

# Intermediate-Level Visual Representations and the Construction of Surface Perception

**Paul Sajda**

David Sarnoff Research Center

**Leif H. Finkel**

University of Pennsylvania

## Abstract

■ Visual processing has often been divided into three stages—early, intermediate, and high level vision, which roughly correspond to the sensation, perception, and cognition of the visual world. In this paper, we present a network-based model of intermediate-level vision that focuses on how surfaces might be represented in visual cortex. We propose a mechanism for representing surfaces through the establishment of “ownership”—a selective binding of contours and regions. The repre-

sentation of ownership provides a central locus for visual integration. Our simulations show the ability to segment real and illusory images in a manner consistent with human perception. In addition, through ownership, other processes such as depth, transparency, and surface completion can interact with one another to organize an image into a perceptual scene. ■

## INTRODUCTION

One of the basic characteristics of visual perception is our tendency to see the world as composed of objects and surfaces. In a complex scene, such as a mountain landscape or a crowded shopping mall, there are a variety of cues to surface segmentation, including discontinuities in luminance, color, motion, and disparity. There are also a number of pervasive impediments to surface segmentation, such as occlusion, transparency, camouflage, shadows, and specular reflections. The problem of surface segmentation has traditionally been viewed as taking advantage of the available cues to overcome the impediments.

Segmenting a scene into surfaces implies the existence of a neural representation of a surface, and since Marr (1982), such representations have been thought to be developed at the level of intermediate vision. Marr's view was that representations are constructed in a hierarchical fashion by parallel modules operating on different cues. However, recent evidence seems to suggest that integration begins in early vision and continues through a rich, multidirectional interaction between intermediate-level visual representations. For instance, both Shimojo and colleagues (1989) and Stoner and Albright (1993) have shown that the manner in which a scene is segmented into surfaces can affect other visual properties such as perceived motion. High-level processing, such as recognition, also depends on the particular

way in which a scene is parceled into surfaces (Nakayama, Shimojo, & Silverman, 1989). Even brightness perception, often equated with early visual processing, may be affected by surface segmentation (White, 1979).

Given the central role surface segmentation plays in visual processing, it is important to consider, from a computational viewpoint, how the cortex develops a representation of a surface. In this paper we consider how intermediate-level cortical processes, presumably within striate and extrastriate cortex, might construct such a neural representation. We formulate a model within the context of a specific problem in intermediate-level vision, namely, determining depth-from-occlusion. Occlusion pervades real scenes and is a consequence of projecting continuous matter (i.e., surfaces) from a three-dimensional world to a two-dimensional image. Our visual system is able to utilize cues to occlusion to organize surfaces in relative depth. Using computer simulations we investigate how these representations are constructed and how they interact.

The first section of this paper outlines the development of our model, beginning with a discussion of the *binding problem*, since the representation of elements belonging to the same surface must be “bound” together. In the second section we propose a set of network processes for binding. We show how these binding mechanisms can construct visual surfaces using an example real image (Fig. 1). The third section presents results of computer simulations for a range of stimuli,

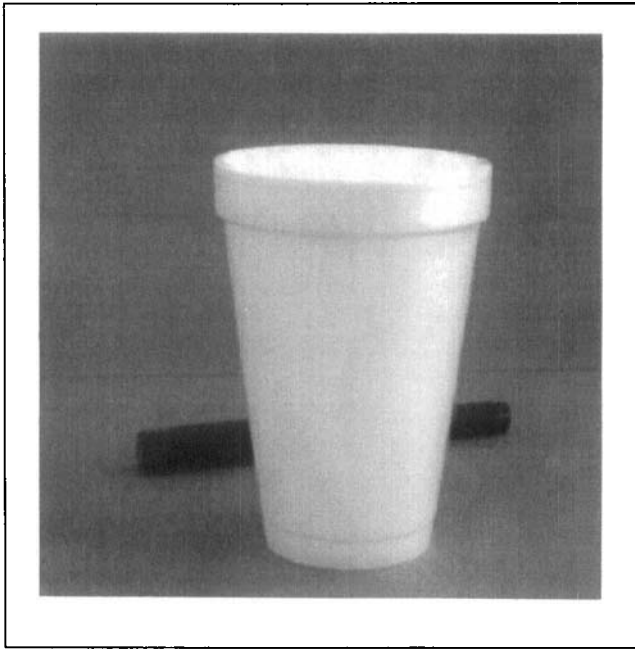


Figure 1. Image of a cup occluding a pen.

both natural and artificial. We compare the representations developed by the model with responses by human observers and identify, within the model, neural correlates to the psychophysics.

## DEVELOPMENT OF THE MODEL

### The Binding Problem

When we look at a simple scene, such as the one shown in Figure 1, we are able to effortlessly organize it into its constituent surfaces. For instance, we construct a representation of the “cup” and “pen” and stratify the two in relative depth. However, underlying this seemingly simple task is the computationally profound “binding problem” (Barlow, 1981; Sejnowski, 1986). The binding problem is essentially a problem of integration, requiring that fragmented visual representations be selectively grouped together to construct the coherent representations that we perceive. For instance, how is it that neuronal activities representing responses to the surface of the cup can be grouped or “bound” together while remaining distinguishable from those activities corresponding to the pen?

The binding problem is ubiquitous across visual perception and appears in a variety of forms. For example, the activities of orientation-selective neurons may serve as a cortical representation of line segments in the scene. However, the representation of a *surface contour* implies that these individual neuronal activities are somehow bound together—representations of individual points are grouped to create a representation of a con-

tour. Similarly, the construction of a surface implies a need for binding these surface contours to two-dimensional regions. The emergence of a “binding hierarchy,” where increasingly more complex representations are constructed by binding more primitive representations, is an important characteristic of our visual system since it enables us to construct abstractions usable for high-level vision (e.g., recognition and navigation).

In addition to constructing hierarchical representations, binding is also critical for integrating representations across space, time, and visual cue. For example, the stratification of the cup and pen in depth is dependent on occlusion cues localized at the cup/pen boundary. However, we perceive the entire surface of the cup to be in front of the pen, implying that the depth ordering established at the cup/pen borders must be selectively integrated into the representations of the two surfaces. Not only depth, but color, motion, texture, and other visual properties must ultimately be bound to the representation of the cup. Understanding the cortical mechanisms responsible for the multiple levels and multiple forms of neural binding emerges as the critical computational problem characterizing visual perception.

In the following sections we outline those features of our model that address how fragmented visual representations are bound together in the construction of a surface representation. These features serve as the foundation for our computational model, and result from consideration of the neuroanatomy, neurophysiology, and visual psychophysics as well as from the computational nature of the problem at hand.

### *Complementary Mechanisms for Binding and Segmentation*

In the model, the binding of attributes or spatially segregated features can occur in two ways: by a low-resolution mechanism based on cortical topographic organization or a high-resolution mechanism based on temporal synchrony between neuronal activities. The low-resolution binding mechanism takes advantage of ordered cortical topography, a hallmark of both striate and extrastriate areas (Van Essen, 1985), to link active neurons based on retinotopic proximity and functional similarity (e.g., similar orientation and directional selectivity). Active neurons in different areas are bound together if they lie in rough topographic register. Such a binding scheme could be accessed by attentional mechanisms that conjoin features (Treisman & Gelade, 1980; Desimone, Wessinger, Thomas, & Schneider, 1990) and can be considered a form of spatial binding—units that are spatially proximate tend to be bound and those that are spatially distant tend to be segmented. An advantage of this approach is that binding can occur in parallel for all objects across the visual field. However, a major problem with this mechanism is that objects that are overlapping or occluded, and that should otherwise be

segmented, will be grouped together. An alternative scheme is therefore necessary to bind and segment overlapping surfaces.

Temporal binding is the second form of binding in the model. It differs from spatial binding in that it binds and segments surfaces in the temporal domain, allowing for the coexistence of multiple objects in the same topographic region. A disadvantage of operating in the temporal domain, however, is that the biophysical properties of cortical neurons (e.g., membrane time constants) together with the rapid processing times required for real-time vision imply that such a mechanism is resource-limited—only a small number of surfaces can be bound and segmented simultaneously. Spatial binding and temporal binding therefore act as complementary mechanisms—spatial binding acting as a resource-free, spatially based process operating across the entire visual field and temporal binding acting as a resource-limited, temporally based process operating across select regions of the visual field.

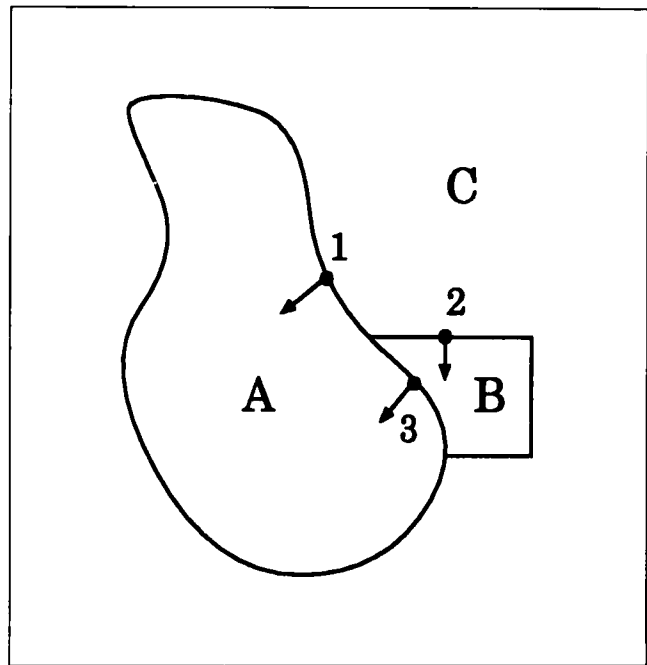
#### *Direction of Figure as a Surface Representation*

The fundamental problem in surface discrimination is the determination of contour “ownership” (Koffka, 1935; Nakayama & Shimojo, 1990; Finkel & Sajda, 1992). Contours and borders, resulting from luminance, chromatic, depth, or motion discontinuities, locally dichotomize an image into regions. If a contour is an occluding contour (Marr, 1977) then one of the regions represents the surface of the object and that surface “owns” the contour and locally constitutes the “figure.” Thus, contour ownership specifies figure-ground relationships, serves as a local representation for determining relative depth, and ultimately has demonstrable effects on recognition (Nakayama et al., 1989).

In the model, ownership along a contour is represented as a vector that we call *direction of figure*, illustrated in Figure 2. This vector is perpendicular to the tangent of the occluding contour segment, and points toward the surface that owns the segment—direction of figure *binds* the contour to the region that is the surface. Since the computation of direction of figure is context-dependent, the neural circuits that compute this vector must have access to information across the image. In the model this access is mediated via the complementary binding mechanisms, allowing direction of figure circuits to selectively integrate information in both spatial and temporal domains.

#### *Ownership as a Medium for Integration*

Junctions between surfaces are important since they create cues that can be used to complete surfaces, determine relative depth, and ascribe transparency. In the surface representation constructed by our model, such junctions are sites where ownership changes. We refer

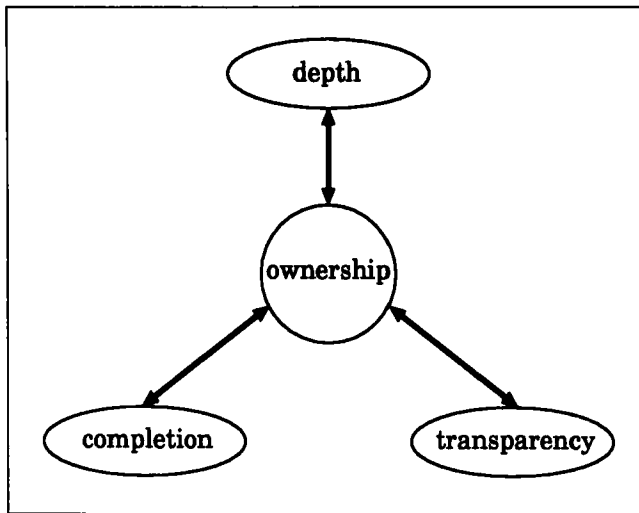


**Figure 2.** Example illustrating the “direction of figure” representation of contour ownership. At point (1) the contour forms the occluding contour of surface A and therefore A owns the contour. In the model this relationship is represented as a direction of figure vector pointing toward region A. Direction of figure is also shown for points (2) and (3).

to these changes in ownership as “ownership junctions.” Since Helmholtz (1909) it has been accepted that T-junctions serve as the primary cue to occlusion and relative depth (Guzman, 1968; Waltz, 1975). However, T-junctions are *context-independent* features, formed by local intersections of contours. Ownership junctions are *context-dependent* features and their appearance is directly linked to the representation of a surface.

Clearly T-junction detectors could be constructed by an appropriate wiring of cortical neurons selective for different orientations. The simplicity by which such circuits could be constructed makes more remarkable the fact that there is little neurophysiological evidence for cells in visual cortex responding exclusively to “T” intersections. However, preliminary experimental evidence shows neurons selective for figure-ground direction in area V2 (Peterhans & von der Heydt, 1992). Together with the psychophysical findings of Nakayama and colleagues (Nakayama et al., 1989; Nakayama & Shimojo, 1990), showing the effect of ownership on depth, transparency, and completion, these results imply that instead of T-junctions, the relevant visual feature may be a context-dependent construct within intermediate-level vision.

Though changes in ownership may serve as cues for organizing a visual scene, we believe that ownership in fact plays a more fundamental role by acting as a *central locus of integration*, as indicated in Figure 3. In the model, depth, direction of figure, and local contrast are



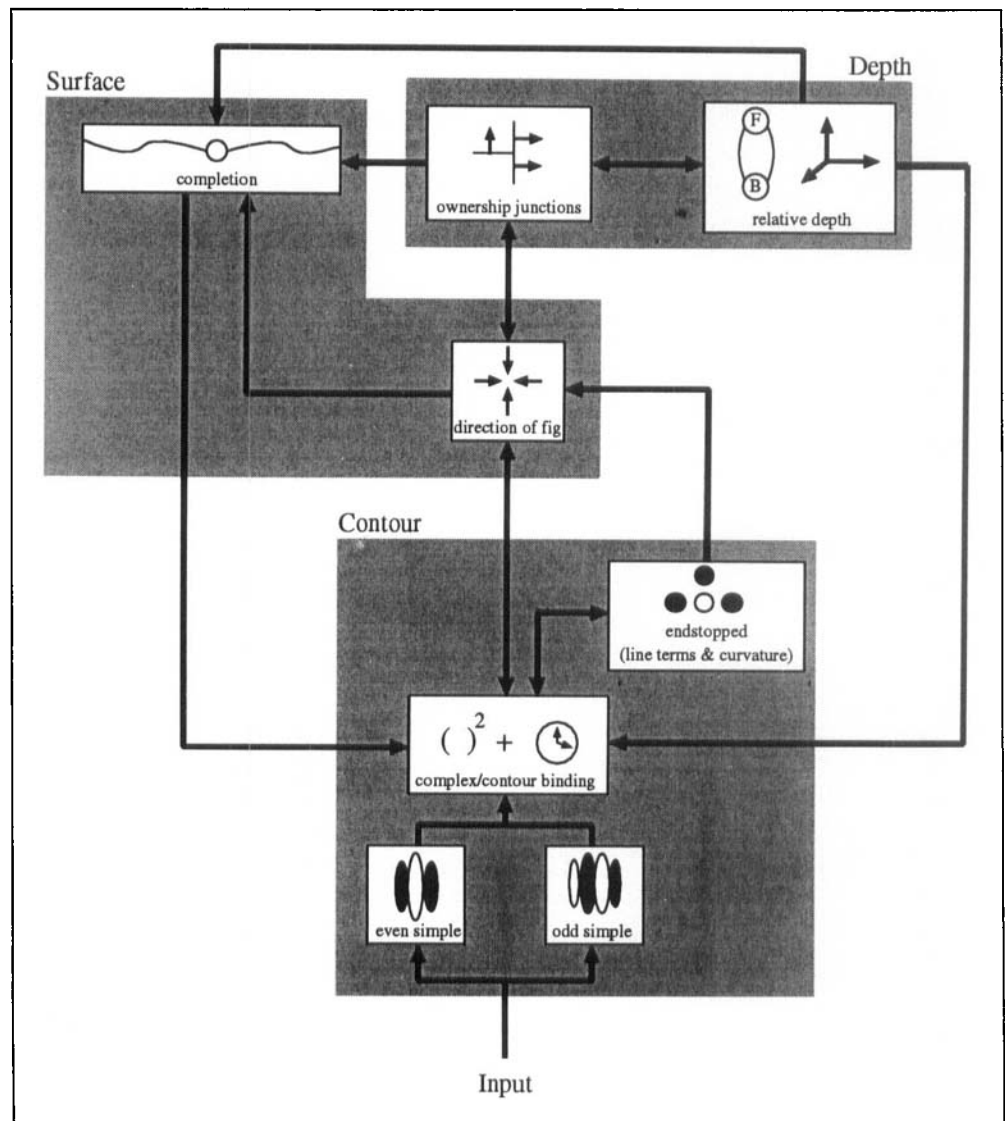
**Figure 3.** Ownership serves as the central locus of integration in determining depth, transparency, and completion.

combined to produce a response that represents the integration of global and local information. Additionally, processes can interact through ownership to affect depth, transparency, and completion (and most likely other properties such as motion), hence the two-way arrows in Figure 3. For instance, depth determined via binocular disparity can effectively constrain ownership and thereby change perceived completion and/or transparency.

### Overview of the Model

A schematic representation of our computational model is shown in Figure 4. The model consists of eight functionally distinct network processes, each belonging to one of three stages; *Contour*, *Surface*, and *Depth*, which interact through feedforward and feedback pathways. The processes in these stages organize input into visual scenes, with contours constructed in the *Contour* stage, surfaces discriminated in the *Surface* stage, and finally

**Figure 4.** Schematic of the model. Eight network processes (white boxes) are grouped into three main stages (*Contour*, *Surface*, and *Depth*) that deal with 1-D, 2-D, and 3-D organization of the input, respectively. Critical interactions between processes are illustrated with arrows.



three-dimensional scene organization established in the *Depth* stage. By placing *Depth* hierarchically above *Contour* and *Surface* we do not mean to imply that all processes typically associated with determining depth occur at such a late stage. For example, the *Contour* stage is assumed to be form-cue invariant (Stoner & Albright, 1993) in that it ultimately represents the processing of all types of contours or boundaries, including discontinuities in binocular disparity, motion, texture, and color. The main point of the architecture is that surface representations must be constructed before higher level attributes (such as object depth or motion) can be determined, and that these surface representations in turn depend on certain edge or contour-based processes.

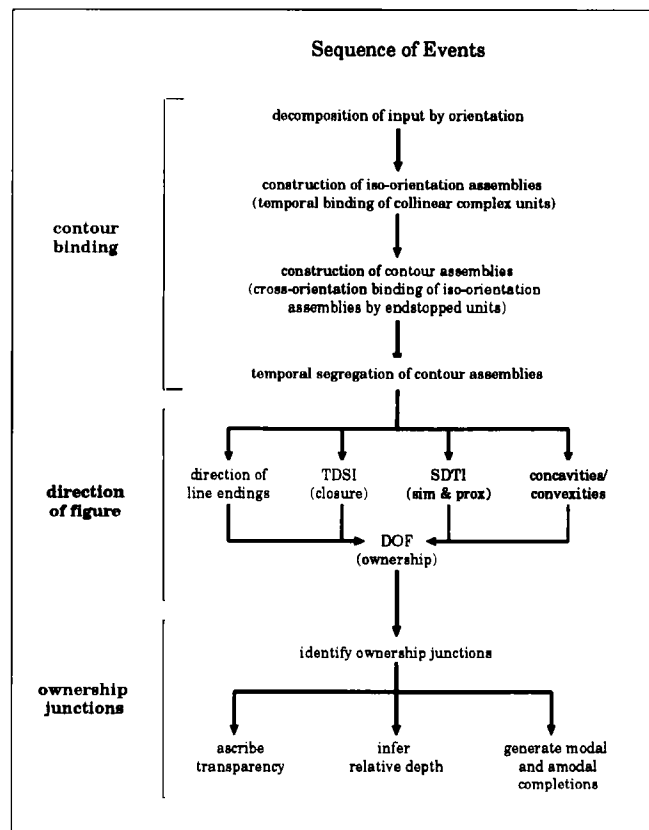
Our model attempts to capture the essence of the neural computations involved in surface perception. Though some of the network processes we propose are in agreement with known neural architectures found in striate and extrastriate cortex, in many instances the details of the architectures are unknown. In these cases we revert to psychophysics to identify critical functions needed in the computation. We implement these functions in network processes given the constraint that they be feasible given known cortical anatomy and physiology. The purpose of the model is not to propose a specific neural architecture but instead to consider how neural representations might interact to organize an image into a perceptual scene.

In the following section we discuss the computational and representational details of the three main network processes in the model: *contour binding*, *direction of figure*, and the use of *ownership junctions* for inferring depth and linking surface elements. A flowchart, illustrating the general sequence of events within these network processes, is shown in Figure 5. To illustrate the basic neural operations carried out in the model we will consider the function of each network process using as an example the stimulus shown in Figure 1.

## IMPLEMENTATION OF THE MODEL: FROM CONTOURS TO SURFACES TO DEPTH

### Contour Binding

Contour binding is a network process that utilizes temporal binding as a means for grouping points belonging to the same surface contour. A number of investigators have put forth the idea that the temporal relationship between neuronal activities can serve as a mechanism for solving particular aspects of the binding problem (von der Malsburg, 1981; von der Malsburg & Schneider, 1986; Eckhorn et al., 1988; Gray & Singer, 1989; Crick & Koch, 1990; Sporns, Tononi, & Edelman, 1991; Hummel & Biederman, 1992). This position is supported by recent experimental findings showing phase-locked oscillations in areas 17 and 18 in cat (Eckhorn et al., 1988; Gray & Singer, 1989; Singer et al., 1990) and temporal synchrony



**Figure 5.** Sequence of events describing the operation of the network processes *contour binding*, *direction of figure*, and the identification of *ownership junctions*. Not shown are feedback pathways between network processes that can influence this event sequence.

in area V1 (Livingstone, 1991; Kreiter & Singer, 1992) and V2 (Bullier, Munk, & Mowak, 1992) of the macaque monkey. However, there is still active debate regarding which of the observed phenomena might play a functional role (Wilson & Bower, 1991; Ghose & Freeman, 1992).

The model we have developed is largely independent of specific biophysical mechanisms and considers temporal binding as simply a grouping process. For this purpose we adopt the concept of a *temporal binding value* to refer to a time-dependent variable in which units having similar values are grouped together while those having different values are segmented from one another. *In vivo* this parameter could potentially be represented by oscillation frequency, phase of neural firing, time-sliced synchronization, or some alternative temporal relationship between neuronal activities. In discussing the network process of contour binding, we focus on how such a temporal variable is selectively disseminated to affect the binding and segmentation of elements in the visual input.

The network events leading to contour binding occur in several stages, as shown in Figure 5. Initially contours are decomposed into their constituent orientations by

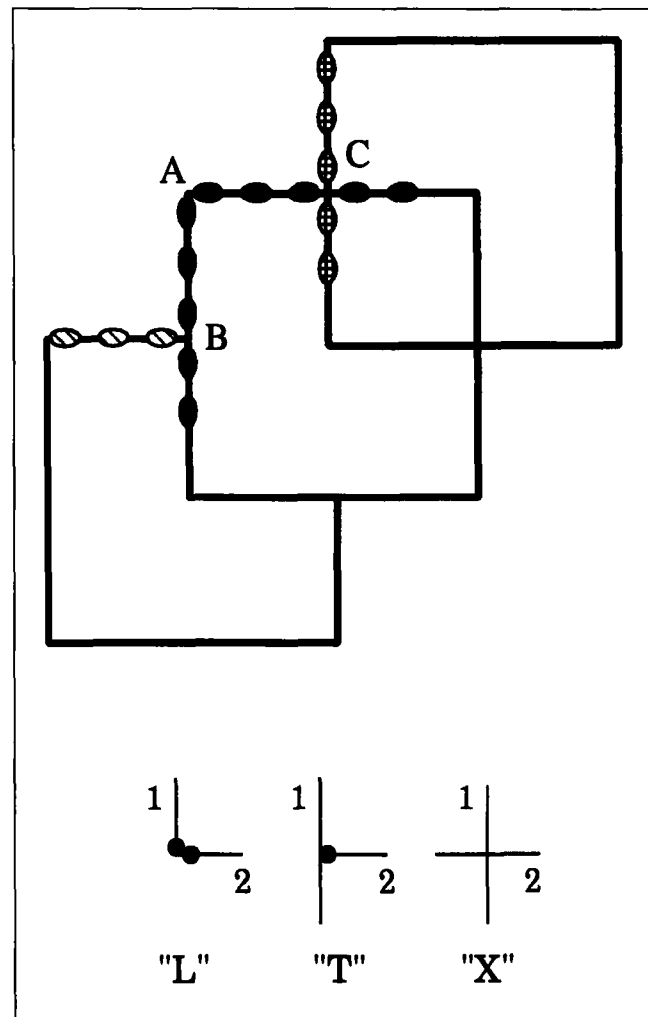
units that capture some of the general properties of the complex cells in visual cortex. For example, their response is nonlinear and phase invariant, with inputs combined in quadrature phase (Adelson & Bergen, 1985; Bergen & Landy, 1991). These “complex-like” units are activated by feedforward input originating from the visual scene or from feedback input generated by “completion” networks in the *Surface* stage. Activation of these units serves as an internal representation of the contours in the scene.

Activated complex units, which are located in close proximity, have the same orientation preference, and whose receptive fields are retinotopically collinear are forced to have similar temporal binding values. The selective dissemination of temporal binding values could be carried out by long-range horizontal connections in striate cortex, which preferentially connect units selective to the same orientation (Rockland & Lund, 1982; Gilbert & Wiesel, 1989). For example, one could envision the temporal binding value representing the phase of neural firing. Long-range horizontal connections could then presumably serve as a medium by which segregated units synchronize. The grouping of complex units based on similar temporal binding values represents the construction of “iso-orientation assemblies.” These assemblies are collectives of units responding to a particular edge of an object’s occluding contour. Iso-orientation assemblies are subsequently grouped at line endings to form “contour assemblies.” Figure 6 illustrates that the model utilizes endstopped unit (Gilbert, 1977) activity as a way of creating cross-orientation binding of iso-orientation assemblies. The linkage of iso-orientation assemblies into contour assemblies serves as the model’s representation of a surface’s occluding contour.<sup>1</sup>

Contour assemblies are segregated from one another by a competitive process that forces their constituent units to have temporal binding values that occupy a unique range in the temporal binding space. For instance, consider simulation results for the cup/pen figure, shown in Figure 7. The 8 bit gray-scale image of Figure 1 is the input to the model. Units responding to the contour of the cup are grouped together in that they tend to have temporal binding values that fall in the range 4.5–5.5. However, this range is different from the range for units responding to the pen (i.e., the assemblies are segmented from one another).<sup>2</sup> Note that at this point the contours of the two halves of the pen remain segmented due to occlusion by the cup. To link the segments of the pen and construct a single compound object requires knowing something about the regions delineated by the contours.

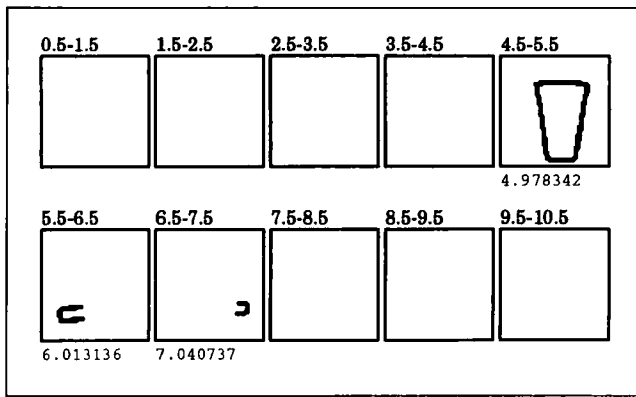
### Direction of Figure

Direction of figure (DOF) is a network process that computes the direction of the region that owns a segment of contour, resulting in the construction of a local

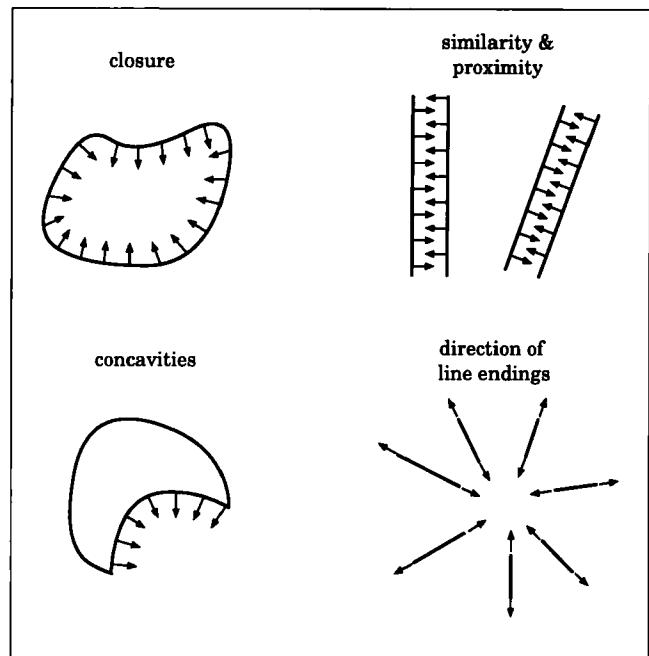


**Figure 6.** Iso-orientation assemblies are linked, based on local patterns of endstopped unit activity, to form contour assemblies. Ovals represent complex units, with their shading indicating their temporal binding value. Units with the same shading are part of the same assembly. For the “L” type junction at A the two iso-orientation assemblies are grouped together since each is associated with a line ending that excites endstopped units. The “T” type junction at B has only a single line ending and the two iso-orientation assemblies remain segregated. Assemblies also remain segregated at an “X” type junction, for example, at C, since no line endings are present and there is no local endstopped activity.

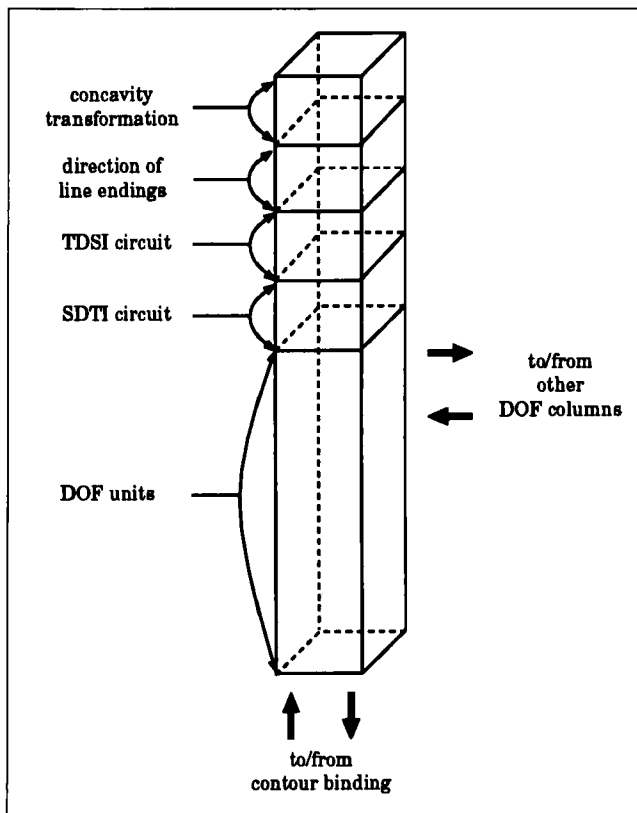
representation of figure-ground. In the model, the DOF circuitry is organized in a columnar structure consisting of four circuits, as shown in Figure 8. There are two primary circuits that operate to compute the direction of figure: one is a temporal-dependent/spatial-independent (TDSI) circuit selective to “closure,” and the other is a spatial-dependent/temporal-independent (SDTI) circuit selective to “similarity and proximity.”<sup>3</sup> There are also two secondary circuits that play a role in determining direction of figure. One is based on the observation that concave segments bounded by discontinuities are a cue for occlusion and ownership (Koffka, 1935 and see Appendix), while the other considers the direction of line endings as a potential cue. Figure 9



**Figure 7.** Initial binding and segmentation of contours constructed by the model for the image in Figure 1. Results are displayed as a 2-D histogram, where each box represents a different range of temporal binding values. The value beneath each box represents the average of the temporal binding values of units within that range.



**Figure 9.** Cues used in determining direction of figure.



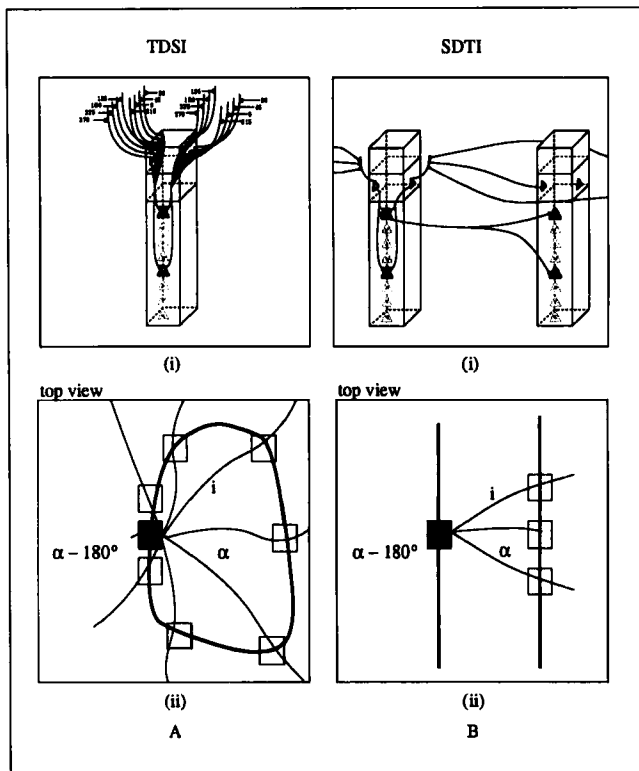
**Figure 8.** Divisions, inputs, and outputs for a DOF column.

summarizes the cues used to determine direction of figure. An important difference between the model's processing of primary and secondary cues is that secondary cues result in only a transient change in direction of figure and this change must ultimately be supported by a primary cue. For example, a transient configuration established by a secondary cue can potentially activate completion processes involved in linking scene elements. Feedback from completion may in turn affect the overall organization of the scene, thereby affecting the

primary cues. After feedback, the primary cues solely determine direction of figure.

Outputs of the circuits selective for the different cues are combined by a set of DOF units whose activation encodes the direction of figure vector. The contribution of each cue to the total activation of a DOF unit is based on empirical observation. Thus, as will be shown below, "closure" has a greater weight than "similarity/proximity." In addition, we assume that "concavities/convexities" and "direction of line endings" can transiently overwhelm "closure" and "similarity/proximity." Concavities and line endings are potential cues to occlusion, but they can also occur in the absence of occlusion (sometimes lines just end). Thus, the influence of these inputs on DOF must be context-dependent. The transient nature of the effect corresponds to a kind of "hypothesis testing" as to whether occlusion is present (see Appendix).

We will focus on the details of the closure (TDSI) and similarity/proximity (SDTI) mechanisms. The TDSI and SDTI circuits both operate at contours, comparing inputs from the two regions separated by the contour. The computation of the TDSI circuit can be thought of as comparing the amount of "closure" exhibited by the two regions while the SDTI circuit can be viewed as computing the difference in "similarity" and "proximity" between the regions. The TDSI "closure" circuit receives inputs from other DOF columns located over a wide region spanning a full 360° in the scene (see Fig. 10A). Input is effective only if it originates from DOF columns having temporal binding values falling within the same range, thus establishing the temporal dependence of the TDSI mechanism. The circuit's function is revealed by



**Figure 10.** The two primary circuits operating to compute direction of figure. (A) (i) Temporal-dependent/spatial-independent (TDSI) circuit architecture. DOF units representing a direction orthogonal to the orientation of the contour can potentially be activated (these units are shown as bold triangles). (ii) Top view showing a network of DOF columns. Filled square represents position of a specific column in the network. Unfilled squares represent other DOF columns serving as input to this column. Bold curve corresponds to a contour in the input. Shown is the pattern of long-range horizontal connections converging on the right side of the column (side  $\alpha$ ). (B) (i) Spatial-dependent/temporal-independent (SDTI) circuit architecture. Recurrent connections bind DOF columns by forcing them to (1) have direction of figures that are antiparallel and (2) have similar temporal binding values. (ii) Top view of network showing pattern of long-range horizontal connections converging on the right side of the column (side  $\alpha$ ).

noting that each set of inputs represents the number of rays intersecting the contour, where a set of rays emanates from within each region. Regions that are enclosed generate more intersections and therefore are discriminated as having more “closure.”

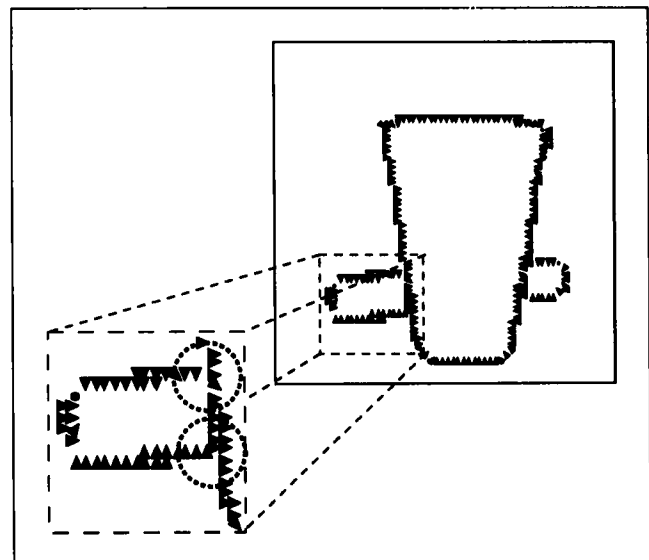
The SDTI circuit organizes elements in the scene based on “proximity” and “similarity” of oriented contours. Unlike the TDSI circuit, which depends on temporal binding, the SDTI circuit uses spatial binding to access information across the network. Activity is integrated via horizontal connections from units with similar orientation tuning that lie in a direction orthogonal to the contour (i.e., from parallel line segments). The efficacy of these connections decreases with distance, implying that columns that are farther apart are less likely to be bound together. Input is integrated for regions on both sides of the contour, and across all ranges

of temporal binding values (i.e., the SDTI mechanism is temporally independent). Neighboring parallel contours generate the greatest activation and the circuit tends to discriminate the region between two close parallel contours as the figure.

The direction of figure computed for our sample image is shown in Figure 11. Each location along the occluding contour maps to a unit that encodes the direction of figure, relative to the local edge. The model’s representation of the cup now includes both the contour of the cup and the region that owns the contour.

### Ownership Junctions

Much of the information about the three-dimensional organization of a scene is localized at the intersection of surfaces, or their projections. Such locations are usually at contour junctions, where contours meet to locally partition an image into regions. However, it is not the junction of contours per se, but the junction of surfaces that provides a cue for surface overlap and hence relative depth. In the model, network processes operate to discriminate ownership junctions, examples of which are shown in the inset of Figure 11. An ownership junction is defined as a junction where each contour is owned by a different surface (i.e., a junction where the ownership of contours changes). As we have shown, the model uses pictorial information to compute the direction of figure and hence determine contour ownership. However, a region in an image may represent the overlap of transparent surfaces and, therefore, more than one contour at the junction could be owned by the same



**Figure 11.** Direction of figure computed by the model for the cup/pen image. Direction of figure is shown as an oriented arrow-head, where the orientation represents the preferred direction of the DOF unit that is most active (inset). Direction of figure near the cup/pen junction. Highlighted are two ownership junctions serving as indicators for occlusion.



region. Similarly, ownership computed using pictorial cues may not be consistent with depth established by other cues (e.g., stereopsis and motion parallax). Therefore to discriminate an ownership junction one must consider pictorial cues, together with local contrast and depth. Figure 12 shows several examples illustrating our definition of an ownership junction.

The role of ownership junctions in the stratification of surfaces in relative depth is illustrated in Figure 13. Relative depth is represented in the model by the activity of two networks: *foreground* and *background*. In some sense the units in these networks are analogs to near and far tuned disparity selective neurons. Foreground and background units are mutually inhibitory and each retinotopic location is mapped to a foreground/background pair. Discrimination of an ownership junction forces an increase in foreground unit activity and a decrease in background unit activity for the occluding contour at the ownership junction. The remainder of the surface contour is stratified in depth by utilizing temporal binding. The relative depth of the segments at the ownership junction is "broadcast" to

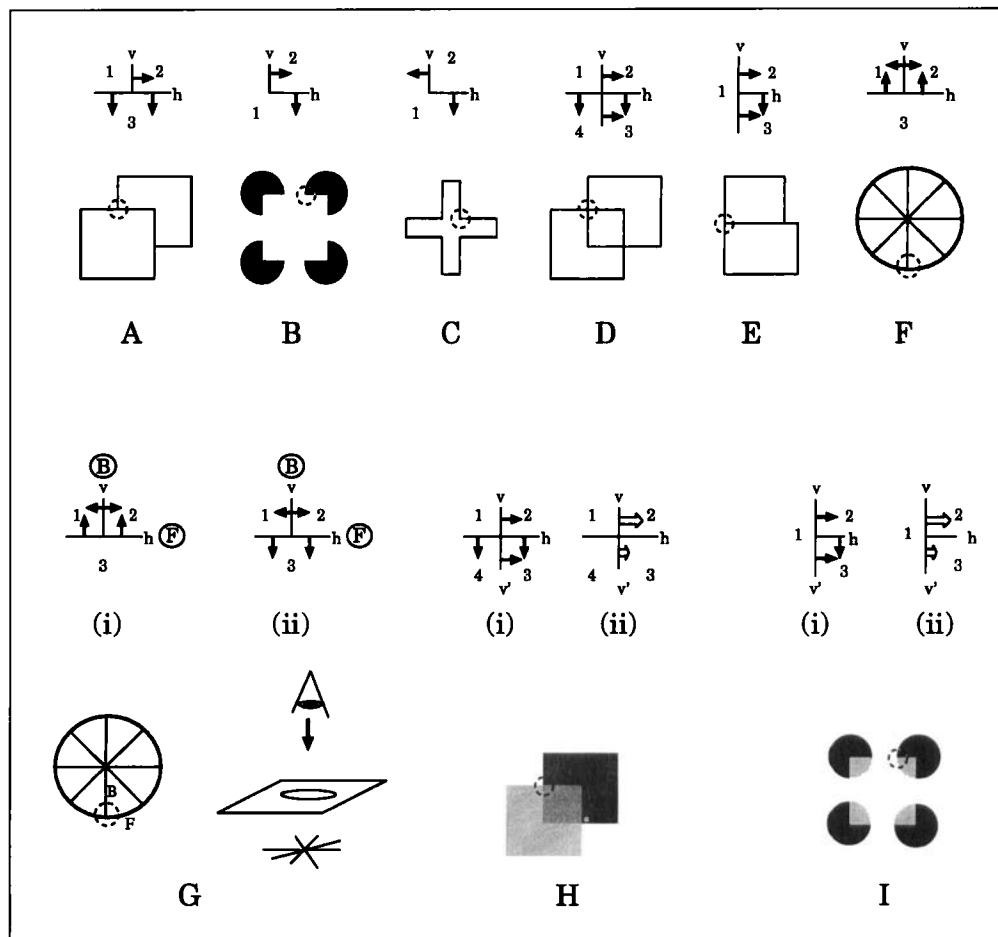
other units having the same temporal binding value, enabling the model to integrate the attribute of depth into the representation of the surface.

Together with the role they play in determining relative depth, ownership junctions are also cues for interpolation and completion of contours and surfaces. As shown in Figure 14, ownership junctions can excite completion units in three different directions: two associated with completion of the occluding contour and one with the occluded contour, representing the construction of modal and amodal contours (Kanizsa, 1979). Both modal and amodal contours are fed back to contour binding where they are integrated with other contours. Modal completion affects both the firing rate and temporal binding value of the complex units, while amodal completion affects only the temporal binding value.<sup>4</sup>

Returning to our sample image, the ownership junctions shown in the inset of Figure 11 force the stratification of the cup into the foreground, where the relative depth determined at the junction is disseminated, via temporal binding, to all units responding to the surface of the cup. The resulting depth ordering is shown in

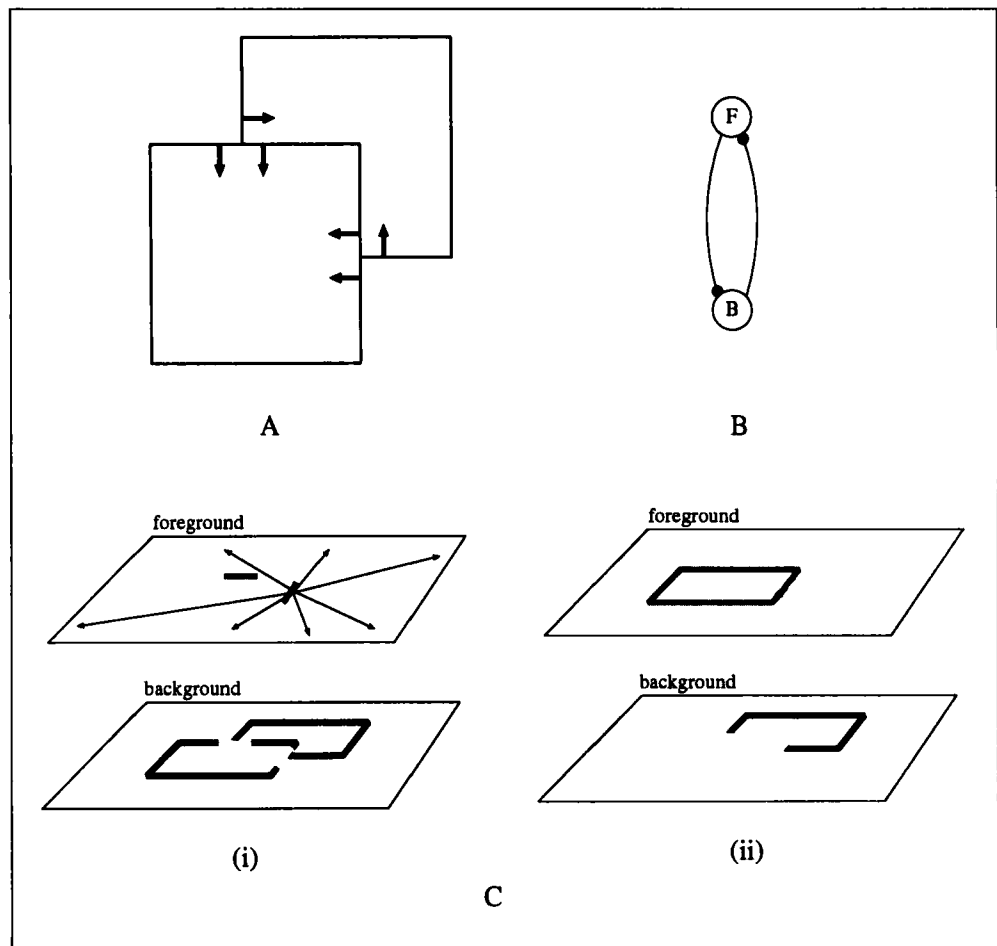
**Figure 12.** Contour junctions versus ownership junctions.

Dashed circle in each figure denotes location of the junction. (A,B) Examples of ownership junctions. Both the horizontal (h) and vertical (v) contour segments are owned by different surfaces. (C-F) Examples of contour junctions that are not ownership junctions (double arrows in F indicate contour can be owned by either region). (G) Depth and ownership junctions. If the depth of the "wheel" in F is changed so that the rim of the wheel is made to be closer to the viewer than the spokes, an ownership junction is created, and the figure is seen as an aperture or "hole" [i.e., the integration of depth forces the ownership at the junction to change from (i) to (ii)]. (H,I) Contrast and ownership junctions. A region may represent the overlap of multiple surfaces, due to transparency. In this case it is valid for different contours to own the same region. To determine whether a region is a region of overlapping transparent surfaces, one must consider contrast at the junction. For example, as one moves along the contour  $v-v'$  the sign of the contrast is preserved and the magnitude decreases [shown in (ii) with direction of arrow representing sign and length representing magnitude]. This condition indicates transparency (Metelli, 1974; Kersten, 1991). Both H and I can therefore be considered ownership junctions since, due to contrast information, the horizontal (h) and vertical (v) segments in these figures may own the same region (in these cases region 3).

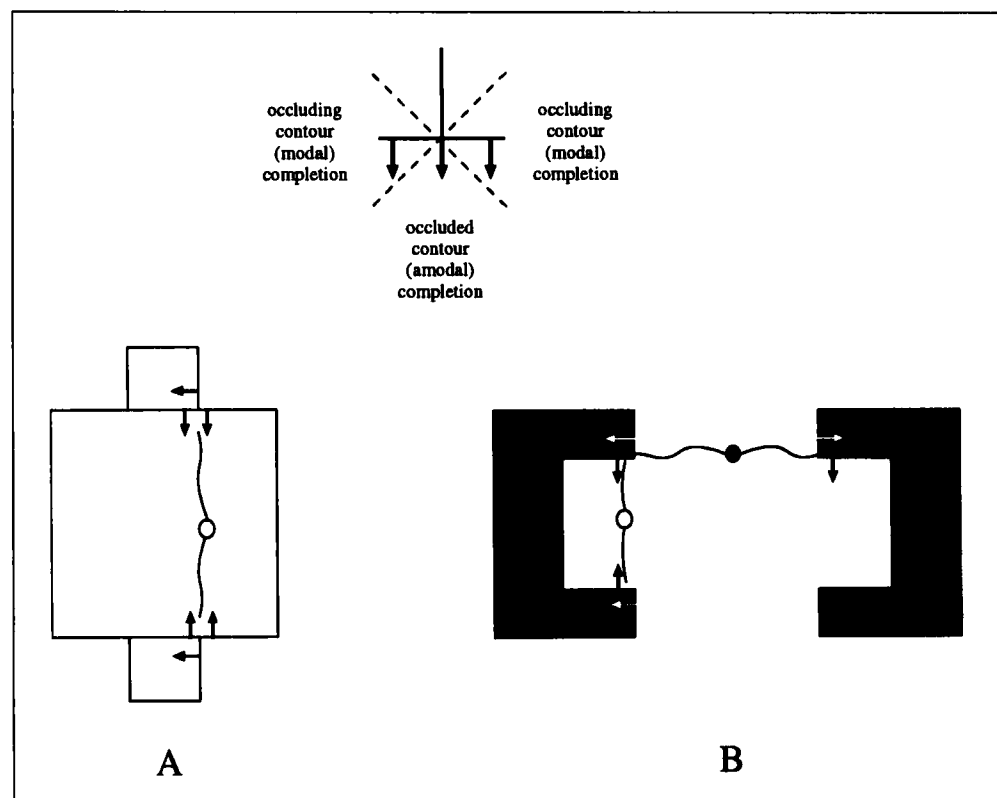


example, as one moves along the contour  $v-v'$  the sign of the contrast is preserved and the magnitude decreases [shown in (ii) with direction of arrow representing sign and length representing magnitude]. This condition indicates transparency (Metelli, 1974; Kersten, 1991). Both H and I can therefore be considered ownership junctions since, due to contrast information, the horizontal (h) and vertical (v) segments in these figures may own the same region (in these cases region 3).

**Figure 13.** Ownership junctions force the stratification of surfaces in depth. (A) A pair of ownership junctions signaling occlusion. (B) Mutually inhibitory foreground/background pair. Each retinotopic location is mapped to such a pair. (C) (i) Ownership junctions force local segments of contour to change their relative depth within the foreground/background maps. This local depth change is "broadcast" to all units having the same temporal binding value, resulting in (ii) the stratification of the entire surface in depth.



**Figure 14.** Ownership junctions serve as the primary cue responsible for completion (in A and B, completion units are shown as small circles with bipolar connections—filled units represent occluding/modal completion, unfilled units represent occluded/amodal completion). Occluded contour (amodal) completion occurs in a direction parallel to the orientation of the ownership junction while occluding contour (modal) completion is in the perpendicular direction. (A) Amodal completion of an occluded contour and surface. Activity of completion units is fed back to complex units, allowing for segregated elements in the scene to be bound together via contour binding. (B) Stimulus with both modal and amodal completion.



**Figure 15.** Results of the cup/pen simulation. (A) Depth of surfaces. Plot shows activity of units in the foreground network (as % maximum firing rate)—higher activity indicates that the surface is perceived as closer to the viewer. (B) Binding and segmentation after feedback. Feedback of completion activity forces the grouping of the two halves of the pen into a single surface.

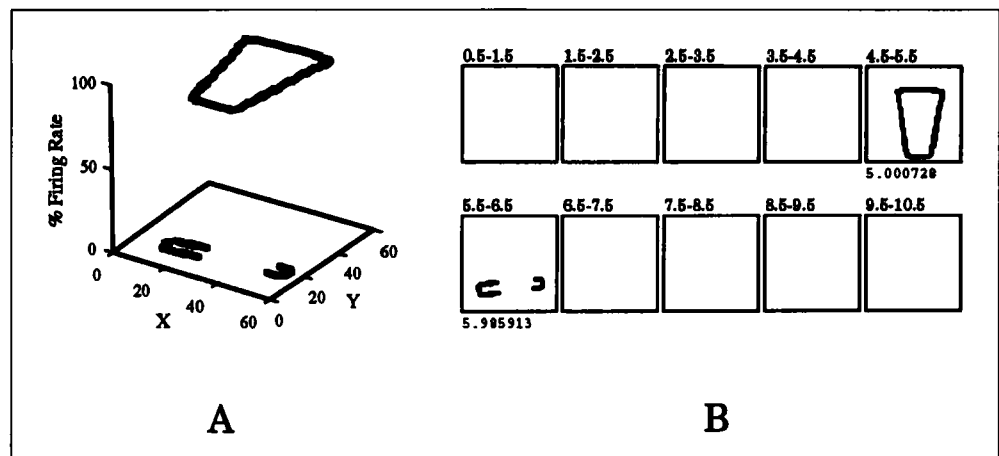


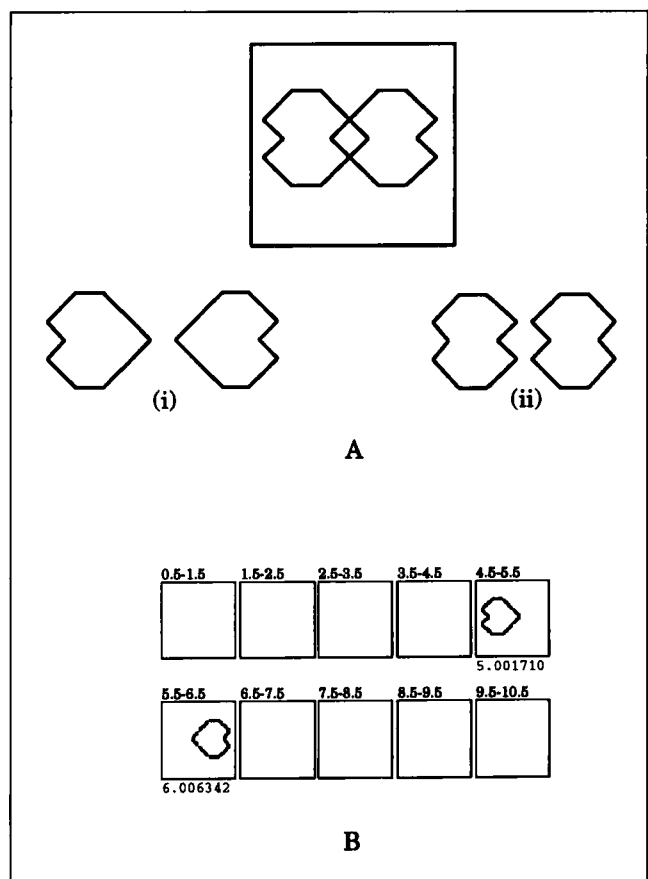
Figure 15. Simultaneously, completion units are activated, with amodal contours generated and subsequently fed back to the *Contour* stage. The two halves of the pen that remained segmented on the first simulation epoch are thus bound together as a single surface by the second epoch. The model's representation of the cup and pen has now come to include the relative depth relationships between the different surfaces.

## RESULTS

The simulation results of the cup and pen image illustrate the basic operations and representations used by the model. In this section we expand our repertoire of images and investigate the ability of the model to emulate human psychophysics by testing it in relation to the gestalt laws, figure-ground discrimination, depth-from-transparency, completion, and a particular neuropsychological case study.

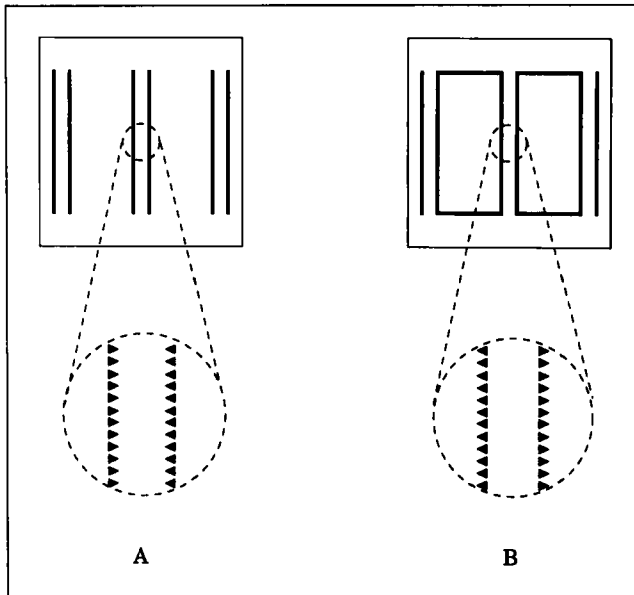
### Network Correlates of the Gestalt Laws

The gestalt psychologists developed a theory of perception based on the observation that particular cues or factors determine how we organize elements in a scene. These cues included "closure," "continuity," "proximity," and "similarity" (Koffka, 1935; Wertheimer, 1938). In our model, these cues are discriminated by specific neural architectures. For example, Figure 16 shows a figure with two possible perceptual interpretations, one exhibiting "good continuation" and the other preserving figural "symmetry." The dominant interpretation for human subjects corresponds to the preservation of local continuity. Simulations of the model show that it constructs the same interpretation, as represented by the temporal segmentation of the two contours. The model's favoring of this particular grouping can be attributed to the connection architecture of contour binding, with complex units having the same orientation tuning being strongly coupled and therefore tending to have similar temporal binding values.



**Figure 16.** (A) Figure having two perceptual interpretations, one preserving local continuity (i), the other preserving global symmetry (ii). Human subjects tend to segment the figure as (i). (B) Binding and segmentation by the model. The interpretation is attributable to the pattern of connections between complex units involved in temporal binding.

Figure 17A is a stimulus that human observers tend to interpret as three thin vertical bars. The model organizes the figure in a similar fashion. The TDSI circuits integrate activity from both sides of the line segments, however, they do not discriminate either region as the surface since both regions generate an equal amount of "clo-



**Figure 17.** (A) Input stimulus interpreted by human subjects as consisting of three thin bars. (Inset) Magnified view of direction of figure in the center of the stimulus as determined by the model. For this case the direction of figure is primarily determined by the SDTI circuit. (B) Stimulus interpreted as two rectangular surfaces. (Inset) Magnified view showing the direction of figure is reversed from that in A.

sure." SDTI circuits tend to favor groupings exhibiting "proximity" and "similarity," with respect to orientation, and hence discriminate as the surface the thin region between neighboring contours.

The second stimulus, shown in Figure 17B, is identical to the first except for the addition of horizontal segments connecting distal pairs of vertical segments. For human subjects this change drastically affects the perceptual organization of the scene—subjects now interpret the image as consisting of two wide rectangles. The addition of the segments can be seen as adding closure to the figure, where the wide regions forming the surfaces of the two rectangles have greater closure than the thin regions. Though the SDTI mechanism responds as before, the TDSI circuits, responding to this added closure, force the model to parcel the figure into two large rectangles.

### Surface Discrimination without Contour Closure

In most computational theories, contours play a crucial role in the discrimination of surfaces and the determination of figure-ground relationships. For example, theories based on diffusion (Grossberg & Mingolla, 1985; Grossberg & Wyse, 1992) or "spreading activation" (Ullman, 1984) use contours as boundaries for "coloring-in" regions. Regions bounded by closed contours are discriminated as "figure" if they tend to maintain their coloring while those that are open tend to be seen as

"ground" since they allow coloring to dissipate. Such a scheme implies a necessity for complete contour closure. However, psychophysical evidence points to the existence of a perceptual "closure continuum" (Elder & Zucker, 1993), with strict topological closure not a requirement for determination of two-dimensional shape.

Figure 18 illustrates how incomplete contours affect our model's construction of surfaces. A real image (8 bit gray-scale) of two occluding seagulls, with the neck of the foreground gull fading into the surface of the background gull, was filtered using a simple non-optimal edge detector ( $3 \times 3$  Sobel filters), creating an input to the model having incomplete occluding contours (shown in Fig. 18B). In spite of the fact that the occluding contours are incomplete, the network is still able to determine the direction of figure and distinguish figure-ground over most of the contour. For example, the direction of figure along 95% of background gull's contour is correctly classified with 75% correct classification for the foreground gull. This performance is attributable to the model's use of relative "closure" on either side of the occluding contour. The preservation of the surface representation in spite of incomplete occluding contours allows the system to organize the scene in relative depth.

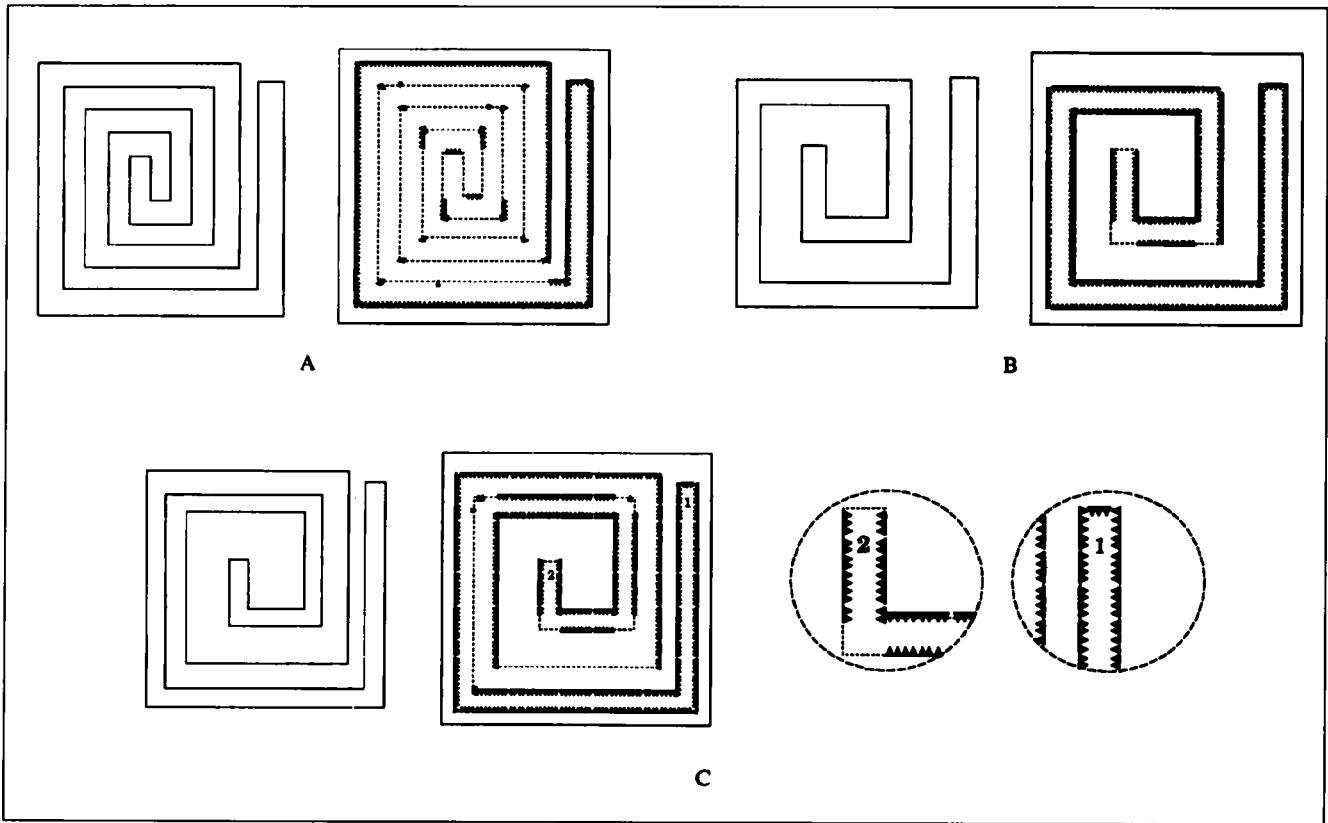
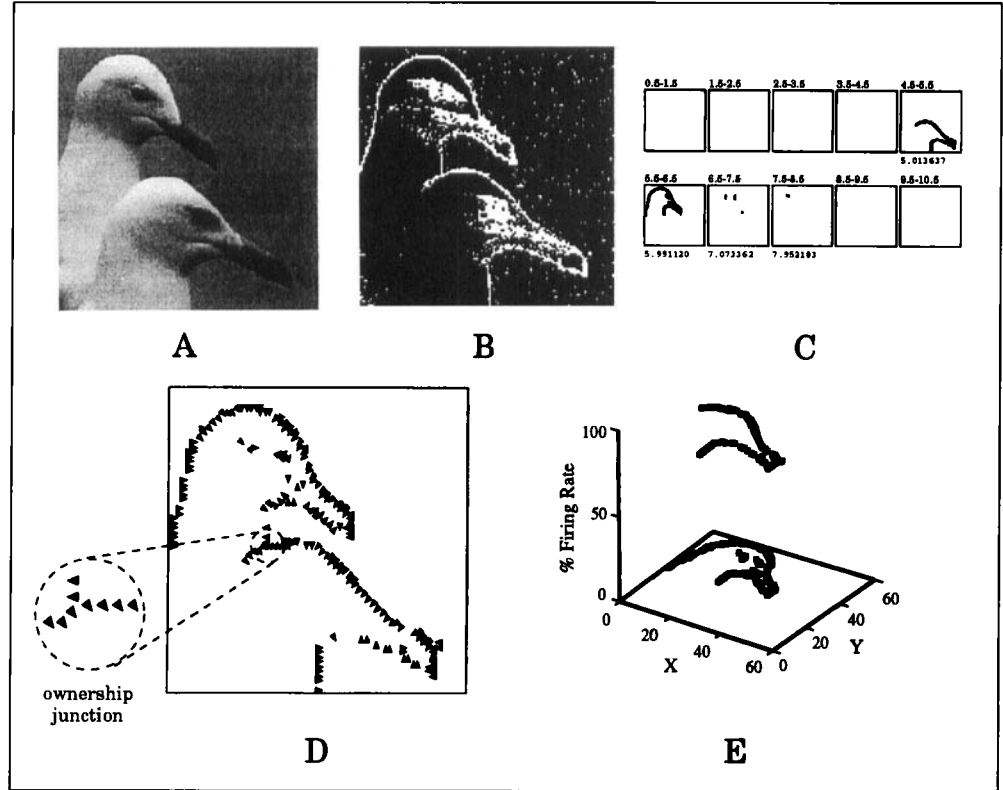
### Ambiguous and Inconsistent Figure-Ground

Certain classes of figures prove difficult in determining an unambiguous figure-ground relationship. Figure 19A is a stimulus for which human observers report difficulty instantaneously discriminating the surface, or inside, of the object without serially tracing the contour. This is in contrast to Figure 19B, where the surface of the object can be determined immediately.

Figure 19A shows simulations of the model's response for the ambiguous case. In this figure all contours are connected and therefore all units are temporally bound. Near the periphery of the object the figure is discriminated unambiguously since TDSI circuits detect the difference in closure across the boundary. However, near the center there is no closure difference. The parallel segments in the central region, which could presumably excite SDTI circuits, are roughly equidistant, so factors such as similarity or proximity also cannot be utilized. The DOF columns in the interior region of the spiral therefore remain inactive or oscillate between the two possible directions of figure, both of which represent ambiguous interpretations by the model.

If the interior line segments in the figure are not equidistant (for example, in the case of a spiral loosely wrapped around itself) then one would expect that the break in symmetry would enable the model to disambiguate figure from ground. This is indeed the case, shown in the simulations of Figure 19B, with the model maintaining a stable and consistent interpretation of the inside surface.

**Figure 18.** (A) 128x128 pixel gray-scale image. (B) Results of filtering image with Sobel filters. This edge image is first down-sampled by a factor of 2 and then input to the model. (C) Binding and segmentation. (D) Direction of figure computed by the network. (Inset) An ownership junction that is detected even though occluding contours are incomplete. (E) Resultant computed depth of surfaces.



**Figure 19.** (A) Figure in which it is difficult to discriminate figure-ground without serially tracing the contour. Simulations show the model has similar difficulties. Dashed segments represent points where the direction of figure is ambiguous (DOF units at these locations are either inactive or oscillate between two possible directions). (B) Spiral in which figure-ground relationship can be discriminated immediately. (C) Figure constructed by varying thickness of the spiral. The model predicts that such a figure would fool subjects into making an incorrect (and inconsistent) figure-ground determination. Simulation results (insets) show that the model discriminates region 1 in the periphery and region 2 in the interior as the figure, though the two are globally inconsistent.

These results lead to a specific prediction, namely, that varying the width of the spiral as it is wound around itself would create asymmetries that are inconsistent within the figure. These asymmetries should fool subjects into misclassifying figure-ground in the interior. Figure 19C shows an example. Both human subjects and the model tend to discriminate the central thin region as the inside of the spiral when, in fact, it corresponds to the outside—thus figure and ground are globally inconsistent.<sup>5</sup> A lack of global consistency checking in human visual perception has been noted by Kanizsa (1979), the consequences of which are observable in a multitude of “impossible” objects (e.g., Penrose’s triangle, Escher’s staircase, and Cheney’s unpatentable machine).

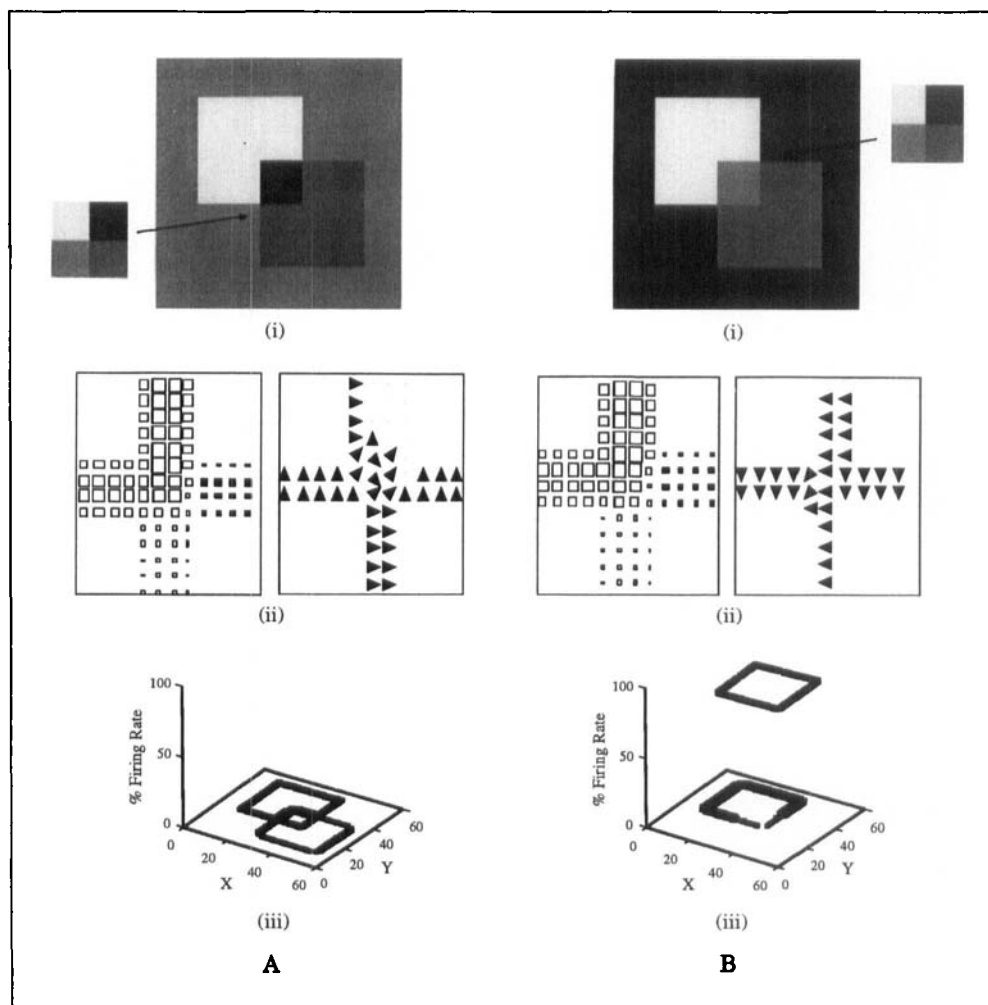
### Depth-from-Transparency

A major determinant of phenomenal transparency is the contrast relationship at surface junctions, for example, whether a given surface is contrast reversing or preserving and/or contrast reducing or enhancing (Kersten, 1991). Contrast reversing surfaces are not usually perceived as transparent, while contrast preserving surfaces are if they are also contrast reducing. Though contrast

reversal at a junction can be determined by simply considering contrast polarity, contrast reduction or enhancement requires the integration of both figural and chromatic factors (Metelli, 1974).

Figure 20 shows two stimuli that illustrate the need for integrating both figural and contrast information in determining transparency. For most observers, the figure on the right appears to contain a transparent square, that on the left does not (but see below). However, the lower junction in Figure 20A is identical, in terms of local contrast, to the upper junction in Figure 20B. One edge represents an edge owned by a contrast reversing surface and the other a contrast enhancing or reducing surface, depending on the surface segmentation. To discriminate between the two cases, the model combines local contrast with ownership, both of which are shown in the middle of Figure 20. In Figure 20A, ownership at the junction specifies the surface bound to the vertical edge as contrast reversing and the surface bound to the horizontal edge as contrast preserving and enhancing. This configuration is not perceived as transparent. (iii) Neither surface is stratified in depth. (B) (i) Input stimulus [luminance at junction same as A (i)]. (ii) Contrast at junction. Note that the local contrast is the same as in A (ii). (Right) Direction of figure at junction [compare to A (ii)]. This configuration specifies the horizontal edge as contrast preserving and reducing, forcing the lower square to be perceived as transparent. (iii) Transparent surface is stratified in depth.

**Figure 20.** Surface representation and transparency. (A) (i) Input stimulus (inset) luminance at junction. (ii) (Left) Activity at junction for units selective to contrast (simple cell units in *Contour* stage). Size of square is proportional to unit firing rate. Sign of contrast is represented as filled/unfilled square. (Right) Direction of figure at junction. Integration of contrast and direction of figure specify the vertical edge as contrast-reversing and the horizontal edge as contrast-preserving and enhancing. This configuration is not perceived as transparent. (iii) Neither surface is stratified in depth. (B) (i) Input stimulus [luminance at junction same as A (i)]. (ii) Contrast at junction. Note that the local contrast is the same as in A (ii). (Right) Direction of figure at junction [compare to A (ii)]. This configuration specifies the horizontal edge as contrast preserving and reducing, forcing the lower square to be perceived as transparent. (iii) Transparent surface is stratified in depth.



therefore transparent and will tend to be perceived as being closer to the viewer (simulation results shown at the bottom of Fig. 20). This example illustrates that ownership can serve as a local representation, which, when integrated with local contrast, can be used to discriminate between contrast enhancing and contrast reducing surfaces, thereby determining depth-from-transparency.

As an aside, consider the effect high-level processing can have on intermediate-level representations. If a subject is presented with Figure 20A and told to view it as if the lighter square is a hole in a transparent sheet, then the scene tends to reorganize—one perceives a transparent aperture, behind which is a dark square. Within the context of the model, the verbal instructions to the subject could force the “flipping” of the direction of figure along the contour of the white square, changing the surface from contrast-enhancing to contrast-reducing and resulting in the construction of a hole.

### Construction of Illusory Surfaces

The amodal completion of occluded objects, as demonstrated in the simulation of the cup occluding a pen, is a distinctive property of human visual perception. Modal completion is the complement of amodal completion in that it corresponds to the construction of *occluding* objects, and the two processes operate within the model as similar mechanisms. The commonality between modal and amodal completion has been pointed out by Kellman and Shipley (1991) and is exemplified in several of their psychophysical demonstrations. Illusory contours are the classic perceptual demonstration of modal completion, an example of which is the Kanizsa triangle (Kanizsa, 1979), shown in Figure 21. For this figure,

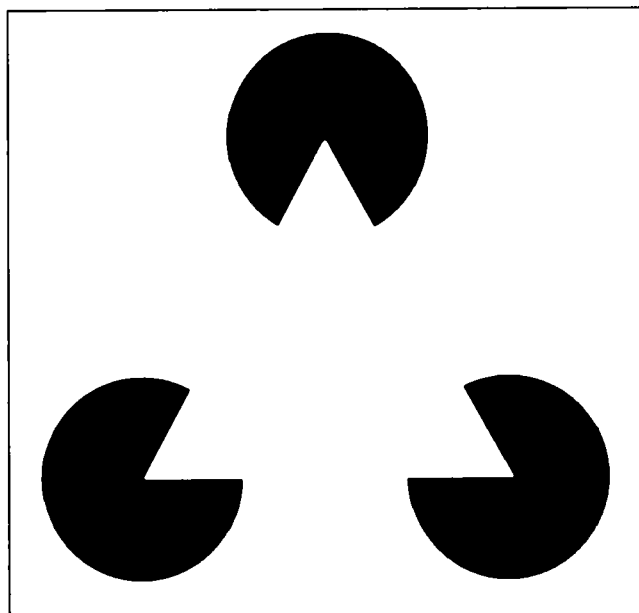


Figure 21. Kanizsa triangle.

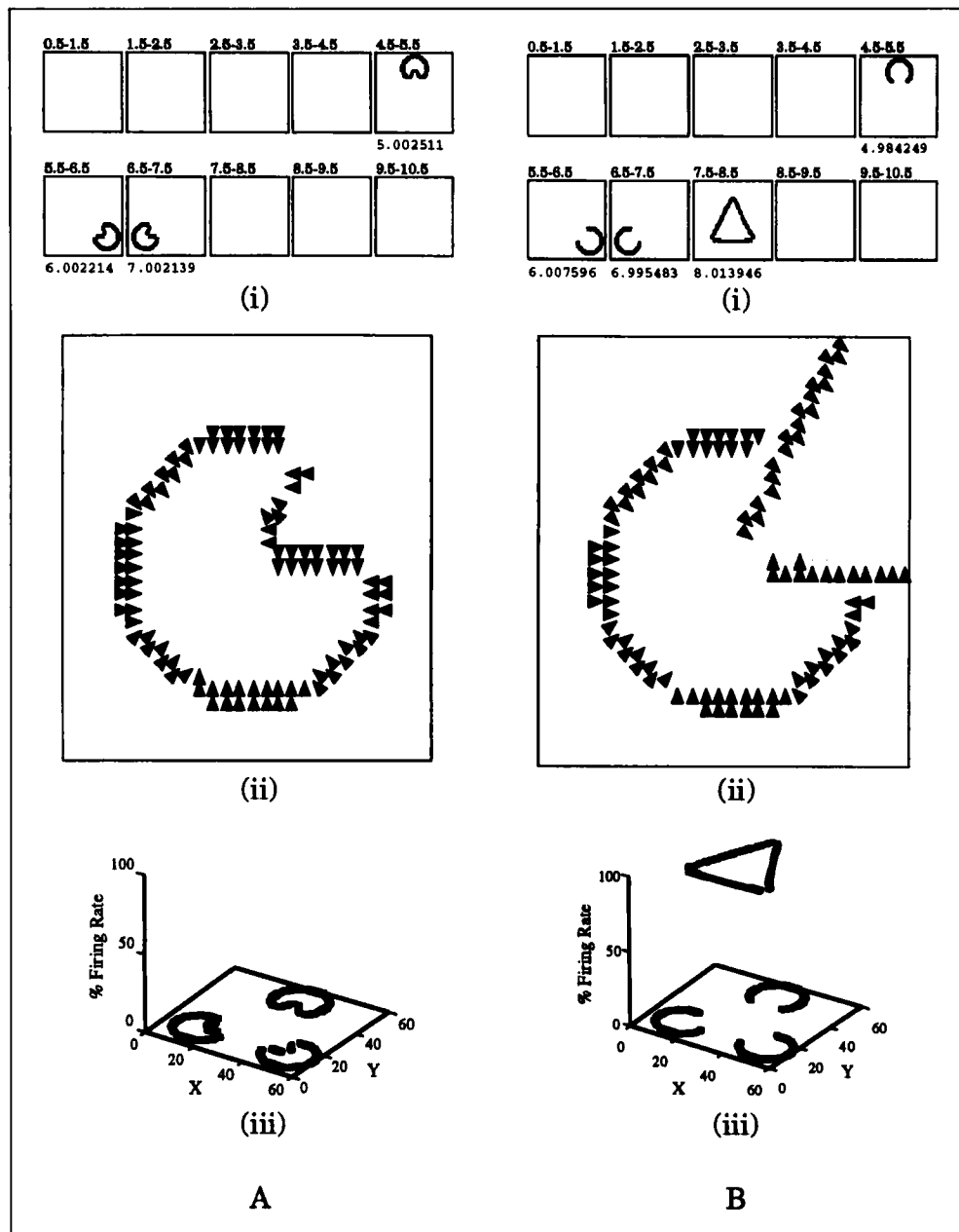
subjects perceive a white triangle with sharply defined contours. The triangle is seen as in front of and occluding three black discs.

In this section we present simulations that argue that the perception of illusory contours is a result of intermediate-level processes involved in the construction of illusory surfaces. The Kanizsa triangle serves as the input stimulus, with simulation results shown in Figure 22A and B. Initially the “pacmen” figures are bound and segmented as individual surfaces. However, the L-shaped contours of the pacmen represent contour concavities, and, as mentioned earlier, such concavities are used by the model as secondary, or transient, cues for direction of figure and ownership. Therefore in the model, concavities bounded by discontinuities are transiently transformed into convexities by “flipping” direction of figure. This transformation overwhelms the TDSI and SDTI circuits, with the net result being the formation of an ownership junction (see Fig. 12B). These transient ownership junctions excite completion units that generate the illusory contours. The illusory contours are reentered, via feedback, into the complex units of the *Contour* stage. This reentry is critical in that it enables the system to construct illusory surfaces—the illusory contours are grouped and segmented, their direction of figure is determined, and they are stratified in relative depth. Figure 22B shows simulation results after the illusory contours have been reentered through the system. Note that the direction of figure of the L-shaped contour has “flipped,” a result of the TDSI (closure) circuits operating on the illusory contours, and therefore the surface of the illusory triangle now owns the contour.

In considering the perception of illusory figures, an important observation is that changes in inducer shape can dramatically affect the perceived strength or vividness of the completion. For example, subjects tend to rank the vividness of the illusory figures in Figure 23 from left to right, with the figure on the left being the most vivid and that on the right the least. We believe our model can account for this effect in terms of the strength of binding between the contours and illusory surface.

Figure 24 shows the individual components contributing to this binding: the TDSI direction of figure circuits, the SDTI direction of figure circuits, and the transient change of the concavity into a convexity. For a typical inducer, such as the pacman in Figure 24, the TDSI and SDTI circuits tend to bind the L-shaped segment to region 1 while the concavity/convexity transformation tries to bind the segment to region 2. As we have said, this transient transformation overwhelms the TDSI and SDTI responses, so that the L-shaped segment is bound to region 2. However, the TDSI and SDTI activation will affect the magnitude of the direction of figure. For example, the weaker the activation of the TDSI and SDTI circuits, the stronger the activation of the DOF units binding the L-shaped segment to region 2.

**Figure 22.** The construction of an illusory surface. The Kanizsa triangle is the input to the model. (A) Simulation results before reentry of illusory contours. (i) Binding and segmentation of pacman figures. (ii) Direction of figure for the lower left pacman inducer—the L-shaped contour segment is initially owned by the pacman. (iii) All three pacmen are confined to the same background depth. (B) L-shaped contour segments are transiently transformed into convexities, resulting in the creation of ownership junctions, which in turn initiate completion. Completion is reentered at the *Contour* stage, resulting in the construction of an illusory surface. (i) Binding and segmentation of real and illusory surfaces. (ii) Direction of figure has “flipped” at the L-shaped contour segment—the illusory triangle now owns the contour. (iii) Illusory triangle is pushed forward in depth.



Referring back to the inducer types in Figure 23, one can see that though the TDSI component is the same for all three inducers (i.e., all three generate the same amount of closure) the SDTI contribution differs—the inducers differ in the amount of “similarity” and “proximity” that exists between the L-shaped segments and the rest of the object. This SDTI component for the three inducers is shown in Figure 23A. The contribution of the SDTI circuit is greatest for the inducers of Figure 23 (iii) and least for those in Figure 23 (i)—the L-shaped segments in (i) are more strongly bound to the surface of the illusory square than in (iii) since the magnitude of the direction of figure is larger in (i), as shown in Figure 24B. This result can be interpreted as the model’s ordering of perceived vividness, which in fact is consistent with that of human observers.

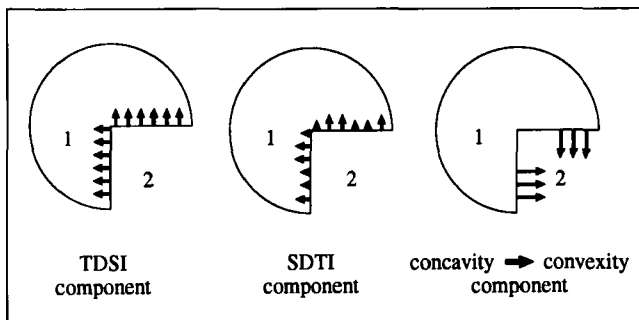
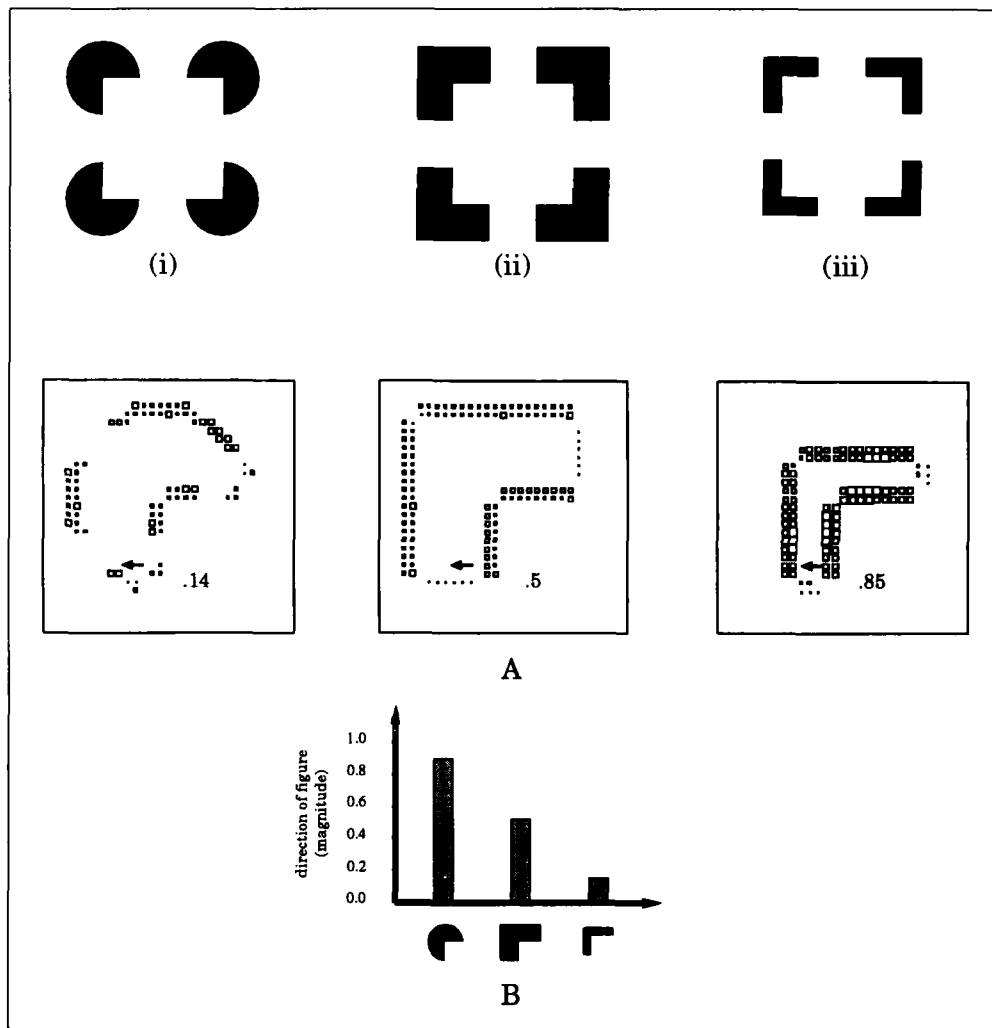
A prediction of the model is that a stimulus such as Figure 25 should generate a vivid illusory surface since the SDTI component for this inducer is very small ( $\approx 0$ ). Clearly this stimulus generates vivid illusory contours, more vivid than either Figure 23 (ii) or (iii). This prediction is a distinguishing feature of the model in that it illustrates how subtle aspects of illusory contour perception can be explained by the interaction between intermediate-level grouping processes.

### Comparison with a Case Study

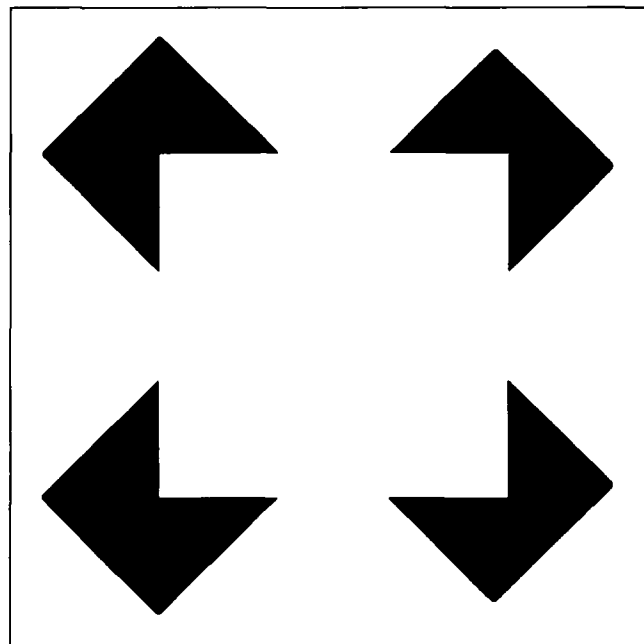
The final simulation compares the model’s performance with that of a patient described in a neuropsychological case study. Stevens (1981) reported that J. R., an elderly male who suffered a cerebral vascular accident, was



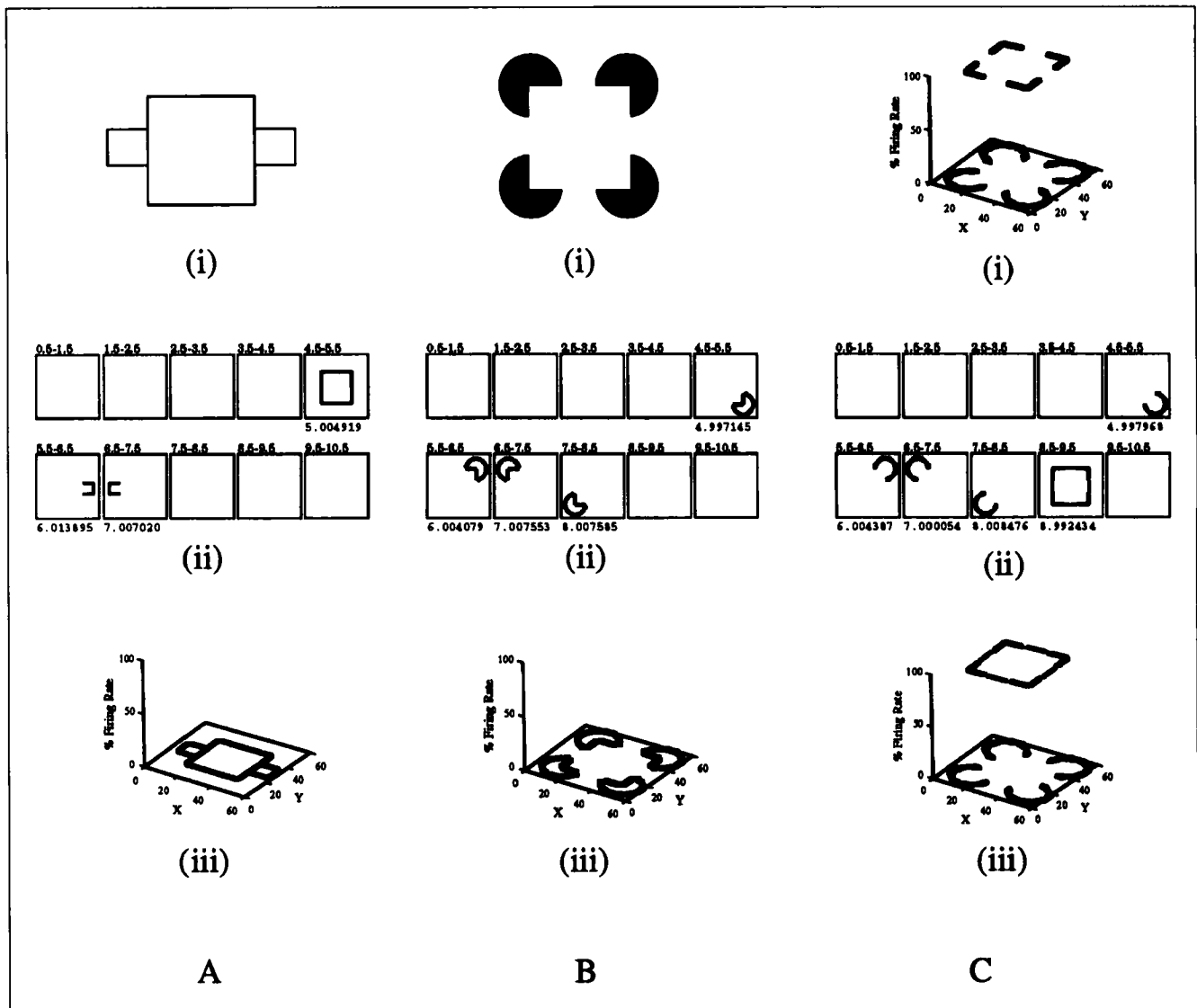
**Figure 23.** Illusory contour vividness as a function of inducer shape. Three types of inducers are arranged to generate an illusory square. (i) pacman inducer, (ii) thick L inducer, and (iii) thin L inducer. Subjects rank the vividness of the illusory squares from left to right [(i) > (ii) > (iii)]. (A) Activity of SDTI units for the upper left inducer of each stimulus, where the area of each square is proportional to unit activity. The SDTI units try to bind the L-shaped segment to the region of the pacman. Numerical values indicate the magnitude of the SDTI effect. (B) Magnitude of direction of figure along the L-shaped segment as a function of inducer shape. The direction of figure in all cases is toward the region of the illusory square. This implies the vividness ranking [(i) > (ii) > (iii)], consistent with human observers.



**Figure 24.** Processes contributing to the binding of the L-shaped contour segment to a region (i.e., processes involved in determining the segment's ownership). The TDSI and SDTI direction of figure circuits bind the contour to region 1, while the transient change of the concavity to a convexity forces the binding of the segment to region 2. The total of these three components determines the "strength" of binding of the segment to the illusory surface and, therefore, illusory contour vividness. Length of arrows represents the magnitude of the binding between the local segment of contour and the region.



**Figure 25.** Stimulus predicted by the model to generate vivid illusory contours and a distinct illusory surface.



**Figure 26.** Simulated cerebral trauma for comparison of the model with a clinical case study. Model was traumatized by transecting connections between the *Surface* and *Depth* stages. (A) The transected model is unable to amodally complete objects or (B) generate illusory contours given only monocular cues, consistent with the patient's performance. (C) Simulations showing that when stimulus *B* (i) is presented stereoscopically, with L-shaped contour segments forced to the foreground by binocular disparity, the model generates illusory contours.

unable to make the usual occlusion interpretations, although he showed normal performance on a set of standard illusions such as the Necker cube, Müller-Lyer, and Hering illusions. For example, stimuli such as Figure 26A (i) were consistently interpreted as two rectangles juxtaposed with a square, not as a square occluding a rectangle. J. R. was also unable to see illusory contours given only monocular cues. However, if the same stimuli were presented stereoscopically J. R. immediately reported seeing illusory contours.

We compare the model against this case study by attempting to incorporate similar processing deficiencies in the system and testing to see if the network produces analogous behavior. Given the extent of J. R.'s cerebral damage, it is difficult to identify the precise processes, or even areas, that were affected. However, based on his

responses to the test stimuli it appears that J. R. retained much of his ability for localized 2-D surface perception. For example, when shown a figure of a Kanizsa triangle he clearly described the figure as consisting of three black shapes (notched discs), though always rejecting the suggestion that the notches formed the corners of a white triangle. Since J. R. could discriminate the individual pacmen as figures, one would assume that he maintained his ability to compute direction of figure. In addition, the fact that he could complete stimuli if they were presented stereoscopically suggests that completion processes remained intact. Therefore, within the context of the model, J. R.'s primary deficiency appears to be in detecting ownership junctions based on pictorial cues. To simulate this deficiency, we lesioned the model and eliminated the input from pictorial direction

of figure cues to the formation of ownership junctions. Figure 26A, B, and C shows simulation results using this “transected” version. The first simulation tests the model’s ability to complete figures given only monocular stimuli. Figure 26A and B shows the transected model is incapable of both amodally completing occluded surfaces and generating illusory contours, both of which are consistent with J. R.’s performance. The model interprets the simple occlusion figure as composed of three objects, all of which are at the same relative depth. Likewise, the monocular version of the Kanizsa square is interpreted as four pacman figures, with no generation of illusory contours.

We next tested the model’s ability to construct illusory contours given binocular input. Though the model does not compute depth from binocular disparity, we simulate the effect of stereo input by artificially changing the activity of units encoding relative depth. Figure 26C shows the simulation results where relative depth units are clamped. Feedback from relative depth networks creates ownership junctions and drives the construction of an illusory square. Thus as with J. R., the model is able to form illusory contours when depth is available from another source, such as stereopsis.

## DISCUSSION

The apparent ease with which our visual system is able to organize a scene into objects and surfaces raises the question of how these percepts are defined within the nervous system. An answer to this question requires consideration of how neural representations are constructed. In this paper we describe the development of a model that serves as a test bed for considering how different representations may arise and how they influence one another. Developed within the framework of the problem of depth-from-occlusion, the model consists of three main stages representing a progression from one-dimensional to two-dimensional to three-dimensional scene organization, with feedback between stages facilitating the interaction between different neural representations. Utilizing computer simulations, we have illustrated how the model accounts for a variety of perceptual phenomena related to surface perception.

### Predictions and Comparisons to Other Approaches

Several predictions and particular features of this model distinguish it from previous models.

#### *Spatial and Temporal Mechanisms for Binding and Segmentation*

The model proposes the existence of two complementary mechanisms for binding and segmenting elements in a visual scene: spatial binding and temporal binding.

Temporally based binding has recently received much attention and, as we have mentioned, a number of models have considered temporal events, either in the form of coherent oscillations or phase-locked firing, as a possible substrate for binding neuronal activities. A temporal-based mechanism is particularly attractive for segmenting spatially juxtaposed elements. One difference between our model and those previously proposed is that most other models treat temporal binding as a “tag” for labeling different parts of a scene—an additional parameter by which units in a distributed representation can be bound together. As Hummel and Biederman (1992) pointed out, any parameter could serve such a role. However, in our model temporal binding is more than just a tag—the value itself is used to compute figure-ground relationships, complete surfaces and contours, and determine relative depth given occlusion and transparency.

It is interesting to speculate, within the context of the model, on the relationship between spatial and temporal binding and visual attention. Given its high resolving power and resource limitations, temporal binding might be equated with early processes related to attention. If the effectiveness of temporal binding were to be reduced, for example, by presenting stimuli tachistoscopically so that such mechanisms could not “engage,” spatial binding would operate exclusively. However, the low resolving power of spatial binding and the lack of an exact retinotopic mapping between visual areas imply a “free-floating” of features—spatial binding will tend to result in a greater number of binding errors. Experimental evidence in this direction comes from Treisman and her colleagues (1980, 1982), showing that for tachistoscopic presentations, illusory conjunctions are created.

#### *Direction of Figure as a Surface Representation*

The model argues that ownership is ideally represented as a vector, which we term direction of figure. Though evidence is sketchy on where direction of figure might be represented *in vivo*, an obvious candidate seems to be area V2, given the previously mentioned findings of von der Heydt and colleagues (1984). One possible paradigm for discovering cortical cells selective for direction of figure would be to use the stimuli of Figure 17. One would expect that though the inputs to the receptive fields are identical for the two stimuli, direction of figure cells would respond differently due to the difference in contour ownership. For example, if the cell in question were tuned to a leftward direction of figure (180°) then as stimulus A passed over the cell’s receptive field (from left to right) one would expect to see excitation followed by inhibition. Conversely, passing stimulus B over the cell’s receptive field would result in inhibition followed by excitation, indicating the cell’s preference for a leftward direction of figure. A closely related experi-

ment has recently been carried out by von der Heydt and colleagues, with results suggesting that cells in V2 may be sensitive to the direction of figure-ground (Peterhans & von der Heydt, 1992)

Other neural computational theories of surface discrimination have been largely based on "coloring" processes such as diffusion or region growing (Grossberg & Mingolla, 1985; Ullman, 1984), with neural activity initially "deposited" on both figure and ground and allowed to spread or diffuse via network connectivity. We have already mentioned that these systems have difficulty for cases in which there is not complete contour closure. Also problematic is the question of how multiple and overlapping surfaces are represented in such "coloring" models. One solution is to have different neuronal activations represent different surfaces. However, a diffusion-based mechanism cannot guarantee, a priori, a unique activity value for each surface. Another solution is to have a new set of neural units dynamically allocated for each surface (Grossberg & Wyse, 1992), though there is little psychological or neurophysiological evidence for such a representation. However, a vector representation of ownership lends itself well to discriminating multiple and overlapping surfaces and is biologically plausible given known neural encoding schemes, for example, the representation of motion.

The model currently includes two primary mechanisms and two secondary mechanisms that cooperate to compute direction of figure. Since we have considered only rather simple, static scenes, we do not wish to imply that these four mechanisms operate exclusively. Almost certainly there are additional neural circuits that could utilize other cues, such as differences in motion velocity, to compute direction of figure. In fact one could imagine augmenting the columnar architecture we have described by adding layers that compute direction of figure via cues such as motion and disparity.

#### *Role of Ownership in Cortical Integration*

An important tenet of the model is that ownership serves as a medium for cortical integration. Table 1 summarizes its role in integrating depth, transparency, and completion. For example, stratification of illusory surfaces is done by way of ownership. An illusory contour is constructed, resulting in a change of ownership at the border (Table 1, completion to ownership). This new ownership configuration is detected and utilized to stratify surfaces in relative depth (ownership to depth). Depth-from-transparency follows a similar path (transparency-to-ownership, ownership-to-depth). As we have mentioned there is evidence that other cues, such as motion, are also linked to ownership. For example, psychophysical demonstrations have shown that motion can affect ownership (Kersten, Bulthoff, Schwartz, & Kurtz, 1992) and ownership can likewise affect perceived motion (Shimojo et al., 1989; Stoner, Albright, &

Ramachandran, 1990). Thus ownership may be a general site of integration, enabling the construction of a common surface representation.

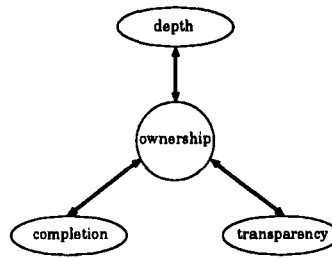
The concept of ownership distinguishes our model from other approaches to surface segmentation, such as the Grossberg BCS/FCS theory (Grossberg & Mingolla, 1985; Grossberg, 1993). Grossberg was the first to develop a neural model of how the visual system might construct surface representations. He provided explanations for a number of Gestalt phenomena, and attempted to account for a wide range of psychophysical phenomena with his models. The BCS/FCS model differs from ours in most of its mechanistic details: the use of diffusion rather than relative closure to determine figure-ground, and the relatively early generation of illusory contours at endstopped lines by the boundary contour system as opposed to our generation of illusory contours as a result of defining a surface. More important than these technical details, however, the models differ in the nature of how surfaces are represented. We predict that the cortex maintains an *explicit* representation of ownership, and that this representation is the key to both the construction of surfaces and the integration of visual processing. Failure to find an explicit representation of ownership would falsify our approach.


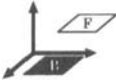



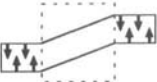









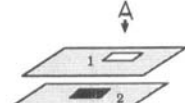
#### **Limitations of the Current Model and Areas for Future Research**

The model we have proposed has several limitations. Some limitations are due to our choice of a macroscopic, nonbiophysical implementation. For example, our model cannot be compared to other models in terms of processing time since we have implemented temporal binding as a network abstraction, independent of the neurophysiology. Similarly, our model of a complex unit lacks many of the details found in more sophisticated models of complex cells. Complex cells may, in fact, represent one of the earliest stages of integration, combining motion, luminance, and disparity information to construct a single representation of form discontinuities. A more complete model might, therefore, include a detailed instantiation of a complex cell, which accounts for selectivities to multiple cues.

The model includes one example of how complex units might integrate form discontinuities, indicated in the model by their response to both veridical and illusory contours. It has been recognized for some time that neurons in V2 respond to both real and illusory contours (von der Heydt et al., 1984). It may, therefore, be incorrect to consider our complex units as the analogs to complex cells in V1 since they may be more representative of neural populations in extrastriate areas such as V2. There is, however, some recent evidence from macaque monkey that indicates that cells in V1 also respond to illusory contours (Grosz, Shapley, & Hawken, 1993). It is not known whether these responses are a

**Table 1.** Summary of the Role Ownership Plays in Integrating Depth, Transparency, and Completion



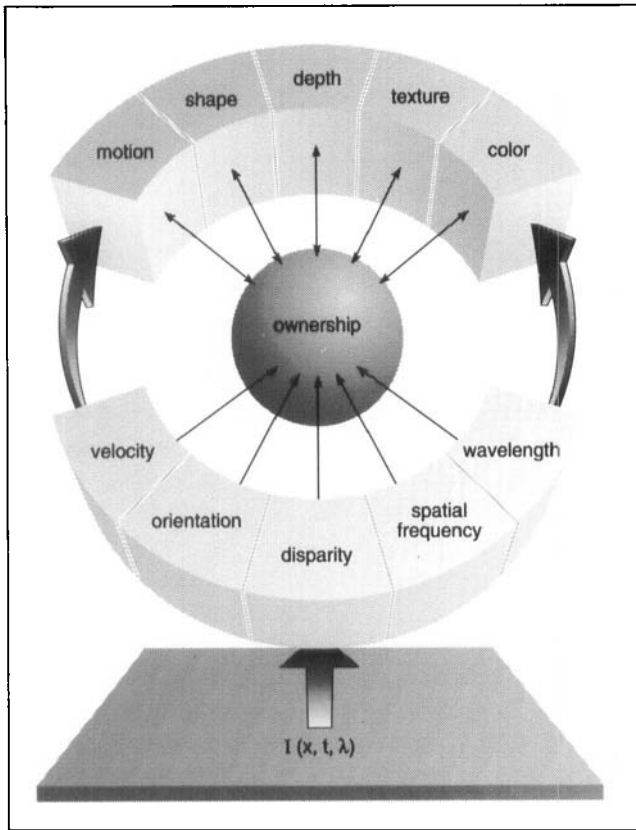
Integration Path	Example
ownership to depth	  <span data-bbox="1273 646 1384 693">stratification of surfaces</span>
depth to ownership	  <span data-bbox="1252 810 1397 836">opaque aperture</span>
ownership to completion	  <span data-bbox="1273 944 1367 991">amodal completion</span>
completion to ownership	   <span data-bbox="1273 1081 1367 1127">subjective contours</span>
ownership to transparency	    <span data-bbox="1245 1210 1405 1291">contrast enhancing vs contrast reducing surfaces</span>
transparency to ownership	   <span data-bbox="1268 1374 1372 1421">transparent aperture</span>

result of feedback from higher cortical centers. Some support for at least the possible existence of feedback comes from psychophysical experiments showing that illusory contours persist longer than real contours (Meyer & Yu Ming, 1988).

We have taken for granted that much of early-level processing, such as edge extraction and detecting line endings, can be done quite reliably. However, it has been repeatedly shown (Malik & Perona, 1992; Freeman, 1992; Nitzberg, Mumford, & Shioh, 1993) that it is difficult to design a reliable set of filters selective for low-level features relevant to occlusion. Indeed, determining

whether an edge is part of an occluding contour remains a difficult problem. Further investigation is needed into how occluding contours are reliably discriminated and whether this discrimination can proceed strictly bottom-up or requires top-down influences.

Our model has concentrated on a small set of pictorial cues and has not explicitly considered how visual properties, such as motion, are integrated into the system. A generalization of the model would explicitly incorporate a cue-invariance within the different stages. For example, a generalized *Contour* stage binds and segments boundaries, regardless of their origin. Likewise direction of



**Figure 27.** Proposed role of ownership in intermediate-level vision. A schematic illustration of the central role of ownership in transforming the stimulus-based representations of early vision (velocity, orientation, disparity, etc.) into object-based representations (motion, shape, depth, etc.).

figure, as part of a generalized *Surface* stage, is computed using both pictorial cues, such as “closure,” and dynamic cues, such as accretion/deletion of texture and relative motion. Since visual cues are integrated at multiple stages, such a system is absent of well-defined parallel modules and is instead characterized by a less rigid, demodularized organization (Stoner & Albright, 1993).

The central theme of this paper has been the construction of visual representations of surfaces within intermediate-level vision. Though our model maintains several levels of representation, ultimately they become unified in the representation of *ownership*. We would expect ownership to be directly represented in the visual cortex. As indicated in Figure 27, ownership would function as a medium for integration between different cues and as a key component in the computation of shape, depth, texture, color, and motion. Marr argued that in studying the information processing of the brain, representation can be decoupled from computation and implementation. However, in this paper we have set out to show how the choice of visual representation may be dependent on both the overall computational strategy and the particular neural mechanisms. All three aspects

together provide a useful set of constraints in understanding how the cortex constructs a scene.

## APPENDIX

### Computing Direction of Figure

#### Temporal-Dependent/Spatial-Independent (TDSI) Circuit

The activity of the TDSI circuit selective for a direction of figure  $\alpha$  is computed by comparing the amount of closure on either side of a contour. Closure is computed by summing the temporal dependent inputs over all directions  $t$ , where units on the same surface contour have similar temporal binding values,  $t_i$ , due to contour binding

$$TDSI^\alpha = \left[ \sum_t s_t^\alpha(t_i) - \sum_t s_t^{\alpha-180^\circ}(t_i) \right]_0^1 \quad (1)$$

The brackets ( [ ] ) indicate an implicit thresholding (if  $x < 0$  then  $[x] = 0$ , if  $x > 1$  then  $[x] = 1$ , otherwise  $[x] = x$ ) and  $s_t^\alpha(t_i)$  and  $s_t^{\alpha-180^\circ}(t_i)$  are the temporal dependent inputs, computed as

$$s_t^\alpha(t_i) = \begin{cases} 1 & \text{if } \begin{cases} s_j > S_T \\ \text{and} \\ (t_i - \Delta t) < t_j < (t_i + \Delta t) \end{cases} \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where  $\alpha$  and  $\alpha - 180^\circ$  represent the regions on either side of the contour,  $s_j$  is the activation of a unit along the direction  $t$  ( $t$  varies between  $0^\circ$  and  $315^\circ$  by increments of  $45^\circ$ ),  $\Delta t$  determines the range of temporal binding values over which the column will integrate input ( $\Delta t = 0.5$ ), and  $S_T$  is the activation threshold.

#### Spatial-Dependent/Temporal-Independent (SDTI) Circuit

The activity of the SDTI circuit selective for a direction of figure  $\alpha$  is computed by comparing input from similar orientations on either side of a contour:

$$SDTI^\alpha = \frac{1}{S_{\max}} \left[ \sum_t s_t^\alpha(\theta_i) - \sum_t s_t^{\alpha-180^\circ}(\theta_i) \right] \quad (3)$$

where  $S_{\max}$  is a constant for normalizing the SDTI activity between 0 and 1 ( $S_{\max} = 2.8$ ) and  $s_t^\alpha(\theta_i)$  and  $s_t^{\alpha-180^\circ}(\theta_i)$  are spatial-dependent inputs selective for an orientation  $\theta$ , computed as

$$s_t^\alpha(\theta_i) = \begin{cases} c_{ij} s_f(x, y, \theta_j) & \text{if } \begin{cases} s_f(x, y, \theta_j) > S_T \\ \text{and} \\ \theta_i = \theta_j \end{cases} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

$$c_{ij} = K \left( \frac{1}{2\pi\sigma^2} e^{-\frac{(x^2 + y^2)}{2\sigma^2}} \right) \quad (5)$$

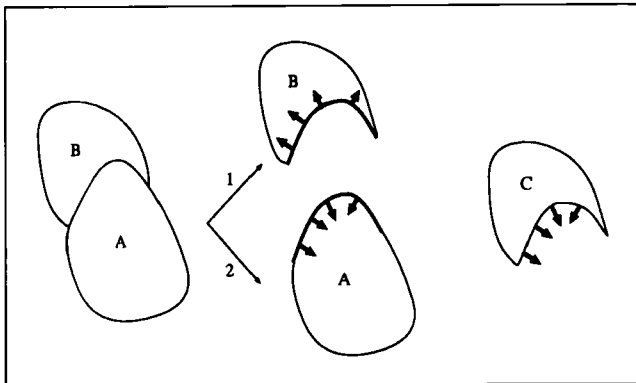
where  $\alpha$  and  $\alpha - 180^\circ$  represent the regions on either side of the contour,  $\theta$  is the orientation of the contour,  $i$  is the direction from which the unit receives input,  $c_{ij}$  is the connection strength, and  $s_j(x, y, \theta_j)$  is the activation of a unit along the direction  $i$ , which is mapped to retinotopic location  $(x, y)$  and selective for an orientation  $\theta_j$ . For simulations  $i$  varies between the following three angles:  $[\perp\theta_i, \perp(\theta_i - 45^\circ), \perp(\theta_i + 45^\circ)]$  and  $c_{ij}$  falls off as a gaussian with distance ( $K = 100, \sigma^2 = 100$ ).

### Secondary Cues—Concavities and Direction of Line Endings

Two secondary cues transiently affect direction of figure: concavities and direction of line endings. In the model a concavity is detected by comparing the direction of curvature (DOC) to the direction of figure (DOF). If they are the same then the contour segment is a convexity, if they are different it is a concavity. Concave regions, bounded by discontinuities (see below), are forced to flip their direction of figure for one cycle. The contribution of the concavity cue for a direction of figure  $\alpha$  is computed as

$$CON_\tau^\alpha = \begin{cases} 1 & \text{if } \begin{cases} DOF = \alpha - 180^\circ \\ \text{and} \\ DOF \neq DOC \end{cases} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where  $\tau$  indicates that this component is transient. Similarly, the direction  $\alpha$  of line endings (DLE) is computed as



**Figure 28.** Concavities as a cue to occlusion. The contour between surfaces A and B is either a concavity and owned by surface B (1) or is a convexity and is owned by surface A (2), where the second interpretation is the one that is consistent with occlusion. For a figure such as surface C, the concavity can be considered a cue for an occluder that owns the concave segment.

$$DLE_\tau^\alpha = \begin{cases} 1 & \text{if } s_{ES}^\alpha > S_T \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where  $s_{ES}^\alpha$  is the activation of an asymmetric endstopped unit selective for a line ending in the direction  $\alpha$ .

### DOF Unit Activity

The activity of a direction of figure unit representing a direction  $\alpha$  is given by the sum of the four components:

$$DOF^\alpha = C_1(TDSI^\alpha) + C_2(SDIT^\alpha) + C_3(CON_\tau^\alpha) + C_4(DLE_\tau^\alpha) \quad (8)$$

where the constants define the contribution of each cue (for all simulations reported here we used the identical set of coefficients:  $C_1 = 1.0, C_2 = 0.9, C_3 = 2.0, C_4 = 2.0$ ). The DOF activation is computed for all directions  $\alpha$  ( $\alpha$  can vary between  $0^\circ$  and  $315^\circ$  by increments of  $45^\circ$ ) with the direction producing the largest activation representing the direction of figure. We are currently investigating more sophisticated schemes in which the contribution of each cue is variable and context-dependent.

### Concavities as Cues to Occlusion

Concavities in a scene can serve as cues to occlusion. For example, Figure 28 is a simple illustration of surface occlusion. There are two ownership interpretations for the contour between surface A and surface B; the contour is a concavity and is owned by surface B or the contour is a convexity and is owned by surface A. Given a nonaccidental viewpoint (Koenderink & van Doorn, 1979; Richards, Koenderink, & Hoffman, 1987) the second interpretation is more likely. Now consider a surface C (right-most figure in Fig. 28) that has the same shape as B but now there is no occluder. The concavity bounded by discontinuities implies the presence of an occluder and likewise is a cue for direction of figure. In the model, discrimination of this cue results in a transient (one cycle) change in direction of figure to point toward the convex region (i.e., convexities are favored over concavities).<sup>6</sup>

### Simulations

Input to the model is in the form of  $64 \times 64$  image arrays. All images are binary valued, except for the two camera images and the transparency stimulus, which are 8 bit grayscale. Input images are normalized to the range  $[0.0-1.0]$ , representing the activation of the input array. Networks in the model are simulated as  $64 \times 64$  2-D arrays of units. All simulations are done using the NEXUS Neural Simulation Environment (Sajda & Finkel, 1992). The parameters used in all simulations described were identical. Simulations typically took 20 min per cycle running in parallel on three Sun SPARCstations. Most simulations

required one or two cycles, where a cycle is defined as one iteration of feedback. Thus stimuli in which feedback plays a crucial role, such as illusory contours, tend to require more processing time.

### Acknowledgments

This work is supported by grants from The Office of Naval Research (N00014-90-J-1864, N00014-93-1-0681), The Whitaker Foundation, and The McDonnell-Pew Program in Cognitive Neuroscience. Paul Sajda was supported by an ONR-NDSEG Fellowship.

Reprint requests should be sent to Dr. Paul Sajda, David Sarnoff Research Center, CN5300, Princeton, NJ 08543-5300.

### Notes

1. We have simplified our model and considered decomposing contours only via orientation selective units. One could generalize this mechanism and assume that the endstopped units that are used to link iso-orientation assemblies also extract curvature and can themselves be linked to form "iso-curvature" assemblies.
2. Contour binding is made to be a resource-limited process by allowing for only a finite number of discriminable ranges. One would expect this number to be a function of the temporal properties of the units (e.g., membrane time constants and synaptic delays). Since we do not explicitly model these properties, we approximate a limit given evidence from psychology. Crick and Koch (1990) have suggested that the binding limit is related to the capacity of working memory, which is taken to be roughly seven objects (Miller, 1956). Therefore, for our current model and subsequent simulations we will assume that the binding limit is of this order of magnitude.
3. The TDSI and SDTI circuits are particular examples within the model of the general temporal and spatial binding mechanisms discussed earlier.
4. It is interesting to speculate on a relation between these two types of feedback and the diffuse versus focal patterns of anatomical feedback found between higher and lower cortical areas (Zeki, 1990). For example, diffuse feedback could serve to synchronize, or temporarily blind, neural populations while more focal feedback could directly drive neuronal activity.
5. If the spirals are shaded so that there is contrast between the figure and the background then the direction of contrast can be used to disambiguate figure-ground (i.e., the relationship between direction of figure and direction of contrast in unambiguous regions, such as the periphery, can be used to constrain ambiguous regions in the interior).
6. Note that this is true only for concavities that are bounded by discontinuities. The conjunction of concavities and discontinuities tends to differentiate concavities resulting from occlusion from those that are simply intrinsic to the object's shape.

### REFERENCES

Adelson, E. H., & Bergen, J. R. (1985). Spatiotemporal energy models for the perception of motion. *Journal of the Optical Society of America A*, 2, 284-299.

Barlow, H. B. (1981). Critical limiting factors in the design of the eye and visual cortex. *Proceedings of the Royal Society (London) B*, 212, 1-34.

Bergen, J. R., & Landy, M. S. (1991). Computational modeling of visual texture segregation. In M. S. Landy & J. A.

Movshon (Eds.), *Computational models of visual processing* (pp. 253-271). Cambridge, MA: MIT Press.

Bullier, J., Munk, M. H. J., & Mowak, L. G. (1992). Synchronization of neuronal firing in areas V1 and V2 of the monkey. *Society for Neuroscience Abstracts*, 18, 11.

Crick, F., & Koch, C. (1990). Towards a neurobiological theory of consciousness. *Seminars in Neuroscience*, 2, 263-275.

Desimone, R., Wessinger, M., Thomas, L., & Schneider, W. (1990). Attentional control of visual perception: Cortical and subcortical mechanisms. *Cold Spring Harbor Symposium on Quantitative Biology*, 55, 963-971.

Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., & Reitboeck, H. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? *Biological Cybernetics*, 60, 121-130.

Elder, J., & Zucker, S. (1993). The effect of contour closure on the rapid discrimination of two-dimensional shapes. *Vision Research*, 33(7), 981-991.

Finkel, L., & Edelman, G. (1989). Integration of distributed cortical systems by reentry: A computer simulation of interactive functionally segregated visual areas. *Journal of Neuroscience*, 9, 3188-3208.

Finkel, L. H., & Sajda, P. (1992). Object discrimination based on depth-from-occlusion. *Neural Computation*, 4(6), 901-921.

Freeman, W. T. (1992). *Steerable filters and local analysis of image structure*. Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, MA.

Ghose, G. M., & Freeman, R. D. (1992). Oscillatory discharge in the visual system: Does it have a functional role? *Journal of Neurophysiology*, 68(5), 1558-1574.

Gilbert, C. D. (1977). Laminar differences in receptive field properties of cells in cat primary visual cortex. *Journal of Physiology (London)*, 268, 391-421.

Gilbert, C. D., & Wiesel, T. N. (1989). Columnar specificity of intrinsic horizontal and corticocortical connections in cat visual cortex. *Journal of Neuroscience*, 9, 2432-2422.

Gray, C. M., & Singer, W. (1989). Neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Science USA*, 86, 1698-1702.

Grosz, D. H., Shapley, R. M., & Hawken, M. J. (1993). Macaque V1 neurons can signal illusory contours. *Nature (London)*, 365, 550-552.

Grossberg, S. (1993). A solution of the figure-ground problem in biological vision. *Neural Networks* 6, 463-483.

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

Grossberg, S., & Wyse, L. (1992). Figure-ground separation of connected scenic figures: Boundaries, filling-in, and opponent processing. In G. A. Carpenter & S. Grossberg (Eds.), *Neural networks for vision and image processing* (pp. 161-192). Cambridge, MA: MIT Press.

Guzman, A. (1968). Decomposition of a visual scene into three-dimensional bodies. *Fall Joint Computer Conference*, 1968, 291-304.

Helmholtz, H. (1909). *Treatise on physiological optics* (Vol. 3). New York: Dover; translated from the third German edition; reprinted 1962.

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

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

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



- Kersten, D. (1991). Transparency and the cooperative computation of scene attributes. In M. S. Landy & J. A. Movshon (Eds.), *Computational models of visual processing* (pp. 209-228). Cambridge, MA: MIT Press.
- Kersten, D., Bulthoff, H. H., Schwartz, B. L., & Kurtz, K. J. (1992). Interactions between transparency and structure from motion. *Neural Computation*, 4, 573-589.
- Koenderink, J. J., & van Doorn, A. J. (1979). The internal representation of solid shape with respect to vision. *Biological Cybernetics*, 32, 211-216.
- Koffka, K. (1935). *Principles of Gestalt psychology*. New York: Harcourt, Brace.
- Kreiter, A. K., & Singer, W. (1992). Oscillatory neuronal responses in the visual cortex of the awake macaque monkey. *European Journal of Neuroscience*, 4, 369-375.
- Livingstone, M. S. (1991). Visually-evoked oscillations in monkey striate cortex. *Society of Neuroscience Abstracts*, 17, 176.
- Malik, J., & Perona, P. (1992). Finding boundaries in images. In H. Wechsler (Ed.), *Neural networks in perception* (pp. 315-344). San Diego: Academic Press.
- Marr, D. (1977). Analysis of occluding contour. *Proceedings of the Royal Society (London) B*, 197, 441-475.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W. H. Freeman.
- Metelli, F. (1974). The perception of transparency. *Scientific American*, 230, 91-95.
- Meyer, G. E., & Yu Ming, C. (1988). The visible persistence of illusory contours. *Canadian Journal of Psychology*, 42(4), 479-488.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63:81-97.
- Nakayama, K., & Shimojo, S. (1990). Toward a neural understanding of visual surface representation. *Cold Spring Harbor Symposia on Quantitative Biology*, LV, 911-924.
- Nakayama, K., Shimojo, S., & Silverman, G. H. (1989). Stereoscopic depth: Its relation to image segmentation, grouping, and the recognition of occluded objects. *Perception*, 18, 55-68.
- Nitzberg, M., Mumford, D., & Shiota, T. (1993). *Filtering segmentation and depth*. Lecture Notes in Computer Science. Berlin: Springer-Verlag.
- Peterhans, E., & von der Heydt, R. (1992). Selectivity for figure-ground direction at occluding contours in monkey area V2. *Society for Neuroscience Abstracts*, 18, 1275.
- 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, *Visual neuroscience* (pp. 343-352). New York: Cambridge University Press.
- Richards, W. A., Koenderink, J. J., & Hoffman, D. D. (1987). Inferring three-dimensional shapes from two-dimensional silhouettes. *Journal of the Optical Society of America*, 4(7), 1168-1175.
- Rockland, K. S., & Lund, J. S. (1982). Widespread periodic intrinsic connections in the tree shrew visual cortex. *Science*, 215, 1532-1534.
- Sajda, P., & Finkel, L. (1992). NEXUS: A simulation environment for large-scale neural systems. *Simulation*, 59(6), 358-364.
- Sejnowski, T. J. (1986). Open questions about computation in cerebral cortex. In J. L. McClelland & D. E. Rumelhart (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (pp. 372-389). Cambridge, MA: MIT Press.
- Shimojo, S., Silverman, G. H., & Nakayama, K. (1989). Occlusion and the solution to the aperture problem. *Vision Research*, 29, 619-626.
- Singer, W., Gray, C., Engel, A., Konig, P., Artola, A., & Brocher, S. (1990). Formation of cortical cell assemblies. *Cold Spring Harbor Symposia on Quantitative Biology*, LV, 939-952.
- Skrzypek, J., & Ringer, B. (1992). Neural network models of illusory contour perception. *Proceedings IEEE Computer Vision and Pattern Recognition*, 681-683.
- Sporns, O., Tononi, G., & Edelman, G. M. (1991). Modeling perceptual grouping and figure-ground segregation by means of active reentrant connections. *Proceedings of the National Academy of Sciences USA*, 88, 129-133.
- Stevens, K. (1981). *Evidence relating subjective contours and interpretations involving occlusion*. Technical Report A. I. Memo 637, MIT Artificial Intelligence Laboratory.
- Stoner, G. R., & Albright, T. D. (1993). Image segmentation cues in motion processing: Implications for modularity in vision. *Journal of Cognitive Neuroscience*, 5(2), 129-149.
- Stoner, G. R., Albright, T. D., & Ramachandran, V. S. (1990). Transparency and coherence in human motion perception. *Nature (London)*, 344, 153-155.
- Treisman, A., & Gelade, G. (1980). A feature integration theory of attention. *Cognitive Psychology*, 12, 97-136.
- Treisman, A., & Schmidt, N. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107-141.
- Ullman, S. (1984). Visual routines. *Cognition*, 18, 97-159.
- Van Essen, D. C. (1985). Functional organization of primate visual cortex. In *Cerebral Cortex*, Vol. 3: *Visual Cortex* (pp. 259-320). New York: Plenum.
- von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science*, 224, 1260-1262.
- von der Malsburg, C. (1981). *The correlation theory of brain function*. Technical Report Internal Rep. No. 81-2, Max-Planck-Institute for Biophysical Chemistry, Department of Neurobiology, Göttingen, Germany.
- von der Malsburg, C., & Schneider, W. (1986). A neural cocktail-party processor. *Biological Cybernetics*, 54, 29-40.
- Waltz, D. L. (1975). Understanding line drawings of scenes with shadows. In P. H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill.
- Wertheimer, M. (1938). *Laws of organization in perceptual forms*. London: Harcourt, Brace.
- White, M. A new effect of pattern on perceived lightness. *Perception*, 8, 413-416.
- Wilson, M. A., & Bower, J. M. (1991). A computer simulation of oscillatory behavior in primary visual cortex. *Neural Computation*, 3(4), 498-509.
- Zeki, S. (1990). Parallelism and functional specialization in the human visual cortex. *Cold Spring Harbor Symposia on Quantitative Biology*, LV, 651-661.