36

Segmentation, Grouping, and Shape

Some Hochbergian Questions

Philip J. Kellman and Patrick Garrigan

It is a privilege to contribute a chapter to this volume honoring Julian Hochberg. Such a volume is deserved, and overdue, simply on the basis of Hochberg's contributions as one of the great perception scientists of the 20th century. What many contributors to this volume have in common, however, goes beyond that. We have benefited from Hochberg's gifts as a teacher and mentor. In the process that allows the questions, tools, and insights in a discipline to be handed off from one generation to another, he has had a special and profound influence on his students, not only from his one-on-one interactions with them but also as they have been able to observe his approach to scholarship and science.

We feel especially fortunate to be included here. Although neither author of this chapter was officially his doctoral student, Kellman was fortunate during his third year of graduate school to be part of a year-long seminar with Hochberg. Some enlightened faculty in the Psychology Department at the University of Pennsylvania at that time realized that, although their department was strong in research on basic visual mechanisms, it was not strong in perception, a topic of increasing interest among its students. Asking "Why not the best?" they arranged for Hochberg to travel to Penn every second week to teach a graduate perception seminar in 1978. This amazing seminar had an enormous impact on Kellman and on a whole group of students who had not initially come to graduate school to study perception. These included Dan Reisberg, David Smith, and Denise Varner in that seminar and others who would benefit later from the climate in the department that Hochberg's seminar helped to create, such as Tim Shipley and Felice Bedford. The idea that one seminar could actually solidify an enduring focus of interest in a department sounds inflated, but that's pretty much what happened.

With regard to key issues in perceptual science, it is striking how well Julian Hochberg's influence endures. In this chapter, we note a few of these issues and describe how they loom large in our current concerns. The chapter is brief and in parts speculative; what we are most certain will persist in the long run are not elements of our models but the basic Hochbergian questions that motivate them.

Some Hochbergian Problems

In current work, we are attempting to connect what is known about grouping and segmentation processes with issues of shape perception and representation. Three problems define much of this work, and all involve issues that Julian Hochberg worked on and in some cases helped to define. We briefly describe these three areas below.

The Problem of Shape Perception and Representation

What is shape and how do we represent it? Hochberg and his colleagues made significant contributions on these issues, such as attempts to quantify figural goodness (Hochberg, Gleitman, & MacBride, 1948; Hochberg & McAlister, 1953). He introduced these problems to a generation of investigators through his book on perception (Hochberg, 1964) and his landmark chapters, which occupy 155 pages in the 1971 edition of *Woodworth and Schlosberg's Experimental Psychology.*

Problems of Abstraction in Perception

In the classic article "In the Mind's Eye," Hochberg (1968) made the case that perception advances beyond local activations by constructing "schematic maps" from successive fixations and that these maps are abstract in that they contain selective information about shape, but not every detail of the local activations that produce them.

Connecting Early Visual Encodings With Higher Visual Representations

How do we get to schematic maps from local activations? Early cortical units respond to local areas of oriented contrast. If we think of the response to a visual pattern as an ensemble of activations like these, how do we get to the perception of contours, unity, objects, and shape?

Contemporary Issues of Shape, Schematic Maps, and the Relation of Early and Middle Vision

Although there has been progress, all of these Hochbergian questions remain important in current research.

Even within visual segmentation and grouping processes, shape issues arise. For example, interpolation processes that connect visible contours across gaps, as in occluded and illusory contour formation, appear to contribute definite contour shapes in regions that are not locally specified by stimulus

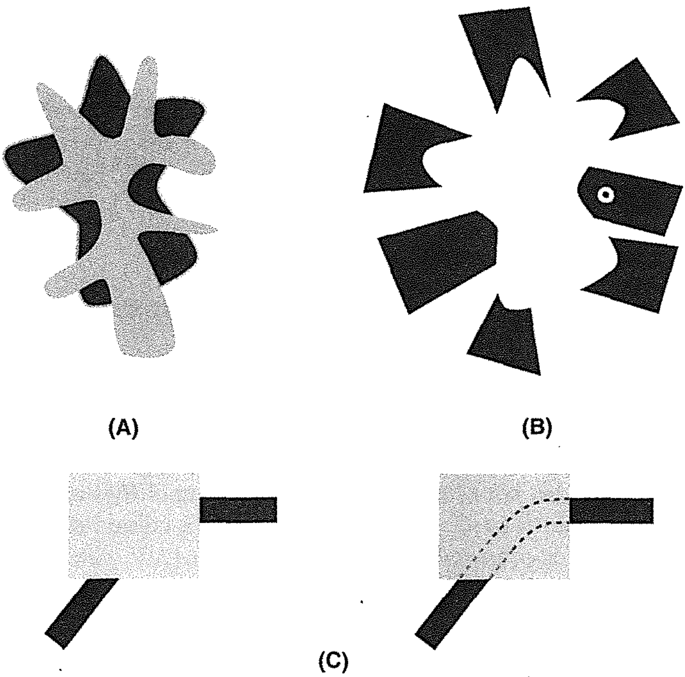

Figure 36.1. Shape and interpolation. Occluded object (A) and illusory object (B) displays with the same set of physically specified contours and gaps. Interpolation leads to similar perceived shapes in both cases. C. The display on the right indicates the perceived shape that may be produced by interpolation in the occlusion display on the left. Note that curved edges may appear in the percept despite their absence in the display presented.

information. Figure 36.1 shows illusory and occluded figures that share the same extents and locations of physically specified contours and the same gaps. Phenomenologically, they also appear to share the same completed shapes, an observation that fits with evidence and arguments that they share a common underlying interpolation process (Shipley & Kellman, 1992; Ringach & Shapley, 1996; Yin, Kellman, & Shipley, 1998; Kellman, Garrigan, & Shipley, 2005). Figure 36.1C illustrates more locally how interpolation contributes to shape of unspecified regions.

Issues of shape and abstraction must be resolved if we are to link middle vision—including contour, surface, and object perception—to known facts about early visual processing, such as the kinds of neural units located in early cortical areas (Heitger et al., 1992, 1998; Hubel & Wiesel, 1968). The typical outputs of neural-style models of these processes (e.g., Grossberg & Mingolla, 1985; Grossberg, Mingolla, & Ross, 1997; Heitger et al., 1998; Yen & Finkel, 1998). are *images*, showing the locations of activation resulting from grouping or interpolation operators (but see Yen & Finkel, 1998). It is easy to forget when viewing these outputs that the models themselves do not "know" anything about what things go together. Issues of identifying connected contours, segmenting objects, etc., remain. This point is not meant as a criticism of these models, which address a number of important issues, but only to make clear that certain crucial problems are not addressed.

The point is relevant, not only for interpolated contours, but for "real" ones as well. A viewed contour produces a large ensemble of responses in early cortical areas, but how do we arrive at a contour "token," a higher-level unit that has properties such as shape and can receive labels such as boundary assignment? One effort to assemble local activations into contour tokens is the model of Yen and Finkel (1998), which uses the synchrony of firing of active local edge detectors to combine them into a unit. Although attempting to address the next step toward object perception, the model says little about how some attribute of this token, in particular its shape, might be extracted.

This issue is crucial, because higher-level shape representations are both more and less than the early neural activations that contribute to them. They are *more* because an abstract representation of shape can be used to recognize, compare, equate, or distinguish other stimuli that differ in many particulars. They are *less* because the higher level representation necessarily omits much of the details of particular local activations.

One might say that we are concerned with some unsolved problems of schematic maps, specifically how to extract and encode contour shapes and how to get from an ensemble of local neural responses to shape tokens that allow object shape to be represented. This effort is only beginning, so we offer only sketches of the problems along with some possible elements of their solutions.

Some Guiding Principles

A few basic ideas about shape representations guide our work. Here, we describe them briefly.

Unity in Shape Representation

We assume that the outputs of perception include higher-level tokens that are larger and more abstract than the responses of orientation-sensitive units in early cortical areas. Contour tokens would appear necessary to explain ordinary observations, such as the fact that contour shapes can be discriminated and matched, and that in reversible figure-ground displays with smooth contours (as in Figure 36.2), boundary assignment (indicating which side of a contour is figure) ordinarily changes for the whole contour at once.

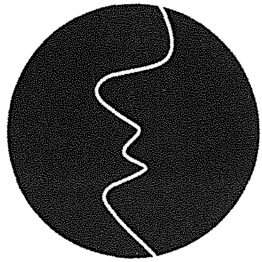

Figure 36.2. Contour tokens and boundary assignment. Although boundary assignment (figure-ground organization) may change as the figure is viewed, it always changes as a unit. This regularity suggests that boundary assignment operates on unitary contour tokens.

Our understanding of what these tokens are is poor. Such higher-level representations may characterize shapes of whole objects, but they also probably include descriptive units that are not so high-level, such as parts (Hoffman & Richards, 1984) and contours.

Limited Resolution in Shape Representation

Another principle guiding our work is that shape representations do not have infinitely fine grain. This is likely to be a principled difference between shape representations in biological systems and machines. In a computer vision application, if a space curve is specified by an equation, the system is able to compute the slope or any other derivative of the curve at any point on the curve. It is unlikely that human representations are equally detailed at all points, and we probably have little or no sensitivity to derivatives higher than the first (slope) (Kellman, Garrigan, Kalar, & Shipley, 2003). In short, human shape representation is far more limited, but also geared to certain tasks. An equation for a curve would not naturally indicate salient parts or their contribution to the perceived complexity of a contour. It also would not easily capture similarities of contour shapes.

More generally, we suspect that the ordinary processing of shape contours utilizes representations that usually simplify the actual contour shape. Doing so not only produces a more economical record; it also makes feasible the detection of shapes as similar or different across changes in position, orientation, scale, and the elements composing the shape. Figure 36.3 illustrates some of these points.

Recently, we have carried out experiments in which subjects make same/different judgments about successively presented smooth, curved contours. Results suggest both that complexity matters a great deal and that human shape representations are simplified relative to real shapes. Nevertheless, across a range of conditions, human judgments are quite good. After considering how a shape representation system might connect with object segmentation and interpolation processes, we suggest how economy in shape representation may be achieved, without much loss of the ability to represent and distinguish different shapes.

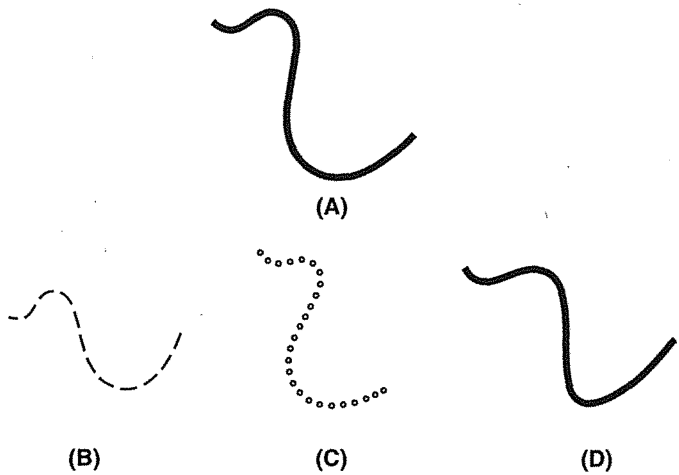

Figure 36.3. Abstract shape. The contour at the top (A) has a distinct shape, a property independent of its position, orientation, or local characteristics. Objects B and C are composed of different local elements and have different sizes and orientations, yet they are seen as having the same overall shape as in A. Object D is most similar to A in terms of overall size, orientation, and type of line, yet it is clearly the only contour in the figure with a different shape.

Constraining Interpolated Contour Shape

One specific issue that has prodded our thinking on shape representations is the shapes of interpolated contours. Evidence suggests that the visual system creates well-defined contour links across gaps in the perception of occluded and illusory objects. Data on the exact shapes of interpolated contours are sparse and conflicting. The premise above regarding the spatial grain of shape representations may provide some theoretical guidance about interpolated contours. Some models propose outputs that are unconstrained by consideration of the complexity of the interpolated contour's shape. For example, Fantoni and Gerbino (2003) suggest an interpolation scheme that progresses from two endpoints by combining two inputs (whose weights change at each point along the path), a straight-line connection between the endpoints of the visible inducing edges and their collinear extensions. This scheme produces curved interpolations that often have changing curvature at every point. Recently, Fulvio, Singh & Maloney (in press) suggested that interpolated shapes might best be described by quintic (5th order) polynomials. Even if some neural basis in terms of interactions of oriented units were able to achieve such contour connections, they would be unlikely to survive into our representations of shape. We believe that shape representations are unlikely to represent different curvatures at every point, even when the stimulus contains them. Two lines of work have suggested more constrained schemes. Analysis by J. Edward Skeath (see appendix II of Kellman & Shipley, 1991) suggested a simple shape for interpolated contours that is naturally compatible with the geometry of contour relatability, a description of the relations of real contours that support interpolation. Interpolated contours can always be composed of a constant curvature segment and a zero-curvature (straight) segment. Earlier, Ullman (1976) proposed that illusory contours can always be composed of two constant curvature segments. The Skeath/Kellman/Shipley (1991) proposal has the virtue of providing a unique solution for the path of interpolation without an additional step in which a unique solution is chosen from a set of possible ones, as in the model proposed by Ullman. However, both proposals are compatible with economy in subsequent coding.

Aspects of a Theory Connecting Early Vision to Shape Representations

Arclets

We are developing a scheme that uses the simplicity of the circle as the link between low- and higher-level vision. We propose that neural circuits exist that combine small groups of oriented units that are linked by constant turning angles, e.g., they encode constant curvature segments (including zero curvature) of shapes. We call these *arclets* (see Figure 36.4). These are likely to include both 2D and 3D groups of oriented units, as recent work suggests that object formation is a 3D process, based on 3D positions and orientations of oriented units (Kellman, Garrigan, & Shipley, 2005).

In their application to interpolation, activation initiated by real contours spreads along restricted paths in a network of oriented units; these paths consist of arclets. This restriction, combined with one simple, additional constraint, provides a unique path of interpolation connecting any relatable edges. In their application to shape coding, the activations of arclets—units that are activated by signals in chains of several oriented units—allow a natural means of handing off the information from real and interpolated contour positions to higher-level shape representations.

A Shape Representation Scheme

We conjecture that at the level at which middle- and higher-level shape representations receive information from early visual activations (clusters of oriented units), shape representations are made up of chunks of constant curvature. The neural basis for extracting these chunks are detectors that sense patterns of activation in a layer of oriented units, such as those found in the early cortical visual areas. A given arclet is activated if a chain of oriented units forming a collinear or cocircular path is simultaneously activated. Different arclets code different curvatures. Activation of a single

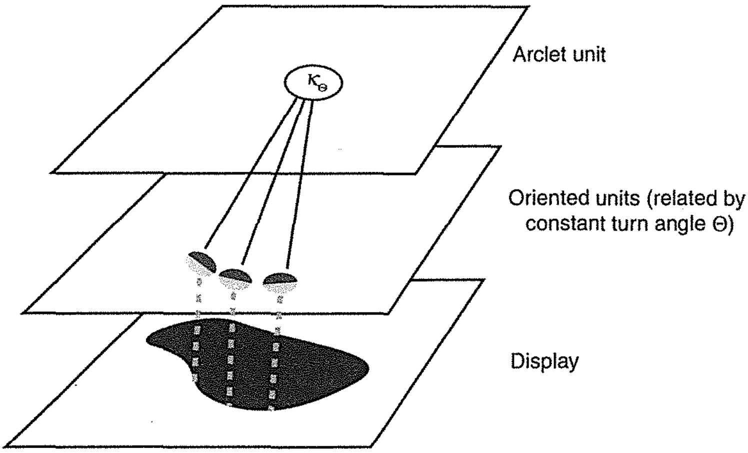

Figure 36.4. From local activations to abstract shape: proposed arclet circuits. Arclets may represent the first step toward a nonlocal description of contour shape. Oriented units simultaneously activated by a stimulus are grouped according to their relative positions and orientations. If the geometry of the contour is within the tolerance of a higher-order arclet grouping unit, a simplified, abstract representation for this part of the contour emerges from the complex array of local activations. For interpolated contours, activation spreads from the ends of visible edges only through collinear or cocircular arclet paths, leading to a unique interpolation path for relatable contours.

arclet indicates the presence of that curvature in a certain position, and activation of a series of adjacent arclets of the same curvature value signals an extended contour region having that curvature value. The encoding of a constant curvature segment extends along a contour until a transition zone, at which arclets of that curvature no longer exceed a certain threshold of accurately matching the contour (or are less well activated than some arclet having a different curvature value). A shape representation consists of a set of constant curvature values (scaled to achieve size invariance; see below) characterizing segments along a contour, along with some marking of transition zones between constant curvature segments.

One implication of this scheme, which we are currently working out and testing, is that shapes of curved contours that have continuously varying curvature must somehow be fit into the representation we are proposing. Obviously, in the limit, if a very large number of constant curvature segments can be used to represent a contour, there would be little difference between our proposal and one that invokes representations of arbitrary space curves. Therefore, one difficult task we are pursuing is to discover the representational scheme that determines how a complicated contour is parsed into segments of nearly constant curvature, as well as how these several pieces are knit together in a representation. Two early indications make us optimistic about such a scheme. First, in simple closed forms of smooth but arbitrary shape, there is an obvious complementary relationship between certain features of contours, such as the concave minima that determine the decomposition of objects into parts (e.g., Hoffman & Richards, 1984; Singh & Hoffman, 1997), and the contour segments between part boundaries. Full representations of such shapes may well be combinations of part boundaries along with constant and zero curvature segments between them. Second, although we have not devised a complete representational scheme, we have obtained preliminary data indicating that shapes with constant curvature chunks are processed more efficiently and support better discrimination performance than similar shapes with nonconstant curvatures.

Shape Invariance

Hochberg (1968) offered arguments as to why schematic maps are abstract rather than literal copies of the original input. Similar arguments not only apply to shape representation but mark its most important characteristics. As the Gestalt psychologists emphasized (and as we illustrated in Figure 36.3), two things can have the same form yet differ in the local elements comprising them. One mandate for a successful shape representation scheme is that it should make explicit what Figures 36.3A, B, and C have in common, as well as how they differ from Figure 36.3D.

This ideal has been most closely accomplished in models of high-level object recognition (e.g., Hummel & Biederman, 1992). Most models of this sort, however, focus on so-called basic-level object recognition, which refers to shape categories at the basic level of conceptual hierarchies. (For example, chairs, airplanes, and cups are basic level, whereas my favorite easy chair or the concept of vehicles are subordinate or superordinate level concepts, respectively.) In this kind of scheme, both a single-engine Cessna and a Boeing 747 will activate the final representational category "airplanes." Likewise, other research (e.g., Attneave, 1954) suggests that specific curved edges may be replaced by straight ones and still allow recognition of a shape category ("Attneave's cat").[1] Yet the invariances useful for detailed object naming are too abstract to account for human perception of the shapes of contours and objects, such as what is the same or different among the displays in Figure 36.3.

At this level of shape representation, there are important and interesting problems. One is that standard mathematical notions of curvature do not capture shape invariance. A large circle and a small circle obviously have the same shape, but they have very different curvatures (where curvature is given by the change in contour orientation per unit arc length). Use of relative curvatures or normalization by some overall size measurement (e.g., length along longest axis for a closed shape, or normalization by chord length for a curved segment of a contour) is a standard operation in computer vision (Costa & Cesar, 2000) and one that might be used to equate a shape characteristic at differing sizes in human vision (cf. Singh & Hoffman, 1997).

Arclets may offer a means of achieving scale invariance in a more natural way. Because orientation-sensitive units in early visual areas exist across a range of spatial scales, arclets would similarly span this range. There is an interesting invariant for arclets related by the same turn angle, but made of different size elements. As long as all elements within each arclet are of equal size, all arclets based on the same turn angle between oriented elements and having the same number of elements represent the same scale-invariant shape. That is, shape pieces that receive the same encoding in terms of arclet turn angle and element number differ only by a scale factor (see Figure 36.5). Thus, activating an arclet at any scale could signal two unique values (turn angle and the number of elements in the arclet) that specify scale-invariant shape for that part of the contour. Two circles of different sizes, for

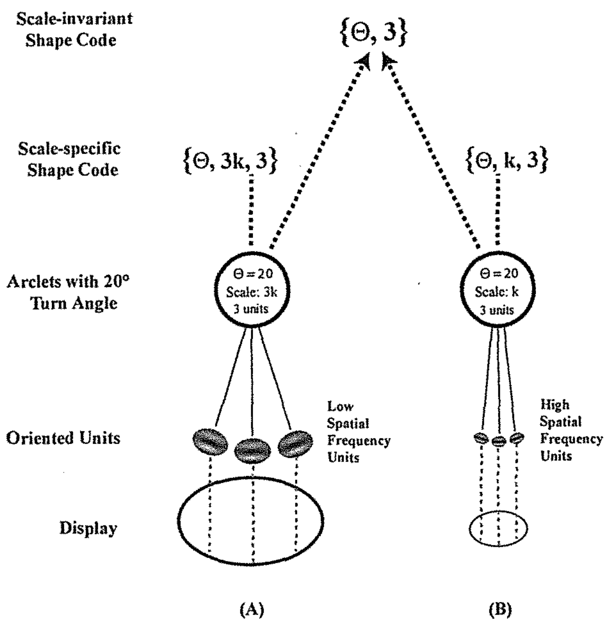

Figure 36.5. Representations consisting of sets of arclets and their relations along a contour are inherently scale-invariant. A. Large ellipse. A contour fragment is detected by an arclet made of coarse orientation-sensitive units related to each other by a 20° turn angle. B. Small version of ellipse in A. The corresponding contour fragment is detected by an arclet made of smaller orientation-sensitive units related to each other by 20° turn angles. Scale-specific representations of shape can derive from three characteristics of the best-activated arclet: turn angle, scale, and number of oriented units. Arclets made from oriented units of a given scale (spatial frequency) are scalar transforms of arclets at other scales having the same turn angle. This property makes available a scale-invariant representation of the shape of the contour fragment based on two parameters of an activated arclet: turn angle and number of oriented units activated.

example, will have contours that best match arclets at different scales, but both arclets will have the same number of elements and the same relative turn angle between them. Some constraints will be necessary in contour encoding for selecting the proper scale of arclet. For example, a slow-turning, small-scale arclet and a faster-turning, large-scale arclet could both match a large curve. It appears, however, that simple constraints, such as selecting the matching arclet of highest turn angle as the best descriptor, would resolve this ambiguity. Similarity of forms and parts across orientations may be straightforwardly given in this scheme. If curvature chunks are encoded around the border of a shape beginning at some recognizable feature, then representations will be orientation invariant. These relatively simple ingredients would allow two contours or closed shapes having the same form but differing in orientation and size to produce the invariant representations needed to support shape perception and classification.

The efforts to understand shape perception and abstraction in perceptual representations and the efforts to connect early visual activations with higher-level tokens are challenging and ongoing. We have much left to learn, but these proposals offer a sketch (a schematic map?) of how we might use certain shape tokens that begin with simple relations of oriented units to bridge the gap from low-level neural activations to higher-level representations. In pursuing these problems, we owe a great debt to Julian Hochberg, who was among the earliest researchers to grapple with the notion of representation in perception, to pinpoint key issues, and to impress his students with their importance. Although we are not yet in possession of clear answers, we are convinced, in no small part due to Hochberg's insights, that these are the crucial questions.

Note

1. Another interesting implication of Attneave's (1954) work and later research (e.g., Norman, Phillips, & Ross, 2001) for the present approach is that points of maximum curvature may be important for some aspects of shape encoding. Yet, in our scheme, such points may not be marked explicitly. One way to reconcile these ideas is to acknowledge that multiple encoding schemes exist in human shape perception. Points of maximum curvature may be extracted during early fitting of arclets (and used for various purposes) even if the final arclet representation sometimes subsumes a curvature maximum into a larger region of relatively constant curvature. These issues are under investigation.

References

ATTNEAVE, F. (1954). Some informational aspects of visual perception. *Psychological Review, 61*, 183–193.

COSTA, L. DA F. & CESAR, R. M. (2000). *Shape analysis and classification.* Boca Raton, FL: CRC Press.

FANTONI, C., & GERBINO, W. (2003). Contour interpolation by vector-field combination. *Journal of Vision, 3*(4), 281–303.

FULVIO, J.M., SINGH, M. & MALONEY, L. T. (in press). Consistency of location and gradient judgments of visually-interpolated contours. *Computer Vision and Pattern Recognition*, Proceedings, 2006.

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

GROSSBERG, S., MINGOLLA, E., & ROSS, W. D. (1997). Visual brain and visual perception: How does the cortex do perceptual grouping? Trends in *Neurosciences, 20*(3), 106–111.

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.

HOCHBERG, J. (1964). *Perception.* Englewood Cliffs, NJ: Prentice-Hall.

HOCHBERG, J. (1968). In the mind's eye. In R. N. Haber (Ed.), *Contemporary theory and research on visual perception* (pp. 309–331). New York: Holt, Rinehart, & Winston.

HOCHBERG, J. (1971). Perception: I. Color and shape. In J. W. Kling & L. A. Riggs (Eds.), *Woodworth and Schlosberg's experimental psychology* (pp. 395–474). New York: Holt, Rinehart & Winston.

HOCHBERG, J. E., GLEITMAN, H., & MACBRIDE, P. D. (1948). Visual threshold as a function of simplicity of form. *American Psychologist, 3,* 341–342.

HOCHBERG, J. E., & MCALISTER, E. (1953). A quantitative approach to figural "goodness." Journal of Experimental Psychology, 46, 361–364.

HOFFMAN, D. D., & RICHARDS, W. A. (1984). Parts of recognition. *Cognition, 18,* 65–96.

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

HUMMEL, J. E., & BIEDERMAN, I. (1992). Dynamic binding in a neural network for shape recognition. *Psychological Review, 99,* 480–517.

KELLMAN, P. J., GARRIGAN, P., KALAR, D. J., & SHIPLEY, T. F. (2003). Good continuation and relatability: Related but distinct principles. *Journal of Vision, 3*(9), 120a.

KELLMAN, P. J., GARRIGAN, P., & SHIPLEY, T. F. (2005). Object interpolation in three dimensions. *Psychological Review, 112* (3), 586–609.

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

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.

NORMAN, J. F., PHILLIPS, F., & ROSS, H. E. (2001). Information concentration along the boundary contours of naturally shaped solid objects. *Perception, 30,* 1285–1294.

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

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

Shipley, T. F., & Kellman, P. J. (2001). From fragments to objects: Segmentation and grouping in vision. Amsterdam: Elsevier.

Singh, M., & Hoffman, D. D. (1997). Constructing and representing visual objects. *Trends in Cognitive Sciences, 1*(3), 98–102.

Ullman, S. (1976). Filling in the gaps: The shape of subjective contours and a model for their generation. *Biological Cybernetics, 25,* 1–6.

Wertheimer, M. (1921). Untersuchungen zur Lehre von der Gestalt: I. *Psychologische Forschung, 1,* 47–58.

Wertheimer, M. (1958). Principles of perceptual organization. In D. C. Beardslee & M. Wertheimer (Eds.), *Readings in perception* (pp. 115–135). Princeton, NJ: D. Van Nostrand.

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. (1998). A common mechanism for illusory and occluded object recognition. *Journal of Experimental Psychology, 24*(3), 859–869.