the location of the shadow. Reflectance edges are edges caused by differences in the light-reflecting properties of two surfaces. Often these differences mark an object boundary, but sometimes they do not. A textural marking on a surface is a reflectance edge, yet it is not a boundary of the surface. Most important for object segmentation is a subset of reflectance edges that are also *occluding edges.* Occluding edges mark locations in the optical projection at which one object or surface ends and another passes behind it.

Classification of edges depends on a number of sources of information. Depth and motion information can be decisive because the existence of depth discontinuities (in a stereoscopic depth map, a map of image velocities, or given by accretion and deletion of visible texture) reliably indicates the presence of an occluding edge.

Contour Junctions

Going hand in hand with edge detection and classification are processes for detecting and classifying contour junctions. Whereas edges may be localized in terms of high rates of change in surface properties, junctions may be characterized as locations in which edges have high rates of change (Heitger et al., 1992; Kellman & Shipley, 1991). One indication of the importance of junctions for a variety of tasks in middle and high-level vision is that investigators have suggested numerous names for them. In models of human and artificial object recognition, they are usually called *corners* or *junctions* (e.g., Barrow & Tenenbaum, 1986; Biederman, 1995). In models of segmentation and grouping, they have been referred to as *tangent discontinuities* (Kellman & Shipley, 1991) or as *key points* (Heitger et al., 1992). Contour junctions can be formally defined as points along a contour that have no unique orientation. More intuitively, a junction is an intersection of two or more contours in the optical projection. Contour junctions include the projections of the sharp corners of objects as well as points of contour intersection in the world. Neurally, detecting junctions may be based on operators that use end-stopped cells (e.g., Heitger et al., 1992; Wurtz & Lourens, 2001) or perhaps high rates of curvature.

Classification of contour junctions plays an important role in segmentation and grouping. Among other things, it provides information for edge classification. For example, a *T* junction usually indicates that one of two intersecting edges passes behind the other (i.e., the latter is an occluding edge). An *X* junction is an important cue for transparency. (See Fig. 5.2.) Obviously, to help with edge classification, junctions themselves must be classified (e.g., into *T*s or *X*s). Little is known about the location or operation of neural mechanisms for classifying junctions.

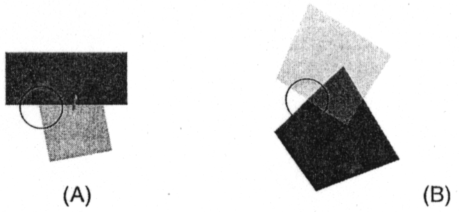


FIG. 5.2. Examples of effects of junction classification on edge classification. (A) A *T* junction indicates an occluding and occluded edge. (B) An *X* junction indicates transparency.

Boundary Assignment

Closely related to edge classification is boundary assignment. Edges classified as occluding edges have the property that the Gestalt psychologist Kurt Koffka called "the one-sided function of contour" (Koffka, 1935). That is, an occluding edge bounds a surface on only one side. On the other side, the visible surface continues behind the contour. Said a different way, only one of two visible surfaces meeting at a contour owns the contour. This assignment of boundary ownership at occluding edges is what changes in reversible figure-ground displays. Shimojo, Silverman and Nakayama (1989) proposed the useful terms *intrinsic,* to refer to a contour that belongs to (bounds) a certain region, and *extrinsic* to refer to a contour that does not.

Boundary assignment is accomplished from information sources similar to edge classification: Depth and motion discontinuities indicate both that a contour is an occluding edge and which side is nearer (the nearer side necessarily owns the contour). Junctions also contribute to boundary assignment. For example, the roof of a *T* junction is a boundary of the surface opposite the stem of the *T* (see Fig. 5.2, part A). Another class of boundary assignment cues was identified by Rubin (1915) in his classic treatment of figure and ground. This class involves relations between visible areas that influence which area is seen as figure and which as ground. What is at stake in figure-ground assignment is simply boundary assignment. Rubin's factors include the following: enclosing areas tend to be seen as grounds, enclosed as figures; symmetric regions tend to be seen as figures; and convex areas tend to be seen as figures (see also Peterson, this volume). These cues to boundary assignment are relatively weak, in that they are readily overridden by depth information given by stereoscopic or kinematic depth cues or even *T* junctions. Finally, familiarity of contour shape may influence boundary assignment (Peterson & Gibson, 1991; von der Heydt, this volume).

The Visible Regions Representation

In the model in Fig. 5.1, there are numerous processes but few representations. Two are the final outputs of unity and form, and there is one intermediate representation: *the visible regions representation.* This representation explicitly encodes continuous visible areas; that is, each is a region of visible points that belongs to an uninterrupted surface. It uses inputs from the surface stream, namely spatially contiguous locations possessing homogeneous or smoothly varying surface attributes. In locating continuous surface patches, smoothly changing depth values are primary, but when depth information is minimal (as in the viewing of far away scenes) other properties including lightness, color, and texture can determine visible regions. This grouping complements the edge process: It depends on the *absence* of the surface discontinuities extracted by the edge stream. Information from the contour stream defines the boundaries of tokens in the visible regions representation. In short, the visible regions representation labels connected surface regions and encodes the locations and orientations of their edges and corners. It also labels edges in terms of their boundary assignment (see later discussion).

This representation captures several important properties of representations suggested earlier by other researchers. Consistent with the goals of image segmentation algorithms, it partitions the optic projection into distinct, nonoverlapping regions. Unlike the results of region segmentation processes in machine vision, the visible regions representation should *not* be understood as a set of frontoparallel image fragments. Here the regions have three-dimensional positions and orientations; in this respect, the visible regions representation resembles Marr's (1982) 2.5-dimensional sketch, which assigned to each point an observer-relative depth. Another related proposal is the *uniform connectedness* idea of Palmer and Rock (1994; see also Palmer, this volume): The visual system encodes closed regions with homogeneous surface properties as a single unit. There are important differences from this notion, however. Palmer and Rock treated common surface lightness, color, and texture as the primary determinants of uniform connectedness. By contrast, in the visible regions representation, depth relations, given by stereoscopic and motion parallax cues, take precedence over the commonalty of lightness and color. A surface that is continuous in space but contains discontinuities in surface coloration would be encoded as a single token in the visible regions representation. Conversely, the visible regions representation would encode as *separate* two adjacent, homogeneously textured regions with an abrupt change of depth between them.

What motivates the idea that human perception incorporates an intermediate representation of visible regions? One is the dual nature of human scene perception (c.f. Rock's, 1979, discussion of "dual aspects" in perception). The primary role of perceiving is to produce representations of whole objects and their arrangement in space. Yet we also have an awareness of which parts of objects are occluded and which reflect light to the eyes. These may be considered "distal" and "proximal" modes of perceiving (Rock, 1979). It is the proximal mode that suggests that we have access to a visible regions representation. Artists can hone their awareness to depict only the light reflecting surfaces in their paintings. (Children do much worse, often attempting to put in a drawing the three or four sides of a house in their object representation, despite only two being visible.)

One issue that helps make clear the difference between visible regions and objects, as well as the differences between this proposed representation and other schemes, is the labeling of edges. As noted previously, at each occluding edge the surface on one side of the contour is bounded, whereas the surface on the other side continues behind. A unit in the visible regions representation is labeled as to whether its defining contours are intrinsic or extrinsic to it. Fig. 5.3 shows an example of a display containing two objects and one visible region from that display. The bottom edge is labeled intrinsic, and the two side edges are labeled extrinsic. These are not the only possibilities. The top edge is labeled as a *crease,* a contour that divides two regions that are contiguous and part of the same object (Barrow & Tenenbaum, 1986). Because of edge labeling, units in the visible regions representation are not confusable with bounded entities in the world (e.g., in Fig. 5.3, there are 10 visible regions, but we see only two objects).

As Fig. 5.3 indicates, the visible regions representation also labels tangent discontinuities (contour junctions, shown by circles) as well as the orientation of edges leading into them (shown by arrows). By explicitly labeling these various features of regions, (or implicitly distinguishing their roles in further computations) this representation provides important inputs into further processes, such as contour interpolation and unit formation.

Significant work has been done by the time the visual system obtains a visible regions representation. Yet this representation does not make explicit objects in the physical world (or the objects we perceive). In normal scenes, the chief source of this discrepancy is occlusion. If we define visible to mean "reflects light to the eyes," most objects are only partly visible. Because scenes are three-dimensional, objects are opaque, and light moves in straight lines, most objects are partly occluded. Any three-dimensional object occludes parts of itself when viewed from a particular vantage point. An even greater complication is that most objects are partly occluded by other objects. Visible parts of a single object very often appear in noncontiguous locations on the retina. In a scene partially obscured by foliage (e.g., tree branches), a single object, such as a building, may project to literally hundreds of separate retinal locations. When the observer moves, the problem of spatial fragmentation is exacerbated by the continuously changing projections in which different patches of a partly occluded object are available at different times.

Interpolation Processes: Contours and Surfaces

As mentioned previously, because of occlusion, the relationship between visible regions and the objects in the world is complex. Research suggests that two complementary processes of *interpolation* are used by the visual system to overcome fragmentation (Kellman & Shipley, 1991; Yin, Kellman, & Shipley, 1997).

Contour interpolation connects oriented edges across gaps. It operates on edges satisfying certain geometric relationships. Its products are most often occluded contours, but depending on context and depth relationships in a scene they may also

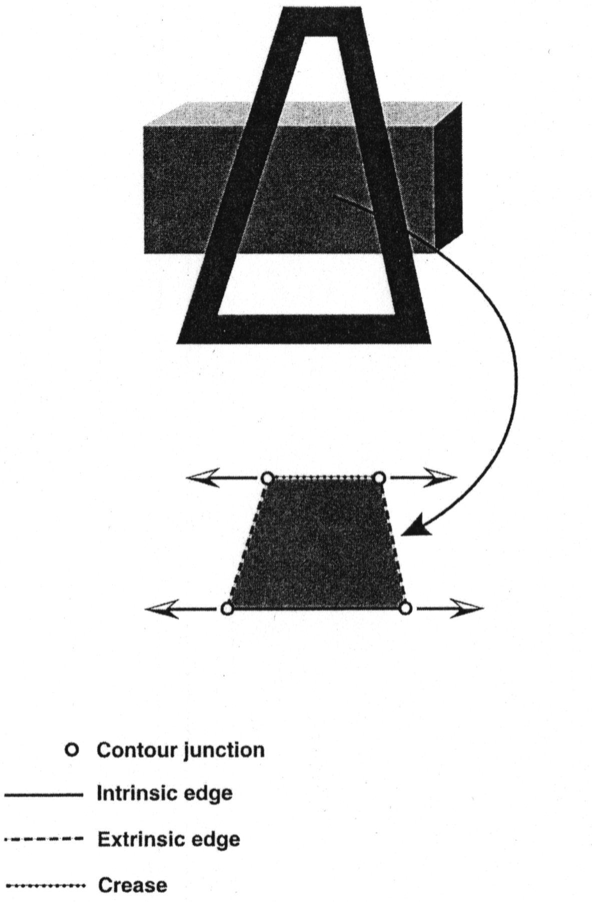


FIG. 5.3. Example of a visible region and the information encoded in the visible regions representation. A visible region from the center of the upper display is shown in the lower display. Some of the information made explicit in this representation is indicated. Contour junctions are indicated by circles; orientation of edges leading into the junction is indicated by arrows. Solid lines indicate boundaries owned by the region. Straight dotted lines indicate boundaries owned by another region. Dotted lines with circle elements indicate a crease—a contour lying within an object.

be illusory contours. We elaborate significant aspects of the boundary interpolation process later because it forms the focus for much of the rest of this chapter.

The *surface interpolation process* complements the boundary process. It leads to perceived connections among visible regions, even when the object's boundaries are not well specified. Whereas boundary interpolation depends on spatial relations between visible edges, surface interpolation (or spreading under occlusion) depends on similarity of visible surface regions. Although investigation of the surface process is a recent endeavor (Yin, Kellman, & Shipley, 1997, 2000), several aspects are clear. Surface fragments of similar color, luminance, or texture, or all three, can join behind an occluder, despite the absence of contour interpolation. Analogous to phenomena described by Yarbus (1967) in image stabilization experiments, surface quality appears to spread within real contours and also within interpolated contours (Yin et al., 1997). The process can influence boundary assignment and perceived depth (Yin et al., 1997, 2000).

Fig. 5.4 shows an example of the surface spreading process. Because the circles in the display have no tangent discontinuities, they do not participate in contour interpolation processes. Note that the incorporation of the circular areas into the

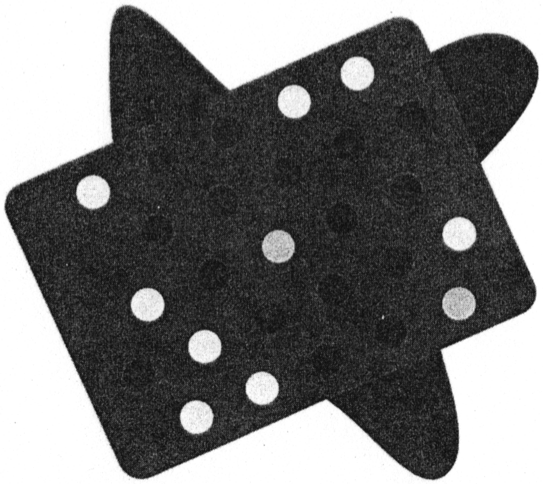


FIG. 5.4. Example of the surface spreading process. The blue circles that fall within relatable edges (or linear extensions of nonrelatable edges) appear as holes, connecting them with the blue surface pieces outside of the gray occluder. In contrast, a blue circle outside of relatable edges is seen as a spot on the gray surface, as are yellow circles. White circles are also seen as holes revealing the white background behind. (See color panel.)

partly occluded figure is selective for areas of matching lightness and color and restricted to operating within the boundaries given by real or interpolated contours (or in some cases, extended tangents of partly occluded contours).

The Units Representation

Regions connected by the interpolation processes feed into two output representations. The *units representation* encodes explicitly the connectedness of visible regions under occlusion. When the surface interpolation process alone has given all or some of these connections, overall shape may be vague. One reason for believing in a units representation separate from a shape representation is that in some circumstances, surface interpolation operates without much support from visible contours. An example would be seeing the sky through numerous branches of a tree. The sky appears to be a connected, but unshaped, unitary surface.

The Shape Representation

More often, boundary interpolation accompanies surface interpolation, and a determinate shape is encoded in the *shape representation.* This representation serves as the primary input to object recognition. This need not be the case, however. As we will consider later, recognition processes may often use shortcuts—activating higher level representations based on even a single feature in a stimulus or, in any case, using far less than a complete and well-defined representation of shape.

Bottom-Up and Top-Down Processes

The model as described is a feedforward or bottom-up view of object perception processes. It does not so far incorporate any feedback from higher levels to earlier ones. Clearly, perception can proceed without such feedback; it must do so in cases in which objects are unfamiliar and asymmetric. Whether there are top-down influences on basic segmentation and grouping processes, as opposed to recognition from partial input, remains controversial (e.g., Kellman, 2000; van Lier, 1999); we take up some of these issues later. One valuable aspect of the current framework is that it allows us to consider explicitly and distinguish possible top-down effects in terms of which higher level processes they come from and which earlier processes they affect. To take one example, Peterson and her colleagues (e.g., Peterson, 1994; Peterson & Gibson, 1991, 1994) found evidence that figure-ground segregation in otherwise ambiguous stimuli can be influenced by the familiarity of a shaped region, such that the familiar shape is preferentially seen as figure. Such an effect could be incorporated into the model, as shown in Fig. 5.5. Boundary assignment is the process that determines figure-ground relationships in the model. If familiar shape influences boundary assignment, the shape of some contour or

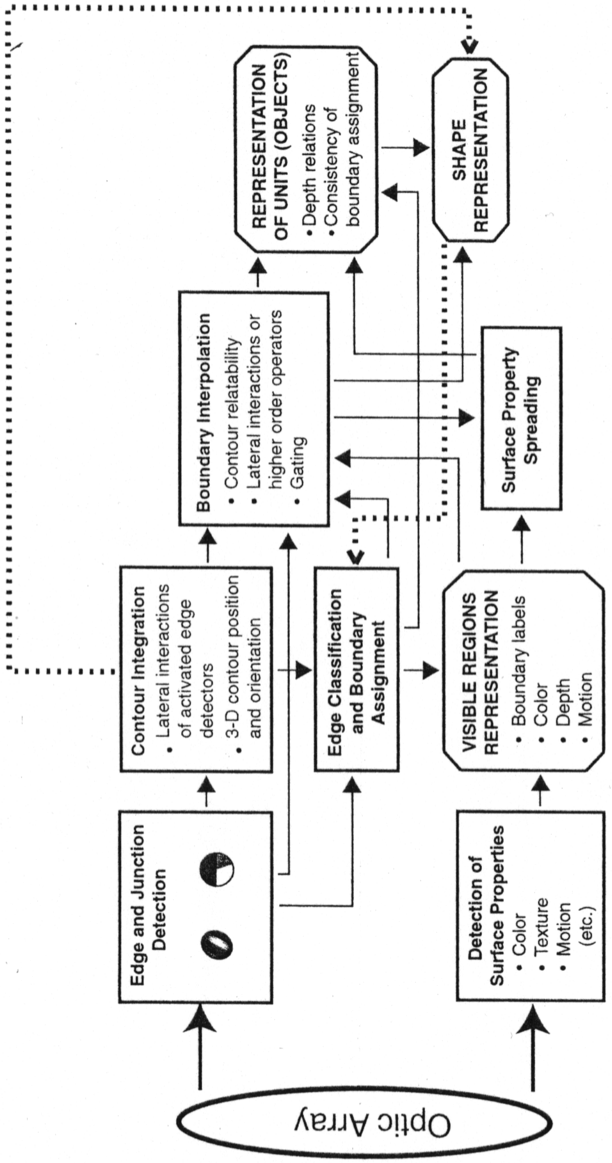


FIG. 5.5. Illustration of top-down effects in the framework of Fig. 5.1. Peterson's effect is shown as an influence of shape on boundary assignment. (See text.) From Kellman, Guttman, and Wickens, "Geometric and Neural Models of Object Perception. In *From Fragments to Objects: Segmentation and Grouping in Vision.* (p. 189), by T. F. Shipley and P. J. Kellman (Eds.), Amsterdam: Elsevier. Copyright 2001 by Elsevier Science. Reprinted with permission.

region must be encoded and recognized, that is, matched to some stored memory representation. Finding a match leads to feedback from the shape representation to the computation of boundary assignment.

This example shows how top-down effects can be incorporated into the model. Placing them in this framework may help to sharpen empirical tests. I consider selected ideas and some data about top-down processing later.

CONTOUR INTERPOLATION IN OBJECT PERCEPTION

So far, I have taken a wide-angle view of the processes of object perception. In the remainder of this chapter, I zoom in on the particular problem of *unit formation*—how visible regions get connected—with particular emphasis on the primary process involved in deriving object representations from fragmented input: contour interpolation. I first develop an argument about the generality of contour interpolation processes and then examine the key components of a model of spatial interpolation of contours. These considerations are important in their own right and will provide needed scaffolding for my discussion of three-dimensional and spatiotemporal interpolation.

The Identity Hypothesis in Contour Interpolation

In the model in Fig. 5.1, the boundary interpolation process is intended to apply to several interpolation phenomena that have often been considered distinct. Specifically, it applies to the completion of contours across gaps in partly occluded objects (amodal completion), illusory contours (modal completion), so-called self-splitting figures, and certain transparency phenomena. Specifically, my colleagues and I (Kellman & Loukides, 1987; Kellman & Shipley, 1991) proposed the *identity hypothesis* in contour interpolation: The same underlying process extends contours across gaps in these several different-looking phenomena.

Although this claim is now widely accepted, it is not universally accepted. In this section I elaborate the arguments for the identity hypothesis, indicating its logical and empirical bases, and address some concerns that have been raised about it.

Both empirical evidence and several logical considerations lead to the idea that the contour interpolation process in object formation—the specific process in which visible edges become represented as connected by interpolated contours—is common to occluded contours, illusory contours, and related phenomena. This claim by no means implies that all aspects of the processing of occluded and illusory objects are identical. Indeed, if that were the case, the output representations—that an illusory object is the nearest object to the observer in some visible direction and an occluded object is not—could not differ. Issues of depth relations with surrounding surfaces surely differ for occluded and illusory objects. This difference has sometimes been used to claim that the interpolation process itself differs in these cases (Anderson, Singh, & Fleming, 2002) a claim that does not follow and is almost surely incorrect.

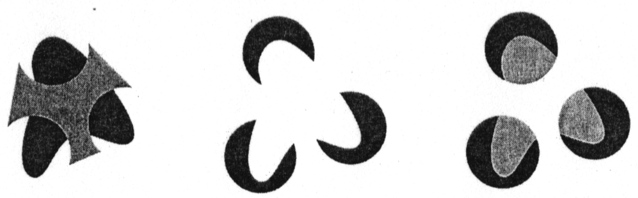

(A) (B) (C)

FIG. 5.6. Equivalent illusory, occluded, and transparent objects. (A) partly occluded object, (B) illusory object, (C) transparent object. Although they appear quite different, these three images formally are similar in that the same physically specified edges define the central figure in each case. According to the identity hypothesis (see text), the same interpolation process produces contour connections across gaps in these cases.

A good beginning is to air out some theory (as some of it is a bit dusty) connected with these central cases of perceptual organization. Consider the displays in Fig. 5.6. These are partly occluded, illusory, and transparent shapes having equivalent physically specified contours and gaps. Phenomenologically, the shapes of the interpolated contours are the same in the three images, yet they differ in appearance. Illusory contours and surfaces appear to cross in front of other surfaces in the scene; they have a robust sensory presence, which led Michotte, Thines and Crabbe (1964) to label their formation as *modal completion* (possessing sensory attributes or modes). Occluded contours pass behind other surfaces. (In the terminology of Michotte et al., they result from *amodal completion,* meaning that they exist perceptually but do not have sensory attributes.) Clearly, occluded contours are out of sight, a fact that leads to perplexities. If a contour or surface is out of sight, in what sense can it be said that we *see* it? If one believes there is a problem here, two resolutions may seem appealing. We could say that we do not really represent hidden contours or surfaces but that our perceptual system simply groups the visible parts that belong to a common object. A second idea is that we do represent the hidden contours, but we cognitively infer, rather than perceive, them.

Both of these ideas, and the line of reasoning that leads to them, need to be reconsidered. The relevant arguments are not new (Gibson, 1966, 1979; Kanizsa, 1979; Koffka, 1935; Michotte et al., 1964) yet they seem to have incompletely penetrated many discussions. The source of the problem is that to see does not necessarily mean to have local sensory responses based on local physical data. Perceptual descriptions are functions of incoming information, but they need not be restricted to some narrow class of functions, such as a one-to-one correspondence between represented properties of objects and local sensations or local stimuli.

The point that seeing is not a summing up of local sensations is a very old one, made convincingly by the Gestalt psychologists, but it is also a tenet of contemporary computational views of perception. Either view can be used to untangle certain perplexities regarding perception of occluded contours. On the computational/representational theory of mind, there is simply no principled difference between representing some contour as going behind another surface and representing some contour in front of another surface. One is not more privileged or real than the other. The argument that one must be a sensory effect and the other a cognitive effect lingers from an epistemology that did not work out. Occluded contours are no more and no less "inferred" than are illusory ones.

Substantial experimental evidence suggests that occluded and illusory contours are processed and represented similarly. They exert similar or identical effects on speed and accuracy in perceptual tasks (Kellman, Yin, & Shipley, 1998; Ringach & Shapley, 1996; Shipley & Kellman, 1992a). Likewise, procedures geared to assessing the precise locations of perceived boundaries lead to the conclusion that interpolated contours in both illusory and occluded cases are represented as being in very specific locations (Kellman, Temesvary, Palmer, & Shipley, 2000).

If there is no theoretical barrier separating represented contours that are occluded or illusory, why do they *look* so different? What is the cause of the robust difference in phenomenology? The answer is that this aspect of our phenomenology—the modal/amodal distinction—appears simply to code whether some surface or contour is nearest to the observer in some visual direction. This piece of information is an important aspect of scene perception that depends on the vantage point of the observer. The parts of an object that are nearest in a visual direction may be reached, grasped, or touched, whereas those equally real edges and surfaces behind some other object may not be reached without going around or through some nearer object.

The understanding of modal/amodal in terms of nearest (or not nearest) in some visual direction (or, equivalently, in terms of a surface area reflecting light directly to the eyes or not) brings up a related theoretical idea. Modern computational analyses as well as many results in perceptual research suggest that it is often advantageous for perceptual systems to represent information in terms that are viewpoint independent. Constancies in perception exemplify this idea. For example, in size constancy, a description of an object's physical size is obtained that does not vary as the observer moves closer or farther from an object. In unit formation, support ratio (Shipley & Kellman, 1992b)—a property that determines strength of contour interpolation—is scale invariant; this has the consequence that the strength of a contour connection does not vary with viewing distance. Representing modal and amodal contours as fundamentally different (or as one being seen and the other inferred) would tend to violate this general tendency in perception. The reason is that what is nearest to the observer from one vantage point may not be from another. Although it may be important to also encode the information about what is nearest in some visual direction, the core representation of contours and surfaces that span gaps in the input should not change as the viewer moves around and experiences changing patterns of occlusion.

The foregoing are rather philosophical arguments. In recent years, more specific logical considerations have emerged, forming essentially a proof that the contour interpolation step in visual processing is shared in common by illusory and occluded contours. I describe these considerations briefly.

Quasimodal Objects. Kellman, Yin, and Shipley (1998) showed that illusory and occluded edges can join. More accurately, contour interpolation occurs in cases in which the stimulus arrangement fulfills neither the normal conditions for illusory contour nor occluded contour formation. Such contours have been called *hybrid* or *quasimodal* contours. An example is shown in Fig. 5.7. All of the interpolated contours in this display extend between an illusory contour-inducing element on one end and an occluded section on the other. The existence of quasimodal objects highlights the point made previously. The visual system represents interpolated contours and surfaces; sometimes these are in front of other surfaces, and sometimes they are behind. In a quasimodal case, a *single contour* is in front of some surface along part of its length and behind some surface along another part. Although it is a useful piece of information that the edge passes behind and in front of other objects in the scene, the basic representation of contours across gaps seems to incorporate easily segments of both kinds. It is not clear how quasimodal completions could occur from distinct modal and amodal interpolation processes.

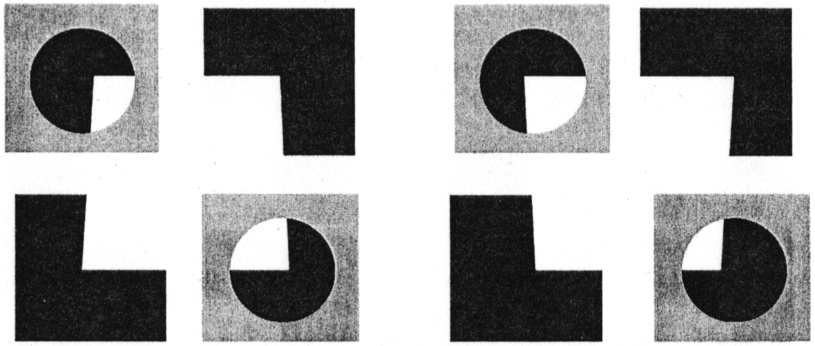


FIG. 5.7. Example of a quasimodal object. The display is a stereo pair that can be free-fused by crossing the eyes. All of the interpolated contours in the central figure are occluded on one end and illusory on the other. (See text.)

Petter's Effect. The existence of quasimodal contours does not prove the identity hypothesis. After all, it is possible that there are three separate contour completion processes: modal, amodal, and quasimodal. A more conclusive logical argument about the nature of interpolative processes has been implicit in phenomena that have been known since the 1950s. Consider the display in Fig. 5.8, part A, an example of a class of displays studied by Petter (1956). The display has several salient properties. First, although it is a contiguous, homogeneously colored and textured area, it is not seen as a single object. Instead it resolves into two objects, a triangle and a quadrilateral. Second, wherever parts of the two objects lie in the same visual direction, one object is seen as crossing in front of the other. Although all parts of this pictorial display lie in the same depth plane, the visual system appears to obey a constraint that two objects may not occupy the same space.

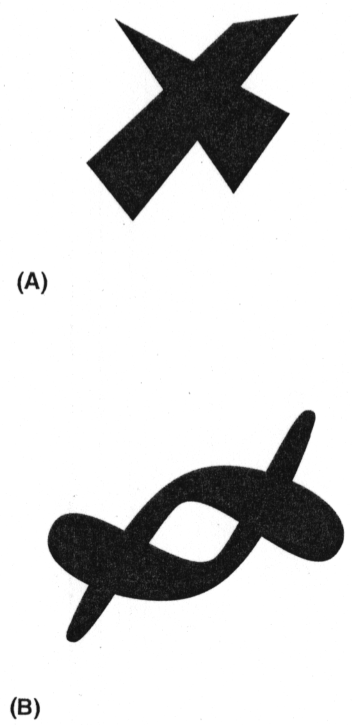


FIG. 5.8. (A) Self-splitting object, (B) Example of Petter's effect. Where interpolated contours cross, the one spanning the smaller gap appears in front as an illusory contour, and the one spanning the larger gap appears behind, as an occluded contour.

Where the two objects cross, their physically specified boundaries have gaps, and these gaps are spanned by interpolated edges. In accord with the depth ordering of the two objects, the nearer display has illusory contours, whereas the farther display has occluded contours. In Fig. 5.8A, if the triangle appears in front, it has illusory contours, and the quadrilateral has occluded contours. However, the depth ordering of the two objects is unstable over time; which object appears in front fluctuates. When the triangle switches from being in front to behind, its contours switch from being illusory to being occluded (and vice versa for the quadrilateral).

Self-splitting objects do not always possess this instability of depth order. The display in Fig. 5.8, part B, similar to one devised by Petter (1956), appears more stable. It is seen as containing two interlocking bent objects. The stimulus itself does not have luminance contours defining two objects; the partitioning of the array into two bounded objects is the result of contour interpolation processes.

Implicit in the idea of interlocking, the perceived depth ordering of the two objects in Fig. 5.8B varies across the image; on each side, the thicker object passes in front of the thinner one. Petter (1956) discovered that this perceptual outcome follows a rule, which we can state as follows: *Where interpolated boundaries cross, the boundary that traverses the smaller gap appears to be in front.* Thus, the thicker parts of the objects appear to lay on top of the thinner parts because the former have smaller gaps in their physically-specified contours. Petter's rule also helps us to understand the reversibility of depth order—and of occluded and illusory contours—in Fig. 5.8A: Because the contours of both objects span roughly equal gaps, there is no consistent information about depth order.

The relevance of Petter's (1956) effect to the identity hypothesis may now be apparent. According to Petter's rule, the perception of each interpolated contour as in front or behind—and, in turn, as illusory or occluded (or modal vs. amodal)—depends on its length relative to the interpolated contours that cross it. Logically, this statement implies some sort of comparison or competition involving the crossing interpolations. To accomplish this comparison, the visual system must first register the various sites of interpolation. Comparing the lengths of the crossing interpolations precedes the determination of whether an interpolated contour ultimately will appear as in front of or as behind other contours (and, thus, as illusory or occluded). Therefore, the registration of interpolation sites and lengths must precede the determination of depth ordering. That is, at least in some cases, *contour interpolation processes must operate prior to the processes that determine the final depth ordering of the constructed contours.* This, in turn, implies that there cannot be separate mechanisms for the interpolation of contours in front of and behind other surfaces. At least the steps of locating sites of interpolation and the extents of interpolation must be in common.

Depth Spreading and Object Formation. Kellman, Yin, and Shipley (1998) introduced a new type of display that also provides a proof that interpolation cannot be from separate modal or amodal processes because, as in Petter's (1956)

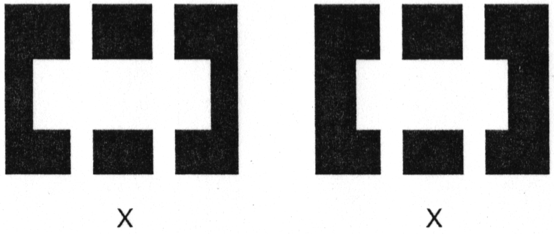


FIG. 5.9. Display in which interpolation occurs prior to determination of modal or amodal appearance. (See text.)

cases, interpolation must occur before amodal or modal appearance of contours and surfaces can be determined. Fig. 5.9 shows an example of this type of display. It is a stereo pair that may be free-fused by crossing or diverging the eyes. For the analysis here, I assume that the reader has free-fused by crossing the eyes; for diverging, all of the depth relations will be opposite to those in this description.

When fused, the central rectangle appears as the nearest object in the display at its left edge and farthest at its right. With reference to the two white columns in the display, the rectangle appears to pass in front of the one on the left and behind the one on the right. The importance of this simple appearance for the identity hypothesis rests on three elements. First is the mere existence of the rectangle: It is a product of interpolation across two gaps. Second is the notion of depth spreading. Stereoscopic disparity in this display is given only by the positions of the vertical edges of the central rectangle. The middle section of the display (two black squares separated by a white gap) is identical in the images given to the two eyes. There is no information that says the central square is in front or behind either column—this information must come from somewhere else. Depth spreading occurs *within objects* or continuous surfaces. It does not spread unconstrained across the whole scene. Therefore, the rectangle must be a unified object to partake of depth spreading from disparity of the remote vertical edges in the left and right sides of the display.

Now consider the modal or amodal appearances of various contours. The slant in depth of the rectangle's horizontal edges is a consequence of depth spreading. This causes the rectangle to pass in front of the left vertical column; thus, the rectangle here has *illusory* contours and the left column has *occluded* contours where these surfaces cross. On the right, the rectangle passes behind the white column. Opposite to the situation on the left, the rectangle's contours are here occluded and the right column's contours are illusory. Note that not just the modal/amodal appearance of the rectangle's contours but those of the white columns are consequences of depth spreading.

The crucial point is that the amodal or modal appearance of several contours necessarily *follows* from depth spreading, which necessarily requires that interpolation has already occurred. Contour interpolation necessarily precedes determination of the appearance of contours as illusory or occluded in this display. The following reviews the logic:

1. The rectangle as an object is the result of interpolation.
2. Depth spreading presupposes the rectangle as a unified object.
3. Modal or amodal appearance of various contours is the result of depth spreading.

Therefore, interpolation *necessarily* precedes determination of modal or amodal appearance.

In sum, both empirical studies and logical arguments support the idea that contour interpolation relies on a common process that operates without regard to the final determination of illusory or occluded appearance. The phenomenology of an interpolated contour—its appearance as illusory or occluded—relates to the issue of whether it is nearest to the observer in some visual direction or not. Representation of this property depends in turn on mechanisms responsible for assigning relative depth, which lie outside and sometimes operate subsequent to the interpolation process itself.

Some findings have led investigators to suggest that different processes or mechanisms are at work in producing illusory and occluded contours. Peterhans and von der Heydt (1989), for example, reported that single cells in area V2 of the macaque responded to an illusory bar but not to a display that could be interpreted as an occluded bar. This finding can be viewed as evidence for separate mechanisms of illusory and occluded interpolation. There are two specific concerns with connecting the observed data to such a conclusion. First, Peterhans and von der Heydt did find cells in V1 that responded to both their illusory contour and the related occlusion displays, although they suggested that these results may have been artifactual. Another possibility is that the equivalent responses of V1 cells were not artifactual; perhaps the V1 responses indicate a common contour interpolation step in V1, whereas the nonequivalent responses of V2 cells indicate that depth information relevant to the final contour appearance has come into play by that point. Second, as is evident in the foregoing speculation about interpolation in V1 and V2, we currently lack a sufficiently detailed mapping of computations onto cortical layers to use this sort of data to rule out a common interpolation step. Given the logical considerations following from Petter's effect (1956), quasimodal displays and depth spreading displays, as described, putative data suggesting separate visual processes of amodal and modal interpolation must be treated with skepticism. No claim of separate processes can make much sense without giving a new account of the three phenomena and their logical implications. To my knowledge, no such account has been proposed.

The identity hypothesis has several salutary consequences. It has helped in separating the crucial determinants of interpolation from extraneous aspects. It allows convergent evidence about the interpolation processes to be derived from studies of occlusion, illusory contours, and some other contour-connection phenomena. In some cases, one or another of these display types is more convenient. For example, tangent discontinuities may be removed from illusory contour-inducing elements but not from occlusion displays. Both illusory and occluded interpolation displays are used in the research discussed later.

A Model of Contour Interpolation

In this section and the next I present a model of contour interpolation. The model rests on two complementary notions: *tangent discontinuities* and *relatability*. The initiating conditions and geometric relations that govern the connecting of visible fragments across gaps depend on these two kinds of optical information.

Tangent Discontinuities. An important regularity may be observed in Fig. 5.6. In all of the displays, interpolated edges all begin and end at points of tangent discontinuity (TD). In occlusion cases, these are usually *T* junctions. For illusory contours, they are *L* junctions. Formation of the illusory contour turns *L* junctions into what might be called *implicit T junctions,* that is, *T*s whose roof is formed by the interpolated edge.

That TDs provide the starting and ending points of interpolation is not a coincidence. Contour interpolation overcomes gaps in the input caused by occlusion. It can be proven that instances of occlusion generically produce TDs in the optical projection (Kellman & Shipley, 1991). This ecological fact underlies the role of TDs in initiating contour interpolation processes: TDs are potential indicators of the loci of occlusion. Not all TDs are points of contour occlusion, but all points of contour occlusion produce TDs.

The fact that occlusion displays always have TDs makes it difficult to test their necessity for interpolation processes; one cannot compare occlusion displays with and without TDs. This situation is one in which the equivalence of interpolation processes in occluded and illusory displays helps the researcher. In illusory contour displays, tangent discontinuities *can* be removed, by rounding off corners, for example. Research shows that this manipulation greatly reduces or eliminates contour interpolation (M. Albert, cited in Hoffman, 1998; Shipley & Kellman, 1990). Likewise, in homogeneous displays that split into multiple objects (sometimes called self-splitting objects or SSOs), rounding the tangent discontinuities eliminates the splitting (Kellman, 2000).

Because only some TDs are loci of occlusion, other information comes into play to determine when interpolation occurs. One influence is the type of contour junction. Some types (e.g., so-called *Y* junctions) indicate that the boundary has come to an end and does not continue. In other cases, a contour may be seen as passing behind an occluder but does not link up perceptually with any other visible contour.

What determines when contours link up behind occluding surfaces to form objects? In recent years, it has become clear that the visual system uses contour interpolation processes that decide connections between visible areas depending on certain geometric relations of their edges. The most important properties of edges are their positions and orientations leading into tangent discontinuities. The relevant spatial relations bear some close relations to the classic Gestalt notion of *good continuation,* but they also differ from it. These relations have been formally defined as the notion of contour *relatability* (Kellman & Shipley, 1991). The requirements can be summarized intuitively as follows: Two edges separated by a gap or occluder are relatable if they can be connected with a continuous, monotonic (singly inflected) curve. The relatability criterion embodies the constraints that interpolated contours are smooth, monotonic (bending in only one direction), and bend through no more than 90°. Except for the 90° constraint, these conditions can be expressed as the idea that two edges are relatable if their linear extensions meet in their extended regions. Fig. 5.10 shows some examples of relatable and nonrelatable edges in occluded and illusory object displays.

Formally, relatability may be defined with reference to the construction shown in Fig. 5.11. E_1 and E_2 in the drawing represent the edges of surfaces. R and r indicate the perpendiculars to these edges at the point where they lead into a tangent discontinuity, with R defined as the longer of the two. The angle of intersection

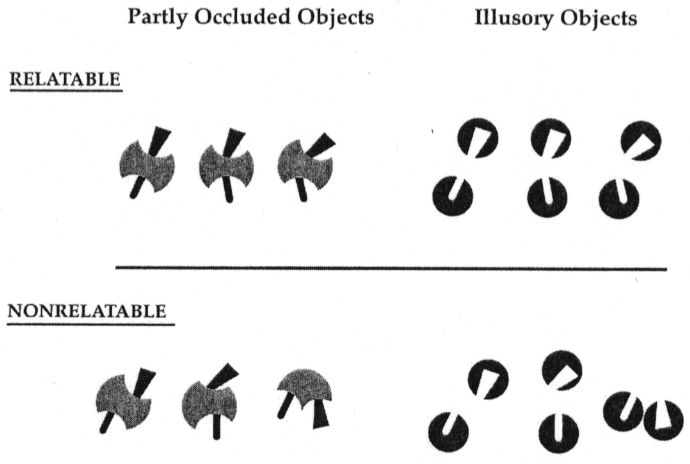


FIG. 5.10. Examples of relatable and nonrelatable edges in occluded and illusory object displays. From *The Cradle of Knowlege: Development of Perception in Infancy* (p. 143), by P. J. Kellman and M. A. Arterberry, 1998, Cambridge, MA: MIT Press. Copyright 1998 by MIT Press. Reprinted with permission.

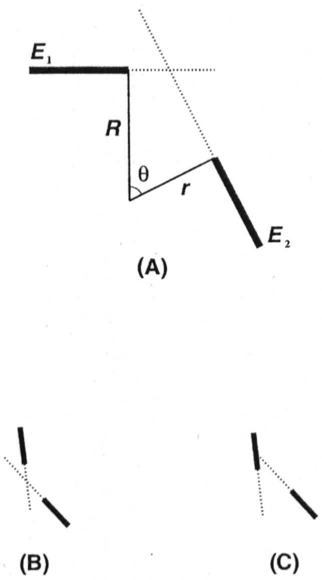


FIG. 5.11. Construction used to define relatability. E1 and E2 are edges of surfaces; R and r are perpendiculars to E1 and E2 at points of tangent discontinuity. E1 and E2 are relatable iff $0 \leq R\cos\theta \leq r$.

between R and r is termed θ. Relatability holds whenever a smooth, monotonic curve can be constructed starting from the endpoint of E_1 (and matching the slope of E_1 at that point) and proceeding through a bend of not more than a 90° to the endpoint of E_2 (and matching the slope of E_2 at that point). More formally, E_1 and E_2 are relatable if and only if

$$0 \leq R\cos\theta \leq r. \tag{1}$$

This inequality can be explained in two steps. The part on the left side expresses the limitation that the curve constructed to connect $E1$ and $E2$ cannot bend through more than 90°; cos θ is negative for θ greater than 90°. The right-hand side of the inequality states that the projection of R onto r ($R\cos\theta$) must fall within the extent of r. If this inequality is violated (i.e., $R\cos\theta > r$), then any connection between $E1$ and $E2$ would have to be doubly inflected to match the slopes at the TDs or would have to introduce sharp corners where the interpolated edge meets the physically specified edge. According to this model, boundary interpolation does not occur in such cases.

Although the definition gives the limits of relatability, it is not intended as an all-or-none concept. Kellman and Shipley (1992) described contour relatability as decreasing monotonically with deviations from collinearity, falling to 0 at a relative angle of 90°. Singh and Hoffman (1999) proposed a specific measure for this graded decrease.

The notion of contour relatability is often referred to interchangeably with the Gestalt idea of good continuation (Wertheimer, 1923/1958) but differs from it in important ways. For one, the original Gestalt principle has no precise formulation. Wertheimer's original discussion of it contained no definition but gave examples in which the idea was said to be obvious. Nor did later work indicate which of various possible mathematical notions of smoothness or other notions of goodness capture the phenomena. Kellman (2000) argued that most of Wertheimer's examples of good continuation implicate smoothness in the first derivative of functions describing contours, whereas first-derivative discontinuities (tangent discontinuities) indicate possible breakpoints in continuous contours. In the respects in which relatability resembles good continuation, it is a more precise, formal statement of it.

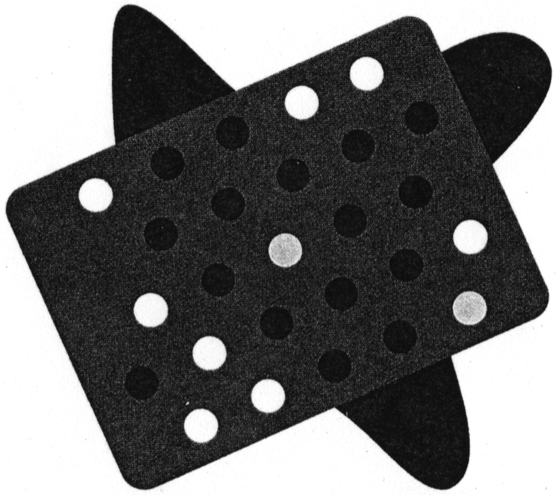
This is not the main issue, however. It appears that a precisely formulated notion of good continuation and relatability are cousins. Although the same principle has been invoked to explain segmentation in visible contiguous lines and figures (Wertheimer, 1923/1958) and in perception of partly occluded objects (Michotte et al., 1964), it is becoming clear that the geometric principles required to explain each of these perceptual tasks are different (Kellman, 2003). Both contexts have in common the notion that tangent discontinuities are significant indicators of segmentation points. The notions of good continuation and relatability are different, however, with the latter being much more restrictive. Two ways in which relatability is a more constrained notion are the monotonicity constraint and the 90° limit on interpolated contours. As mentioned previously, the monotonicity constraint specifies that interpolated edges are singly inflected, (i.e., they bend in only one direction). The 90° constraint specifies that interpolated edges do not bend more than 90°. These theoretical notions have received substantial empirical confirmation (e.g., Field, Hayes, & Hess, 1993; Kellman & Shipley, 1991). They are crucial ingredients in the relatability account of contour interpolation, but neither of these constraints has any counterpart in good continuation. Many excellent demonstrations of good continuation, for example, contain smooth, multiply inflected contours. Likewise, when a continuous contour bends through more than 90°, no break is seen; good continuation has no 90° constraint. These considerations and related research will be described in a forthcoming paper. What is common to both relatability and good continuation is exploitation of contour smoothness, and deviations from it, as information for more than one aspect of visual segmentation and grouping.

Evidence About Contour Relatability. Results of a number of experiments support relatability as a description of the spatial relationships that support contour interpolation. Some of the best evidence comes from an elegant paradigm introduced by Field et al. (1993) for the study of contour integration. The stimuli in these experiments consisted of arrays of spatially separated, oriented Gabor patches (small elements consisting of a sinusoidal luminance pattern multiplied by a Gaussian window; a Gabor patch closely approximates the ideal stimulus for the oriented receptive fields of V1 simple cells). In some arrays, 12 elements were aligned along a straight or curved path, constructed by having each element in the sequence differ by a constant angle from its neighbors (e.g., 0° for a straight, collinear path; ±15° to create a curved path). The remainder of the array consisted of elements oriented randomly with respect to one another and the path, creating a noisy background. In the experiments, observers judged which of two successively and briefly presented arrays contained a path.

The results of Field et al.'s (1993) experiments strongly supported the geometry of relatability as a description of the conditions under which contour segments connect across gaps. When the positional and angular relations of successive path elements satisfied the relatability criterion, observers detected the stimulus efficiently. Contour detection performance declined gradually as the orientation difference between elements increased, falling to chance at around 90°. Moreover, complete violations of relatability, accomplished by orienting the elements perpendicular to the path rather than end-to-end along it, resulted in drastically reduced task performance. Together, these data suggest that interpolated contour connections require specific edge relationships, the mathematics of which are captured quite well by the notion of relatability. Moreover, interpolated contours become salient, allowing them to play a meaningful role in higher level object perception processes.

I noted previously that relatability expresses a certain notion of contour smoothness. Indeed, its value as a principle of unit formation (connecting separate visible areas into unified objects) may derive from facts about the objects in the world to be perceived. As has been noted in other contexts (e.g., Marr, 1982), objects and surfaces in the world are not random collections of points and attributes but are generally smoothly varying on a number of dimensions, especially spatial position. On the other hand, boundaries between objects may be sites of abrupt change. Relatability intuitively embodies the notion that objects tend to be smooth, such that when gaps are imposed by occlusion, visible edge fragments that can be connected by smooth contours are likely to be parts of the same object.

Recently, Geisler, Perry, Super, and Gallogly (2001) attempted to put these intuitions on a firmer quantitative footing. Their work suggests that relatability captures certain spatial relationships between visible contours that have a high probability of belonging to the same object. Through an analysis of contour relationships in natural images, Geisler et al. found that the statistical regularities governing the probability of two edge elements cooccurring correlate highly with the geometry of relatability. Two visible edge segments associated with the same contour meet the mathematical relatability criterion far more often than not. The success of relatability in describing perceptual interpolation processes appears to derive from ecological regularities in the natural environment.



Color Panel for FIG. 5.4

Implementation of the Relatability Geometry. How are the relationships captured by the geometry of relatability implemented in neural mechanisms of vision? This is an active area of research. Several neural-style models of contour interpolation have been proposed (Grossberg & Mingolla, 1985; Heitger, von der Heydt, Peterhans, Rosenthaler, & Kübler, 1998). None of these fully implements what is known about the geometry of relatability or certain other influences on contour interpolation, such as support ratio (Banton & Levi, 1992; Shipley & Kellman, 1992b) or possible gating effects on interpolation caused by junction types or boundary assignment (Nakayama, Shimojo, & Silverman, 1989).

Nevertheless, existing neural-style models have introduced a number of valuable concepts. Most share the assumption that contour interpolation mechanisms are built on the outputs of orientation-sensitive cells in early cortical visual areas. These cells, sensitive to retinal position, contrast, orientation, and spatial frequency typically provide the inputs into edge and junction detection mechanisms (e.g., Heitger et al., 1992). In turn, contour integration (e.g., Yen & Finkel, 1998) and interpolation processes (e.g., Heitger et al., 1998) may use the outputs of edge and junction analyses. (For a detailed account of the relation of geometric and neural models in object formation, see Kellman, Guttman, & Wickens, 2001.)

THREE-DIMENSIONAL INTERPOLATION IN OBJECT PERCEPTION

The foregoing discussion indicates that, although much remains to be learned, progress has been made in understanding the early and middle vision processes in object perception. From edge detection through interpolation, we know the tasks that visual processes must accomplish, a great deal about the information for the task, and a bit about processes and mechanisms.

Our analysis has been artificially constrained, however. Most of it has addressed object perception in static, two-dimensional displays. This may be an important aspect of the problem, but it is not nearly the whole problem. In ordinary perception, objects are three-dimensional and arranged in three-dimensional space. The inputs to the visual system consist of visible contours oriented not only in frontoparallel planes but also in three-dimensional space. In recent years, these considerations have led us to broaden our investigations of object formation into all three spatial dimensions (as I describe in this section) and to information given over time via motion (as I describe in the next section).

Although the research literature provides some examples of three-dimensional illusory contours (e.g., Gregory & Harris, 1974), there has been little systematic investigation of three-dimensional information in unit formation, its geometry and processes. Over the past few years, my colleagues and I have posed two main

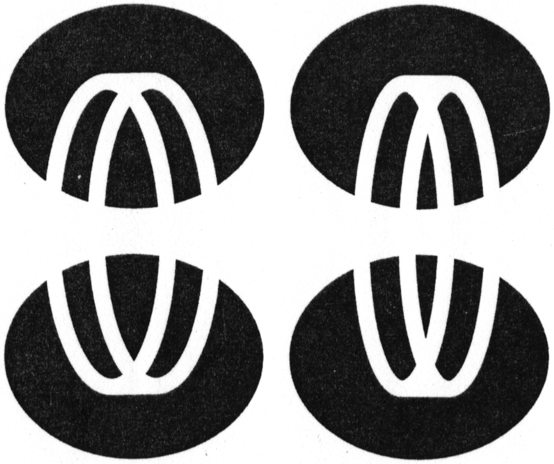


FIG. 5.12. Example of three-dimensional contour interpolation. The display is a stereo pair, which may be free-fused by crossing the eyes. When cross-fused, vivid three-dimensional illusory contours are seen on the left, and occluded contours are seen arcing behind the white surface on the right. (If the display is fused by diverging the eyes, the positions of the illusory and occluded contours are reversed.) From Kellman, P. J. "An Update on Gestalt Psychology." In *Essays in Honor of Henry and Lila Gleitman*, by B. Landau, J. Jonides, E. Newport, and J. Sabini (Eds.), 2000, Cambridge, MA: MIT Press. Copyright 2000 by MIT Press. Reprinted with permission.

questions about three-dimensional aspects of interpolation:

1. Do visual interpolation processes use as their inputs three-dimensional relationships of contours and surfaces in the world?
2. Do visual interpolation processes produce as their outputs representations of three-dimensional contours and surfaces?

The answers to both questions appear to be yes. As an illustration, consider the demonstration in Fig. 5.12. If this display is viewed stereoscopically (free-fuse by crossing or diverging the eyes), it gives rise to three-dimensional illusory contours on one side and three-dimensional occluded contours on the other. Binocular disparity places the inducing edges at particular three-dimensional orientations, and contour interpolation processes build the connections, smoothly curving through three dimensions, across the gaps. The demonstration suggests that interpolation processes take three-dimensional positions and relations as their inputs and build connections across all three spatial dimensions.

How can these phenomena be studied experimentally and objectively? If the visual system sometimes creates connections of contours and surfaces across gaps, the resulting representations may have important consequences for performance on certain tasks. By discovering such a task, we could use performance measurements on that task to determine the conditions under which such object representations form.

This strategy has been used with two-dimensional displays. Ringach and Shapley (1996), for example, devised the fat-thin method for studying two-dimensional interpolation processes. Illusory squares were created from four partial-circle inducing elements. By rotating the circular inducers around their centers, displays whose width was bulging outward (fat) or compressed inward (thin) were created. The subject's task was to classify displays as fat or thin. Performance on this task turns out to be facilitated by contour interpolation, relative to control displays in which contour interpolation does not occur (Gold, Murray, Bennett, & Sekuler, 2000; Kellman, Yin, & Shipley, 1998; Ringach & Shapley, 1996).

My colleagues and I devised a three-dimensional performance task that is in some ways analogous to the fat-thin task (Kellman, Machado, Shipley & Li, 1996). A full report will appear elsewhere (Kellman, Yin, Garrigan, Shipley, & Machado, 2003); here I note some of the main results.

We used three-dimensional illusory object stimuli such as the one shown in Fig. 5.13, part A. Such displays appear to produce vivid three-dimensional illusory contours and surfaces. We hypothesized that these occur when the physically given contours satisfy a three-dimensional criterion of relatability. The extension from the two-dimensional case is this: Bounding contours are three-dimensional relatable when they can be joined in three dimensions by a smooth, monotonic curve. This turns out to be equivalent to the requirement that, within some small tolerance, the edges lie in a common plane (not necessarily a frontoparallel plane), and within that plane the two-dimensional relatability criterion applies. Another way of saying the same thing is that the linear extensions of the two edges meet in their extended regions in three-space (and form an angle greater than 90°).

Three-dimensional relatability can be disrupted by shifting one piece in depth, as shown in Fig. 5.13B. Another relatable display and a corresponding shifted, nonrelatable display are shown in Fig. 5.13, parts C and D.

The experimental paradigm used these displays as follows. On each trial, subjects were shown a stereoscopic display. Stereoscopic disparities were produced by outfitting the subject with liquid-crystal-diode (LCD) shutter glasses, in which the left and right eyes' shutters were alternately opened, synchronized with alternating computer images. Subjects made a speeded judgment on each trial about a particular relationship of the upper and lower parts of the display. Displays like those in Fig. 5.13, parts A and B, were said to be in intersecting or *converging* planes. Those in Fig. 5.13, parts C and D, were said to be in *parallel* planes

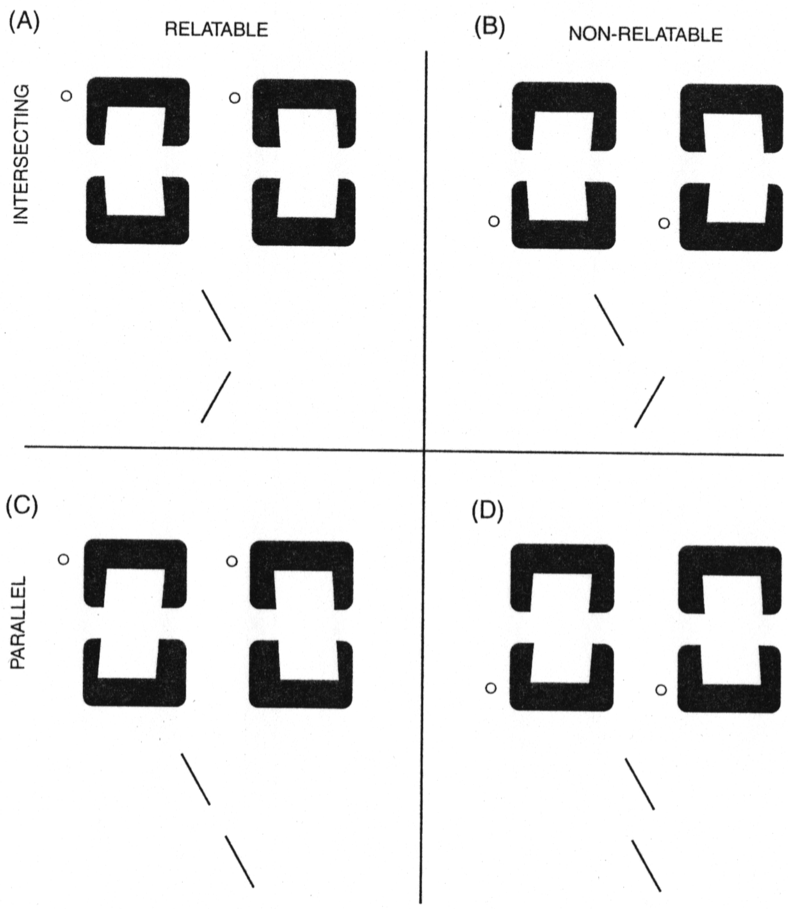


FIG. 5.13. Displays used to study three-dimensional interpolation. Displays are stereo pairs, which may be free-fused by crossing the eyes. (The two small circles in each display can be used as a guide for fusing.) Lines underneath displays indicate side views; orientation is correct for a viewer to the right. (A) three-dimensional relatable display; converging planes; (B) three-dimensional nonrelatable display; converging planes; (C) three-dimensional relatable display; parallel planes; (D) three-dimensional nonrelatable display; parallel planes. (See text.) From Kellman, Guttman & Wickens "Geometric and Neural Models of Object Perception." In *From Fragments to Objects: Segmentation and Grouping in Vision* (p. 236), by T. F. Shipley and P. J. Kellman (Eds.), 2001, Amsterdam: Elsevier. Copyright 2001 by Elsevier Science. Reprinted with permission.

(including coplanar). Note that the classification required from the subject on each trial was orthogonal to the display's status as relatable or nonrelatable. The key predictions were that perception of a unified object would facilitate classification performance, and perceived unity would depend on relatability. The former was expected based on results in two-dimensional displays showing that object completion produces an advantage in detecting boundary orientation. One advantage of this task is that, unlike the two-dimensional analogue, it requires use of a *relation* between the visible parts, which may encourage dependence on interpolation.

Results of one experiment (Kellman, Yin, Garrigan, Shipley, & Machado, 2003) are shown in Fig. 5.14. The graph shows discrimination sensitivity (d') by condition. Two values of depth displacement (used to disrupt relatability) were used, corresponding to about a 5-cm and a 10-cm shift in depth of one of the pieces from

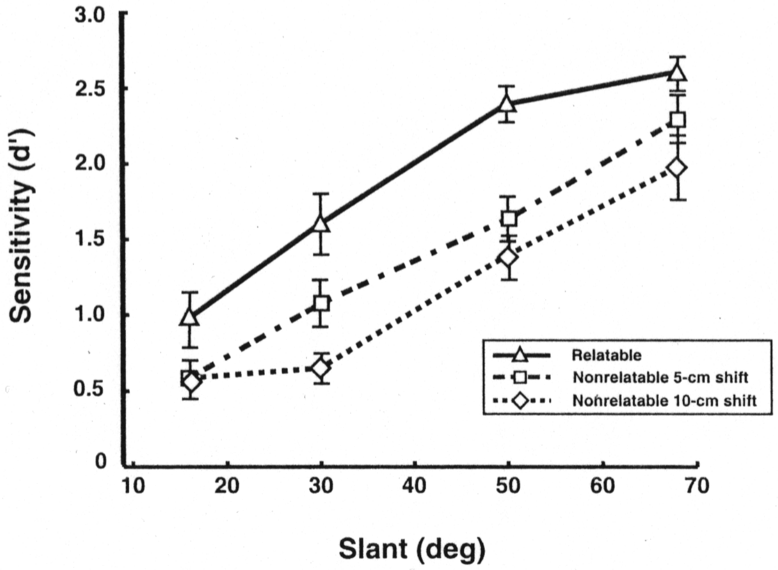


FIG. 5.14. Sensitivity results from a three-dimensional interpolation experiment. Sensitivity is plotted as a function of slant for relatable displays, nonrelatable displays in which one piece was shifted approximately 5 cm in the virtual display, and nonrelatable displays in which the shift was approximately 10 cm. Relatable displays were more accurately and rapidly classified, suggesting that the upper and lower inducing areas were processed as a connected unit. From Kellman, Yin, Garrigan, Shipley, & Machado, 2003.

the observer's viewing distance (100 cm). Participants did the task at four different values of slant; increasing slant made the classification of converging or parallel planes easier. It can be seen that the relatable displays showed a clear superiority over the nonrelatable displays at all slant values. Response times reflected the same advantage: Both parallel and converging relatable displays produced faster, as well as more accurate, responding.

These results suggest that performing the classification of displays as parallel or converging was made easier when the two surface patches were perceived as a single, connected object. Moreover, in most respects a three-dimensional version of contour relatability seems to predict the conditions under which tabs where connected into single objects. Even the smaller value of depth shift disrupted performance markedly.

These data are consistent with the notion that interpolation processes are truly three-dimensional and are described by a notion of three-dimensional relatability. At the same time, there is substantial bootstrapping in this initial set of results. In the same experiment, we validated the paradigm (as being sensitive to object completion) and revealed some determinants of object completion (e.g., it is disrupted by depth shifting of one tab relative to the other).

Accordingly, we must consider alternative explanations for the data. First, it is possible that performance in our task might not really require object completion. Perhaps relatable displays were better processed because their pieces were more nearly equidistant from the observer. Comparing two parts' orientations might be easier when the parts are equidistant. Our design allowed us to check this hypothesis using a subset of the data. As Fig. 5.13 illustrates, a subset of parallel displays used a shift away from the canonical (relatable) stimulus that actually made the two parts more nearly equidistant. We compared these displays (which had roughly 0- or 5-cm depth differences) with relatable parallel displays having parts that differed substantially in depth (10 cm for the largest slant condition). Results showed that relatability, not similarity in depth, produced superior accuracy and speed.

More recently we have used other control groups to test the idea that the effects in this paradigm are due to object completion. Our strategy has been to keep the three-dimensional geometry of the stimulus pieces as similar as possible to that of the original experiment, while using other manipulations that should disrupt object completion. In one control experiment, for example, tangent discontinuities were rounded, a manipulation that disrupts interpolation processes but should have had no effect if comparisons of tabs is coincidentally facilitated by certain geometric arrangements. In this case, the advantage of relatable over nonrelatable displays was completely eliminated. In general, all of the control experiments support the idea that object completion caused the observed performance advantage.

We also examined another important concern. The observed effects appear to be object completion effects, but are they truly *three-dimensional effects*? Introducing

binocular depth differences involves monocularly misaligning contours in each eye. Perhaps these monocular effects, not true depth effects, cause the performance decrement. It is known that misalignment of parallel or nearly parallel contours disrupts two-dimensional object completion (Kellman, Yin, & Shipley, 1998; Shipley & Kellman, 1992a).

In designing the original study, we aimed to produce significant depth shifts using misalignments that remained within the tolerances for two-dimensional completion. It has been estimated that contour completion breaks down at about 15 min of misalignment of parallel edges (Shipley & Kellman, 1992a). Our misalignments were on the order of about 10 min in the maximum depth shift condition. To check the effect of monocular misalignment, we carried out a separate experiment. In our binocular, depth-shifted displays, each eye had the same misalignment with opposite sign. In this experiment, we used the same displays but gave misalignment of the same sign in both eyes. Thus, the amount of monocular misalignment was exactly identical in every display as in the original experiment. Because both members of each stereo pair had misalignments of the same sign, shifted displays appeared to be at the same depths as relatable displays but with some lateral misalignment. Results showed no reliable accuracy or speed differences between shifted and relatable displays in this experiment. This outcome is consistent with the idea that the basic experimental effect is a three-dimensional object completion effect. The effects are not explainable by monocular misalignment.

These several experiments make a clear case for contour interpolation as a three-dimensional process. Human visual processing in object formation takes as inputs the positions and orientations of contours in three-dimensional space. Interpolation processes produce as outputs contours and surfaces that extend through all three spatial dimensions. Although this line of research is just beginning, there is considerable support for the notion that a simple piece of geometry—contour relatability—can provide not only an account of two-dimensional spatial interpolation but can be extended to three-dimensional cases.

Even the initial data on three-dimensional interpolation are very challenging for existing models of interpolation. Few if any current models take as their inputs three-dimensional positions and orientations of edges. It is possible that a three-dimensional analogue of the kinds of interactions of oriented units envisioned in neural models of two-dimensional interpolation may offer a plausible route for understanding the neural mechanisms underlying three-dimensional interpolation. Alternatively, it is possible that early interpolation processes are actually done in a two-dimensional substrate, but depth information is used to correct the results somehow. The latter possibility seems unlikely, but it would be more consistent with neural models that depend heavily in their initial stages on two-dimensional orientation information in early visual cortical areas. These possibilities remain to be explored. What is clear is that human perception produces interpolations spanning all three dimensions, whereas current models do not.

SPATIOTEMPORAL INTERPOLATION

Complex perceptual systems are exclusively the property of mobile organisms. Most research on interpolation processes in object formation has emphasized static, two-dimensional displays, but object perception processes must serve the moving observer and may need to use information available over time. When looking through dense foliage, for example, an observer may see bits of light and color from the scene behind but may be unable to detect specific objects or spatial layout. However, if the observer moves parallel to the occluding foliage, the scene behind may suddenly be revealed. This ordinary experience suggests that humans can perceive the objects in a scene from information that arrives fragmented not only in space but in time.

In recent years, my colleagues and I have sought to understand visual processes that assemble objects from information fragmentary in time and space. Early work showed that illusory figures may arise from inducing elements that are partially occluded in sequence (Kellman & Cohen, 1984; see also Bruno & Bertamini, 1990; Kojo, Liinasuo, & Rovamo, 1993). Perception under these circumstances requires that the visual system not only integrate information over time but also interpolate because some parts of the object never project to the eyes.

Existing models of visual interpolation are not set up to handle inputs that arrive fragmented over time (but see Shipley & Kellman, 1994). One issue in broadening both our understanding and our models is the difficulty of characterizing the stimulus relationships in both space and time that lead to the perception of complete, unified objects. The extra degrees of freedom given by motion make the problem formidable.

My colleagues and I approached this problem by assuming that spatiotemporal interpolation in object perception may build on the same basic geometry that underlies static, spatial interpolation. Palmer, Kellman, and Shipley (1997, 2000) proposed two hypotheses that connect the geometry of spatial relatability to the problem of spatiotemporal interpolation. Fig. 5.15 illustrates these ideas. The *persistence hypothesis* (Fig. 5.15A) suggests that the position and edge orientations of a briefly viewed fragment are encoded in a buffer, such that they can be integrated with later-appearing fragments. In Fig. 5.15A, an opaque panel containing two apertures moves in front of an object, revealing one part of an occluded object at time t_1 and another part at time t_2. If information concerning the part seen at t_1 persists in the buffer until the part at t_2 appears, then the standard spatial relatability computation can be performed to integrate the currently visible part with the part encoded earlier.

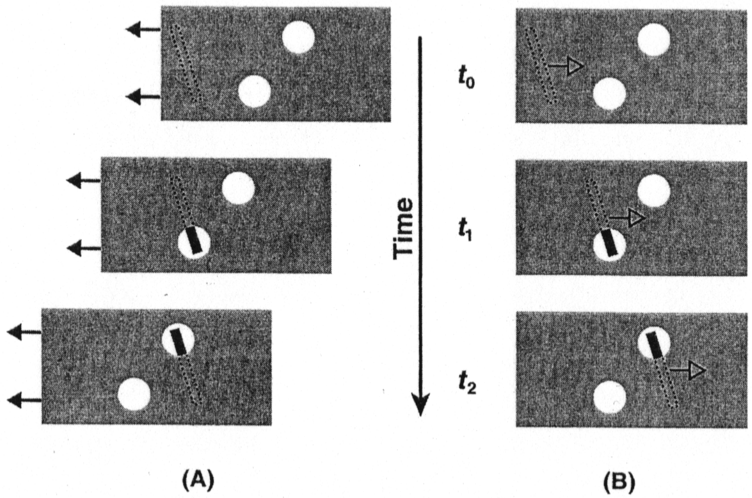


FIG. 5.15. Illustrations of (A) the persistence hypothesis and (B) the spatial updating hypothesis in spatiotemporal interpolation. (See text.) (A) The moving occluder reveals relatable parts of the rod sequentially in time (t_1 and t_2). Perceptual connection of parts requires that the initially visible part persists over time in some way. (B) Parts of the moving rod become visible through apertures sequentially in time. Perceptual connection of the parts requires not only persistence of the initially visible part but also positional updating based on velocity information. From Kellman, Guttman & Wickens "Geometric and Neural Models of Object Perception." In *From Fragments to Objects: Segmentation and Grouping in Vision* (p. 238), by T. F. Shipley and P. J. Kellman (Eds.), 2001, Amsterdam: Elsevier. Copyright 2001 by Elsevier Science. Reprinted with permission.

In Fig. 5.15B, the object moves behind a stationary occluder, again revealing one part through the bottom aperture at t_1 and a second part through the top aperture at t_2. This figure illustrates the *spatial updating hypothesis*. According to this idea, the visual system encodes a velocity signal of any moving objects or surfaces, in addition to their positions and edge orientations; once these surfaces become occluded, the visual system uses the velocity signal to update their spatial position over time. Thus, when a later-appearing object part (upper aperture at t_2) becomes visible, it can be combined with the updated position of the earlier-appearing part (lower aperture at t_1) using the standard spatial relatability computation. Together with the spatial notion of relatability, the persistence and spatial updating hypotheses comprise the notion of *spatiotemporal relatability*.

Palmer, Shipley and I (1997) developed an experimental paradigm to test spatiotemporal relatability. On each trial, an object passed once back and forth behind

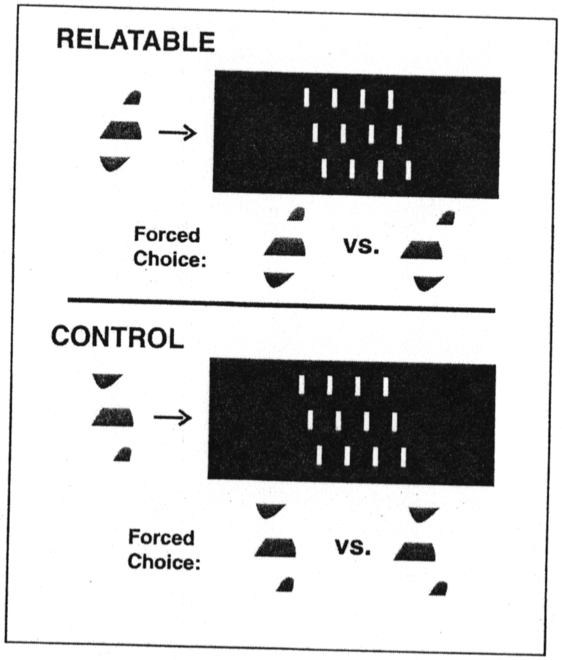


FIG. 5.16. Design of a spatiotemporal interpolation experiment. (See text.) From Kellman, P. J. "An Update on Gestalt Psychology." *Essays in Honor of Henry and Lila Gleitman* by B. Landau, J. Jonides, E. Newport, and J. Sabini (Eds.), 2000, Cambridge, MA: MIT Press. Copyright 2000 by MIT Press. Reprinted with permission.

an occluder with several narrow slits, vertically separated so that some parts of the object never projected to the eyes. This feature makes the task a completion or interpolation task as opposed to only an integration task (where visible parts are integrated over time). Subjects made a forced choice between two test displays, choosing which matched the moving target display. The design is illustrated in Fig. 5.16.

Three display conditions were used. Relatable displays (apart from the shift manipulation; see later text) met the criterion of spatiotemporal relatability. The upper test display in Fig. 5.16 is an example. The other test display differed from the first by having one of the three fragments shifted by some amount. Five different amounts of shift (ranging from 1.67 arcmin to 8.33 arcmin of visual angle) were used. The target matched unshifted and shifted test displays each equal numbers trials.

We predicted that relatability would facilitate encoding of the visible parts in the target display. If three parts moving behind slits were grouped into a single, coherent object, this might lead to more economical encoding and memory than for control displays, in which three detached pieces were encoded. For simplicity, I will consider here only the cases in which either a test display or both the target and a test display were relatable. In these cases, it was predicted that the greater ease of encoding a relatable display would lead to better performance.

Displays in two control conditions were compared with the first. Nonrelatable displays consisted of the identical three pieces as in the relatable condition, but the top and bottom pieces were permuted. (See Fig. 5.16.) It was hypothesized that visual completion would not occur with these nonrelatable displays. Because each nonrelatable target might have to be encoded as three distinct pieces, it would lead to greater encoding demands and lower sensitivity to the relative spatial positions of the three parts. The third condition left the overall geometry of the relatable displays but sought to disrupt interpolation by rounding the tangent discontinuities of the visible parts. To make the rounded corners visible, the occluder was moved away slightly from visible parts.

A series of experiments has confirmed the usefulness of spatiotemporal relatability in describing relations over space and time that produce representations of connected objects. Fig. 5.17 shows accuracy data (discrimination d') from two experiments. In Fig. 5.17, part A, the data show means for 16 subjects for relatable and nonrelatable displays as a function of shift. Relatable displays were far more accurately discriminated than displays made of the identical physical parts but placed in nonrelatable (permuted) positions. The control group in which the corners or visible parts were rounded to eliminate tangent discontinuities provided additional confirmation that the initial results were due to interpolation effects. As predicted, rounding of tangent discontinuities weakened performance, despite the figure fragments appearing in the same positions and relations as in the relatable displays. Performance was not reduced as much as in the nonrelatable condition, however; this result may indicate that junction detectors at low spatial frequencies still signal junctions in these rounded displays.

Results of later experiments showed that illusory contour versions of these displays produced nearly identical speed and accuracy data as in the occluded case (Fig. 5.17B). This suggests that the identity hypothesis in contour interpolation applies to spatiotemporal interpolation.

The results support the notion of spatiotemporal relatability as a description of the geometry of object formation over time. There are numerous issues yet to be addressed. These and other studies in the spatiotemporal domain, however, are beginning to reveal the details of the persistence and spatial updating aspects of spatiotemporal object formation. The results suggest that connecting dynamic object perception to previous work with static displays is plausible in terms of the

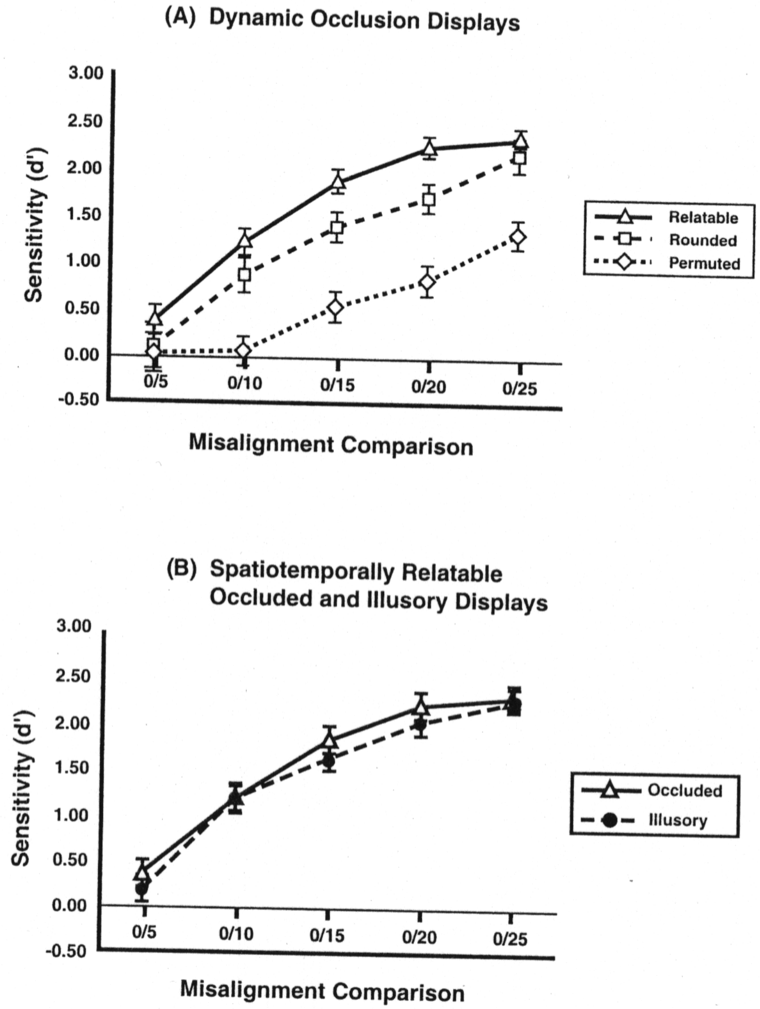


FIG. 5.17. Results of spatiotemporal interpolation experiments. Sensitivity (d') is plotted as a function of the amount of shift distinguishing the aligned and misaligned choices presented to the subject. (A) Results from dynamic occlusion displays with spatiotemporally relatable and nonrelatable (permuted) displays, as well as relatable displays in which interpolation was disrupted by rounding off tangent discontinuities. (B) Results from dynamic occlusion and illusory contour versions of the experiment. (Displays had identical physically specified edges and gaps.) From Palmer, Kellman & Shiplay (2003).

underlying geometry, over space and time, that allows fragments to be connected perceptually.

GLOBAL AND LOCAL PROCESSING IN INTERPOLATION

Most of our discussion to this point has focused on relatively local determinants of object perception. It has often been argued that in addition to these, there are more global influences. Perhaps the original idea of this type is the Gestalt principle of *good form,* which suggests that perceptual processes tend toward simple or regular (symmetric) outcomes (e.g., Koffka, 1935). Subsequent research attempted to formalize this notion (e.g., Buffart, Leeuwenberg, & Restle, 1981), and some provided empirical support for a role of symmetry in object completion (Sekuler, Palmer, & Flynn, 1994).

Both the role of global factors and the relationship between global and local factors in object formation remain unsettled. In object completion, Kanizsa (1979) argued that that global symmetry is a questionable or weak determinant, based on demonstrations that pitted global factors against local edge continuity. Evidence from priming studies, however, tends to show global effects (Sekuler, Palmer, & Flynn, 1994), along with local ones (Sekuler et al., 1994; van Lier, van der Helm, & Leeuwenberg, 1995). Van Lier et al. interpreted their results in terms of dual or multiple representations activated by partly occluded displays.

An example of a display for which global completion has been claimed is shown in Fig. 5.18, part A. Global completion entails seeing a fourth articulated part behind the occluder, making the display radially symmetric. The argument is that the process of contour interpolation somehow takes into account the potential for a symmetric shape in cases like this and represents boundaries consistent with that shape. In this example, the local completion (via relatability) would consist of a smooth, monotonic, constant curvature connection. (See Kellman, Guttman, & Wickens, 2001, for a discussion of several variants of claims about global completion.)

Does the visual system really interpolate edges to construct symmetric object representations? Here is a case in which the identity hypothesis can shed light on interpolation processes (or at least raise an interesting question). Recall the argument that a common contour interpolation process underlies occluded and illusory contours. If so, we might learn something by examining the idea of global completion not just under occlusion but in the illusory contour domain. Fig. 5.18B shows an illusory contour display that has physically given edges equivalent to the occlusion display in Fig. 18A. If completion operates to produce symmetric perceived objects, the four-lobed figure should be obvious, complete with clear illusory contours, in Fig. 5.18B. In fact, however, no global completion is evident in Fig. 5.18B. To my knowledge, there is not a single report of completion based on global symmetry of this sort in illusory object displays among all the hundreds of published reports on this topic. What might account for subjective reports and priming data indicating at least some tendency toward global completion under occlusion, but none in illusory object cases?

I have suggested the hypothesis that the discrepancy indicates the operation of two distinct categories of processing (e.g., Kellman, 2001). One is a bottom-up, relatively local process that produces representations of boundaries according to the relatability criterion. This process is *perceptual* in that it involves a modular process that takes stimulus relationships as inputs and produces boundaries and forms as outputs. The other process is more top-down, global, and cognitive, coming into play when familiar or symmetric forms can be recognized. For lack of a more concise label, we call it *recognition from partial information (RPI).*

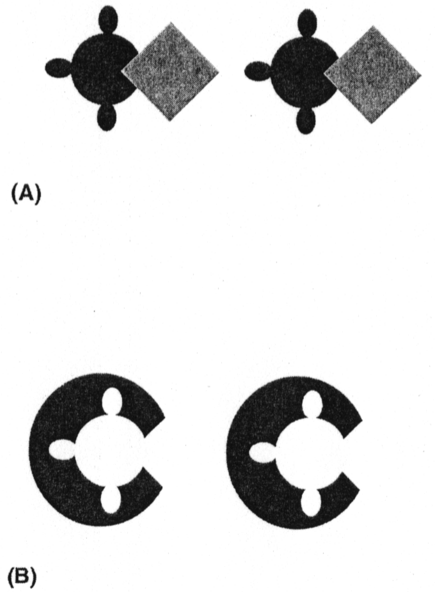


FIG. 5.18. (A) Example of an occlusion display used to study global completion. (B) Illusory contour version of the occlusion display in (A). Displays are stereo pairs, which may be free-fused by crossing the eyes. (The stereoscopic effect enhances the effect but is not necessary, as can be seen by viewing any single image.) According to the recognition-from-partial information hypothesis (see text), contour interpolation processes do not produce global (symmetric) completion (appearance of a fourth lobe of the central figure) in either display. In the occlusion display, observers perceive that part of the figure is out of sight, allowing the possibility of reports of symmetric completion.

If the identity hypothesis is true, why should global completion occur in occluded but not illusory object displays? The answer may be that the displays are the same in terms of the perceptual process of contour interpolation but different in terms of RPI. An occluded surface is an interpolated surface that is not nearest to the observer in some visual direction (i.e., there is something in front of it). An illusory surface is nearest to the observer among all surfaces in a certain visual direction. The crucial consequence of the difference is this: An observer viewing an occluded display is aware that part of the object is hidden from view. This allows certain kinds of reasoning and responses that are not sensible when no part of an object is occluded. In particular, despite any local completion process, *the observer can notice what parts are visible (unoccluded) and whether they are consistent with some familiar or symmetric object.*

Consider a concrete example. If you see the tail of your calico cat protruding from under the sofa, you may easily recognize and report that the cat is present, even though the particular contours and surfaces of the hidden parts of the cat are not given perceptually. A stored representation of the cat may be activated and a belief about its presence may be formed. But RPI differs from perceptual processes that actually specify the positions of boundaries and surfaces behind an occluder.

This separation of processes might explain the continuing disagreements about global and local processing. First, objective data supporting global outcomes have come solely from priming studies. It is well-known that priming occurs at many levels, from the most basic representation of the stimulus to higher conceptual classifications involving the stimulus (e.g., Kawaguchi, 1988). To my knowledge, no attempts have been made to specify the locus of priming influences in occlusion studies. Studies reporting global completion have typically used large numbers of trials with a small set of familiar and symmetric figures, such as circles and squares. Even if the subjects start out with little familiarity or do not notice the possibility of symmetry under occlusion, repeated exposure may produce familiarity or symmetry responses.

The observation that different levels have not been teased apart in priming studies on occlusion is not meant as a criticism. Priming may not be suitable for separating perceptual processes from more cognitive influences in this realm. Testing the possibility of different processes may require a different experimental paradigm. My colleagues and I developed such a paradigm, focusing on the idea that perceptual boundary completion processes lead to relatively specific representations of boundary locations; RPI will not do so, as in our occluded cat example. We measured the precision of boundary location by showing an occluded display and briefly flashing a probe dot in front of the occluder. Subjects were instructed to respond on each trial whether the probe dot fell inside or outside the occluded object's boundaries (i.e., whether the projection of the occluded object to the eye would or would not encompass the dot).

Adaptive staircase procedures were used. The stimulus value for each trial changed depended on the subject's responses. Systematic changes allowed estimation of particular points on a psychometric function. For each display, we used both a two-up, one-down and a one-up, two-down staircase to estimate two points: the .707 probability of seeing the dot as outside the boundary and .707 probability of seeing the dot inside the boundary (= .293 probability of outside). We took the *difference* between these estimates as a measure of the *precision* of boundary perception, and the *mean* of these estimates as an estimate of the perceived *location* of the boundary. Staircases for several stimulus patterns were interleaved, (i.e., patterns appeared in a random order, and screen position was varied randomly).

We realized that competing perceptual and recognition processes might lead to different strategies across subjects. Therefore, we gave subjects explicit strategy instructions. In the *global instruction condition,* we told subjects that they should see the display as symmetric; for the display in Fig. 5.18A, for example, they were told that there was a fourth protrusion behind the occluder identical to the three visible protrusions around the circle. In the *local instruction condition,* we told them that we wanted them to see the display as containing a simple curve connecting the two visible edges. In this manner, we sought to find the subjects' *best* ability to localize boundaries under a global or local set.

A number of interesting findings emerged (Kellman, Temesvary, Palmer, & Shipley, 2000). Localization of boundaries in displays where completion is predicted by relatability is extremely precise. This is true for both straight (collinear) and curved completions. The outcomes in cases where completion is predicted to follow global symmetry are quite different. Here, the range (difference between out and in thresholds) is far greater (approaching an order of magnitude greater). Not only is localization imprecise in the global cases, but it is also inaccurate. The estimate of subjects' perceived position of the boundary is consistently and substantially different from its theoretically predicted location. This result has shown up consistently in a variety of displays testing symmetry and related global notions of object completion. Many questions are still being investigated in this paradigm. What is already clear is that global influences do not lead to specification of precise boundary position in the way local perceptual completion does. These outcomes are consistent with the idea of separate perceptual completion and more cognitive RPI processes. For a more extensive discussion of these ideas, see Kellman (2001).

CONCLUSIONS

We have examined a wide range of phenomena, theories, and issues in object perception. Perhaps the most conspicuous conclusion of our tour is that this fundamental and active area of research poses a vast array of challenges to scientists interested in vision and cognition. It will take a great deal of work just to fill in

the details of the framework given in Fig. 5.1. Moreover, it is a certainty that some of its processes, representations, and relations will be altered or reconfigured as progress occurs.

The other conclusion that pulls together most threads of our discussion is the idea that object perception is a four-dimensional process, in the sense that it depends on all three spatial dimensions as well as time. Researchers have come a long way in terms of understanding the phenomena of two-dimensional segmentation, grouping, and object formation. Our models of these processes have progressed in specifying their geometry, computations, and, to some degree, their neural implementation.

Yet the frontiers beckon. Truly realistic and comprehensive explanations will incorporate the ideas of the perceiver and the objects to be perceived existing in a three-dimensional world and moving relative to each other. The data are already clear in showing that we have capacities to detect relationships and produce representations that involve three-dimensional space and time. Achieving a detailed understanding of the information, processes, and neural circuitry involved will require considerably more work.

In exploring these frontiers, however, there is reason for early optimism. Investigations suggest that a simple piece of geometry, the notion of contour relatability, may account for many of the cases in which the visual system connects fragments into objects, not only in static, two-dimensional arrays but also, with simple extensions, in the three- and four-dimensional space-time in which we must perceive and act.

ACKNOWLEDGMENTS

Preparation of this chapter was supported in part by research grants from the National Science Foundation (SBR 9496112) and the National Eye Institute (R01 EY13518) to P. Kellman. I thank Patrick Garrigan, Sharon Guttman, and Evan Palmer for helpful discussions, Marlene Behrmann and Rutie Kimchi for useful comments on an earlier draft, and Jennifer Vanyo for general assistance. Correspondence should be addressed to Philip J. Kellman, UCLA Department of Psychology, 405 Hilgard Avenue, Los Angeles, CA 90095-1563 or via email (Kellman@cognet.ucla.edu).

REFERENCES

Anderson, B. L., Singh, M., & Fleming, R. W. (2002). The interpolation of object and surface structure. *Cognitive Psychology, 44,* 148–190.

Banton, T., & Levi, D. M. (1992). The perceived strength of illusory contours. *Perception and Psychophysics, 52,* 676–684.

Barrow, H. G., & Tenenbaum, J. M. (1986). Computational approaches to vision. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance: Vol. 2. Cognitive processes and performance* (pp. 38.1–38.70). New York: Wiley.

Biederman, I. (1995). Visual object recognition. In S. M. Kosslyn & D. N. Osherson (Eds), *Visual cognition: An invitation to cognitive science* Vol. 2, 2nd ed., (pp. 121–165). Cambridge, MA: MIT Press.

Bruno, N., & Bertamini, M. (1990). Identifying contours from occlusion events. *Perception & Psychophysics, 48,* 331–342.

Buffart, H., Leeuwenberg E., & Restle, F. (1981). Coding theory of visual pattern completion. *Journal of Experimental Psychology: Human Perception & Performance. 7*(2), 241–274.

Campbell, F. W., Cooper, G. F., & Enroth-Cugell, C. (1969). The spatial selectivity of the visual cells of the cat. *Journal of Physiology, 203,* 223–235.

Field, D., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local "association field." *Vision Research, 33,* 173–193.

Fischer, B., & Poggio, G. F. (1979). Depth sensitivity of binocular cortical neurons of behaving monkeys. *Proceedings of the Royal Society of London B, 1157,* 409–414.

Geisler, W. S., Perry, J. S., Super, B. J., & Gallogly, D. P. (2001). Edge co-occurrence in natural images predicts contour grouping performance. *Vision Research, 41,* 711–724.

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

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton-Mifflin.

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

Gillam, B., Chambers, D., & Russo, T. (1988). Postfusional latency in stereoscopic slant perception and the primitives of stereopsis. *Journal of Experimental Psychology: Human Perception and Performance, 14*(2), 163–175.

Gold, J. M., Murray, R. F., Bennett, P. J., & Sekuler, A. B. (2000). Deriving behavioural receptive fields for visually completed contours. *Current Biology, 10* (11), 663–666.

Gregory, R. L., & Harris, J. P. (1974). Illusory contours and stereo depth. *Perception and Psychophysics, 15*(3), 1974, pp. 411–416.

Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception and Psychophysics, 55,* 48–120.

Grossberg, S., & Mingolla, E. (1985). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review, 92,* 173–211.

Guttman, S. E., & Kellman, P. J. (2000). Seeing between the lines: Contour interpolation without perception of interpolated contours. *Investigative Ophthalmology and Visual Science, 41*(4), S439.

Guttman, S. E., & Kellman, P. J. (2003). *Contour interpolation: Necessary but not sufficient for the perception of interpolated contours.* Manuscript in preparation.

Heitger, F., Rosenthaler, L., von der Heydt, R., Peterhans, E., & Kübler, O. (1992). Simulation of neural contour mechanisms: From simple to end-stopped cells. *Vision Research, 32,* 963–981.

Heitger, F., von der Heydt, R., Peterhans, E., Rosenthaler, L., & Kübler, O. (1998). Simulation of neural contour mechanisms: Representing anomalous contours. *Image and Vision Computing, 16,* 407–421.

Hoffman, D. D. (1998). *Visual intelligence: How we create what we see.* New York: W. W. Norton.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *Journal of Physiology, 195,* 215–243.

Humphreys, G. W., Cinel, C., Wolfe, J., Olson, A. & Klempen, A. (2000). Fractionating the binding process. Neuropsychological evidence distinguishing binding of form from binding of surface features. *Vision Research, 40,* 1569–1596.

Johansson, G. (1970). On theories for visual space perception: A letter to Gibson. *Scandinavian Journal of Psychology, 11*(2), 67–74.

Julesz, B. (1971). *Foundations of cyclopean perception.* Chicago: University of Chicago Press.

Kanizsa, G. (1979). *Organization in vision.* New York: Praeger.

Kawaguchi, J. (1988). Priming effect as expectation. *Japanese Psychological Review, 31,* 290–304.

Kellman, P. J. (2000). An update on Gestalt psychology. In B. Landau, J. Jonides, E. Newport, & J. Sabini (Eds.), *Essays in honor of Henry and Lila Gleitman* (pp. 157–190). Cambridge, MA: MIT Press.

Kellman, P. J. (2001). Separating processes in object perception. *Journal of Experimental Child Psychology, 78*(1), 84–97.

Kellman, P. J. (2003). *Relatability and good continuation.* Manuscript in preparation.

Kellman, P. J. & Arterberry, M. (1998). *The Cradle of Knowledge: Perceptual Development in Infancy.* Cambridge, MA: MIT Press.

Kellman, P. J., & Cohen, M. H. (1984). Kinetic subjective contours. *Perception and Psychophysics, 35,* 237–244.

Kellman, P. J., Guttman, S., & Wickens, T. (2001). Geometric and neural models of contour and surface interpolation in visual object perception. In Shipley, T. F. & Kellman, P. J. (Eds.) *From fragments to objects: Segmentation and grouping in vision.* (pp. 183–245). Elsevier Press.

Kellman, P. J., & Loukides, M. G. (1987). An object perception approach to static and kinetic subjective contours. In S. Petry & G. E. Meyer (Eds.), *The perception of illusory contours* (pp. 151–164). New York: Springer-Verlag.

Kellman, P. J., Machado, L. J., Shipley, T. F., & Li, C. C. (1996). 3-D determinants of object completion. *Investigative Ophthalmology and Visual Science, 37,* S685.

Kellman, P. J., & Shipley, T. F. (1991). A theory of visual interpolation in object perception. *Cognitive Psychology, 23,* 141–221.

Kellman, P. J., & Shipley, T. F. (1992). Visual interpolation in object perception. *Current Directions in Psychological Science, 1*(6), 193–199.

Kellman, P. J., Shipley, T. F., & Kim, J. (1996, November). *Global and local effects in object completion: Evidence from a boundary localization paradigm.* Paper presented at the 37th Annual Meeting of the Psychonomic Society, Chicago, IL.

Kellman, P. J., Temesvary, A., Palmer, E. M., & Shipley, T. F. (2000). Separating local and global processes in object perception: Evidence from an edge localization paradigm. *Investigative Ophthalmology and Visual Science, 41*(4), S741.

Kellman, P. J., Yin, C., Garrigan, P., Shipley, T. F., & Machado, L. J. (2003). *Visual interpolation in three dimensions.* Manuscript in preparation.

Kellman, P. J., Yin, C., & Shipley, T. F. (1998). A common mechanism for illusory and occluded object completion. *Journal of Experimental Psychology: Human Perception and Performance, 24,* 859–869.

Koffka, K. (1935). *Principles of Gestalt Psychology.* NY: Harcourt.

Kojo, I., Liinasuo, M., & Rovamo, J. (1993). Spatial and temporal properties of illusory figures. *Vision Research, 33,* 897–901.

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

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

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

Morrone, M. C., & Burr, D. C. (1988). Feature detection in human vision: A phase-dependent energy model. *Proceedings of the Royal Society of London B, 235,* 221–245.

Nakayama, K., Shimojo, S., & Silverman, G. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception, 18,* 55–68.

Palmer, E. M., Kellman, P. J., & Shipley, T. F. (1997). Spatiotemporal relatability in dynamic object completion. *Investigative Ophthalmology and Visual Science, 38*(4), S256.

Palmer, E. M., Kellman, P. J., & Shipley, T. F. (2000). Modal and amodal perception of dynamically occluded objects. *Investigative Ophthalmology and Visual Science, 41*(4), S439.

Palmer, E. M., Kellman, P. J., & Shipley, T. F. (2003). *Contour relatability in spatiotemporal object perception.* Manuscript in preparation.

Palmer, S., & Rock, I. (1994). Rethinking perceptual organization: The role of uniform connectedness. *Psychonomic Bulletin & Review, 1,* 29–55.

Peterhans, E., & von der Heydt, R. (1989). Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. *Journal of Neuroscience, 9,* 1749–1763.

Peterhans, E., & von der Heydt, R. (1991). Subjective contours: Bridging the gap between psychophysics and physiology. *Trends in Neurosciences, 14,* 112–119.

Peterhans, E., & von der Heydt, R. (1993). Functional organization of area V2 in alert macaque. *European Journal of Neuroscience, 5,* 509–524.

Peterson, M. A. (1994). Object recognition processes can and do operate before figure-ground organization. *Current Directions in Psychological Science, 3*(4), 105–111.

Peterson, M. A., & Gibson, B. S. (1991). The initial identification of figure-ground relationships: Contributions from shape recognition processes. *Bulletin of the Psychonomic Society, 29,* 199–202.

Peterson, M. A., & Gibson, B. S. (1994). Must figure-ground organization precede object recognition? An assumption in peril. *Psychological Science, 5,* 253–259.

Petter, G. (1956). Nuove ricerche sperimentali sulla totalizzazione percettiva. *Rivista di Psicologia, 50,* 213–227.

Ringach, D. L., & Shapley, R. (1996). Spatial and temporal properties of illusory contours and amodal boundary completion. *Vision Research, 36,* 3037–3050.

Rock, I. (1977). In defense of unconscious inference. In W. Epstein (Ed.), Stability and constancy in visual perception. (pp. 321–373), NY: Wiley.

Rubin, E. (1915). *Synsoplevede figurer.* Copenhagen: Gyldendal.

Sekuler, A. B., Palmer, S. E., & Flynn, C. (1994). Local and global processes in visual completion. *Psychological Science, 5,* 260–267.

Shimojo, S., Silverman, G. H., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem for motion. *Vision Research*, 29, 619–626.

Shipley, T. F., & Kellman, P. J. (1990). The role of discontinuities in the perception of subjective figures. *Perception and Psychophysics, 48,* 259–270.

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 and Performance, 18,* 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 and Psychophysics, 52,* 97–106.

Shipley, T. F., & Kellman, P. J. (1994). Spatiotemporal boundary formation. *Journal of Experimental Psychology: General, 123,* 3–20.

Singh, M., & Hoffman, D. D. (1999). Completing visual contours: The relationship between relatability and minimizing inflections. *Perception and Psychophysics, 61,* 943–951.

van Lier, R. (1999). Investigating global effects in visual occlusion: From a partly occluded square to the back of a tree-trunk. *Acta Psychologica, 102,* 203–220.

van Lier, R. J., van der Helm, P. A., & Leeuwenberg, E. L. J. (1995). Competing global and local completions in visual occlusion. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 571–583.

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

Wurtz, R. P., & Lourens, T. (2000). Corner detection in color images through a multiscale combination of end-stopped cortical cells. *Image and Vision Computing, 18,* 531–541.

Yarbus, A. L. (1967). *Eye movements and vision.* New York: Plenum Press.

Yen, S. C., & Finkel, L. H. (1998). Extraction of perceptually salient contours by striate cortical networks. *Vision Research, 38,* 719–741.

Yin, C., Kellman, P. J., & Shipley, T. F. (1997). Surface completion complements boundary interpolation in the visual integration of partly occluded objects. *Perception, 26,* 1459–1479.

Yin, C., Kellman, P. J., & Shipley, T. F. (2000). Surface integration influences depth discrimination. *Vision Research, 40,* 1969–1978.