In Shipley, T.F. & Kellman, P.J. (Eds.). (2001). *From fragments to objects: Segmentation and grouping in vision*. Oxford, UK: Elsevier Science Publishers.

7

Geometric and Neural Models of Object Perception

Philip J. Kellman, Sharon E. Guttman, and Thomas D. Wickens
Department of Psychology, University of California, Los Angeles, CA 90095-1563 USA

Introduction

It is an exciting time to study visual object perception. Although object perception research has a long tradition, lately its visibility in cognitive science and neuroscience has greatly increased. One reason for heightened interest is that diverse areas of research now suggest a central role for objects in many aspects of human cognition, including the organization of attention, perception, knowledge representation, and language.

Meanwhile, approaches to studying object perception have expanded and matured. Since the Gestalt psychologists first framed basic questions, significant progress has been made in identifying key principles and describing important phenomena. For the most part, however, these ideas have not coalesced into a coherent structure. A textbook in perception is more likely to offer a catalog of phenomena on "perceptual organization" than a systematic account of how objects are perceived.

The situation is changing. We can glimpse, if only schematically, an interrelated set of information processing tasks that enable us to perceive objects. Moreover, we are beginning to understand both the computations and the neural mechanisms that accomplish these tasks. We owe this current good fortune to several developments, not least of which is an expanding body of research on contour, surface, and object perception. More formal computational analyses of these problems have also progressed substantially. Accompanying these recent developments are comparatively mature psychophysical and neurophysiological accounts of the earliest stages of visual cortical processing. Together, these converging areas of research provide a strong foundation for understanding functions of the visual system that depend on, but go far beyond, basic sensitivity to contrast and orientation.

In this chapter, we have several goals. One is to present an overall theoretical picture of the processes of object perception, extending from the early extraction of edges and junctions to the higher-order tasks of unit formation and shape perception. This framework sets the stage for our second goal: to emphasize several issues that confront object perception researchers and indicate specific questions for continuing research. As we will see, some tasks of object perception are relatively well understood, whereas others remain vague. Much of the value of any overall framework lies in highlighting areas where more work is needed.

Our final goal in this chapter is to enlarge the domain. The study of visual object perception has tended to focus on static, two-dimensional (2-D) images. In the natural environment, human perception both grapples with and benefits from information in three dimensions and information given over time through object and observer motion. The research that we will consider on three-dimensional (3-D) and kinematic object perception falls well outside the scope of existing models. Nonetheless, we will indicate points of continuity in the constraints and processes that may, in time, lead toward a unified account of two-dimensional, three-dimensional, and dynamic processing in object perception.

Geometric, Process, and Neural Models

Efforts at modeling the processes of object perception have evolved in two different directions. Investigations of one kind have addressed the stimulus relationships that govern the perception of objects. The aim of these efforts is to specify precisely the spatial and temporal relations of contours and surfaces that determine perception of an object's unity and shape. We will label the theoretical accounts derived from these investigations as *geometric models*. One such approach — Kellman and Shipley's (1991) model and its extensions — will be considered in detail.

Other efforts, somewhat independent from the first type, have focused on devising *neural models* that perform the kinds of computations necessary for various aspects of object perception. In the section of this chapter on neural models, we will examine some of the psychophysical and neurophysiological research on which these models are based, and spotlight three models of object perception processes: the model of Heitger, Rosenthaler, von der Heydt, Peterhans, and Kübler (1992) for edge and junction detection; the model of Yen and Finkel (1998) for contour integration; and a later elaboration of the Heitger et al. model (Heitger, von der Heydt, Peterhans, Rosenthaler, & Kübler, 1998) as an example of a neural-style model of contour interpolation. These computational models have been chosen both for the success of their simulations and, more importantly, their use of biologically plausible mechanisms in their implementation.

One might think that detailed proposals concerning the neural interactions subserving object perception would await a precise understanding of relevant stimulus relationships. That is, neural models might presuppose complete geometric models. In practice, this has not been the case. Rather, concepts for implementing contour, surface, and object processes via neural circuitry have co-evolved with experimental and theoretical work on the stimulus relationships that govern object perception.

This co-evolution has important implications. On the positive side, work on issues of implementation in neural circuitry need not await a finished geometry and psychophysics of object perception. On the negative side, existing neural models do not implement all that is known even now about the geometry of object formation, nor, obviously, can they encompass factors that are not yet determined. Our task, then, after describing the different types of models, will be to assess their relationships: how geometric and neural models can each advance the other type, and, indeed, how they can merge into a complete picture of object perception. As this complete picture is still beyond reach, we use the opportunity to highlight issues for future research.

A Framework for Object Perception: Tasks, Geometry, and Processes

Much of what we know about object perception can be captured in the framework displayed in Figure 1. This process model combines established findings with several hypotheses about the representations and information processing tasks involved with visual object perception.

In the model, rectangular boxes indicate functions or processes, and octagonal ones indicate representations. Note that the model is rather conservative regarding representations. Aside from output representations of shape and unity (the specification of which regions belong to a single form), there is only one intermediate representation: the *visible regions representation*. The nature and evidence supporting the existence of these representations will be discussed below.

Overview of the Model

The processing scheme described by Figure 1 begins with the input to the model: the optic array itself. Although much of our discussion will focus on a single static image as the input, a complete model should utilize the time-varying optic array, sampled by two eyes of an observer. Both depth and motion play important roles in segmentation and grouping, as we will consider later.

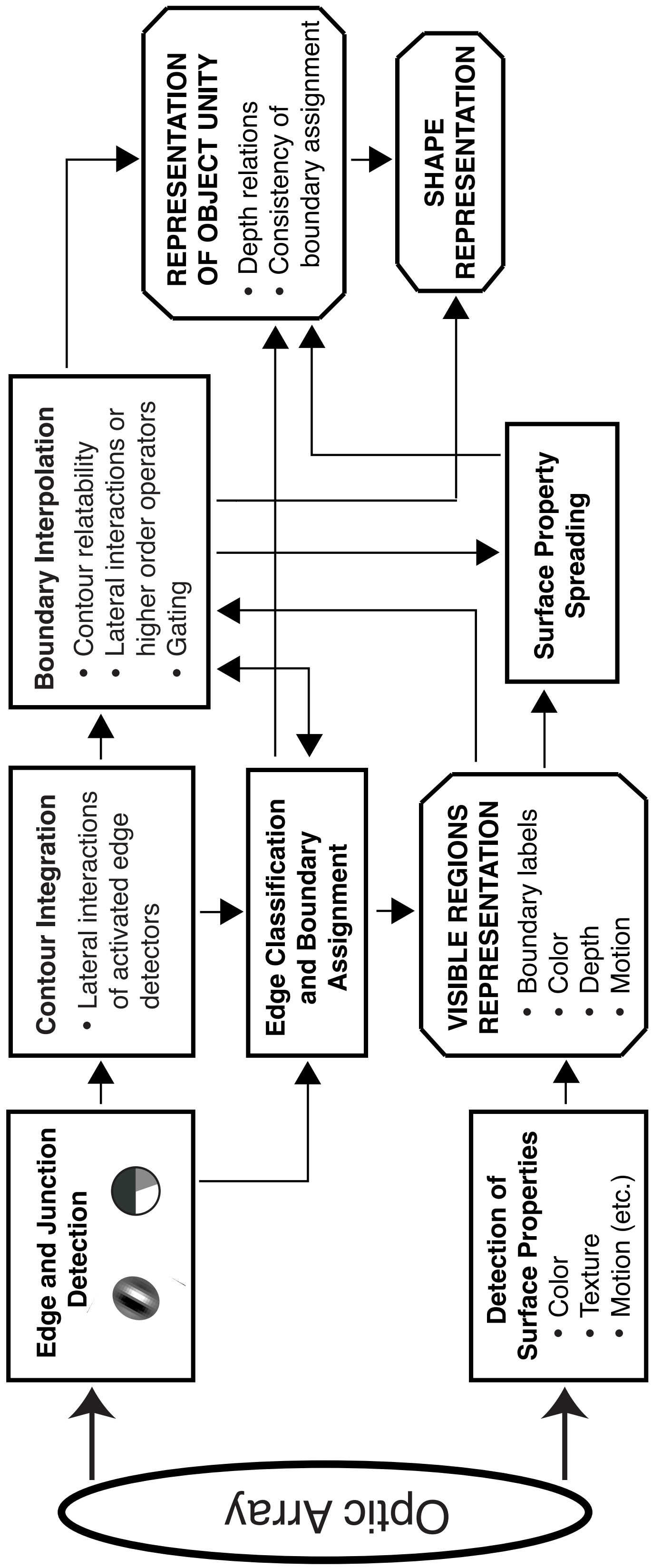


Figure 1. A Framework for Object Perception. Rectangles indicate functions or processes and octagons indicate representations. See text for details.

The visual system extracts two types of information from the optical input. Characteristics of luminance, color, texture, depth, and motion enter a surface processing stream, which represents these properties in relation to their surface locations; this information later will be used to help determine connections among spatially-distinct visible regions. *Discontinuities* in luminance, color, texture, depth, and motion enter a separate stream concerned with the detection of edges and junctions, and, later, the processing of meaningful contours.

In the contour stream, local activations of orientation-sensitive units are integrated according to their spatial relations to form visible contours. Some of these contours are classified as occluding edges, based in part on junction information. At occluding edges, the direction of boundary assignment is determined, thus indicating which of two adjacent surfaces "owns" the contour.

In the surface stream, spatially contiguous locations possessing homogeneous or smoothly varying surface attributes, including depth, become grouped together. The grouping in this stream complements the edge process, in that it depends on the *absence* of the surface discontinuities extracted by the edge stream. Together, the contour and surface streams define tokens in the visible regions representation. This representation labels relatively homogeneous regions as connected surface areas, encodes the locations and orientations of edges and corners of these regions, and specifies for each edge whether it is owned by that region or by another region.

Because of occlusion, visible regions are not objects. In fact, they bear a complex relationship to the objects in the physical world. For the objects of perception to correspond to meaningful objects in the world, interpolation processes must operate to connect visible regions under occlusion.

Evidence suggests the existence of two such interpolation processes (Kellman & Shipley, 1991; Yin, Kellman, & Shipley, 1997). The *boundary interpolation process* connects oriented edges across gaps, according to the geometry of contour relatability. These interpolated boundaries most often appear as occluded contours, but given certain depth relationships, may also be perceived as illusory contours. The *surface interpolation process* complements the boundary process, in that it can lead to perceived connections among visible regions even when the object's boundaries are not well specified. Unlike the boundary interpolation process, however, operation of the surface interpolation process requires that two visible surfaces match or fall along a smooth gradient. When one of these criteria is satisfied, surface qualities spread under occlusion within real and interpolated boundaries.

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. 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.

The framework described operates “bottom-up.” That is, the basic process model does not incorporate any feedback from higher levels to earlier ones. Object perception undoubtedly can proceed without such feedback, and likely does so in cases where there is no obvious involvement of familiarity or symmetry. Whether there really are top-down influences in basic segmentation and grouping processes, as opposed to recognition from partial input, remains controversial (e.g., Kellman, 2000; van Lier, 1999), as we will see below. One valuable aspect of the current framework is that it allows us to consider explicitly the loci and nature of putative top-down effects.

As an example, Peterson and her colleagues (e.g., Peterson, 1994; Peterson & Gibson, 1991, 1994) have argued that figure-ground segregation in otherwise ambiguous stimuli can be influenced by the familiarity of a shaped region; the familiar shape is more likely to be seen as figure. Such an effect could be incorporated into the model as shown in Figure 2. Figure-ground determination corresponds to boundary assignment in the model. For familiar shape to influence boundary assignment, the shape of some contour or region must be encoded and recognized — matched to a representation stored in memory. As a result of the match, the boundary assignment of the stored representation feeds back via the current shape representation to determine the boundary assignment of the stimulus.

We present this example only to illustrate how top-down effects could, in principle, be incorporated into the model. Other, perhaps more controversial, ideas about top-down processing are considered later in this chapter.

Having completed our overview of the model, we now look more closely at the constituent processes and representations, beginning with the extraction of edges and junctions.

Edge and Junction Detection

For our purposes, stimulus encoding in the earliest cortical visual areas — V1 and V2 — represents the starting point for the computations leading to edges, contours, surfaces, and objects. Individual cells in these areas respond to luminance contrast in particular areas of the visual field, with selectivity for specific orientations and spatial frequencies (e.g., 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.

The framework in Figure 1 assumes that these early cortical responses form the inputs into processes that detect meaningful edges and contour junctions in the optical

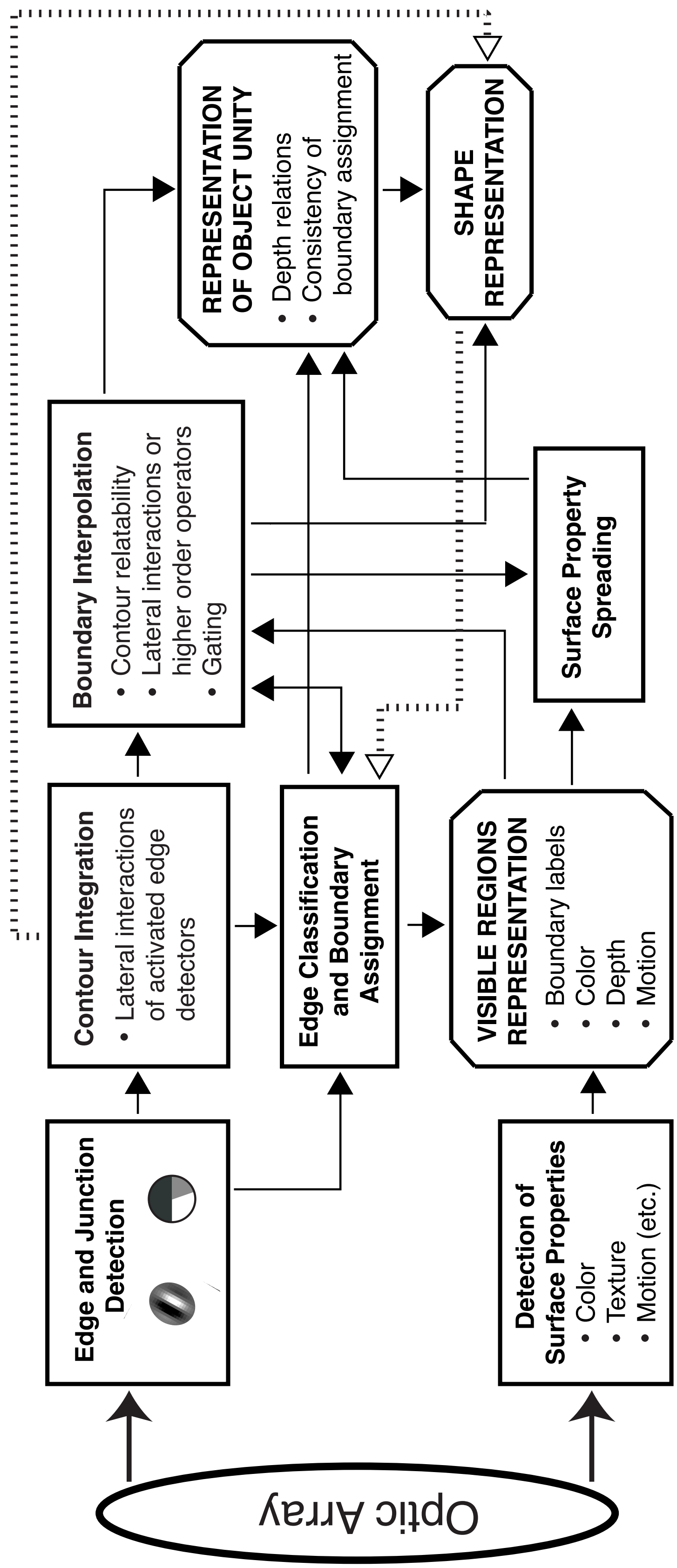


Figure 2. Illustration of a possible top-down effect of contour shape on boundary assignment. See text for details.

projection (Heitger et al., 1992; Marr & Hildreth, 1980; Morrone & Burr, 1988; but see Watt, 1994, for an alternative approach). Although it appears as a simple box in the model, there actually are several complexities even at this stage.

Multiple Edge Inputs. Object perception utilizes several types of edge inputs, including luminance and chromatic changes, but also discontinuities in texture, stereoscopic depth, and motion (Gibson, Kaplan, Reynolds, & Wheeler, 1969; Julesz, 1971; Shipley & Kellman, 1994). To complicate matters further, both object motions and image displacements given by observer motion contribute to edge processing.

Though luminance discontinuities receive the most attention in discussions of edge detection, some of these other edge inputs actually may be more important. The usefulness of luminance discontinuities for edge detection rests on certain ecological facts — facts about the physical world and the information it makes available for perception. Luminance and chromatic edges provide meaningful information because separate objects and surfaces in the world tend to be made of different materials that interact with light differently. Thus, significant boundaries in the world often correspond to locations at which the amount or spectral composition of light changes abruptly.

The same logic applies to edges detected from discontinuities in texture, depth, and motion. Adjacent surfaces in the optic array often project from different objects, but these will, in general, be at different depths. Thus, a stereoscopic depth map will have discontinuities at surface boundaries. The same is true of the velocity field, in which optical change is registered at each visible feature: velocity discontinuities will appear at boundaries during object or observer motion.

Whereas luminance, color, and texture edges may correspond to markings on a surface, motion and depth discontinuities rarely arise in the absence of a boundary between two objects. The high ecological validity of motion and depth edges suggests that they play a primary role in the detection of meaningful edges in the world. More research is needed to determine how the visual system integrates these various sources of edge information.

Junctions. Contour junctions can be 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 sharp corners of objects, as well as the points where contours of separate objects intersect.

Contour junctions of all types play an important role in the segmentation and grouping of objects. Kellman and Shipley (1991) observed that the contours interpolated between two visible regions invariably begin and end at junctions, which they labeled “tangent discontinuities” (TDs). In fact, Kellman and Shipley presented a proof that *all* instances of occlusion produce TDs in the optical projection. This ecological invariant may be the reason that junctions figure prominently in initiating contour interpolation processes: TDs are a

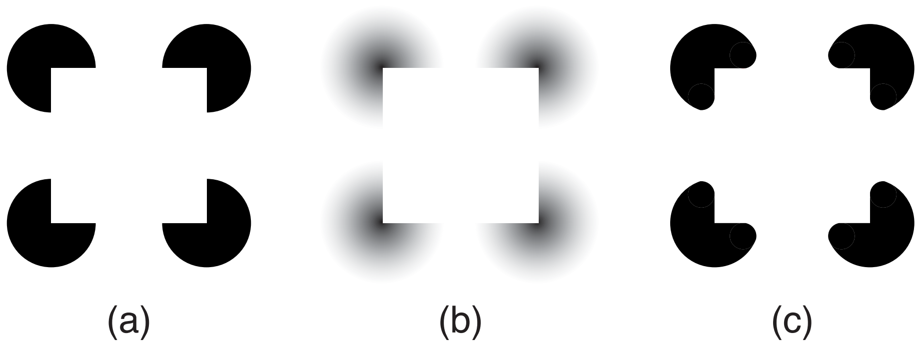


Figure 3. Three variations of the Kanizsa square. (a) Illusory contours clearly are visible in the classical display, which contains uniformly colored inducers and sharp tangent discontinuities. The illusory contours remain visible in (b), despite the elimination of luminance discontinuities from the inducers. The illusory contours become weak or absent with the rounding of tangent discontinuities (c). Figure B redrawn from *Vision Research*, *33*, Lesher, G. W., & Mingolla, E., The role of edges and line-ends in illusory contour formation, pp. 2253-2270, Copyright 1993, with permission from Elsevier Science. Figure C redrawn from *Perception*, *27*, Tse, P. U., & Albert, M. K., Amodal completion in the absence of image tangent discontinuities, pp. 455-464 (Fig 8, p. 460), Copyright 1988, with the permission of Pion Ltd., London.

potentially rich source of information about the loci of occlusion. In illusory contour displays, for which tangent discontinuities *can* be eliminated, the rounding of TDs eliminates or vastly reduces the perception of interpolated contours (Shipley & Kellman, 1990).

Some researchers have questioned the generalization that interpolated contours must begin and end at TDs. Lesher and Mingolla (1993) presented a Kanizsa-style illusory contour display in which the inducers changed gradually in luminance, thus blending into the background (Figure 3b). Phenomenologically, this display supports the formation of illusory contours, even though, formally speaking, the inducers contain no bounding edges and thus cannot have any discontinuities in edge direction.

Although this display raises interesting issues, it does not bear directly on the role of TDs. Close inspection suggests that removing the luminance discontinuities that normally define edges, as in Lesher and Mingolla’s display, does not entail the removal of tangent discontinuities — sudden changes in the slope of an edge. Despite the luminance gradient, the edges remain reasonably clear and well-localized. In all likelihood, contrast-sensitive neurons selective for low spatial frequencies respond to the gradients much as they do to “real” edges, giving rise to perceived contours. The relevant tangent discontinuities are defined by the perceived edges.

Other researchers have pointed out that weak illusory contours sometimes arise from inducers with slightly rounded corners (Figure 3c; Hoffman, 1998; Shipley & Kellman, 1990; Tse & Albert, 1998). Neurons selective for low spatial frequencies may, once again, be responsible for this phenomenon. Low spatial-frequency operators cannot discriminate between regions of high curvature and sharp junctions. In essence, rounded corners *are* tangent discontinuities to neurons that encode information at larger spatial scales. The conflict between these coarse-coding neurons, which register TDs, and cells that respond to higher spatial frequencies, which detect the rounded corners, likely explains why the illusory contours in these displays appear quite weak. When high spatial frequencies — and this conflict — are eliminated by squinting or by increasing viewing distance, the perceived strength of the illusory contours increases substantially.

Available evidence supports the crucial role of contour junctions in boundary interpolation (e.g., Shipley & Kellman, 1990), which in turn is crucial for unit formation and shape perception, as we consider shortly. To support these and other perceptual tasks, both edge and junction information may be extracted early in cortical processing and at multiple spatial scales. Although the mechanisms of junction detection have, as yet, received little attention, a promising approach may be found in the model of Heitger et al. (1992). This model proposes a specific operator that extracts "key points," which includes both contour ends and junctions. We examine this model in more detail below.

Contour Integration

The problem of contour integration involves representing a visible contour as continuous. Meaningful contours extend well beyond the receptive fields of oriented neurons in early cortical visual areas. Therefore, representation of a contour as a connected unit requires the integration of information from neurons with receptive fields tuned to different regions of space. Several recent studies address possible mechanisms for linking the separate responses of orientation-sensitive units (e.g., Moulden, 1994; Pettet, McKee, & Grzywacz, 1998; Yen & Finkel, 1998).

These investigations of contour integration ultimately may help to answer another important question: whether a similar mechanism governs *boundary interpolation* — the connection of edges across gaps in the input. Field, Hayes, and Hess (1993) investigated the spatial relations necessary for contour integration in a series of elegant empirical studies. The relations they uncovered mirror the formal requirements of Kellman and Shipley's (1991) model of contour interpolation across gaps. This finding suggests a close relationship between contour integration and contour interpolation, an idea that will be further explored later in this chapter.

Other important issues about contour integration involve the nature of the representation and underlying neural mechanisms. Explicit contour representations presumably are important for assigning contours to objects and for encoding shape. Whereas earlier stages of visual processing may be viewed as more or less direct responses to energy variables in the input, contour integration entails a symbolic representation (i.e., an explicit description about a specific aspect of an object or surface). Reversible figure-ground displays are among the many visual phenomena implying that contours are, indeed, explicitly represented as unitary entities. In these displays, a contour separating two regions switches its boundary assignment (which region it "belongs to") as a unit; we do not experience switching of parts of contours. Thus, it appears that by this stage of visual processing, contours, and not smaller fragments, are the units to which boundary assignment applies.

Edge Classification and Boundary Assignment

To perceive objects, it is not sufficient merely to detect edges. Some edges in the visual array delimit the boundaries of objects, whereas others represent shadows or textural markings on a surface. The process of distinguishing these possibilities is termed *edge classification*. Edge classification, most importantly, results in the identification of *occluding edges*: locations where an object or surface comes to an end. As the name suggests, these edges also mark places at which one surface continues behind another.

A critical piece of information for edge classification may be the convergence of different kinds of evidence. Shadows and textural markings typically produce only one or two types of discontinuity in the visual input (luminance and color for textural markings; luminance and perhaps motion for shadows). By contrast, occluding edges often involve discontinuities in multiple perceptual properties, including luminance, color, texture, motion, and stereoscopic depth. Therefore, the visual system may, in part, define occluding edges by the convergence of several discontinuities in the same spatial location.

Closely related to edge classification is *boundary assignment*, or figure-ground perception. Many visible edges are occluding edges, which, as the Gestalt psychologist Kurt Koffka put it, have a "one-sided function." A contour appearing in the optical projection represents the boundary of the surface on one side, whereas the surface on the other side continues behind. Each such edge poses a problem to the visual system insofar as it defines the shape only of the surface in front; no local sensory information is received about the location, shape, or other characteristics of the surface continuing behind. Interpolation processes, described below, pursue the task of recovering object unity and shape despite occlusion.

Several kinds of information contribute to boundary assignment. In ordinary perception, stereoscopic and kinematic cues to depth (including motion parallax and accretion-deletion of texture) probably play the most important role. Occluding edges signaled by kinematic and stereoscopic discontinuities carry with them depth-order information that reveals, unambiguously, the appropriate boundary assignment: the nearer surface owns the boundary.

Another source of information comes from contour junctions. In particular, "T" junctions, which underlie the depth cue of interposition, may be helpful in determining boundary assignment. As shown in Figure 4, contour junctions can be classified into different categories (e.g., Barrow & Tenenbaum, 1986), depending on their configurations. The defining feature of a T-junction is a smooth, continuous contour on which another contour

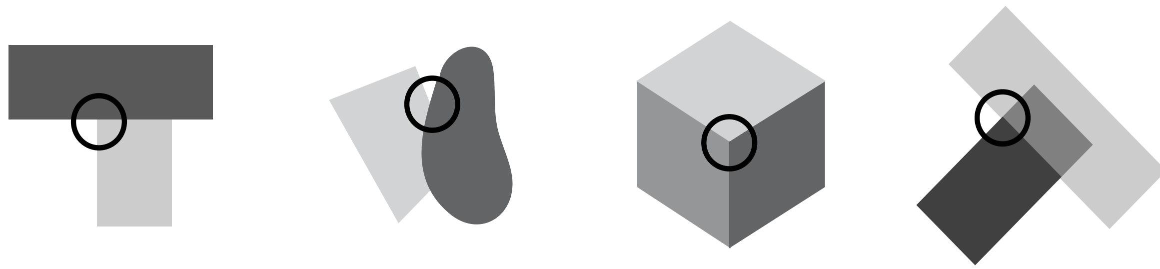


Figure 4. Types of contour junctions. In a classic T-junction (a), the “roof” is seen as passing in front of the “stem.” The junction in (b) also is a “T,” as a terminating contour meets a continuing contour; the relative orientation of the two contours is irrelevant. Other types of junctions assist with image segmentation, but are not immediately relevant to boundary assignment. The Y-junction (c) depicts an object corner, whereas the X-junction (d) indicates transparency.

terminates (Figure 4a); the absolute and relative orientations of the two contours are incidental (Figure 4b). Once the visual system has encoded a T-junction, the depth relation of the two constituent contours follows: the one forming the “roof” of the T appears to be in front of the one forming the “stem.” Therefore, the surface that is uninterrupted by the “stem” owns the contour in question.

In his classic discussion of figure-ground organization, Rubin (1915) emphasized a third class of boundary assignment cues: the relations among visible areas. Rubin noted, for example, that an enclosed area tends to be seen as figure, whereas the enclosing area is perceived as ground. Other factors discussed by Rubin involve orientation, convexity, and symmetry. These relational cues to boundary assignment may be considered relatively weak, as they will readily be overridden by depth information given by stereopsis, kinematic cues, or T-junctions. Finally, as mentioned earlier, the familiarity of a contour shape may influence boundary assignment (Peterson & Gibson, 1991; Rubin, 1915).

The Visible Regions Representation

Apart from the final outputs of unity and form, the process model presented in Figure 1 contains one intermediate representation: *the visible regions representation*. This representation makes explicit the unity of a continuous visible area; that is, it encodes the visible points in certain regions as belonging to a single, uninterrupted surface.

The visible regions representation captures several important properties of earlier proposals. Like the intended results of image segmentation algorithms (e.g., Wang & Terman, 1997), it partitions the optic projection into distinct, non-overlapping regions. Like Marr’s (1982) 2.5-dimensional sketch, it assigns observer-relative depth to these regions. The visible regions representation also resembles the *uniform connectedness* idea of Palmer and Rock (1994): the visual system encodes closed regions with homogeneous surface properties as a single unit. However, Palmer and Rock treated common surface lightness, color, and texture as the primary determinants of uniform connectedness; motion and depth were attributed secondary roles. By contrast, our visible regions representation assumes that depth relations, given by stereoscopic and motion parallax cues, take precedence over the commonality of lightness and color. A surface that is continuous in depth but contains various discontinuities in surface coloration would be encoded as a single unit in the visible regions representation. Conversely, the visible regions representation would tag as separate two adjacent, homogeneously textured regions with an abrupt change of binocular disparity between them.

A number of facts support the idea that human perception incorporates an intermediate representation of visible regions. For example, although complete object representations are one result of perceptual processes, we have little trouble seeing which areas of objects are partly occluded. Moreover, artists can paint or draw the visible regions of objects. Although this ability may take some practice, it might not be possible at all without some explicit representation of visible regions.

The visible regions representation probably should *not* be understood as a set of frontoparallel image fragments (unlike the results of region segmentation processes in machine vision). One of the major problems facing art students is foreshortening — drawing the correct projected size of an object slanted in depth, relative to other objects. For example, if one's hand is rotated away from the vertical (around a horizontal axis), it projects a very small vertical extent to the eye. Novice art students tend to draw the hand's projection as much too tall. This error suggests that the visible regions representation is not a canvas-like or image-like sheet, but a representation that includes depth. The art student can see which parts of the hand are visible, but these are seen at true size, oriented in depth; depicting them as frontoparallel fragments on a canvas presents difficulties.

Boundary Interpolation

As a consequence of occlusion, many objects in ordinary environments project to the eyes as multiple, spatially distinct fragments. As a result, a single object may be represented, at an intermediate stage, as several visible regions. Interpolation processes allow the visual system to assign these spatially distinct visible regions to a unitary object. Ultimately, these processes lead to more complete object representations, crucial for extracting meaningful information about an object's shape.

By definition, interpolation processes connect parts across gaps in the input. There appear to be two types: *boundary interpolation*, which we consider in this section, and *surface interpolation*, which we consider in the next.

<u>Geometric Model of Boundary Interpolation: Kellman and Shipley (1991)</u>. In the Kellman and Shipley model, tangent discontinuities (TDs) mark the possible starting points for

boundary interpolation. However, not all TDs lead to boundary interpolation. Some types of contour junctions indicate that the boundary has come to an end and should not be continued (e.g., a Y-junction; see Figure 4c). In other cases, a contour may be seen as passing behind an occluder, but does not link up perceptually with any other visible contour. Evidence suggests that these boundaries are, nonetheless, represented as continuing behind the junction point, perhaps for some fraction of the visible edge's length (He & Nakayama, 1994; Kanizsa, 1979; Yin, Kellman, & Shipley, 1997). We call this phenomenon amodal *continuation* to distinguish it from interpolation or completion. Yin et al.'s data suggest that amodally continued contours follow the direction of the visible edge's tangent at its point of occlusion.

Boundary *interpolation* occurs when a contour that disappears behind a surface, thus creating a TD, connects to a spatially separated contour on the other side of an occluding object. This interpolation process proceeds only when the visible contours, and their TDs, satisfy certain geometric relationships. These relationships, first suggested by the Gestalt idea of good continuation, have been formalized in the construct of contour *relatability* (Kellman & Shipley, 1991, 1992).

The notion of relatability expresses the conditions necessary for interpolation between two visible edges. Intuitively, 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 constraint that the boundaries of objects tend to be smooth.

Mathematically, relatability can be defined with reference to the construction shown in Figure 5. In this diagram, E_1 and E_2 represent the edges of surfaces. R and r depict the perpendiculars to these edges at the point where they lead into a tangent discontinuity; R is defined as the longer of the two perpendiculars. The angle of intersection between R and r is termed φ. 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).

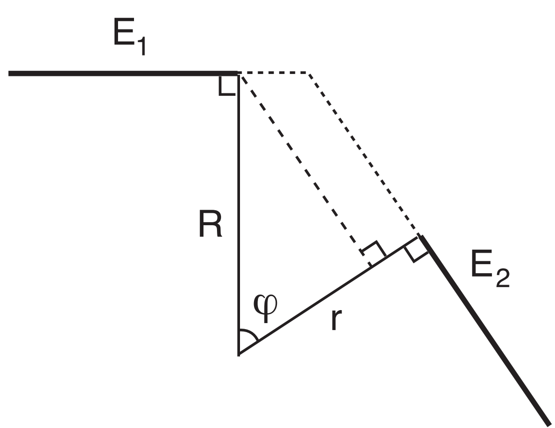


Figure 5. Geometric relationship for defining whether two edges (E_1, E_2) are relatable. See text for details. After Kellman and Shipley (1991).

More formally, E_1 and E_2 are relatable if and only if:

$$0 \leq R\cos\varphi < r.$$

This equation can be unpacked in two steps. The left-hand side of the inequality expresses the limitation that the curve constructed to connect E_1 and E_2 cannot bend through more than 90˚; if φ is greater than 90˚, then cos φ is negative. The right-hand side of the inequality states that the projection of R onto r (i.e., R cos φ) must fall within the extent of r. If this inequality is violated (i.e., $R \cos \varphi \geq r$), then any connection between E_1 and E_2 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 zero at a relative angle of 90˚. Singh and Hoffman (1999) proposed a specific measure for this graded decrease.

The basic notion of relatability may be extended in several ways. First, the relatability criterion, as originally formulated, considers only the tangents at the points of discontinuity. It is possible that the boundary interpolation mechanism also utilizes the curvature of the visible contours (e.g., Guttman & Sekuler, 2001; Takeichi, Nakazawa, Murakami, & Shimojo, 1995). This is not entirely clear, however, as the issue of curvature may be confounded with the issue of how much of a contour is used to determine its slope at the point of tangent discontinuity. Second, it now appears that straightforward extensions of the static, 2-D relatability construct govern both 3-D contour interpolation (Kellman, Yin, Shipley, Machado, & Li, 2001) and dynamic visual interpolation, when visible contours appear sequentially in time (Palmer, Kellman, & Shipley, 1997). We discuss these developments briefly at the end of the chapter.

Relatability and Good Continuation. The notion of contour relatability descends from the Gestalt idea of good continuation. Specifically, it formalizes the original Gestalt principle in some respects and extends it in others.

Max Wertheimer, in his classic (1923/1958) paper *Untersuchungen zur Lehre von der Gestalt* ("Laws of organization in perceptual forms"), presented a number of figures illustrating the principle of good continuation. The displays, which involved the segmentation of line drawings with fully visible contours, were accompanied by this advice:

> On the whole, the reader should find no difficulty in seeing what is meant here. In designing a pattern, for example, one has a feeling how successive parts should follow one another; one knows what a "good" continuation is, how "inner coherence" is to be achieved, etc.; one recognizes a "good Gestalt" simply by its own "inner necessity."

Though the demonstrations were compelling, neither Wertheimer nor his successors offered a real definition of "good continuation." Michotte, Thines, and Crabbe (1964) extended the idea of good continuation to the problem of partly occluded objects; again, the notion remained intuitive, illustrated by compelling displays, rather than formally characterized.

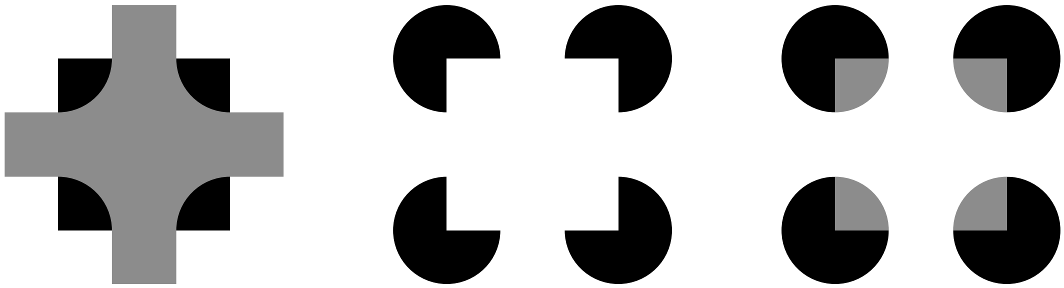
So just what is the "good" in good continuation? An obvious candidate involves mathematical notions of smoothness. But which notion of smoothness? A number of possibilities exist (e.g., Prenter, 1989). Kellman and Shipley's (1991) model defines smoothness in reference to the first derivatives, or slopes, of contours. Formally, a contour is smooth if there are no discontinuities in its first derivative; both sharp corners and contour intersections violate this description of smoothness. As Kellman and Shipley pointed out, this definition has the benefit of producing a complementary relationship between first-order contour continuity as the basis for interpolation and first-order discontinuities (i.e., TDs) as the basis for image segmentation. As previously discussed, the latter indicate possible loci of occlusion and mark the beginning and end points of interpolated edges.

In addition to specifying the relevant smoothness notion, relatability imposes additional constraints not embodied in earlier notions of good continuation. One such constraint is monotonicity: interpolated contours bend in one direction only. A general notion of smoothness permits double inflections, whereas relatability excludes them (apart from a small threshold tolerance for misalignment of relatable edges; see Shipley & Kellman, 1992a). This limitation has not been universally accepted; a number of investigators suggest that doubly-inflected interpolations may occur in object completion (Liu, Jacobs, & Basri, 1999; Takeichi et al., 1995). Another constraint expressed in the notion of relatability is the limitation that interpolated contours cannot bend through more than 90˚, an idea that has received substantial empirical support (Field et al., 1993; Polat & Sagi, 1993, 1994).

In sum, by formalizing and adding constraints to the original notion of good continuation, relatability produces clear predictions that can be empirically assessed. Although some aspects of the model may require elaboration, available evidence, as we will see shortly, supports relatability as a formal account of the basic geometry of human boundary interpolation.

<u>The Identity Hypothesis in Contour Interpolation</u>. The geometry and processes of contour interpolation in the Kellman and Shipley (1991) model apply equally to several interpolation phenomena that, traditionally, have been considered separately. In particular, the *identity hypothesis* suggests that the same contour interpolation process connects contours under occlusion (amodal completion), creates illusory contours (modal completion), and plays a role in other contour-connection phenomena including certain transparency displays.

Figure 6 depicts partly-occluded, illusory, and transparent shapes having equivalent physically-specified contours and gaps. Phenomenally, the shape of the interpolated contours is the same in the three images. More importantly, however, we believe that the same interpolation process gives rise to the interpolated contours in all cases. The perceptual differences among these phenomena reside only in the depth ordering of the interpolated contours and other surfaces; the processing of relative depth may *interact* with the boundary interpolation process, but relies on a different mechanism.

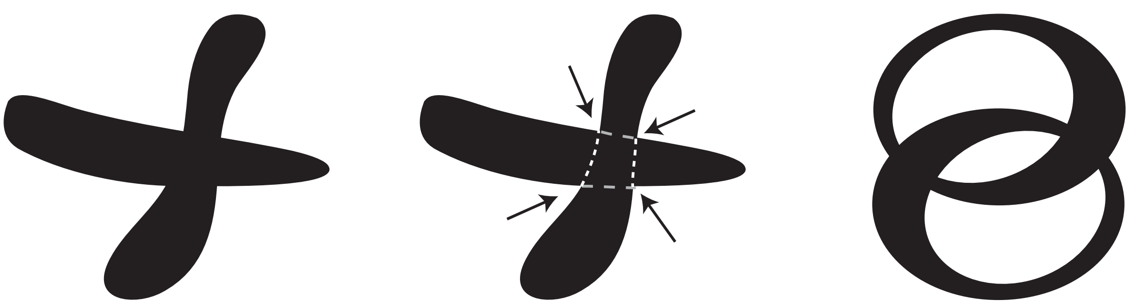


(a) (b) (c)

Figure 6. An illustration of the identity hypothesis: (a) partly occluded square; (b) illusory square; (c) transparent square. 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, the process that interpolates edges across gaps also is the same in these cases.

A large body of evidence supports the idea that a common boundary interpolation process serves modal and amodal completion (Kellman, Yin, & Shipley, 1998; Ringach & Shapley, 1996; Shipley & Kellman, 1992a). For example, Kellman et al. found that partly occluded contours can join illusory contours to produce a shape with clearly defined boundaries. (An example appears below in Figure 15.) This merging suggests that illusory and occluded contours arise from a common mechanism.

Several other theoretical arguments support the identity hypothesis. A compelling one relates to some interesting phenomena that we have termed *self-splitting objects*; examples appear in Figure 7. Remarkably, in these cases, homogenous areas split into two perceived objects. The Kellman and Shipley (1991) model explains this effect as follows. The display in Figure 7a contains four TDs (marked with arrows in Figure 7b), each of which can initiate contour interpolation processes. Four pairs of edges lead into the TDs; as each pair satisfies the relatability criteria, contour interpolation connects all four edges (indicated in Figure 7b



(a) (b) (c)

Figure 7. Self-splitting objects. (a) This shape typically is seen as two elongated black forms, even though the entire figure possesses identical surface properties. (b) Arrows indicate tangent discontinuities and dotted lines mark the resulting interpolated contours of the same self-splitting object. (c) Despite uniform surface qualities, most observers describe this image as two interlocking rings.

by dotted lines). Closed contours, comprised of physically-specified and interpolated contour segments, define each of the two perceived objects.

For our present purpose, the relevant question is: How do we interpret these objects' interpolated boundaries? That is, do they appear as illusory contours or as partly occluded contours? At any time, one of the two objects in Figure 7a appears to cross in front of the other. The boundaries of the object in front appear as illusory contours, whereas those of the object in back appear as occluded contours. However, the depth ordering of the two objects is unstable over time; which object appears in front may fluctuate. When one object switches from being in front to being in back (or vice versa), its contours switch from being illusory to being occluded (or vice versa).

Self-splitting objects do not always possess this instability of depth order. The display in Figure 7c appears to be more stable. Most observers describe the figure as containing two interlocking rings. However, the "two rings" possess identical surface qualities; therefore, the boundaries that separate the objects must be attributed to the contour interpolation processes described above.

Implicit in the idea of "interlocking," the perceived depth ordering of the two rings varies across the image; the ring at the top appears to pass in front of its counterpart on the right side of the display, but beneath it on the left. Petter (1956) studied displays of this sort and 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 rings appear to lay on top of the thinner parts, as the former have smaller gaps in the physically-specified contour.[1]

The relevance of Petter's effect to the identity hypothesis becomes apparent when one considers the nature of the rings' perceived boundaries. As in the case we considered above, we perceive illusory contours where the rings pass in front, but partly occluded contours where they connect behind. However, according to Petter's rule, the perception of each interpolated contour as in front or behind — and, in turn, as "illusory" or "occluded" — 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. As stated above, comparing the lengths of the 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 also precedes 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

---

[1] Looking back at Figure 7a, we can now see that Petter's rule can explain the instability of perceived depth order in Figure 7a: the two overlapping objects have interpolated contours of very similar length.

cannot be separate mechanisms for the interpolation of contours in front of versus behind other surfaces; illusory (i.e., in front) and occluded (i.e., behind) contours arise from the same contour interpolation process.

In sum, both empirical studies and logical arguments indicate that contour interpolation relies on a common mechanism that operates without regard to the final determination of illusory or occluded appearance. The subjective appearance of interpolated contours depends on mechanisms responsible for assigning relative depth, which lie outside and sometimes operate subsequent to the interpolation process itself. In the discussion that follows, we use studies of occlusion, illusory contours, and other contour-connection phenomena interchangeably in examining the nature of the boundary interpolation process.

<u>Empirical Studies of Boundary Interpolation</u>. A variety of experimental studies support relatability as a formal description of the boundaries interpolated by the visual system (Kellman & Shipley, 1991; Shipley & Kellman, 1992a). 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, which are 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, twelve 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 (0˚ for a straight, collinear path; ±15˚, for example, 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 support the notion of relatability. 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.

In addition to degree of curvature, the strength of boundary interpolation depends on the relative extents of the physically-specified edges and gaps in a scene. Interpolation strength appears to be a linear function of the *support ratio*, the proportion of total edge length that is physically-specified, as opposed to interpolated. This relationship holds over a wide range of display sizes (Lesher & Mingolla, 1993; Ringach & Shapley, 1996; Shipley & Kellman, 1992b). In essence, the support ratio idea makes precise a version of the Gestalt law of *proximity*: nearer elements are more likely to be grouped together.

Recent work by Geisler, Perry, Super, and Gallogly (2001) further suggests that relatability may capture 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 co-occurring correlate highly with the geometry of relatability. In other words, two visible edge segments associated with the same contour meet the mathematical relatability criterion far more often than not. In sum, the success of relatability in describing perceptual interpolation processes may derive from ecological regularities that characterize the natural environment.

Surface Interpolation

The boundary interpolation process appears to operate without sensitivity to the similarity of surface properties (Kellman & Loukides, 1987; Kellman & Shipley, 1991; Shapley & Gordon, 1987; Shipley & Kellman, 1992a). Using a large sample of randomly generated figures, Shipley and Kellman found evidence that object completion under occlusion proceeds similarly whether relatable pieces are of the same or different luminance and color. These data indicate that contour interpolation does not depend on the Gestalt principle of *similarity*; relatability governs edge completions regardless of whether the connecting regions are similar in luminance, spectral characteristics, or texture.[2]

This characteristic of boundary interpolation does *not* imply that surface similarity cannot influence object completion. Kellman and Shipley (1991) described a surface spreading process that complements boundary interpolation in the completion of partly occluded objects. The process may be related to phenomena described some time ago in retinal stabilization experiments (Yarbus, 1967; cf., Grossberg & Mingolla, 1985). Yarbus presented displays containing a circle of one color surrounded by an annulus of a different color. The boundary between the inner circle and outer ring was stabilized on the retina, which caused it to disappear after several seconds. Following the disappearance of this boundary, the color of the annulus (whose outer boundary was not stabilized) spread throughout the entire circle.

[2] A few phenomena in the literature, especially an intriguing demonstration by He and Ooi (1998), suggest that this conclusion may need to be qualified. Although the contour interpolation process tolerates wide variation in the surface characteristics of the regions being connected, there may be some constraints involving contrast polarity that we do not, as yet, fully understand.

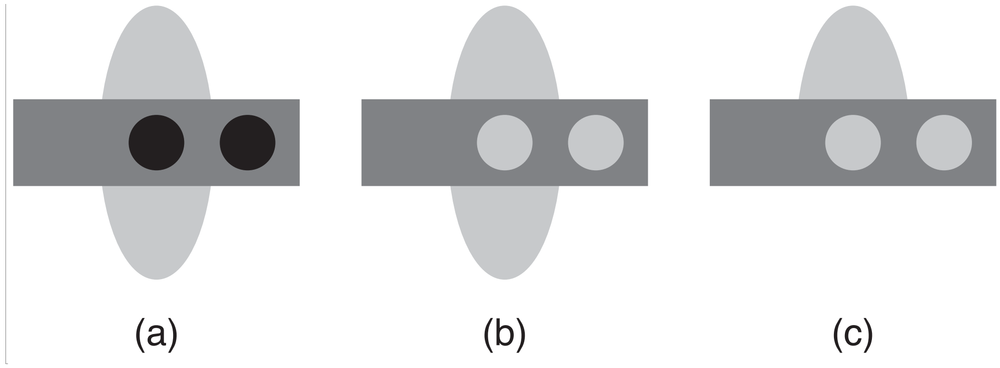


Figure 8. Some effects of the surface interpolation process: (a) no surface spreading; (b) surface spreading within interpolated boundaries; (c) surface spreading within amodally continued boundaries. See text for details.

A similar surface spreading process may operate under occlusion (Kellman & Shipley, 1991). Surface qualities spread behind occluding edges; however, interpolated boundaries confine the spread in much the same way as physically-specified boundaries. Figure 8 illustrates some effects of this surface spreading process.

In Figure 8a, both circles appear as spots on a background because their surface qualities differ from other regions in the image. In Figure 8b, the right-hand circle still looks much like a spot, but the left-hand circle is seen as a hole in the occluding surface. This percept depends on the similarity between the surface lightness of the circle and the partly occluded ellipse. Because the circle lacks TDs, its link to the ellipse cannot be attributed to the boundary interpolation process. Surface spreading alone governs the perceived connection. The fact that the right-hand circle retains its appearance as a spot indicates that the interpolated boundaries of the partly occluded figure confine the surface spreading process. Figure 8c illustrates that, in the absence of relatable edges, surface completion can still occur; the left-hand circle appears as a hole, with its visible area linked to the visible half-ellipse. For such cases, Kellman and Shipley (1991) proposed that the surface spreading process is confined within an area partially bounded by the tangent extensions of the partly occluded object (i.e., lines tangent to the visible contours at the points of occlusion). This idea draws on the notion that, even in the absence of connection to other visible edges, contours continue amodally for some distance behind occluders (He & Nakayama, 1994; Kanizsa, 1979). The right-hand circle of Figure 8c appears as a spot, rather than a hole, because it falls outside the tangent extensions of the half-ellipse.

These observations and hypotheses have been confirmed in a series of experiments. Using displays resembling those in Figure 8, Yin, Kellman, and Shipley (1997) tested whether surface qualities spread within relatable edges and also within extended tangents of non-relatable edges that continue behind occluders. For a number of displays with varying edge and surface similarity relations, observers made a forced-choice judgment of whether a circular area appeared to be a hole in a surface or a spot on top of the surface. If the data suggested a "hole" rather than a "spot" percept, then Yin et al. (1997) assumed that the surface properties of the partly occluded shape spread to the location of the test circle. The results indicated that observers tend to perceive the circle as a hole if and only if its surface properties match those of the partly-occluded shape, and it resides either within relatable edges (left-hand circle in Figure 8b) or within the tangent extensions of non-relatable edges (left-hand circle in Figure 8c); a "spot" percept resulted if the circle's surface properties differed from that of the partly-occluded shape (Figure 8a) and/or if it fell outside of the relatable edges (or non-relatable tangent extensions; right-hand circles in Figures 8b and 8c, respectively). In a subsequent pair of experiments, Yin, Kellman, and Shipley (2000) studied surface completion using an objective performance paradigm that pitted the effects of surface completion against small amounts of binocular disparity in making a circle look like a hole versus a spot. Results indicated that surface spreading inside of relatable edges reduced sensitivity to contradictory disparity information; judgments of depth were not affected by surface similarity in displays lacking relatable edges. Together, these findings strongly suggest that surface spreading alone can lead to perceived connections under occlusion. The surface interpolation process operates within relatable edges and also within the tangent extensions of contours without relatable edges.

The characteristics of the surface spreading process may help to clarify some apparent confusions regarding object completion phenomena. Tse (1999a, 1999b) created a number of pictorial displays in which distinct regions appear to be connected despite a lack of relatable edges. Tse argued that these displays disconfirmed contour relatability as an explanation of interpolation and that a new notion of "volume completion" may be required. However, this argument neglects the surface completion process, which operates separately from and complements contour relatability (Kellman & Shipley, 1991). The rules of the surface completion process, which previously have been experimentally supported (Yin, Kellman, & Shipley, 1997, 2000), predict virtually all of the results reported by Tse (e.g., Experiments 1-5 in Tse, 1999a). Although Tse's data do not adequately separate a volume completion hypothesis from the known effects of the surface spreading process, his question of whether volumes per se (or 3-D contour and surface relations) play important roles in object formation is an important one. We consider some experimental evidence on three-dimensional boundary and surface interpolation (Kellman, Machado, Shipley, & Li, 1996; Kellman, Yin, Shipley, Machado, & Li, 2001) in a later section.

In sum, surface spreading and edge relatability appear to play complementary roles in object perception. Both the edge and the surface interpolation processes can, themselves, specify connections under occlusion. Whereas contour interpolation establishes the shape of an object's boundaries, surface qualities (lightness, color, and texture) spread within real and interpolated boundaries, and along their tangent extensions, to further specify other aspects of the object's appearance.

The Units Representation

One output of object perception processes may be a representation describing which visible regions belong to the same object or surface. This output, the *units representation*, is depicted in Figure 1. The motivation for a distinct units representation, independent of a description of shape, comes from situations in which an object's shape is wholly or partly indeterminate. In Figure 9a, several fragmented regions of a surface are visible, but all of their visible edges belong to occluding regions. The surface appears to have continuity, yet it is shapeless. In the natural environment, the sky seen through trees mirrors this example. Figure 9b shows a case in which there is some visibility of object contours, but poor specification of the overall shape. Nevertheless, the surface interpolation process reveals the unity of the various dark areas.

When the visual system receives adequate shape information, unity and shape may be encoded together. However, we suggest that shape descriptions ordinarily presuppose a representation of unity.

The Shape Representation

Description of object shape is one of the most important results of object perception processes. The exact nature of shape descriptions in the brain remains an unsolved problem. It

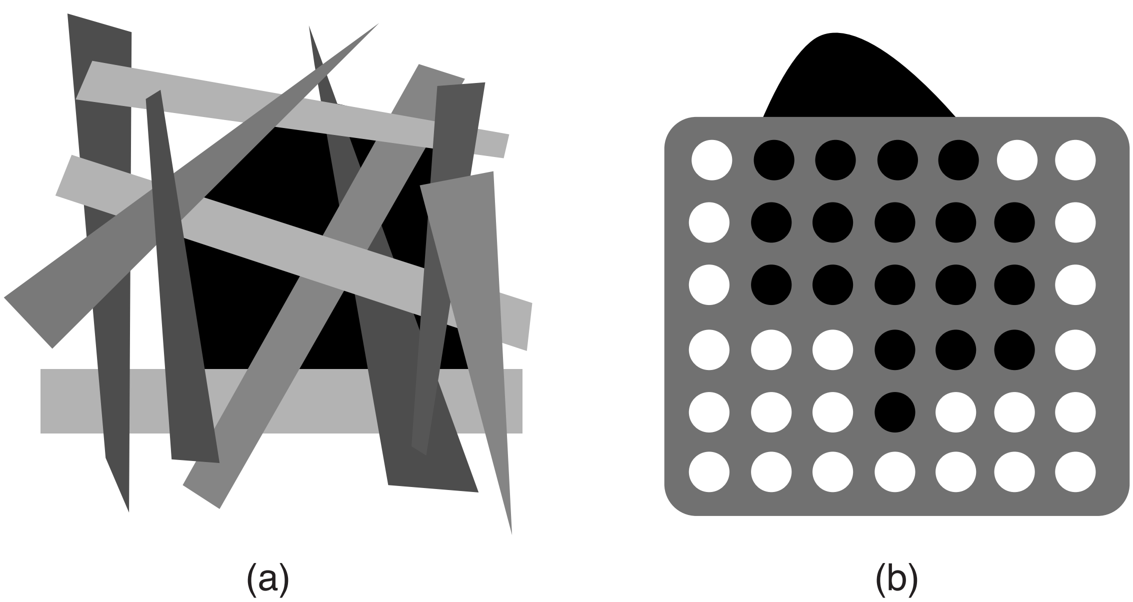


Figure 9. In (a), we perceive a unitary black surface despite a complete absence of shape information. In (b), surface spreading gives the unity of the partly occluded surface.

appears that shape descriptions are true 3-D representations, or at least incorporate viewer-relative variations in depth, as in Marr's 2.5-D sketch (Liu, Kersten, & Knill, 1999; Marr, 1982).

A stronger commitment can be made, we believe, regarding *what* receives a shape description: In ordinary perception, the visual system assigns shape descriptions to unitary objects. A connected surface whose overall boundaries are indeterminate may also receive a limited shape description, as may the natural parts of objects. However, the visual system does *not* naturally assign shape descriptions to regions that encompass parts of separate objects. Moreover, visible regions, as opposed to natural parts of objects, do not automatically receive shape descriptions, although they can be recovered, with effort, from the visible regions representation. These claims — that shape descriptions are automatically given to objects but not to arbitrary arrays — have interesting and testable empirical consequences regarding object recognition and perceptual classification.

Having considered the various processes and representations involved in object perception, we turn now to neural models of three specific processes: contour and junction detection, contour integration, and contour interpolation.

NEURAL MODELS

Whereas geometric and process models aim to capture the stimulus relationships and information processing operations that determine contour and object perception, neural models emphasize how these computations can be carried out by neural circuitry. To date, no comprehensive neural model exists to perform the multitude of tasks involved with object perception, diagrammed in Figure 1 and discussed in relation to geometric and process models. Existing neural models differ in which subset of object perception tasks they address.

In this section, we will describe three neural-style, computational models that perform specific information processing tasks necessary for object perception. The first, by Heitger, Rosenthaler, von der Heydt, Peterhans, and Kübler (1992), extracts edge segments, lines, and junctions from a two-dimensional scene. This model builds upon the one-dimensional feature detection model of Morrone and Burr (1988), extending the edge- and line-detection algorithms into the second dimension, and adding operators that detect discontinuities (junctions) explicitly.

The other two models describe the integration of elements across space to form meaningful edges and contours. The model of Yen and Finkel (1998) implements a contour integration algorithm, by which spatially distinct contrast elements (Gabor patches) are bound together to form an extended contour. The model of Heitger, von der Heydt, Peterhans, Rosenthaler, and Kübler (1998) performs boundary interpolation, constructing a contour across gaps in the physical stimulus. Although these models differ substantially in their structure and architecture, a deeper examination shows that they share many common characteristics, albeit differently implemented.

All of these models employ the functional architecture of the earliest visual cortical areas; their building blocks resemble known neural units. As its starting point, each model postulates a set of linear filters with responses similar to those of the simple cells of primary visual cortex (V1). These units each respond to local stimulus contrast of a particular orientation and spatial frequency. The filters come in pairs, one member of which has an even-symmetric receptive field, and the other an odd-symmetric receptive field. Although slightly different representations of the filters are used by the different authors, they extract essentially equivalent information from the stimulus. At this point the models diverge, combining filter responses in different ways to extract the desired information.

Edge and Junction Detection: Heitger et al. (1992)

The detection and location of contour terminations and junctions represents an important early step in identifying objects in the scene. Although numerous models exist for edge detection, relatively little computational work has addressed the neurophysiological mechanisms involved in the detection and classification of junctions. One promising approach to this problem emphasizes the importance of *end-stopped cells* — neurons triggered by the termination of an edge in their receptive fields (Heitger et al., 1992; Peterhans & von der Heydt, 1991, 1993; Peterhans, von der Heydt, & Baumgartner, 1986). Heitger et al. (1992) constructed their edge and junction detection model to be roughly concordant with physiological evidence from areas V1 and V2 in the monkey. Figure 10 shows the basic architecture of the model, along with its extensions in Heitger et al. (1998).

The model's implementation begins with filters known as *S-operators*, the functional organization of which closely resembles the simple cells of primary visual cortex (Hubel & Wiesel, 1968). The model postulates six pairs of S-operators at each location, oriented 30º apart. These filters approximate odd-symmetric and even-symmetric Gabor functions, but are adjusted to eliminate any response to homogenous fields. These odd- and even-symmetric operators commonly are conceptualized as bar and edge detectors, respectively, but this characterization is oversimplified because they do not give unambiguous responses to lines and edges (Morrone & Burr, 1988). Thus, determining the nature of the detected feature requires an explicit comparison of the S-operator outputs.

In the second stage of the model, *C-operators*, analogous to complex cells, collect the responses of the even and odd S-operators. The C-operators determine the "local energy" within an orientation channel (Morrone & Burr, 1988), calculated as the root-mean-square average of the responses of the S-operators:

$$C = \sqrt{S_{odd}^2 + S_{even}^2}$$

C-operators, like complex cells, respond to any appropriately oriented edge or line within their receptive fields, and do not differentiate bright lines from dark ones. In essence, C-operators localize oriented luminance discontinuities, without regard to their nature. Information about whether a detected feature is a line or an edge can be recovered by comparing the relative responses of the even- and odd-symmetric S-operators that served as inputs.

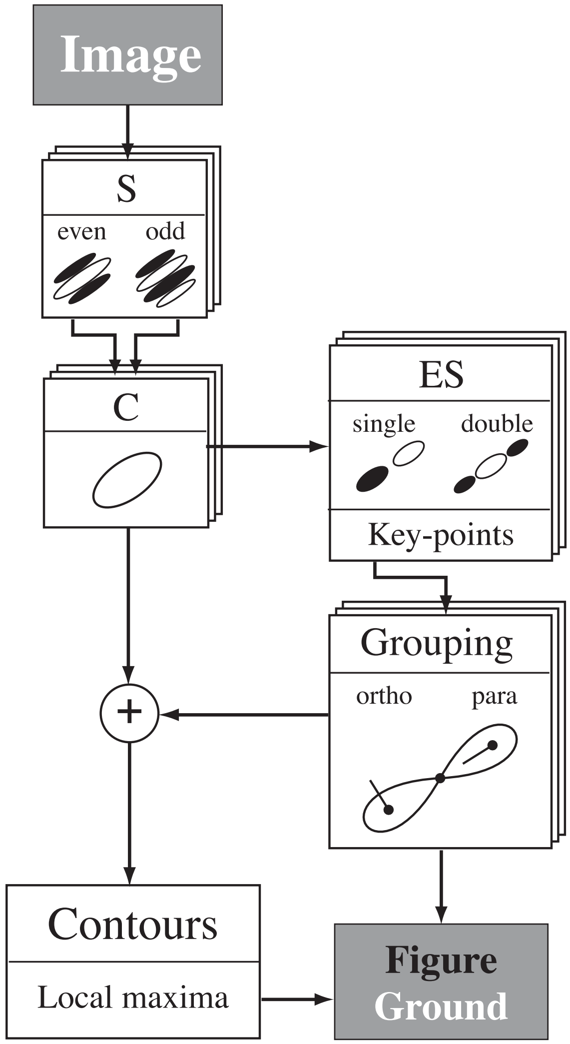


Figure 10. Architecture of the Heitger et al. (1992) and Heitger et al. (1998) models. The early model implements the stages through the end-stopped (ES) operators and the extraction of the key points; the later model describes the grouping of points and the extraction of contours. Redrawn from *Image and Vision Computing*, *16*, Heitger, F., von der Heydt, R., Peterhans, E., Rosenthaler, L., & Kübler, O., Simulation of neural contour mechanisms: Representing anomalous contours, pp. 407-421, Copyright 1988, with permission from Elsevier Science.

The third stage in the model combines the output of the C-operators to form *end-stopped operators*. These operators, analogous to V1 and V2 end-stopped cells, provide an explicit representation of edge and line terminations, corners, and strongly curved contours. End-stopped operators are constructed by taking differences in output between two or three identically oriented C-operators, displaced along the axis of orientation (Figure 11).

There are two types of end-stopped operators. The *single-stopped operators* have one excitatory and one inhibitory zone, constructed by taking the difference in the responses of two identical C-operators, positioned end to end (Figure 11a). These operators respond maximally to a line along the orientation of the operator that terminates between the two zones. The *double-stopped operators* have inhibitory zones on either side of a central

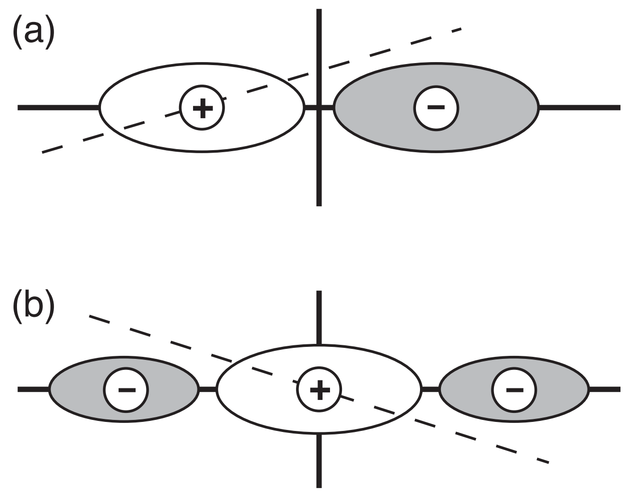


Figure 11. End-stopped operators constructed by taking differences in the responses of identically-oriented C-operators: (a) single-stopped operator; (b) double-stopped operator. Positive input is indicated by "+", and negative input is indicated by "–"; these inputs have balanced weights. Dotted lines depict an off-oriented stimulus that could stimulate the central receptive field without stimulating the inhibitory zones, thus causing the operator to respond to a long line or edge. Redrawn from *Vision Research, 32,* Heitger, F., Rosenthaler, L., von der Heydt, R., Peterhans, E., & Kübler, O., Simulation of neural contour mechanisms: From simple to end-stopped cells, pp. 963-981, Copyright 1992, with permission from Elsevier Science.

excitatory zone; the inputs are weighted so that the flanking C-operators, when summed together, precisely match the central C-operator (Figure 11b). Double-stopped operators respond best to a small disk. As with the S- and C-operators, the model includes end-stopped operators oriented every 30°.

The usefulness of the end-stopped operators for highlighting 2-D image features is limited by their responses to extended lines that pass obliquely through their receptive fields (dashed lines in Figure 11). To minimize these responses, Heitger et al. (1992) proposed a system of horizontal inhibition that suppresses signals to such off-axis stimuli. The end result is a system of operators that respond specifically, with very few false alarms, to endpoints of lines, corners of objects, and regions of strong curvature.

As its final step, Heitger et al.'s (1992) model identifies *key points*, defined as locations where the summed response of the single- and double-stopped operators reaches a local maximum. In simulations, these key points corresponded well to the endpoints and junctions of lines and edges. Heitger et al. hypothesize that the key points, many of which arise from the occlusion of one edge by another, play a critical role in initiating edge interpolation processes (see also Kellman & Loukides, 1987; Shipley & Kellman, 1990).

Once the key points are located, their characteristics are defined by the relative responses of the two types of end-stopped operators. That is, different types of features (end-points, corners, curves) produce different characteristic response patterns across the single- and double-stopped operators. In general, the double-stopped operators identify the tangents to sharply curved edge segments. By contrast, the single-stopped operators indicate the direction of a terminating edge. Thus, at a T-junction, the single-stopped operators pointing in the direction of the T's "stem" would respond, but those oriented along the "roof" would be silent. At a corner, however, the single-stopped operators would signal the orientations of *both* edges converging at that point. Therefore, the relative responses of the various end-stopped operators provide useful information, not just for locating junctions, but also for their classification.

Heitger et al.'s model has several strengths that make it a plausible depiction of contour and junction detection. First, all of the hypothesized operators have clear correlates in the human visual system: S-operators correspond to simple cells, C-operators to complex cells, and end-stopped operators to end-stopped cells. Second, the responses of these operators build upon one another in ways that are similar to hypothesized interactions in early human vision (Gilbert, 1994; Hubel & Wiesel, 1968). Third, casual inspection of simulation output suggests a strong correspondence between perceived features and activity in the various operators. It should be noted, however, that Heitger et al. did not compare the model's output to any psychophysical data, although they presented results for two test images. Further study is needed to assess the model's performance on real-world, noisy images and its agreement with human perceptual processing.

A second limitation of the Heitger et al. (1992) model involves what happens after the filtering described. The model yields maps of local image features: an edge map, comprised of the locations of significant activity in the C-operators, and a map of key points. In effect, it implements the edge and junction detection box depicted in Figure 1. However, it does not describe how this information is combined or integrated into the perception of contours. This issue is addressed in the other models we discuss: that of Yen and Finkel (1998) and Heitger et al. (1998).

Contour Integration: Yen and Finkel (1998)

Perceiving objects requires that, at some level of visual processing, each object's boundary is represented explicitly as a connected, unified entity. Several familiar phenomena support this claim. For example, in reversible figure-ground displays, the contour between two regions tends to switch its boundary assignment as a unit, not piecemeal. The construction of higher-level contour units — beyond local orientation responses — requires linking together the products of many of the kinds of operators we have already considered.

We do not yet fully understand how a contour token (or a unitary object, for that matter) is realized by neural activity. In recent years, however, both neurophysiological and psychophysical studies have provided suggestive clues. Specifically, evidence suggests the existence of extensive horizontal connections among neurons at early stages of visual processing (e.g., Kapadia, Ito, Gilbert, & Westheimer, 1995; Polat & Sagi, 1993, 1994; Ts'o, Gilbert, & Wiesel, 1986). Importantly, a majority of these interactions occur between neurons with similar orientation preferences; this means that neurons responding to different regions of a smooth, continuous contour may "speak" to one another, and thus process the contour as a whole. Neurophysiological research further suggests that the association among like-oriented neurons may be achieved via correlated firing patterns, or neural synchrony, in addition to a simple facilitation of activity (Livingstone, 1996; Ts'o et al., 1986).

Yen and Finkel (1998) proposed a model that accounts for contour integration and perceived contour salience by temporal synchronization of responses in a network of interconnected units akin to oriented cortical cells. The basic processing units of the model, which represent the output of oriented V1 cells, are pairs of linear steerable filters, one with an even-symmetric receptive field and the other with an odd-symmetric receptive field. Although the implementation differs, the combined responses of these units bear a strong similarity to the responses of the C-operators in Heitger et al.'s (1992) model.[3]

The responses of the basic processing units are modified by long-range horizontal connections; the sign and strength of the neural interactions depend jointly on the relative location and orientations of the two interconnected cells. The model's processing utilizes three sets of horizontal connections. The primary facilitatory connections, termed *co-axial*, closely resemble the association field hypothesized by Field et al. (1993), with linkages spreading out in circular arcs from the orientation axis of the neuron (Figure 12). For these connections, association strength decreases rapidly with increasing distance, curvature, and as local orientation deviates from the relevant circular arc. A second set of facilitatory connections entails interaction between cells with parallel receptive fields (see Figure 12); the strengths of these *trans-axial* connections also fall off rapidly with distance and deviation from the preferred orientation. In the model, the co-axial and trans-axial connections compete, so that only one set can be active in a given unit at a given time.

The third set of horizontal interactions operates during a second stage of processing, becoming influential only after the facilitative activity has stabilized. The inhibitory connections suppress the responses of all units whose total facilitation from other active units falls below some threshold. This inhibition helps to distinguish the contour signal from background noise by minimizing the responses of cells that are facilitated by accidental alignment of unrelated contour elements.

In Yen and Finkel's (1998) model, contour integration depends on synchronization of neural units responding to interrelated contour segments. According to the model, the activity of strongly facilitated neurons begins oscillating over time, which allows them to synchronize with other similarly oscillating cells. Initially, these "bursting" cells oscillate with a common temporal frequency but different phases. Over time, the phase of each oscillating neuron is

---

[3] One difference might be the fact that, due to the nature of the steerability computation, only the "preferred" edge orientation for a given location survives in the output representation of Yen and Finkel's (1998) model; in Heitger et al.'s (1992) model, activations in the various orientation channels all contribute to the derived edge map. Freeman and Adelson (1989) offer useful insight into the computational advantages of steerable filters.

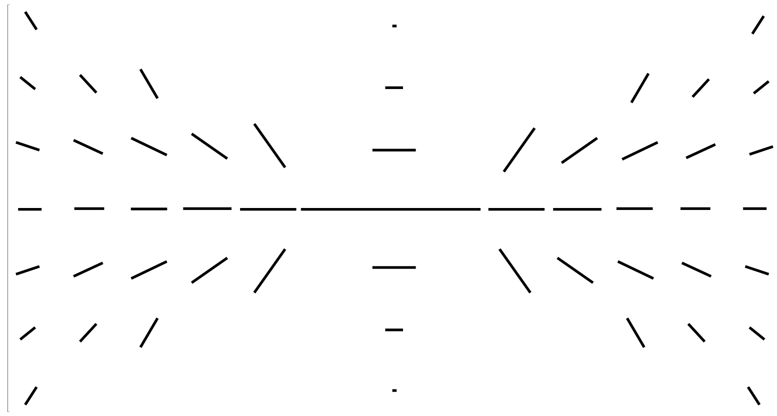


Figure 12. Connectivity pattern of a horizontally-oriented cell, located at the center of the image. At each given location, the “preferred” orientation of the connection is represented by the orientation of the line, while the length of the line is proportional to connection strength. Reprinted from *Vision Research, 38,* Yen, S. C., & Finkel, L. H., Extraction of perceptually salient contours by striate cortical networks, pp. 719-741, Copyright 1998, with permission from Elsevier Science.

modulated by the phase of the oscillators with which they are associated; the strengths of these phase modulations mirror the strengths of the facilitatory connections between the two cells. The oscillations between bursting cells with strong, reciprocal, facilitatory connections rapidly come into phase, and neural synchrony is achieved. The model assumes that a set of commonly oscillating cells leads to the perception of a meaningful contour (although no particular mechanism is provided to extract this information), and that different contours synchronize independently so that they are perceived as separate entities. Yen and Finkel further proposed that the perceptual salience of a contour equals the sum of the activity of all synchronized cells, such that long contours (which activate more units) become more salient than shorter contours.

Yen and Finkel (1998) compared the simulations of their model against a range of psychophysical data, with generally positive results. Importantly, the model achieved all of its successes using a single set of parameters. Among the observations for which the model can account are the findings that closure enhances the visibility of smooth contours (Kovács & Julesz, 1993; Pettet, McKee, & Grzywacz, 1998), and that contrast sensitivity to an oriented target is enhanced by the appropriate placement of like-oriented flanking stimuli (Kapadia et al., 1995; Polat & Sagi, 1993, 1994).

Even more importantly, Yen and Finkel’s model can simulate effectively the contour integration data provided by the experiments of Field et al. (1993). Specifically, the results suggested that the ability to detect a contour consisting of spatially distinct Gabor patches decreases with increasing path curvature, increasing distance between successive elements, and increasing deviation of local orientation from the overall contour path. The model’s performance correlated highly with human data across all of these stimulus manipulations.

Yen and Finkel’s model is distinguished from other contour integration models in its use of temporal synchronization to determine contour salience. An alternative approach, in which salience depends solely on facilitation of neural activity (Pettet et al., 1998), also can explain several psychophysical effects. However, whereas a synchronization model might represent separate contours by synchronizing each neural population independently, activity-based models may experience some difficulty in representing multiple contours as independent entities, especially when the contours in question occupy neighboring regions of space.

In sum, Yen and Finkel's (1998) model provides a dynamic binding scheme for contour integration — combining information across receptive fields sensitive to different regions of space. This model complements nicely the junction detection algorithms proposed by Heitger et al. (1992). Both models begin with the same type of information: activity in units akin to V1 complex cells. However, whereas Yen and Finkel focus on representing connected contours explicitly, Heitger et al. concentrate on the detection of object-relevant junctions (key points). As both of these tasks play important roles in the object perception process, an integration of the two approaches into a single processing scheme would represent a clear advance in the development of a thorough model of early human vision.

It should be noted, however, that this general approach to contour and junction processing limits its focus to *luminance* edges. As discussed earlier, edge detection, in reality, depends on multiple inputs; discontinuities in color, texture, stereoscopic depth, and motion all can lead to the perception of an occluding edge. Our understanding of both junction detection and contour integration may benefit from an attempt to extend existing algorithms to non-luminance edge inputs.

Another limitation of Yen and Finkel's (1998) model is that it is set up to work at one level of spatial scale. In fact, the model not only filters at a single level of scale, but also restricts the stimuli for its simulations to Gabor patches of one size. We consider the issue of single versus multiple-scale filtering in relation to this and other models below.

An additional avenue for future research involves applying the contour integration mechanisms proposed by Yen and Finkel (1998) to the problem of contour *interpolation*. Considerable behavioral and neurophysiological evidence suggests a common representation of physically-defined and illusory contours at some levels of visual processing (e.g., Dresp & Bonnet, 1995; Dresp & Grossberg, 1997; Greene & Brown, 1997; Peterhans & von der Heydt, 1989; von der Heydt & Peterhans, 1989). Moreover, the co-axial facilitatory connections used in the contour integration model correspond nicely to the geometric requirements for edge interpolation proposed by Kellman and Shipley (1991).

In its current state, Yen and Finkel's (1998) model cannot account for the perception of interpolated contours. In short, the model assumes that the facilitatory connections are modulatory in nature, meaning that a neuron cannot become active through lateral interactions alone. Because the facilitatory interactions are effective only for neurons with luminance contrast within their receptive field, the units in Yen and Finkel's model could not respond to an illusory or occluded contour. This assumption is appropriate to a model of V1 interactions, as neurons responsive to illusory contours probably do not exist at this cortical level

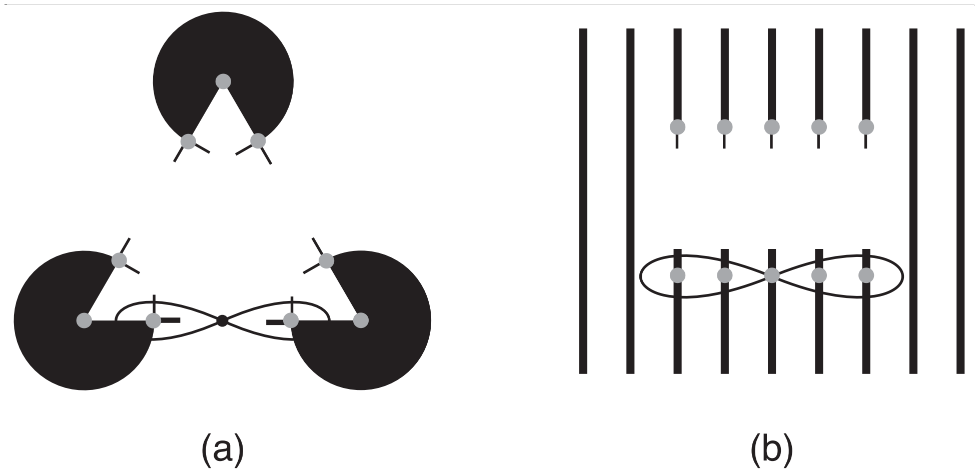


Figure 13. Two examples of the grouping of key points in the Heitger et al. (1998) model: (a) *para* grouping; (b) *ortho* grouping. Gray dots indicate key points, and the short lines extending from the key points indicate the orientation of the associated single-stopped operator. The key points falling within the lobular field produce strong responses along the illusory contour. Redrawn from *Image and Vision Computing, 16,* Heitger, F., von der Heydt, R., Peterhans, E., Rosenthaler, L., & Kübler, O., Simulation of neural contour mechanisms: Representing anomalous contours, pp. 407-421, Copyright 1988, with permission from Elsevier Science.

well-established idea that interpolated boundaries tend to begin and end at contour discontinuities (e.g., Kellman & Shipley, 1991; Shipley & Kellman, 1990). In general, grouping occurs when two single-stopped operators at key points fall in a specific spatial arrangement.[4]

One form of grouping, referred to as *para* grouping, generates illusory contours such as those seen in Figure 13a. When applied to the notched circles, the early stages of the model detect all of the luminance-defined edges and generate key points at the discontinuities. The lines extending from these points indicate the open-ended direction of the single-stopped responses that generated the points. For two key points to become grouped together, the active single-stopped operators must possess characteristics that match a particular *grouping field*. For *para* grouping, the grouping field (depicted in Figure 13a along the lower contour of the illusory figure) decreases in strength with distance from the origin and as orientation deviates from the axis of the field; the lobular shape prevents the grouping of nearby, but unrelated, key points. The grouping algorithm convolves this field with the responses of single-stopped operators at the key points; if a sizeable response emerges, indicating that the single-stopped outputs fell in the appropriate spatial relationship, then the key points become grouped. The numerical value of this computation is assigned to the image point corresponding to the center of the grouping operator. In the example in Figure 13a, each lobe of the *para* grouping field contains a key point with a single-stopped response aligned along its axis, and a second single-stopped response pointing orthogonal to it. The two aligned responses, when convolved

---

[4] The Heitger et al. (1998) model is elaborated in considerable formal and quantitative detail, which makes its predictions explicit for particular stimulus inputs. In our discussion, we have omitted much of this detail.

with the field, produce a significant output and thus a high value at the central location of that grouping operator. According to Heitger et al. (1998), this output forms the basis of illusory contour perception.

The second form of grouping, referred to as *ortho* grouping, generates illusory contours that run orthogonal to the inducing line-ends, as illustrated in Figure 13b. The grouping algorithm parallels that described above, except that the *ortho* grouping field prefers end-stopped responses oriented orthogonally to its axis. The key points in Figure 13b become grouped via convolution with the *ortho* grouping field.

In the model, the grouping operators occupy the points of a grid, and at each point are available at 30° orientation steps. Once the *para* and *ortho* grouping responses have been determined, they are combined with the responses of the C-operators to luminance edges. The final, perceived contours correspond to the maxima in the combined representation of the C-operator and grouping output. For the displays in Figure 13, this combination generates an output image that includes both the real contours and the illusory contours typically perceived by human observers.

In addition to contour interpolation, the grouping mechanisms contribute to the process of boundary assignment. Figure-ground determination arises from the assumption that most key points with end-stopped responses approximately orthogonal to a contour result from occlusion; therefore, the terminating contours that give rise to these key points are assumed to fall on a background surface. Thus, the model compares the responses of end-stopped operators sensitive to one direction of termination with the responses of like operators sensitive to the opposite direction; the side of the contour with the line-ends producing the greater end-stopped response is assigned as "background," and the other side of the contour is deemed "foreground."

To summarize, Heitger et al.'s (1998) model builds from simple units a relatively complex processing scheme that can interpolate contours and determine boundary assignment. When applied to scenes, the model shows some ability to connect the contours of objects that have been occluded by other objects and to distinguish the occluding objects by enhancing contours of low contrast.

The Heitger et al. (1998) model successfully performs contour interpolation using a reasonable architecture for neural processing. The model's basic processing units correspond to neurons known to exist at early stages of cortical visual processing: simple, complex, and end-stopped cells. To accomplish the higher-level tasks (boundary assignment and contour interpolation), the model combines the responses of these basic units in neurally plausible ways. For example, the grouping response depends on a multiplicative "AND" operator. Evidence for such an operator has been found in electrophysiological studies, in which contour-sensitive neurons in area V2 responded when two luminance-defined edge elements flanked its receptive field, but not when either element appeared alone (Peterhans & von der Heydt, 1989).

Interestingly, there is a high degree of similarity between the shape of the grouping fields in Heitger et al.'s (1998) model and the geometry of horizontal interactions proposed by Yen and Finkel (1998) for contour integration. The shape of the *para* grouping field closely mirrors the co-axial connections for integration, whereas the shape of the *ortho* grouping field resembles the trans-axial connections. As previously discussed, several lines of evidence converge on the possibility of a common mechanism for the perception of physically-specified and interpolated contours (e.g., Dresp & Bonnet, 1995; Dresp & Grossberg, 1997; Greene & Brown, 1997; Peterhans & von der Heydt, 1989; von der Heydt & Peterhans, 1989). Therefore, although very different implementations characterize the two models, the similarity in shape of the grouping fields for contour interpolation (Heitger et al., 1998) to the horizontal connections for contour integration (Yen & Finkel, 1998) fits well with our geometric understanding of these perceptual processes.

Heitger et al.'s (1998) model, however, is not without its limitations. For one, some of the operators appear to have been designed to produce a particular result, with little theoretical or empirical justification for either their existence or their nature. Second, the grouping process depends on some complex weighting structures for cross-orientation inhibition (not discussed above), necessary to suppress grouping when signals of multiple orientations exist in a single location. Despite the overall neural plausibility of the model, the proposed weightings may be difficult to implement in simple neural circuitry. Some other issues of more general significance, such as the relation of the Heitger et al. (1998) model to the perception of partly occluded contours, are taken up in the next two sections.

Issues for Geometric and Neural Models

As we have seen, recent work has led to an understanding and quantification of the spatial relations crucial for object perception, as well as the development of neural-style models of some underlying processes. This progress raises a number of important issues for future work. Some involve unsolved problems within the domains addressed individually by geometric or neural models. A number of other issues involve connecting these two types of models. We consider some important examples of each type below. First, we address problems intrinsic to neural models, in their current instantiation. Next, we consider connections between geometric and neural models, emphasizing knowledge about geometry and processes that could be, but have not yet been, implemented in neural-style computational models. In a final section, we examine the frontier in research on geometric models: phenomena that suggest the need for major changes or additions to all current models in order to capture key aspects of object perception.

Neural Models of Contour Processes: Issues and Challenges

<u>Contour Interpolation: Higher-Order Operators vs. Network Models</u>. How does the nervous system carry out contour interpolation across gaps in the input? The model of Heitger et al.

(1998), considered above, addresses this question directly. In this model, real edge inputs on either side of a gap feed into nonlinear, higher-order grouping operators. The activation of these operators, centered over a discontinuity in edge input, may be used to define an illusory contour existing across the gap.

Alternatively, interpolation may be carried out by an interactive network of orientation-signaling units (e.g., Field et al., 1993). According to this idea, luminance-defined edges activate some oriented units, leading to facilitative interactions with other units that do not receive direct stimulus input. Interpolation occurs when a path of units, defined on either end by directly-stimulated units, becomes active as the result of these interactions. Although the model of Yen and Finkel (1998) prohibits the activation of units that receive no direct stimulus input and thus cannot perform contour interpolation in its current state, its network architecture is highly compatible with this general concept of interpolation.

Can existing data determine which approach to interpolation — or perhaps what combination of network interactions and higher-order operators — is correct? We doubt that this issue can be decided from present knowledge. Each approach presents some clear advantages along with some equally obvious difficulties.

One advantage of the higher-order operator approach involves its apparent consistency with certain perceptual and neurophysiological findings. Heitger et al. (1998) note that single-cell recording data from area V2 of the macaque indicate that cells responding to illusory contours combine the inputs of two adjacent areas multiplicatively (Peterhans & von der Heydt, 1989). In other words, neurons signaling illusory contours cannot be activated without real contour activation on both sides of the gap. This finding fits nicely with the perceptual fact that noticeable illusory contours do not arise from a single inducing element. Moreover, psychophysical work suggests that thresholds for detecting low-contrast, oriented stimuli decrease when co-axial stimuli appear on both sides, but not when a single flanking stimulus is presented (Polat & Sagi, 1993).

All of these findings may be explained by the existence of higher-order operators that group like-oriented stimuli and facilitate interpolation. Alternatively, the results could be captured by a network model that requires facilitation from units on both sides to produce an above-threshold response in intermediate neurons. Indeed, in suggesting that the thresholds of intermediate units can be affected by flanking stimuli (Polat & Sagi, 1993), the work on lateral interactions seems more consistent with a network-style model than with a higher-order operator model. Nonetheless, further research is needed to distinguish the approaches.

A potential drawback of higher-order operators involves curved interpolations, which occur readily in perception. In the Heitger et al. (1998) model, interpolation of collinear edges depends on the *para* grouping process. Para grouping utilizes collinear grouping fields to specify whether an interpolation should occur at each image point; non-collinear grouping fields are not envisioned within the Heitger et al. model. As a result, points along mildly curving illusory contours can produce non-zero interpolation responses only if both inducing edges fall at least partially within the grouping fields of a collinear operator. Thus, this scheme could interpolate between two input edges whose relative orientations are approximately 150˚ to 180˚. By contrast, psychophysical research suggests that interpolated contours may be induced by edges whose relative orientations fall between 180˚ (collinear) and 90˚ (Field, Hayes, & Hess, 1993; Kellman & Shipley, 1991). The para grouping mechanism of Heitger et al. cannot easily account for curved interpolations in illusory and occluded contour perception.

The Heitger et al. (1998) model does allow for non-collinearity in the *ortho* grouping process. For this process, recall that illusory contours arise from appropriately oriented line ends, and that interpolation orthogonally to the line ends specifies an occluding contour. At line ends, the authors argue, the exact orientation of the occluding contour may be poorly specified; hence, the model's operators allow 30 degrees of orientation change between adjacent inducers. Though these operators could produce appropriately curved interpolated contours, they seem somewhat *ad hoc* when one considers the context in which they are introduced. Even more problematic, the introduction of similar units into the para grouping process likely would cause an unrealistic proliferation of the number of higher-order operators necessary to account for the varied, possible curvatures of perceptible interpolated contours.

This limitation in accounting for curved interpolations can be overcome by appealing to network-style models. By recruiting intermediate units along various curved paths, network-style models can interpolate curves without necessitating the introduction of an unwieldy number of operators. As previously suggested, Yen and Finkel's (1998) model for the integration of visible contours may be extended to contour interpolation simply by allowing facilitative interactions among intermediate units in the absence of direct stimulus-driven activation.

Nevertheless, network-style models also have drawbacks. Whereas contour integration conceivably could depend solely on facilitative interactions among V1 neurons, the network interactions responsible for illusory and occluded contours arguably must require an additional layer or network. Simply put, interpolated contours can be distinguished easily from visible edges, despite their perceptual status as real contours that connect under occlusion. If real and interpolated contours depended on identical network computations, with neural units activated either by direct stimulus inputs or lateral interactions, then it becomes difficult to see how occluded and "visible" contours could be distinguished. An arrangement in which V1 inputs feed into a separate network layer could handle this problem; information about stimulus-driven activations could be preserved in the earliest layer, whereas interpolated contours could be computed in the next.

Another computational task that may be harder for network models than for higher-order operators involves enforcing monotonicity. Research indicates that interpolated contours bend in one direction only (Kellman & Shipley, 1991); this monotonicity constraint effectively rules out, beyond a very small threshold tolerance, doubly inflected contour interpolations. Evidence suggests that interpolation breaks down when initially collinear edges are misaligned by more than about 10 or 15 minutes of visual angle (Shipley &

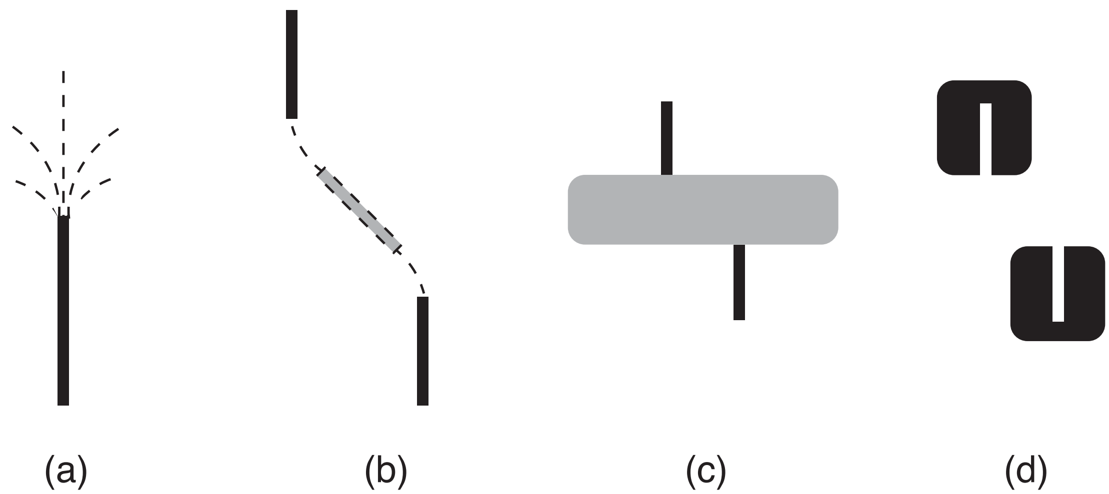


Figure 14. An illustration of the monotonicity problem. A simple network scheme (a and b) cannot easily enforce the monotonicity constraint: the notion that interpolated contours cannot be doubly-inflected (c and d). See text for details.

Kellman, 1992a).

The difficulty with enforcing monotonicity in a simple network of interactions, illustrated in Figure 14, arises from the fact that an intermediate neuron could be facilitated by two flanking neurons that would not facilitate one another. Figure 14a shows a luminance edge with excitatory connections branching outward to approximately ±45˚. Two such edges, displaced horizontally and vertically, are shown in Figure 14b. Note that an intermediate unit (shown as a dotted gray segment) between the two stimulus-driven units should, according to a simple network model, be facilitated by both. Nonetheless, such misaligned inputs do not support interpolation (Figures 14c and 14d). Therefore, network models must use some scheme to exclude formation of an interpolated edge between two such inputs.

The point to be distilled from much of the foregoing discussion is that there are computational tradeoffs between models that use higher-order operators and models that perform interpolation in a network of interacting, oriented units. The ability of units in a network to link up in different ways is a strength: it provides the model with flexibility. On the other hand, the fact that contours are encoded over a number of neurons, each of which "knows" only about its own activity, raises a problem of labeling or identifying the emergent structure. (There is, after all, no homunculus looking at the network's output.) In higher-order operator models, an operator's output can be easily labeled for further processing based on its nature and location. At the same time, the flexibility needed for finding varied image properties, such as contours of various curvatures, may require unreasonable proliferation of operator types. Further research may reveal that both processing strategies, operating at different levels, combine to perform basic computations in object perception.

Multi-Scale Filtering. In the natural environment, images contain significant contrast variations at different spatial scales. Consequently, numerous researchers have suggested that visual processing requires the integration of information across multiple spatial frequency channels (e.g., Marr & Hildreth, 1980; Morrone & Burr, 1988).

The neural-style models we have considered all utilize a single level of scale. By contrast, a number of tasks in object perception appear to require multiscale modeling. For example, consider the problem of distinguishing meaningful edges from noise. Because many occluding edges involve abrupt changes between regions with different properties (e.g., luminance, color, texture, motion and/or depth), they will produce activation at the same location in channels at different spatial scales. Similarly, edge classification (e.g., distinguishing between an occluding edge and a shadow) often may be accomplished using multiscale information. A shadow's penumbra, for example, involves a gradual luminance change over space; this might activate a low spatial frequency detector strongly, but a high frequency detector only minimally.

Similar considerations may apply to the detection and classification of junctions. For example, low spatial frequency filters may explain our ability to recognize a square with rounded edges as a square; by contrast, our ability to distinguish squares with rounded corners from those with sharp corners might depend on high spatial frequency operators. Low spatial frequency filters may miss important details, whereas high spatial frequency filters often cannot distinguish meaningful information from noise.

As discussed earlier, tangent discontinuities play a crucial role in initiating the contour interpolation process. For effective interpolation, filtering on a fine scale seemingly would be crucial for distinguishing actual TDs from regions of high curvature; however, filtering at a coarse scale also is important for excluding the responses of fine junction filters to local noise. By integrating over multiple spatial scales, true junctions may be accurately disambiguated from other possibilities (Marr & Hildreth, 1980; Watt & Morgan, 1985; Witkin, 1983). Incorporating edge and key point detection at multiple levels of scale, as well as elaborating algorithms for integrating information across different spatial frequency channels, are important goals for future work.

<u>Output Representations</u>. Figure 1 set out an overall framework of tasks and processes required for object perception. In this context, it is important to review what actually has been accomplished by the neural models we have considered. Heitger et al.'s (1992, 1998) models seek to locate edges and junctions (key points), and to interpolate illusory contours between visible edges. The models' outputs, which come from convolution operations, consist of numerical values assigned to two-dimensional image coordinates. These output maps may be viewed as images themselves. Upon inspection, one can see the locations of edges, junctions, or illusory contours.

With these sorts of outputs, it is easy to become confused about what a model has and has not accomplished. When viewing an output "image," we bring along our own grouping and segmentation routines, and we may suppose that the image contains bounded objects. Crucially, however, the final representations in the Heitger et al. models (1992, 1998) do not represent explicitly any linked contours, any units, or any shapes in the scene.

How are the maps produced by Heitger et al.'s (1992, 1998) filters used to segment and group contours and objects? These issues remain to be modeled. Yen and Finkel's (1998) work suggests a useful algorithm for one necessary task — that of integrating local edge responses into a meaningful contour. Further research is needed to determine how the visual system might develop the other symbolic representations required for object perception. Besides visible contour tokens, these include representations of visible regions, interpolated edges, units, and shapes.

Connecting Geometric and Neural Models

The Common Interpolation Mechanism for Occluded and Illusory Contours. In discussing a geometric model of contour interpolation, we described findings and phenomena indicating that a common contour interpolation process governs illusory and occluded contours. Most importantly, we presented the argument that, in at least some cases (e.g., self-splitting objects), contour interpolation logically must occur prior to the determination of modal (illusory) or amodal (occluded) appearance. This need not always be the case; clear depth order information is often available prior to interpolation. Cases in which interpolation occurs prior to final depth ordering, however, unmask the fact that there cannot be separate interpolation processes for contours that ultimately appear in front of, or behind, other surfaces. Rather, the differing phenomenology of illusory and occluded contours derives from depth ordering information that arises at various processing stages.

Most neural-style models have addressed either illusory or occluded contours, but not both. For example, the model of Grossberg and Mingolla (1985) described the gap-filling process as a means of surmounting occlusion by retinal veins, but not occlusion by nearer objects. Accordingly, only modal contours could be interpolated by that model. (More recent proposals by Grossberg and colleagues (e.g., Grossberg, 1994), however, may be more compatible with the identity hypothesis.)

The interpolation model we have considered in some detail — that of Heitger et al. (1998) — is interesting in this regard. This model is motivated, in part, by single-cell recordings indicating that some V2 cells respond to illusory contours but not to comparable occlusion displays (Peterhans & von der Heydt, 1989). Thus, the model is designed to account for illusory, but not occluded, contours. For most occluded contours, the *para* grouping process, which connects edge inputs of like orientation, would be blocked by contour junction information on the occluding surface, which contradicts the existence of an illusory edge in the foreground. Nevertheless, Heitger et al.'s simulations do produce interpolation responses for some occluded contours, such as the partly occluded circles in the standard Kanizsa triangle display. The authors consider the latter result to be something of an embarrassment; it is described as "a limitation of the present model" (p. 414).

We believe that the model's responses need not be a source of embarrassment. As described above, evidence implicates a common interpolation process for illusory and

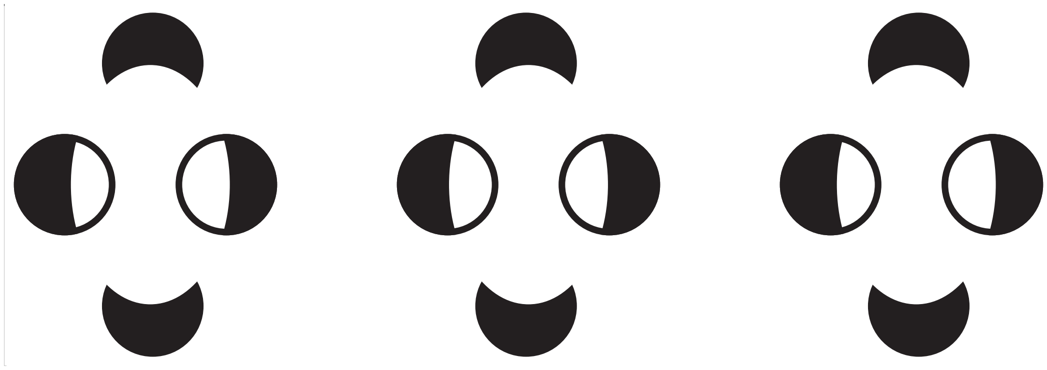


Figure 15. Example of a quasimodal object. The central white ellipse has interpolated contours at four locations. Each connection links an occluded contour with an illusory contour. The effect is visible in each image, but it is more vivid when stereoscopic disparity gives the appropriate depth relations. To obtain the stereoscopic effect by free fusion, either cross-fuse the left and middle images, or diverge the eyes to fuse the middle and right images. After Kellman, Yin, and Shipley (1998).

occluded contours. Moreover, the process seems to be somewhat promiscuous. As we will suggest below, the edge interpolation mechanism appears to create some contour connections that never reach conscious perception; they are deleted based on contour junction and boundary assignment information, as well as processes that require consistent depth order of surfaces in a scene. Accordingly, a better fit to existing data is a relatively unconstrained interpolation process, followed by both "gating" effects (that delete some contours) and depth ordering effects (that lead to the differing illusory or occluded appearance). The model of Figure 1 indicates this idea with an interactive connection between boundary assignment and edge classification processes and the outputs of contour interpolation.

Do any single-cell recording data rule out such a scheme? It is worth noting that Peterhans and von der Heydt (1989) did find cells in V1 that responded to both their illusory contour and the related occlusion displays, although they ventured that these results may have been artifactual. Another possibility is that the equivalent responses of V1 cells to illusory and occluded contours indicate a common contour interpolation step, whereas the nonequivalent responses of V2 cells indicate that depth information relevant to the final contour appearance has come into play. At this point, we lack any sufficiently detailed mapping of computations onto cortical layers to determine the probable location of cells carrying out the common interpolation step.

Hopefully, it is clear why we have argued for a common interpolation step in object formation. One other phenomenon underscores the difficulties of approaching interpolation separately for illusory and occluded contours. Figure 15 illustrates interpolation in a situation that fulfills the requirements for neither illusory nor occluded contours — the case of so-called *quasimodal* objects.[5] At each of four gaps in the figure, interpolation creates an edge

---

[5]The term *quasimodal* is used to describe cases that fall in between amodal and modal completion (the latter terms used by Michotte, Thines, & Crabbe, 1964, to describe

that is occluded on one end and illusory on the other. Kellman, Yin, and Shipley (1998) studied quasimodal displays using an objective performance paradigm (the "fat-thin" method of Ringach & Shapley, 1996) known to be sensitive to contour interpolation and found that they provided virtually identical facilitative effects on classification performance as the equivalent displays containing only illusory or occluded contours.

The model of Heitger et al. (1998) would not interpolate contours in quasimodal displays. However, quasimodal contours present no special problem in our model. An interpolated contour initially is constructed depending on the spatial relations of two visible contours. However, the initial interpolation step does not determine the constructed contour's relation to other contours or surfaces in the scene. Thus, depth ordering processes subsequently can place an interpolated contour either in front of or behind other surfaces, depending on depth information about other objects in the scene and other interpolations (as in Petter's Effect). A quasimodal contour simply is an interpolated contour that, based on these depth ordering influences, finds itself in front of some surfaces and behind others.

In sum, our analysis suggests that neural models must be elaborated in a way that provides a common interpolation step whose outputs are subject to gating and depth ordering effects that lead to a number of different perceptual outcomes. A model of illusory contour formation alone, as proposed by Heitger et al. (1998) would leave open how unit formation under occlusion — arguably the most important function of interpolation — could occur. More importantly, however, our framework is most consistent with known phenomena, and may offer some useful hints in the search for the neural substrates of interpolation.

Support Ratio. One well-established quantitative finding from research on contour interpolation is the notion of *support ratio* (Banton & Levi, 1992; Shipley & Kellman, 1992b). Consider an interpolated edge to consist of two physically-specified segments (e.g., given by luminance contrast) and an intervening gap. Over a wide range of display sizes, up to and perhaps beyond 10 degrees of visual angle, interpolation strength increases linearly with support ratio — the ratio of physically-specified segments to total edge length (physically-specified segments plus gap). Support ratio makes ecological sense as a property influencing object perception in that it is scale invariant. For partly occluded objects, changes in viewing distance leave support ratio unchanged, so long as viewing distance is large relative to the depth separation of the occluding and occluded objects.

Support ratio is an example of a robust finding about the geometry of object perception that is not implemented in current neural models of interpolation. For example, in the Heitger et al. (1998) interpolation model, grouping operators have fixed retinal size, and

---

interpolation of occluded and illusory objects respectively). The displays were initially described as "hybrids," with the alternative term quasimodal mentioned with tongue (pen) in cheek. However, a unanimous vote of the reviewers of Yin et al. (1998) established quasimodal as the preferred technical term.

therefore are not scale invariant. Due to the size and configuration of the operators, grouping does produce similar shapes over a limited range of scales. However, it appears that some changes in the operator would be needed to obtain results fully consistent with human psychophysical data on support ratio.

Surface Interpolation. Earlier we described and illustrated the surface interpolation process that operates under occlusion (e.g., Yin et al., 2000). No current neural-style models incorporate this aspect of processing, which influences unit formation and object perception more generally. Grossberg and Mingolla (1985) described complementary processes involving contours and surface features, including a notion of surface spreading in modal displays (e.g., in "neon" color spreading). The surface spreading process under occlusion may be closely related to modal spreading phenomena, but the specifics remain unclear. Likewise, the specific interactions of contour and surface processes found by Yin et al. (1997, 2000) have not been implemented in terms of underlying neural units.

Geometric Models: Issues and Challenges

In the previous section, we considered discrepancies between the known geometry of interpolation and current neural-style models. To conclude, we raise several challenges for advancing geometric and process models of object perception themselves. These challenges are of several different types. First, we consider one part of the geometry of contour interpolation that is not fully understood: the misalignment of parallel contours. Next, we consider issues regarding the relation of a contour interpolation processing stage to the final appearance of contours. Then, we ask whether the relatively local, bottom-up model we have suggested is adequate to accomplish object perception; in particular, we address the question of top-down influences. Finally, in the last two sections, we consider findings indicating that the domain of object perception models must be radically enlarged; specifically, models must be broadened from the consideration of static, 2-D images to encompass 3-D and spatiotemporal (motion-carried) information for object formation.

Contour Misalignment and Interpolation. An example of an unresolved problem in geometric models is the problem of contour misalignment. Strictly speaking, the geometry of relatability should not tolerate any misalignment of parallel (or co-circular) edges. As with the application of any mathematical concept to perceptual processing, we would expect that the notion of alignment is not absolute; that is, there should be some small tolerance for misalignment. This expectation follows from several factors, including the spatial scale and resolution of the neurons encoding the relevant visual properties, and internal noise in perceptual systems. Indeed, evidence from both perceptual report (Shipley & Kellman, 1992a) and objective performance paradigms (Yin, Kellman, & Shipley, 1998) suggests a small tolerance for misalignment: contour interpolation appears to break down when misalignment exceeds about 10 to 15 minutes of visual angle. This value is much larger than Vernier acuity

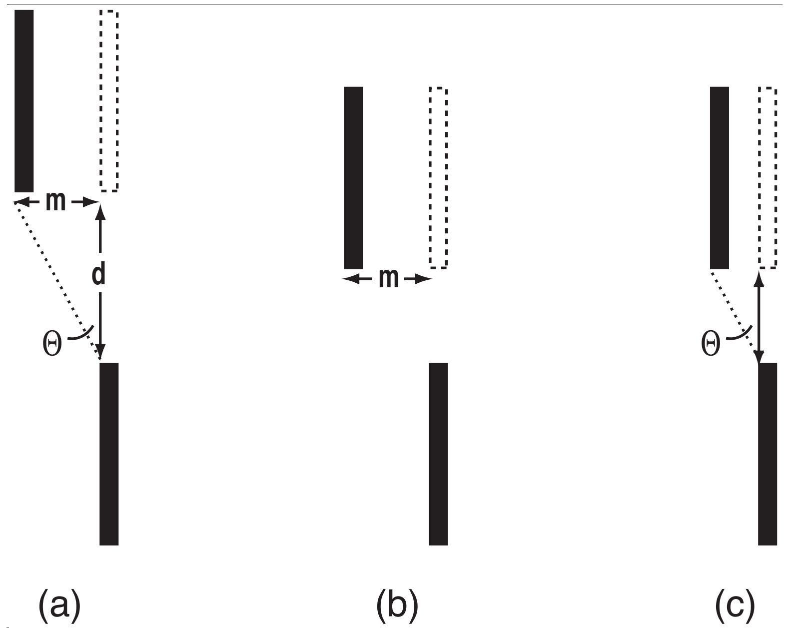


Figure 16. Possible determinants of the threshold for misalignment in contour interpolation. (a) Horizontal misalignment, *m*, and vertical separation, *d*, determine angle Θ. In (b), the edges have the same horizontal misalignment, *m*, as in Figure (a), but with the vertical separation, *d*, reduced. If tolerance depends only on *m*, then the effect of misalignment on this display should be similar to that in Figure (a). In (c), both the horizontal misalignment, *m*, and vertical separation, *d*, have been reduced proportionately from Figure (a), thus preserving angle Θ. If misalignment tolerance depends on *m*, then the effect of misalignment on this display should be smaller than in Figure (a); if misalignment tolerance depends on Θ, then the effect of misalignment on this display should be similar to that in Figure (a).

thresholds, but interestingly, it is about the magnitude of misalignment found in the Poggendorf illusion (Robinson, 1972). The Poggendorf and related illusions produce significant distortions in the perceived alignment and relative orientations of visible contours. Tolerance for an amount of misalignment of the same order of magnitude as in these illusions may allow the visual system to connect appropriate contours despite distortions that occur in spatial vision.

The value of 10 to 15 minutes of misalignment may not be a constant of nature, however. Instead, it may be an artifact of the display sizes used in the few studies that have examined this issue. If a constant value of misalignment determines whether or not contours get interpolated, then interpolation would not be invariant with viewing distance; in other words, some contours that appear connected from far away would appear disconnected when the viewer moves closer. A scale-invariant relationship, such as that depicted in Figure 16, more likely determines tolerable misalignment. The relevant geometry may be the ratio of contour misalignment to contour separation (or, equivalently, the angle Θ; see Figure 16). We currently are carrying out studies to disentangle whether this scale-invariant, angular notion or an absolute retinal misalignment value governs contour interpolation.

<u>Contour Interpolation and Perceived Contours</u>. Our earlier discussion suggested that contour

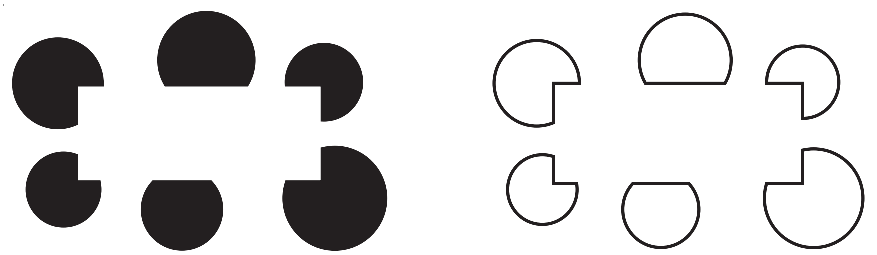


Figure 17. Filled inducers lead to the perception of strong illusory contours, whereas identical figures composed of outlines fail to elicit illusory contour perception. After Guttman and Kellman (2001).

interpolation mechanisms operate between oriented edges whenever they satisfy the geometric constraints of relatability. This idea raises an interesting question: Is relatability a sufficient condition for perceiving interpolated contours (illusory or occluded)? Some exceptions seem readily apparent. For example, it is well known that unfilled, outline figures do not effectively evoke illusory contours, even when these outlined inducers have identical edge relationships to filled inducers that do produce robust interpolated contours (Figure 17). If contour interpolation depends on the relative orientation and position of physically-defined edges, then what explains the different appearances of these two displays?

Guttman and Kellman (2000) recently addressed this question in a series of experiments. In each experiment, observers viewed shapes with edges defined by either filled or outline notched-circle inducers. For each image, observers performed a classification task that could be carried out either by judging the overall shape depicted by the real and interpolated edges, or by examining the edge orientations of individual elements (Figure 18). In some studies, we measured reaction time for making a speeded classification judgment; in other experiments, stimuli were masked after a brief exposure, and sensitivity (d') for discrimination was measured.

The results of these experiments suggested that a low-level contour linking mechanism operates between outline inducers, even though no interpolated contours are perceived (Guttman & Kellman, 2000). As in previous studies, the classification tasks appeared to be sensitive to the process of contour interpolation; manipulations that disrupted the geometry of edge relatability, such as turning the inducers outward, rounding tangent discontinuities, or misaligning the contours (as in Figure 18), dramatically reduced classification performance. However, in all experiments, performance was both accurate and rapid for shapes defined by outline inducers as well as shapes defined by filled inducers. This proficiency would not be expected in the absence of interpolation. Results from a priming experiment provided converging evidence for contour interpolation between appropriately oriented outline inducers (Guttman & Kellman, 2000).

These data are consistent with an interesting theoretical proposal: *Contour interpolation occurs at an early processing stage whenever visible edges satisfy the relatability criteria, but not all interpolated contours ultimately are manifest in perception.* To progress to later stages of representation, interpolated contours must pass certain gating

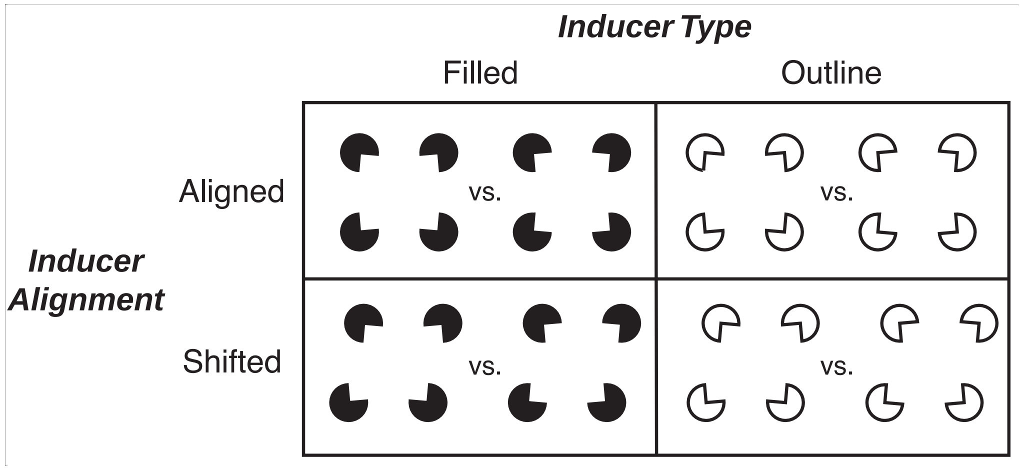


Figure 18. An example of one of the tasks used by Guttman and Kellman (2000) to investigate the operation of interpolation mechanisms on "illusory" figures composed of outline inducers. For these stimuli, observers made a fat-thin judgement based on the rotation of the corner elements — or on the overall appearance of the illusory figure. Control stimuli consisted of misaligned inducers, which contain disrupted edge relationships.

criteria, such as requirements about consistency of boundary assignment. The outline figures in our experiments are not recruited into the final perceptual representations of unity and shape that are available to consciousness. Nonetheless, outlines, like filled inducers, appear to trigger contour interpolation responses at an early stage that influence performance on classification tasks.

If true, the idea that all relatable contours are interpolated at an early stage, then subjected to gating mechanisms, may clarify several issues in object perception. Consider the outline displays depicted in Figures 17 and 18. *Why* do these stimuli fail to support illusory contours? Clearly, both filled and outline inducing patterns will produce edge and junction responses in the models we have considered. These outputs may, in turn, produce interpolation responses linking contours across gaps. Only some of these interpolation responses, however, may be recruited into the final perceptual representation, depending on criteria imposed subsequently. One such criterion may be consistency of boundary assignment (Nakayama, Shimojo, & Silverman, 1989). If interpolated contours are constructed at an early stage but lead to contradictory results in assigning boundary ownership to the objects in the scene, the perception of interpolated boundaries may not occur. In outline displays, boundary assignment considerations likely are the factor preventing illusory contour perception. Thin lines may always be encoded as "owning" their boundaries (see Kellman & Shipley, 1991, for discussion). Moreover, a closed outline form may have an even stronger tendency for its boundary to be intrinsic (i.e., owned by the outline shape). An assignment of boundary ownership to an outline inducer is incompatible with perception of an illusory form, because perception of the form would require a reversal of boundary assignment, such that the illusory figure owns the boundary.

The idea that interpolation mechanisms operate on all relatable contours at some stage is consistent with suggestions that contour interpolation occurs relatively early in cortical visual processing. Although imaginable, it seems unlikely that information regarding boundary assignment is present early enough to prevent “inappropriate” interpolation. Feedback from higher levels to V1 and V2 could, theoretically, provide the necessary boundary assignment information to allow interpolation between some appropriately oriented contours but not others. However, this explanation cannot account for the proficient classification of shapes delineated by outline inducers (Guttman & Kellman, 2000).

More likely, the lateral interactions necessary for contour interpolation are engaged by any appropriately oriented edges, regardless of junction type and whether they ultimately lead to the perception of interpolated contours. By this interpretation, boundary assignment processes subsequently delete any inappropriately interpolated contours at later stages of processing (or through feedback to the early interpolation mechanism), thus preventing their conscious perception. Alternatively, it is also possible that all interpolated contours continue to reside in the visual system, but cannot be seen unless the visual system also constructs a bounded *surface*. Boundary assignment processes may facilitate the perception of a central illusory surface in figures consisting of filled inducers, but prevent this perception in the case of figures consisting of outline inducers. Accordingly, the absence of a surface in the latter case would render any interpolated contours “invisible” to conscious visual perception. The way in which boundary assignment processes interact with interpolation mechanisms is an important question for future research.

<u>Symmetry, Regularity, and Familiarity in Object Perception</u>. The Gestalt notion of Prägnanz describes the tendency of perceptual processing to maximize the simplicity and/or regularity of perceived forms. In Bayesian approaches to computational vision, perceptual outcomes might include information about the prior likelihood of particular objects existing in the environment. What these perspectives have in common is the idea that object perception may not be accomplished through purely feed-forward, relatively local processes. Instead, global and/or top-down influences may affect one or more stages of processing in object perception. Although suggestions about such influences have been perennial (e.g., Kanizsa, 1979), it remains controversial whether and how notions such as object symmetry, regularity, familiarity, and likelihood actually affect the processes of object perception.

As the claims and phenomena in this domain have been diverse (not to mention confusing), we think it is important to distinguish three ways in which such effects might arise, with reference to the model presented in Figure 1. As given, the model is primarily “bottom up,” in that representations of unified objects derive from stimulus information and computations on that input.

One way that object familiarity or regularity might affect object perception processes would be through some feedback loop or “top-down” influence. Top-down processing encompasses any effects in which the possible *outcomes* of processing directly affect the processing itself. That is, some relatively early stage of processing activates a late representation, probably the shape representation; the activated shape representation then feeds back to and influences the outcome of some earlier processing stage(s). We will call this sort of top-down processing a *Type I effect*. As an example of a Type I effect, recall the suggestion given in Figure 2 about how familiar shape might affect boundary assignment (Peterson, 1994; Peterson & Gibson, 1991, 1994). The results from contour integration might activate a shape representation, which then feeds back to influence boundary assignment; this top-down influence makes it more likely that the familiar shape will "own" the contour.

A second kind of effect — let us call it a *Type II effect* — occurs when the products of some fairly early stage of processing take a direct shortcut to a late representation, which can then be used in some cognitive task. Type II effects differ from Type I effects in that activation of the late representation does not influence earlier processing in any way. For example, suppose that one owns a pair of athletic shoes of a unique purple color. Suppose that one such shoe resides on a bed, mostly covered by a blanket, so that only one small purple patch is visible. Viewing this scene, only a small purple patch becomes available in the visible regions representation, and this patch does not own its boundaries; in this case, the visual system lacks sufficient information to perform boundary interpolation and recover shape through the processes given in the model. However, the purple color alone may activate a memory representation of the uniquely colored shoes. Therefore, the observer can "recognize" the shoes, in the sense that some representation of the shoes becomes active. This can be modeled in terms of Bayesian priors — the patch of purple may be highly correlated with the presence of these shoes.

The shoe scenario exemplifies a Type II effect because recognition (a cognitive task) has been accomplished while bypassing several processing stages; in fact, Type II effects may best be described as shortcuts that result in "recognition from partial information" (Kellman, 2000). The example cannot be described as a Type I effect, as activation of the shoe representation need not have any effect on earlier stages of processing, like contour integration or boundary interpolation; these processes may continue to produce indeterminate outputs. We should note, however, that the detailed contour and surface processes of the model can proceed in parallel with these sorts of recognition shortcuts. That is, multiple representations may emerge from processing (van Lier, van der Helm, & Leeuwenberg, 1995), one from the full array of contour and surface processes (integration, interpolation, unit formation, etc.), and the other from a processing "shortcut" whereby some early representation activates a shape representation directly.

Finally, symmetry or regularity information may affect a particular component process in object perception without any feedback from later representations. That is, some functions in the model in Figure 1 may take into account stimulus information from outside the immediate processing area, thus allowing global symmetry or regularity to affect the perceptual outcome. We will refer to these sorts of global influences, which occur without any top-down processing, as *Type III effects*.

Sekuler, Palmer, and Flynn (1994) reported evidence suggesting that symmetry influences object completion in displays like the one shown in Figure 19. In this example, the presence of symmetry, which resulted in "global" completion, could have influenced perceptual processing via any of the three kinds of effects. That is, the three visible lobes (Figure 19a) may have activated a later representation of a fully symmetric object, which then fed back to influence the interpolation process; in this case, the role of symmetry would be a Type I effect. However, symmetry may influence perception without any feedback from a higher-level representation, as a Type I effect requires. If symmetry influences visual interpolation in general, whether the resulting figure is familiar or unfamiliar (and thus whether or not a later shape representation exists), then the effect may be better understood as a Type III effect. That is, the process of contour interpolation itself may take symmetry into account; whereas we have stressed local edge relationships in guiding the boundary interpolation process, the visual system might instead use some algorithm whereby the three visible lobes in Figure 19a trigger the generation of a fourth lobe. Finally, symmetry's influence could be a Type II effect — symmetry may not affect contour interpolation per se, but instead might simply activate some representation of a symmetric object through recognition from partial information. Sekuler et al.'s data were derived from a priming paradigm; during priming experiments, observers view both occluded and unoccluded

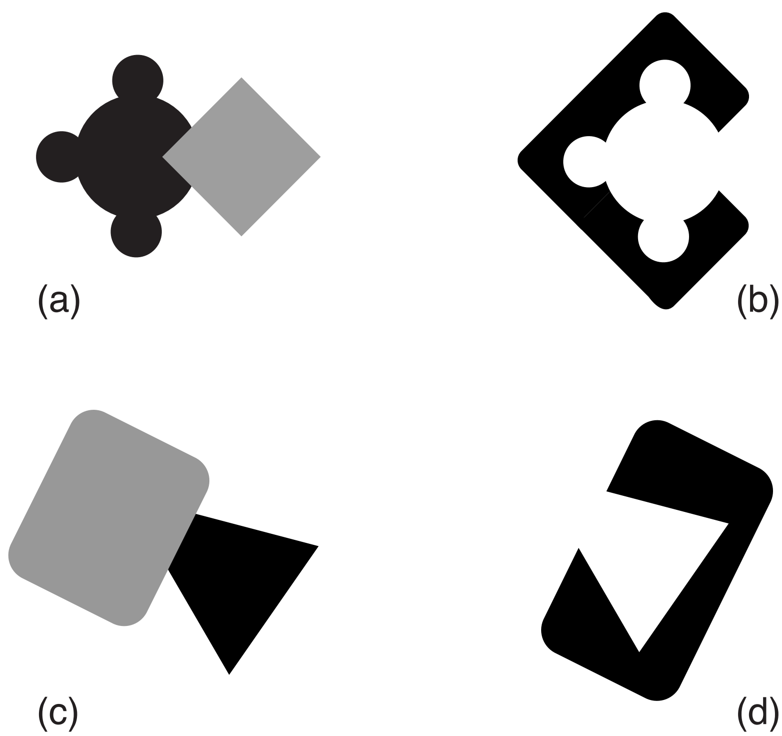


Figure 19. Global completion and the identity hypothesis: (a) Partial occlusion display in which symmetric completion (i.e., interpolation of a fourth rounded lobe behind the occluder) may, theoretically, occur. (b) Illusory contour display in which the central figure has equivalent visible edges to Figure (a); a locally smooth completion, rather than a globally symmetric completion, is seen. (c) Partial occlusion display with non-relatable edges in which interpolation of a triangle has been claimed. (d) Illusory contour display with visible contours equivalent to Figure (c). Observers do not perceive a triangular or other contour completion.

versions of the stimuli over a large number of trials. After a few exposures, a partly occluded, potentially symmetric form might activate directly the stored representation of the symmetric counterpart.

How might we determine the actual locus of the symmetry effect? We have already mentioned one consideration suggesting that symmetry might involve a Type II effect — recognition from partial information. Recall the *identity hypothesis*, the idea that a common contour interpolation process underlies partly occluded and illusory contours. If true, the identity hypothesis sheds light on the nature of the symmetry effect because symmetry-based or global completion phenomena *never* are experienced in illusory figures. As an example, Figure 19b shows an illusory object with physically-defined edges equivalent to those in Figure 19a. The reader may notice that there is no appearance of a fourth lobe in the illusory figure display. In fact, most observers perceive a smooth illusory contour connecting relatable edges. Thus, given the logical arguments and empirical findings supporting the identity hypothesis (Kellman & Shipley, 1991; Kellman, Yin, & Shipley, 1998; Ringach & Shapley, 1996; Shipley & Kellman, 1992a), we would argue that the contour interpolation process underlying both illusory and occluded contour completion is not influenced by symmetry.

If the identity hypothesis is true, then why should global completion occur in occluded but not illusory object displays? The answer may be that the displays trigger the same perceptual processes of contour and surface interpolation, but only occluded figures activate a shape representation from partial information. But why doesn't recognition from partial information occur with illusory figures? In one sense it may. With an illusory figure, an observer certainly could report that the visible lobes "suggest" or "remind them of" the appropriate symmetric figure; however, the observer would not *see* any illusory contours along the figure's boundaries (i.e., the contours have no modal presence). Thus, there exists an obvious discrepancy between what is suggested by partial information and the contours actually interpolated; as a result, the perceptual system rejects the symmetric figure as actually existing in the image. With occluded figures, however, the difference between the representation activated via recognition from partial information and the representation activated through contour interpolation may not be so obvious. By definition, part of an occluded object is hidden from view, and the hidden parts have no sensory presence. Thus, both interpretations — the local representation developed through contour interpolation and the global representation induced by a Type II effect — are possible. Thus, an observer may perceive a partly occluded object as a globally symmetric form, due to a recognition from partial information, whereas this representation is rejected in the case of illusory figures because nothing is hidden.

The idea that Type II effects — recognition from partial information — are responsible for certain "global effects" in contour interpolation may clarify certain issues. First, pointing to an occlusion display similar to the one shown in Figure 19c, Boselie and Wouterlood (1992) argued against Kellman and Shipley's (1991) relatability geometry, presumably because they found the presence of a triangle in the display to be obvious. (The visible edges of the black figure are not relatable, because they violate the 90 degree bending limit for interpolated contours.) However, it is doubtful that contour interpolation occurs in this display, even though the image clearly makes us think of triangles. Consider the illusory contour version, which contains the same visible edges, shown in Figure 19d. No third point of a triangle is visible.

Second, the research literature on the completion of partly occluded objects contains several conflicting reports about global and local processing (e.g., Boselie, 1988, 1994; Sekuler, Palmer, & Flynn, 1994). We suggest that "local" effects derive from actual contour interpolation processes, whereas "global" effects depend on recognition from partial information. Importantly, the only objective data supporting global percepts come from priming studies. Priming occurs at many levels (e.g., Kawaguchi, 1988), from the most basic representation of the stimulus to conceptual interpretations (e.g., a picture of a fork would probably prime the word "knife"). Unfortunately, there have been no attempts to determine the level at which priming occurs for partly occluded objects.

How might we differentiate Type I, II, and III effects experimentally? Type II effects differ from the others in that recognition from partial information may occur without any influence on local contour perception. That is, when shape representations become activated through a "shortcut," there is no need for earlier stages of processing to include specification of precise boundaries in particular locations. When we see the tail rotor of a helicopter protruding from behind a building, we may activate an overall shape representation for "helicopter," but the boundaries of the hidden parts remain poorly specified. By contrast, contour interpolation processes (whether or not influenced by feedback from higher levels), *do* produce well-localized boundaries. Accordingly, it may be possible to distinguish Type II effects from Type I or Type III effects based on the precision of contour representations in partly-occluded or illusory contour displays.

Kellman and colleagues (Kellman, Shipley, & Kim, 1996; Kellman, Temesvary, Palmer, & Shipley, 2000) developed an objective paradigm to measure the precision of boundary localization. In these studies, observers viewed short presentations of partly occluded stimuli during which a small, briefly presented dot was superimposed somewhere on the occluder. On each trial, observers judged whether the probe dot fell inside or outside the occluded object's perceived boundaries. The position of the dot was adjusted on the basis of these responses; two interleaved staircase procedures gave certain threshold points for seeing the dot inside versus outside of the boundary. From these data, Kellman et al. derived estimates of perceived boundary position and precision of localization.

To minimize performance differences based on competing perceptual and recognition processes, Kellman et al. (1996, 2000) provided subjects with explicit strategy instructions. In the *global instruction condition*, subjects were told (with specific examples) that they should see the display as symmetric; in the *local instruction condition*, subjects were told that they should see the display as containing a simple curve connecting the two visible edges. In this manner, Kellman et al. sought to determine our best ability to localize boundaries under a global or local interpretation.

When subjects produced “local” completions, as predicted by relatability, their localization of boundaries was extremely precise (i.e., inside and outside thresholds differed very little). This finding held for both straight (collinear) and curved interpolations in a large range of displays. In contrast, “global” completions resulted in boundary localizations that were nearly an order of magnitude less precise for all displays; moreover, the estimated positions of the contour usually differed markedly from the predicted positions based on symmetry. These results held even when the predicted positions of the relatable (local) and symmetry-predicted (global) contours were equidistant from the nearest visible contours.

For a “triangle” display such as Figure 19d, Kellman et al. (1996) found that observers exhibited a large uncertainty about the location of the triangle’s occluded vertex. Most subjects' best estimate of its position differed from the actual position (determined by extending the visible contours as straight lines) by at least 15% of the height of the triangle.

In sum, global influences like symmetry and familiarity apparently do not produce local contour interpolation. Contrary to Type III models, symmetry does not appear to work within the interpolation process to create precisely localized contours. Contrary to Type I models, although shape representations become active, they do not establish local contours via feedback to interpolation mechanisms. Thus, it appears that effects suggesting global perceptual completion may actually depend on recognition from partial information — Type II effects. Operating in parallel with the object perception processes we have outlined, partial information may activate later cognitive representations directly. Both processing routes play important roles in human object recognition; distinguishing them, however, may clarify the theoretical situation considerably.

<u>Interpolation in Three-Dimensional Object Perception</u>. Until now, most of our discussion of object perception has emphasized information in static, 2-D luminance images. This focus mirrors the research literature: although geometric studies using more complex stimuli have started to emerge, all existing neural-style models possess tight constraints on the kinds of inputs they can accept. Detailed models of the processes that extract information from such 2-D, static images represent an important advance, but even with further elaboration and new data, they cannot be complete.

Object perception is at least a four-dimensional process. It involves information derived not only from spatial variations across a frontoparallel plane, but from depth differences and changes over time given by motion. In fact, more modern views of perception suggest that the basic structure of the visual system includes mechanisms for extracting depth and spatiotemporal relationships (Gibson, 1966, 1979; Johansson, 1970; Marr, 1982). These accounts emphasize that perceptual systems adapted to serve mobile organisms in a 3-D world.

At a few junctures, we have already noted some relevant details. For example, perceived edges may depend on motion and stereoscopic depth discontinuities, as well as luminance contrast. Here we comment more generally on the role of depth and motion in object perception.

Consider the problem of perceiving object unity and boundaries in the 3-D world. Acquiring accurate representations of contour and surface orientations in three dimensions would seem to be of high priority for comprehension of and action within the environment. In terms of the processes of visual segmentation and grouping, it would seem that sensitivity to 3-D relationships among visible parts of contours and surfaces would be important in discovering the connected objects that exist in the world. Research suggests that contour and surface interpolation processes do, indeed, take edge and surface orientations in 3-D space as their inputs and produce interpolated contours and surfaces that extend through depth (Kellman, Machado, Shipley, & Li, 1996; Kellman, Yin, Shipley, Machado, & Li, 2001). Figures 20a and 20c illustrate 3-D interpolations, where the contour positions depend on stereoscopic depth information. Observers tend to perceive smooth contours and surfaces connecting the top and bottom white tabs across the intervening gaps. Figures 20b and 20d illustrate that misalignment in depth *disrupts* the interpolation processes. Although the monocular images fall within the tolerable misalignment for 2-D interpolation, the top and bottom white surface pieces do not appear to connect when the display is seen in depth.

Kellman et al. (1996) proposed that 3-D interpolation is governed by an elaboration of the notion of contour relatability to three dimensions. Informally speaking, two contours extending through depth are relatable if, in 3-D space, there is a smooth, monotonic connection between them. As with the original construct, two contours will be connected by the boundary interpolation process if and only if they are relatable.

For a more intuitive idea of 3-D relatability, the white surface patches given in the stereoscopic displays are shown below in side view (Figure 20). When the surfaces meet the 3-D relatability criteria, the contours in the side view also are relatable in two dimensions. Thus, the examples in Figures 20a and 20c meet the conditions of 3-D relatability, whereas Figures 20b and 20d contain misalignments in depth that disrupt relatability.

To study 3-D interpolation experimentally, Kellman et al. (1996) developed an objective task using the kinds of stimuli displayed in Figure 20. On each trial, observers viewed two surface patches, presented stereoscopically, that could be classified as lying in parallel planes (including the case of coplanarity) or intersecting planes. Observers made a speeded classification judgment by pressing one key for “intersecting” (as in Figures 20a and 20b), and another key for “parallel” (as in Figures 20c and 20d); note that the response did not depend on impressions of interpolation or the notion of 3-D relatability. Kellman et al. predicted that: (1) analogous to certain tasks using 2-D shapes, perception of a unified object would facilitate classification performance; and (2) perceived unity, as indexed by superior speed and/or accuracy, would depend on the relatability criteria.

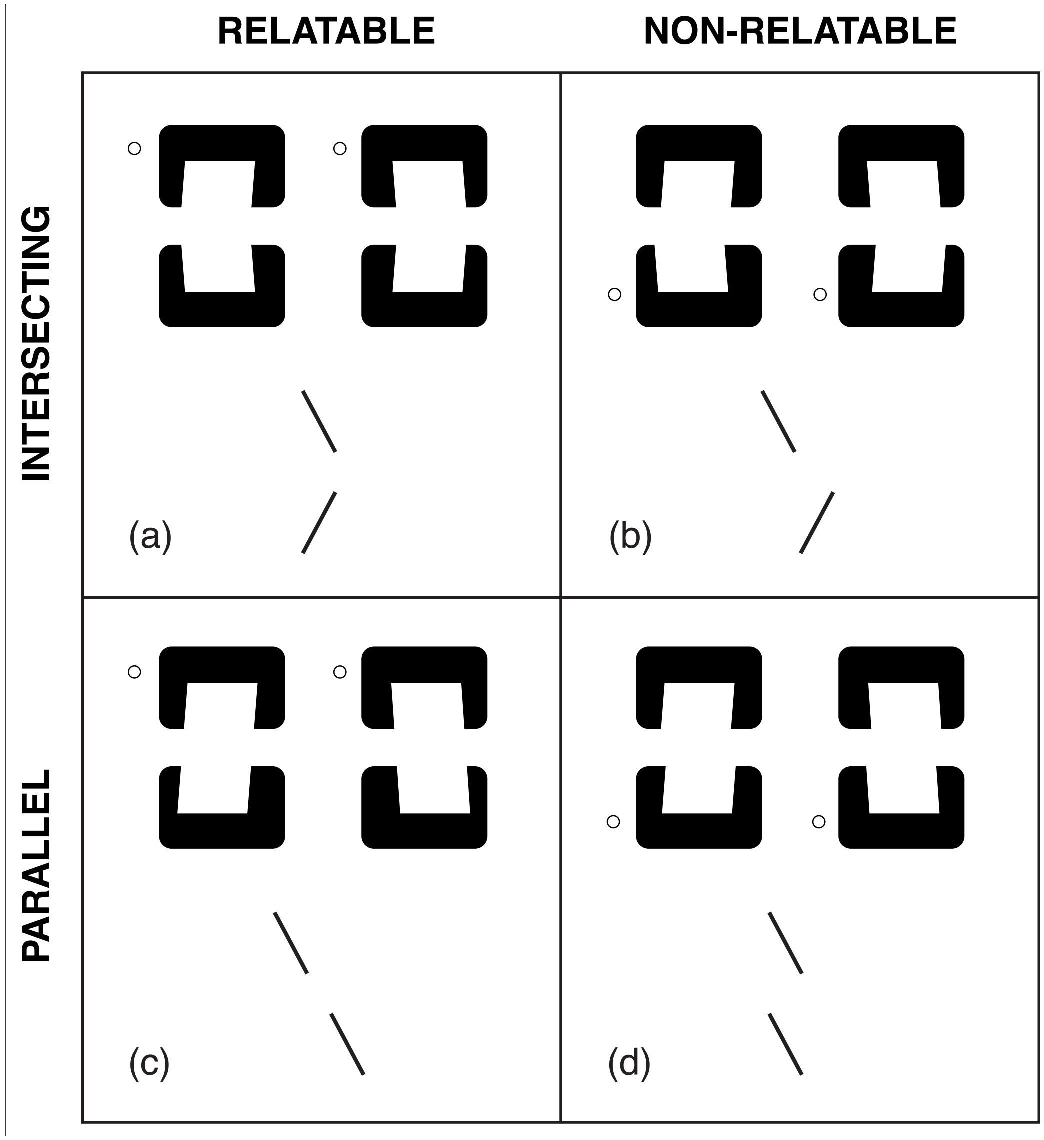


Figure 20. Displays used by Kellman et al. (1996) to test 3-D interpolation: (a) relatable, intersecting surfaces; (b) non-relatable, intersecting surfaces; (c) relatable, parallel (coplanar) surfaces; (d) non-relatable, parallel surfaces. To experience the stereoscopic effect, cross fuse the two images in each pair, using the small circles to focus. A side view of the two white surfaces appears below each stereoscopic pair.

The results of this study, which included a number of control groups to ensure that the results truly depended on depth relationships and interpolation, will be reported in detail elsewhere (Kellman et al., 2001). To summarize, all of the results supported the predictions: observers classified relatable displays more accurately and rapidly than non-relatable displays. These data suggest that relatability describes the conditions necessary for contour interpolation in depth.

The study of three-dimensional interpolation processes using objective tasks is in its infancy. Nonetheless, the results already indicate that models of object perception will need to be broadened to accept as inputs the 3-D positions and orientations of edges, and to produce

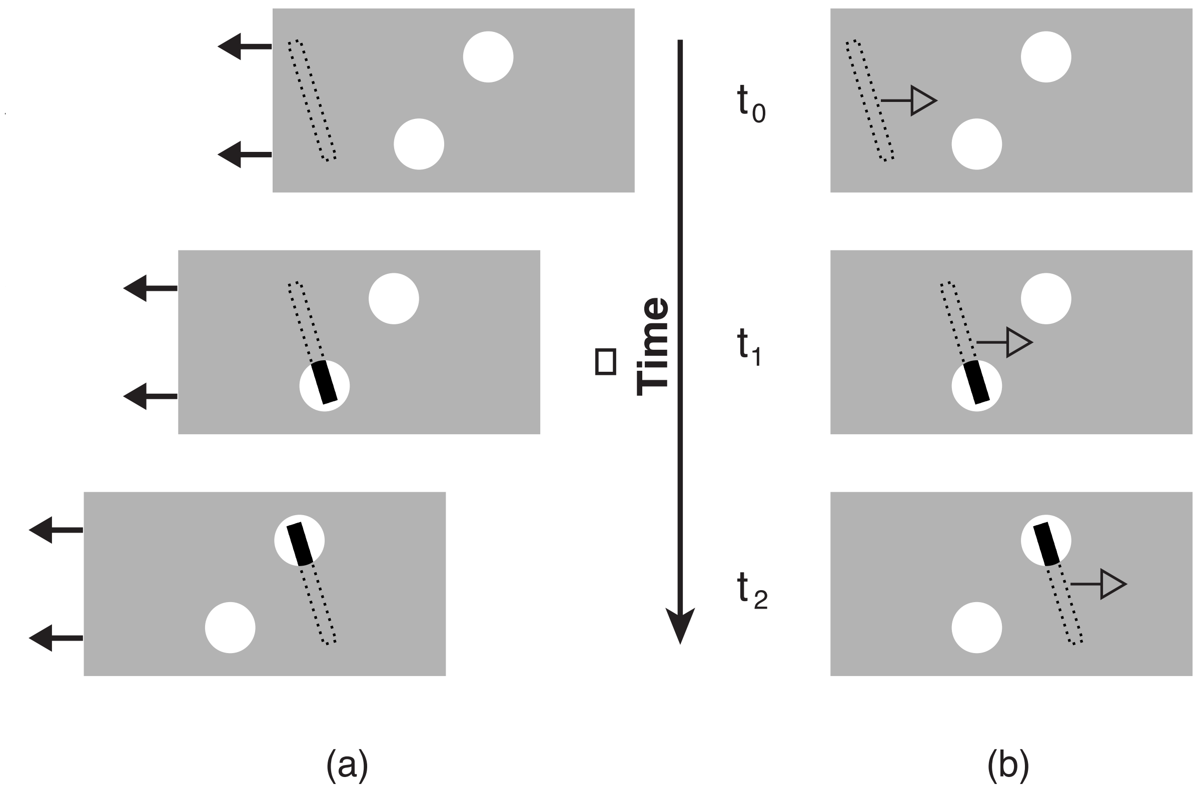


Figure 21. Illustrations of spatiotemporal relatability. (a) The moving occluder reveals relatable parts of the rod sequentially in time (t_1 and t_2). Perceptual connection of parts requires that the initially visible part persists over time in some way. (b) Parts of the moving rod become visible through apertures sequentially in time. Perceptual connection of the parts requires not only persistence of the initially visible part but positional updating based on velocity information. After Palmer, Kellman, and Shipley (1997).

hypothesis (Figure 21a) 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 Figure 21a, 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 relatability computation can be performed to integrate the currently visible part with the part encoded earlier.

In Figure 21b, 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.

Ongoing investigations suggest that this notion of spatiotemporal relatability, based on the persistence and spatial updating hypotheses, may account for a number of phenomena in which observers accurately perceive moving objects that are exposed piecemeal and only partially from behind apertures (Palmer et al., 1997, 2000). Whether or not the current notion of spatiotemporal relatability proves to be an adequate account of dynamic object perception, it is clear that both geometric and neural-style models must be broadened to accept inputs over time, and also to assemble such sequentially available fragments into meaningful units and forms.

CONCLUSION

Research on object perception is a multifaceted enterprise. In this chapter, we have attempted to set out some of the information processing tasks that must be accomplished in the visual perception of objects, as well as our current state of knowledge regarding the underlying processing. One of the clearest themes in our discussion may be that multiple levels of analysis must be undertaken to understand the information, computations, and neural machinery involved in object perception. In particular, geometric models describe the tasks, information, and stimulus relationships necessary to accomplish visual segmentation and grouping, while neural-style models address how the processing might actually be carried out in neural circuitry. These two modeling efforts serve complementary functions, and neither can be fully appreciated without a thorough examination of the other. Although our knowledge about the different component tasks varies widely, today we are closer to a coherent view of object perception than ever before. Our understanding may be expected to advance even further as geometric, process, and neural models co-evolve.

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 PJK. We thank Patrick Garrigan and Evan Palmer for helpful discussions, Tim Shipley for useful comments on an earlier draft, and Stephanie Lau 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

Banton, T., & Levi, D. M. (1992). The perceived strength of illusory contours. *Perception & 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: John Wiley & Sons.
Boselie, F. (1988). Local versus global minima in visual pattern completion. *Perception & Psychophysics, 43,* 431-445.
Boselie, F. (1994). Local and global factors in visual occlusion. *Perception, 23,* 517-528.
Boselie, F., & Wouterlood, D. (1992). A critical discussion of Kellman and Shipley's (1991) theory of occlusion phenomena. *Psychological Research*, *54,* 278-285.
Bruno, N., & Bertamini, M. (1990). Identifying contours from occlusion events. *Perception & Psychophysics, 48,* 331-342.
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.
Dresp, B., & Bonnet, C. (1995). Subthreshold summation with illusory contours. *Vision Research, 35,* 1071-1078.
Dresp, B., & Grossberg, S. (1997). Contour integration across polarities and spatial gaps: From local contrast filtering to global grouping. *Vision Research, 37,* 913-924.
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.
Freeman, W. T., & Adelson, E. H. (1989). Steerable filters. *Image Understanding and Machine Vision, 14,* 114-117.
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 & Psychophysics, 5,* 113-116.
Gilbert, C. D. (1994). Circuitry, architecture and functional dynamics of visual cortex. In G. R. Rock & J. A. Goode (Eds.), *Higher-order processing in the visual system* (pp. 35-62). Chichester, England: John Wiley & Sons.
Greene, H. H., & Brown, J. M. (1997). Spatial interactions with real and gap-induced illusory lines in Vernier acuity. *Vision Research, 37,* 597-604.
Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception & 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 & Visual Science, 41*(4), S439.

Guttman, S. E., & Kellman, P. J. (2001). *Contour interpolation: Necessary but not sufficient for the perception of interpolated contours*. Manuscript in preparation.

Guttman, S. E., & Sekuler, A. B. (2001). *Visual completion of partly occluded objects: An exploration of spatial and temporal limits*. Manuscript submitted for publication.

He, Z. J., & Nakayama, K. (1994). Perceived surface shape not features determines correspondence strength in apparent motion. *Vision Research, 34,* 2125-2135.

He, Z. J., & Ooi, T. L. (1998). Illusory-contour formation affected by luminance contrast polarity. *Perception, 27,* 313-335.

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 & Company.

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

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.

Kapadia, M. K., Ito, M., Gilbert, C. D., & Westheimer, G. (1995). Improvement in visual sensitivity by changes in local context: Parallel studies in human observers and in V1 of alert monkeys. *Neuron, 15,* 843-856.

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., & Cohen, M. H. (1984). Kinetic subjective contours. *Perception & Psychophysics, 35,* 237-244.

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 & 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 & Visual Science, 41*(4), S741.

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.

Kellman, P. J., Yin, C., Shipley, T. F., Machado, L. J., & Li, C. C. (2001). Visual interpolation of contours and surfaces in depth: A 3-D extension of spatial relatability. Manuscript in preparation.

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

Kovács, I., & Julesz, B. (1993). A closed curve is much more than an incomplete one: Effect of closure in figure-ground segmentation. *Proceedings of the National Academy of Sciences of the United States of America, 90,* 7495-7497.

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

Liu, Z., Jacobs, D. W., & Basri, R. (1999). The role of convexity in perceptual completion: Beyond good continuation. *Vision Research, 39,* 4244-4257.

Liu, Z., Kersten, D., & Knill, D. C. (1999). Dissociating stimulus information from internal representation — a case study in object recognition. *Vision Research, 39,* 603-612.

Livingstone, M. S. (1996). Oscillatory firing and interneuronal correlations in squirrel monkey striate cortex. *Journal of Neurophysiology, 75,* 2467-2485.

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

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.

Moulden, B. (1994). Collator units: Second-stage orientational filters. In G. R. Rock & J. A. Goode (Eds.), *Higher-order processing in the visual system* (pp. 170-184). Chichester, England: John Wiley & Sons.

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 & 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 & Visual Science*, *41*(4), S439.
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.
Peterhans, E., von der Heydt, R., & Baumgartner, G. (1986). Neuronal responses to illusory contour stimuli reveal stages of visual cortical processing. In J. D. Pettigrew, K. J. Sanderson, & W. R. Levick (Eds.), *Visual Neuroscience* (pp. 343-351). Cambridge, UK: Cambridge University Press.
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.
Pettet, M. W., McKee, S. P., & Grzywacz, N. M. (1998). Constraints on long range interactions mediating contour detection. *Vision Research, 38,* 865-879.
Polat, U., & Sagi, D. (1993). Lateral interactions between spatial channels: Suppression and facilitation revealed by lateral masking experiments. *Vision Research, 33,* 993-999.
Polat, U., & Sagi, D. (1994). The architecture of perceptual spatial interactions. *Vision Research, 34,* 73-78.
Prenter, P. M. (1989). *Splines and variational methods*. New York: John Wiley & Sons.
Ringach, D. L., & Shapley, R. (1996). Spatial and temporal properties of illusory contours and amodal boundary completion. *Vision Research, 36,* 3037-3050.
Robinson, J. O. (1972). *The psychology of visual illusion*. London, UK: Hutchinson.
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.

Shapley, R., & Gordon, J. (1987). The existence of interpolated illusory contours depends on contrast and spatial separation. In S. Petry & G. E. Meyer (Eds.), *The perception of illusory contours* (pp. 109-115). New York: Springer-Verlag.

Shipley, T. F., & Kellman, P. J. (1990). The role of discontinuities in the perception of subjective figures. *Perception & 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 & 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 & Psychophysics, 61,* 943-951.

Takeichi, H., Nakazawa, H., Murakami, I., & Shimojo, S. (1995). The theory of the curvature-constraint line for amodal completion. *Perception, 24,* 373-389.

Tse, P. U. (1999a). Complete mergeability and amodal completion. *Acta Psychologica*, *102,* 165-201.

Tse, P. U. (1999b). Volume completion. *Cognitive Psychology*, *39,* 37-68.

Tse, P. U., & Albert, M. K. (1998). Amodal completion in the absence of image tangent discontinuities. *Perception, 27,* 455-464.

Ts'o, D. Y., Gilbert, C. D., & Wiesel, T. N. (1986). Relationships between horizontal interactions and functional architecture in cat striate cortex as revealed by cross-correlation analysis. *Journal of Neuroscience, 6,* 1160-1170.

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.

von der Heydt, R., & Peterhans, E. (1989). Mechanisms of contour perception in monkey visual cortex. I. Lines of pattern discontinuity. *Journal of Neuroscience, 9,* 1731-1748.

Wang, D., & Terman, D. (1997). Image segmentation based on oscillatory correlation. *Neural Computation, 9,* 805-836.

Watt, R. J. (1994). A computational examination of image segmentation and the initial stages of human vision. *Perception, 23,* 383-398.

Watt, R. J., & Morgan, M. J. (1985). A theory of the primitive spatial code in human vision. *Vision Research, 25,* 1661-1674.

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)

Witkin, A. P. (1983). Scale-space filtering. *Proceedings of the International Joint Conference on Artificial Intelligence, 8,* 1019-1022.

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.