

Dynamic coding of border-ownership in visual cortex

Oliver W. Layton

Center for Computational Neuroscience and Neural
Technology, Boston University, Boston, MA, USA
Program in Cognitive and Neural Systems, Boston
University, Boston, MA, USA



Ennio Mingolla

Department of Speech-Language Pathology and
Audiology, Northeastern University, Boston, MA, USA



Arash Yazdanbakhsh

Center for Computational Neuroscience and Neural
Technology, Boston University, Boston, MA, USA
Program in Cognitive and Neural Systems, Boston
University, Boston, MA, USA



Humans are capable of rapidly determining whether regions in a visual scene appear as figures in the foreground or as background, yet how figure-ground segregation occurs in the primate visual system is unknown. Figures in the environment are perceived to own their borders, and recent neurophysiology has demonstrated that certain cells in primate visual area V2 have border-ownership selectivity. We present a dynamic model based on physiological data that indicates areas V1, V2, and V4 act as an interareal network to determine border-ownership. Our model predicts that competition between curvature-sensitive cells in V4 that have on-surround receptive fields of different sizes can determine likely figure locations and rapidly propagate the information interareally to V2 border-ownership cells that receive contrast information from V1. In the model border-ownership is an emergent property produced by the dynamic interactions between V1, V2, and V4, one which could not be determined by any single cortical area alone.

Keywords: figure-ground segregation, border-ownership, receptive fields, interareal connections, visual cortex

Citation: Layton, O. W., Mingolla, E., & Yazdanbakhsh, A. (2012). Dynamic coding of border-ownership in visual cortex. *Journal of Vision*, 12(13):8, 1–21, <http://www.journalofvision.org/content/12/13/8>, doi:10.1167/12.13.8.

Introduction

Most physiological studies of the primate visual system in the past half-century have followed the path established by Hubel and Wiesel (1962) and largely focused on the function of individual areas or subpopulations of cells within a visual area. Investigations outside the visual system indicate that the cortex can solve complex problems with networks that span multiple areas and whose functionally equivalent circuits are widely distributed throughout the cortex (Nieder & Miller, 2004). Our computational analysis indicates that the visual system may also rapidly recruit an assembly of cortical areas to determine border-ownership in figure-ground segregation, a single emergent function. Neuroanatomical evidence indicates that early visual areas such as LGN, V1, V2, and V4 are massively interconnected with numerous feedforward and feedback connections (Sincich & Horton, 2005). Feedforward connections are believed to quickly propagate sensory visual information to cortical areas

further up the visual hierarchy to subserve a rich perception of the visual scene. Feedback projections are often said to play a modulatory role with respect to bottom-up sensory visual signals by increasing the gain of neuronal responses in attended regions and performing contextual integration. To date, few studies have hypothesized that feedback projections subserve crucial as opposed to supplementary roles for the functions of early visual cortices. It is not clear whether the simultaneous activation of multiple areas early in the visual system only performs modular functions that are later combined or whether such activation can collectively solve problems that individual cortical areas cannot solve alone. We here introduce a computational model that provides a unified explanation for how several cortical areas act coherently to perform figure-ground segregation.

Distinguishing between an object (figure) and its background (ground) in a visual scene is required for performing important higher-order visual functions, such as object recognition. Although figure-ground segregation is fundamental to visual perception, how

the visual system performs it is not well understood. A direct link between visual figure-ground perception and the responses of certain single neurons has, however, been established in the early visual system. These cell responses may require the simultaneous activation of parts of visual areas V1, V2, and V4 acting as a functional network. Researchers have found that as many as 59% or 53% of sampled cells from primate visual areas V2 and V4, respectively, preferentially respond to borders when they form a certain side of a figure (Zhou, Friedman, & von der Heydt, 2000). This side-of-figure selectivity is known as border-ownership (see Figure 1a). For example, when the receptive field of a border-ownership cell (B cell; Craft, Schutze, Niebur, & von der Heydt, 2007) is centered on a vertical edge of a square, a stronger response (Figure 1d) may be elicited when the square is located to the left (Figure 1b) as compared to the right of the edge (Figure 1c), although the local contrast in the cell's receptive field remains the same. Such side-of-figure selectivity could indicate a neurophysiological correlate of the percept that a border is owned by either the region to

one side or another of that border, but not both. Figure 1a shows a well-known bistable display from Shepard (1990) in which the percept of figure alternates between a saxophone player and a female face. The central black-white border is said to belong to or be owned by whichever portion of the display is perceived as the figure. B cells have been shown to exhibit modulation due to bistable visual scenes.

Although B cells in the monkey V2 may demonstrate a border-ownership preference to a light square (Figure 1b) when the border intersects the cell's receptive field and the light square is positioned to the left compared to a dark square (Figure 1c) positioned to the right of the cell's receptive field (Figure 1d, light gray curve), the presence of a light square to the left and a dark transparent overlay to the right (Figure 1e) diminishes the border-ownership response (Figure 1g, light gray curve). Relative to this particular B cell whose receptive field is indicated by the ellipse, the figure is located to the right. When a border of a dark square appears within the receptive field of the same cell, the dark square is located to the right of the receptive field and a

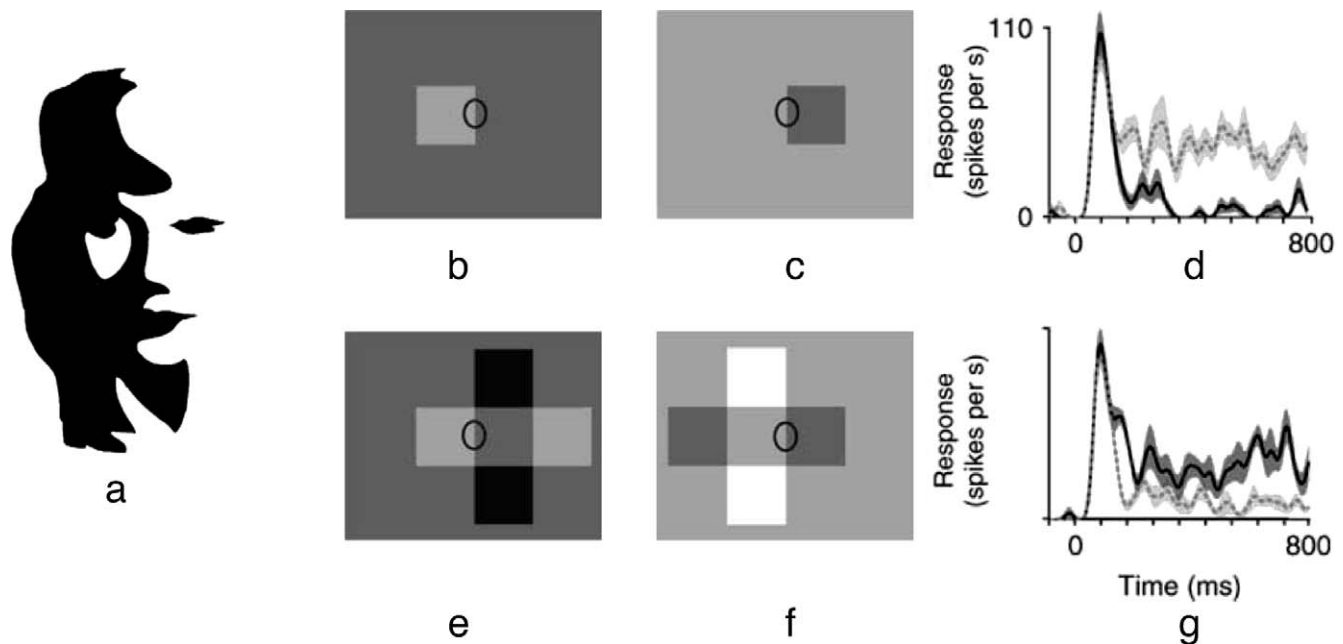


Figure 1. (a) Bistable image in which observers either perceive a saxophone player or a female face; from Shepard (1990). The central white-black borders alternate their direction of ownership with the scene interpretation. (b) The cell has the same light gray and dark gray in the left and right halves of its receptive field (ellipse), respectively, in both (b) and (c), but elicits a larger response (border-ownership preference) when the light gray patch is attached to a figure located to the left of the receptive field (b; d, light gray curve) compared to when the dark gray patch is attached to a figure on the right (c; d, dark gray curve). (e) The same B cell continues to prefer figures whose border enter the cell's receptive field and are located to the left despite the presence of transparent overlays. Although the cell yielded a stronger response when a light square appeared to the left compared to a dark square appearing to the right (d), the presence of a dark transparent overlay to the right (e) diminishes the cell's response (g; light gray curve). (f) When the local luminance configuration remains the same, but a light transparent overlay appears within and to the left of the cell's receptive field, the cell's response increased (g; dark gray curve). The (d) and (g) panels are adapted with permission from figure 1 of Qiu and von der Heydt (2007). The (a) panel is excerpted from *Mind Sights* by Roger N. Shepard. Copyright © 1990 by Roger N. Shepard. Reprinted by arrangement with Henry Holt and Company, LLC. All rights reserved.

light transparent overlay to the left (Figure 1f), the cell exhibits an increased response (Figure 1g, dark gray curve). This particular B cell prefers figures whose borders locally intersect the cell's receptive field and are located to the left of its receptive field center.

B cells demonstrate selectivity to figures in the global scene context far outside the cells' classical receptive fields. By virtue of their prevalence in early primate visual areas, B cells have small receptive field sizes both compared to many other visual cortical areas and the figures to which they respond. For example, Zhou et al. (2000) report that median B cell receptive field sizes at foveal eccentricities are 0.5°, 0.7°, and 3.6° in monkey visual areas V1, V2, and V4, respectively. Despite their small receptive field sizes, B cells respond to a consistent side-of-figure irrespective of the figure size as long as it still perceptually appears as a figure, a property known as *size invariance*. Evidence also indicates that an intercortical network that spans V1, V2, and V4 allows B cells to access global information about the figures in the visual scene irrespective of potentially conflicting local information, such as motion (von der Heydt, Qiu, & He, 2003), luminance (Zhou et al., 2000; Qiu & von der Heydt, 2007; Zhang & von der Heydt, 2010), or disparity (von der Heydt, Zhou, & Friedman, 2000). That is, B cells can register the appropriate side-of-figure response, despite their small receptive field size and potentially ambiguous local information. How B cells do this—which must surely involve interactions with cells in other higher cortical areas and with larger receptive fields—is the point of our model.

The time it takes for border-ownership signals to emerge constrains the type of intercortical network that can plausibly perform figure-ground segregation. Researchers have proposed that B cells access global information either *intra-areally*, i.e., by lateral connections within a single visual cortical area, such as V2 (Zhaoping, 2005), or *interareally*, i.e., where cells with larger receptive fields communicate contextual information about the scene via feedback projections to visual areas with small receptive field cells fewer synapses away from the retina (Angelucci et al., 2002). Intra-areal and interareal axonal conduction velocities have been estimated to be 0.3 m/s (Nowak, Munk, Girard, & Bullier, 1995; Nowak & Bullier, 1997; Angelucci et al., 2002) and 3.5 m/s (Girard, Hupe, & Bullier, 2001) in early visual areas, respectively (Bullier, 2001). Hence, interareal connections can be an order of magnitude faster than intra-areal connections for propagating information across the visual field. Sugihara, Qiu, and von der Heydt (2003) showed that B cell responses to 3° squares did not differ in latency compared to those to an 8° square, which is consistent with the use of interareal connections, but not intra-areal connections, to propagate contextual figure-

ground information. Although a variable amount of time is required to propagate information about a figure within a single cortical area, transmitting the information to another area with large receptive field cells could afford a roughly fixed delay irrespective of the figure size in the visual field. Using published neuroanatomical data (Gattass, Gross, & Sandell, 1981; Gattass, Sousa, Mishkin, & Ungerleider, 1997), we estimate the cortical distance along the horizontal meridian and spanning 0°–5° eccentricity within V2 as 22.25 mm, which agrees with prior estimates (Craft et al., 2007). Traversing such a distance at 0.3 m/s would take approximately 75 ms which cannot account for border-ownership latency of 10–25 ms reported in neurophysiology (Zhou et al., 2000). Hence, it appears that connections within a single cortical area alone could not plausibly account for the fast global scene integration that is observed in B cell border-ownership responses (but see Zhaoping, 2005, who argues otherwise as we describe in the Discussion section).

The visual system connects meaningful properties in the world to mechanisms in cortex. In nature discrete objects that primates interact with tend to be convex. Humans are more likely to interpret convex regions as a figure compared to those that are concave, irrespective of texture, color, and other low-level characteristics (Kanizsa & Gerbino, 1976; Peterson & Salvagio, 2008). One might expect that the natural convexity bias for figures has been mapped directly onto visual cortices. Cells throughout the visual system demonstrate on-center/off-surround selectivity and variants thereof, including off-center/on-surround. V4 neurons exhibit sensitivity to parametrically defined convex and concave curves of various orientations, acutenesses, and partial occlusion of figures (Bushnell, Harding, Kosai, & Pasupathy, 2011). Aside from possessing receptive fields larger than those at equivalent eccentricities in V1, V2, and V3, large numbers of V4 cells appear to have curved, radially symmetric, on-surround receptive fields (Hegde & Van Essen, 2006). With an appropriate receptive field size, neurons with on-surround receptive fields are capable of detecting conjunctions of curved contours. Depending on the number and alignment of the detected contours, the co-occurrence of multiple curved contours within the on-surround region of a receptive field may be important for forming partial shape representations. The interactions between neurons that associate multiple curved contours may be instrumental for the generation of more complete object representations. Neurons in V4 with large, donut-shaped on-surround receptive field organization can respond to the co-occurrence of curved contours that fall within the donut, which suggests the neurons are sensitive to Gestalt properties such as convexity and closure (Pizer, Burbeck, Coggins, Fritsch, & Morse, 1992; Pizer, Eberly, Morse, & Fritsch, 1998). V4 cells

may communicate convexity information about figures in the visual scene with B cells in V2 via fast interareal connections (Craft et al., 2007).

The goal of the present paper is to demonstrate that a simple model of neural competition between convexity-sensitive units with different receptive field sizes in model V2 and V4 connected by fast interareal fibers can account for a number of the properties of B cells reported by neurophysiological studies to date (Zhou et al., 2000; Qiu & von der Heydt, 2007). A successful model of border-ownership and figure-ground segregation in the primate visual system should satisfy the following constraints. (a) Model B cells should demonstrate side-of-figure selectivities toward regions in a visual image that are perceived by humans as figure, which is consistent with known neurophysiological border-ownership signals. For example, in the case of a square occluding a rectangle (Figure 2e), model B cells with receptive fields centered around the square's perimeter should elicit stronger border-ownership toward the square than the rectangle (Zhou et al., 2000). (b) Cortical networks that propagate global contextual information to small receptive field B cell units should bias border-ownership signals toward the inside rather than the outside of figures. Convex (on-surround) receptive fields are essential characteristics of cells extracting information about objects, figures, and surfaces in the environment. However, detecting convexity alone is not sufficient to determine figure-ground relations in a visual scene. The model should not indicate side-of-figure preferences toward concave regions (Figure 2f) that are not perceived by humans as figure (Zhou et al., 2000). (c) The model should demonstrate the B cell size invariance property and have a mechanism to determine border-ownership assignment despite differences in the relative size between the figure and ground. In other words, how does the visual system identify figures that can appear at a wide range of spatial scales? As tested in the present model, we hypothesize that feedback from on-surround model units with different receptive field sizes that undergo normalized competition propagates information about figure locations to B cells irrespective of the particular figure sizes. Our model, the R cell, G cell, B cell (RGB) model, predicts that a substantial portion of the large number of cells in primate V2 and V4 with on-surround receptive fields (Pasupathy & Connor, 1999; Hegde & Van Essen, 2004, 2006; Bushnell et al., 2011) compete and group convexity information interareally to rapidly resolve figure-ground segregation. Akin to the bistable perceptual figure-ground reversals observed in displays such as Figure 1a, the RGB model predicts similar reversal phenomena occur at a local level within the early visual system due to competition between cells with on-surround receptive fields. The dynamics of on-surround

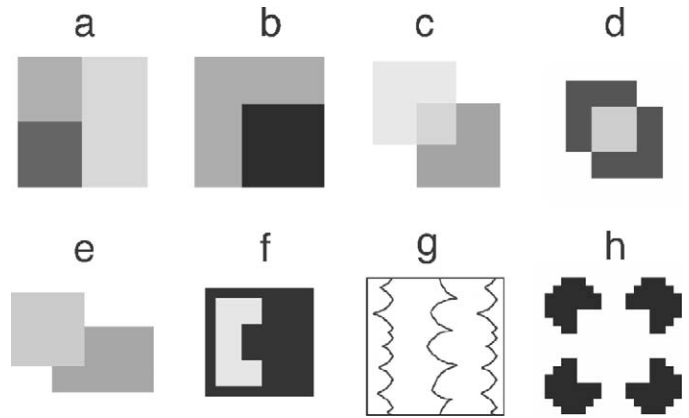


Figure 2. Visual displays that are used in model simulations. Luminance junctions (a–d) represent important tests for the model because similar local junctions appear so frequently within natural and synthetic visual scenes. Visual displays that have been tested on B cells in the electrophysiological literature afford the comparison between model and cell responses (c, e, f). The model makes B cell response predictions in displays that have only been tested psychophysically for border-ownership (g, h). (a) A T-junction that is frequently associated with the occlusion of surfaces (left half; T-junction stem) by another surface (right half; T-junction hat) and the border typically is owned by the occluder. (b) L-junctions do not necessarily implicate occlusion, as the junction may appear at a corner of a figure. (c) The presence of X-junctions that reverse contrast polarity once may elicit the percept of a transparent surface (top left square) occluding another (bottom right square). The borders of the small centrally located square are owned by the transparent occluding surface. (d) When X-junctions reverse in contrast polarity twice, the percept of occlusion vanishes and the borders of the centrally located small square may either be owned by the square or the surrounding L-shapes. (e) A square occludes a rectangle and contains two T-junctions. B cells recorded from in vivo signal border-ownership of the occluding square near the T-junctions (Zhou et al., 2000). (f) C-shape display that contains a concavity. B cells demonstrate side-of-figure preferences to the C-shape, and not to the concave region (Zhou et al., 2000). (g) Convexity display of Peterson and Salvagio (2008). Human subjects are more likely to indicate that the convex region is the figure than the ground compared to the concave region. (h) Kanizsa square. When the pacmen inducers are appropriately aligned, an illusory square is seen in the center that is a brighter white than that on the periphery.

competition (Figure 4b) between units in different cortical areas have not been extensively studied, and we suggest on-surround competition is an integral part of figure-ground segregation in the primate visual system.

Visual displays

Humans responses in psychophysical studies show that T-junctions, which mark the confluence of three

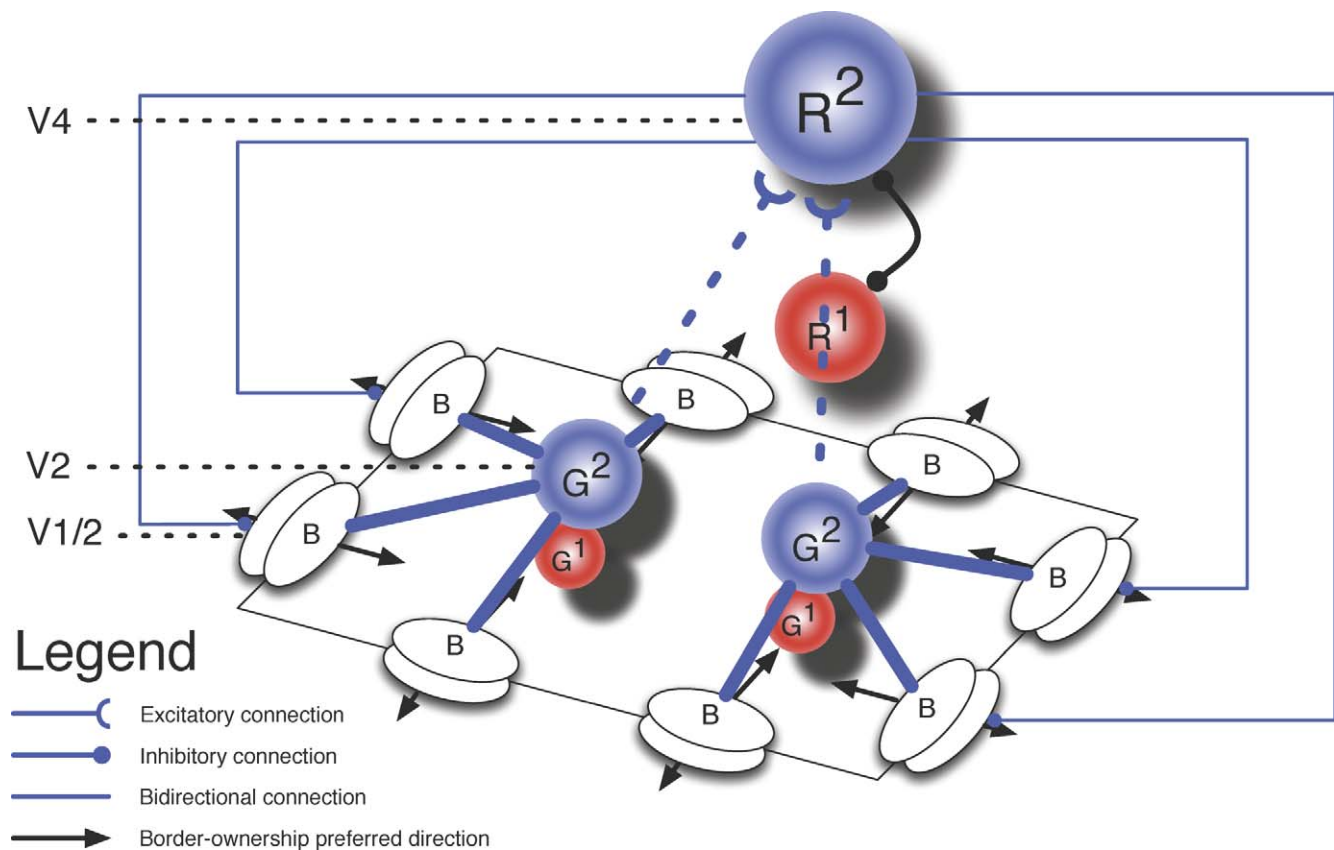


Figure 3. Schematic depiction of the model response to a rectangular visual display. Border-ownership cells (designated by ellipses) with different side-of-figure preferences become active due to the presence of a bottom-up edge signal. Due to the lack of global information about the visual scene, all B cells at each spatial location initially are equally active and do not show a direction of border-ownership bias. B cells interact with small (G1, red) and large (G2, blue) spatial scale grouping cells (G cells) with annular receptive fields, sensitive to convexity and closure. Note, connections for the larger scale (blue) are only shown. While G cells receive feedforward input from B cells within the receptive field, G cells also feedback to inhibit B cells with side-of-figure preferences away from the center of the annulus. The selective feedback biases the border-ownership signals toward the center of the annulus by suppressing the activity of B cells with inconsistent side-of-figure preferences. R cells pool over small (R1, red) and large (R2, blue dashed lines) spatial scale G cells and compete (black solid line) across scale to resolve salient scales in the visual scene. Following the competition, R cells feed back (blue solid lines) to B cells to inhibit B cells with side-of-figure preferences away from the peak R cell activity, which produces a bias in the border-ownership signals toward the figure. Note, for visual clarity not all connections in the model are shown.

luminance values (Figure 2a) in natural and synthetic scenes, indicate the presence of occlusion but the confluence of two luminance values at L-junctions (Figure 2b) do not (McDermott, 2004). Because local junctions appear so ubiquitously in visual scenes and their border-ownership cell response properties are known, these junctions provide important tests for any model of border-ownership. The model should yield border-ownership signals consistent with those obtained in neurophysiological and psychophysical studies. In Figure 2a, the left and right half regions are associated with the occluded and occluding figures, respectively. The central vertical border is owned by the figure on the right half side. For a convex figure, the L-junction in Figure 2b would be owned by the figure attached to the bottom-right region. Moreover, X-junctions represent local luminance constellations

whereby four luminance values converge at a point and may coincide with the presence of a transparent occluding surface (Figure 2c). In Figure 2c, the borders of the central small square region are owned by the top left, not the bottom right, L-shaped region and together the small central square and the top left L-shape form the percept of a transparent filter. Studies have identified heuristic rules about the contrast polarity relations around the junction that generally give rise to the percept of a transparent surface (see Adelson & Anandan, 1990 and Anderson, 1997 for a review). When the contrast polarity at the X-junction reverses twice compared to once (Figure 2c), the percept of transparency is abolished (Figure 2d). In this case, the borders of the small central square may either be owned by the surrounding L-shapes or the square.

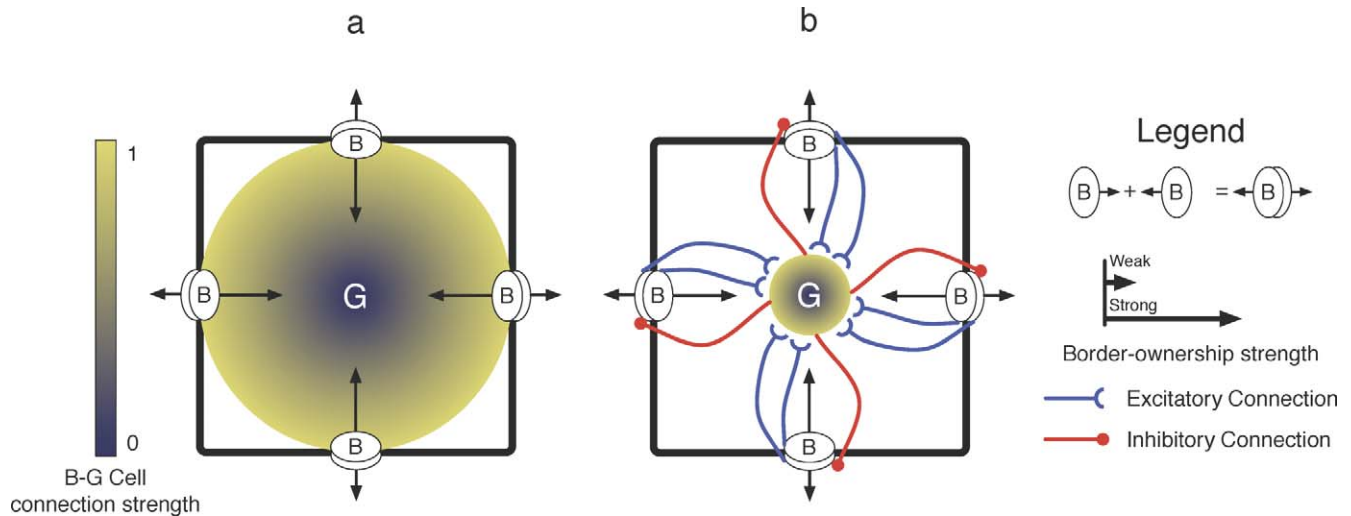


Figure 4. (a) Border-ownership (B) cells in visual cortex preferentially respond to borders when they represent a certain side of a figure. Grouping (G) cells have an on-surround or ring-like receptive field structure (yellow-green) and respond to convexity and bias the competition between two overlapping and similarly oriented but opposite-pointing B cell units (see legend). For example, G cells with appropriately sized receptive fields respond preferentially to a square (the [a] panel; outline of filled square shown) and B cells with border-ownership preferences toward the center of the square are enhanced relative to those with preferences away from the square center. Brighter shades of green indicate stronger connection weights between border-ownership cells (denoted by “B”) and the G cell. (b) G cells bias border-ownership direction by inhibiting B cells with side-of-figure selectivities that point away from the radial center of their receptive fields.

Displays that have been extensively tested in neurophysiological studies also represent important tests for any model of border-ownership. Figure 2e shows the occlusion of a rectangle by a square. Even at the two T-junctions, B cell responses indicate border-ownership by the square and not the rectangle (Zhou et al., 2000). The gray C-shape in Figure 2f provides another test for border-ownership models because B cells provide border-ownership signals toward the C-shape and not toward the concave region to the right of the C.

Although the convexity displays of Peterson & Salvagio (2008) and the Kanizsa square have not been tested neurophysiologically on B cells, border-ownership properties have been identified from human psychophysical studies. Testing the model on these displays will check its psychophysical consistency and predict border-ownership cell responses to inform future experiments. Consider a display (Figure 2g) that contains convex and concave segments (Peterson & Salvagio, 2008). Humans exhibit a bias to report convex regions as the figure more often than concave ones, even when the convex region area equals that of the concave region. Nonrectangular displays that have curvature and do not have T-junctions may pose a challenge to figure-ground segregation models. Although we are not aware of any direct evidence on neural border-ownership data of curved displays or a border-ownership bias for convex or concave shapes, indirect evidence suggests this may be the case. First, B

cells have been shown to produce border-ownership signals in the C-shape displays toward the C-shape and not the concave region (Zhou et al., 2000). Second, when a square region is defined by random dot motion and luminance contrast with its surrounding, the square region may be interpreted as an object or a window. Over 80% of B cells in the sample that had a side-of-figure selectivity to the convex square when it appeared as figure (object) compared to less than 20% when it appeared as the surrounding concave region (window) appeared as figure (von der Heydt et al., 2003). Therefore, we predict that B cells are more likely to elicit side-of-figure preferences to the convex rather than the concave regions of the convexity displays of Peterson & Salvagio (2008) irrespective of region area differences, which is consistent with human psychophysical judgments.

In the Kanizsa square display, which is formed by four pacmen inducers, human subjects see an illusory square in the center that appears brighter than the white luminance to the periphery (Figure 2h). The Kanizsa square is an important test for border-ownership because based on B cell responses to individual convex shapes, such as squares, it is expected that B cells would indicate ownership toward the individual pacmen inducers. We hypothesize that B cells with receptive fields centered along the concave borders of the pacmen demonstrate border-ownership toward the illusory square when the inducers are aligned as in Figure 2h.

Methods

Model overview

Figure 3 shows the architecture of the RGB model. B and grouping (G) cells, as reported by Craft et al., 2007, cannot alone account for the properties of border-ownership cells reported in neurophysiological studies because convexity sensitive G cells in V2 respond to regions that may be correlated with, but often do not represent, figures or surfaces in the visual scene. Because grouping cells respond strongly to convex regions, a C-shape, as shown in Figure 2f, elicits high G cell activity outside of the C due to the exterior convexity, but the C is nonetheless interpreted as the figure. How does the cortex differentiate between regions of convexity that may or may not be associated with a figure? In addition to using well-known LGN and V1 complex cell units, border-ownership (B) cells in V1/V2, and grouping (G) cells in V2, as described by Craft et al. (2007), another unit type is required, called the R cell in our model, that reflects known properties of shape-sensitive cells in primate V4 (Hegde & Van Essen, 2006; Bushnell et al., 2011). The RGB model predicts that competition between R cells with different receptive field sizes plays a fundamental role in figure-ground segregation in concert with B cells by identifying candidate locations for figures in the visual scene. Interareal connectivity between B and R cells is required because of their differing functional properties: R cells, with large receptive fields, can detect which side of an edge the figure is on, but cannot determine the precise location of the boundary; B cells, with small receptive fields, can locate a boundary, but cannot alone determine which side of an edge the figure is on.

B and G cells

Figure 4b summarizes the B-G cell microcircuit. This component of the model is similar to the model of Craft et al. (2007), except the RGB model does not have junction detectors, and we analyze the temporal dynamics of B and G cells. Due to the prevalence of right angles and rectangular objects in the visual displays examined by neurophysiological studies of border-ownership (Zhou et al., 2000; Qiu & von der Heydt, 2007), we simulate four different border-ownership cells that sample each visuotopic location: left, right, up, and down. Graphically, we depict B cells by arrows, whose length indicates the magnitude of response and whose direction indicates the net direction of border-ownership at that visuotopic location (Figure 4a). Directions are determined by a vector, denoted vectorial modulation index (Vmod; Craft et al., 2007;

Mihalas, Dong, von der Heydt, & Niebur, 2011) at each visuotopic location: the x and y components are a difference divided by a sum of the activities of left/right and up/down B cell activations. (See Model equations for details.)

Blue and red connections in Figure 4b signify excitatory feedforward input to and inhibitory feedback from G cells, respectively. By virtue of the radial symmetry and ring-like receptive field shape, each G cell receives input from B cells whose receptive field centers are located at a distance relative to the G cell receptive field center. G cells thus perform on-surround integration of their inputs over certain locations in the visual field. Larger receptive field G cells integrate a larger number of B cells over a wider area in the visual input. With an appropriate receptive field size, G cells can elicit strong responses when figure borders enter their on-surround receptive fields and communicate that information via feedback to border-ownership cells that also have the figure border in their receptive fields. Figure 4b schematically shows how a G cell with the square borders in its receptive field can feed global convexity information back to B cells with a border of the square in their receptive fields to reinforce the presence of the square figure. Note Figure 4b shows that the feedback connections from a G cell to its associated B cells are inhibitory; otherwise an unstable positive feedback loop could occur. We follow the convention of Craft et al. (2007) that G cells monosynaptically integrate spatially offset B cells within their on-surround receptive fields with a preference of ownership toward the center of the ring and feed back to inhibit active B cells with the opposite ownership preference. The RGB model reinforces the presence of a convex figure by having G cells inhibit a subset of B cells that supply it input, but only those with the opposite ownership preference. B cells with active bottom-up input provide a local estimate of the relative figure position.

R cells

R cell units compete in a shunting fashion—R cells receive inhibition from other R cells with different receptive field sizes that are centered at the same visuotopic location proportional to their current activation (Grossberg, 1973). Shunting interactions are well known to occur between cells in visual cortex and have canonical, universal properties, such as divisive normalization present in many places throughout the cortex (Carandini & Heeger, 2011). R cells propagate global information about the location of salient figures in the visual scene interareally to B cells. As dynamics unfold, reversals in the polarity of border-

ownership signals occur toward R cell receptive field centers (left vs. right, up vs. down).

Figure 5 shows how R cells use shunting competitive dynamics between units with different receptive field sizes to produce peaks in the location of the occluding surface in a T-junction and suppress activity in the location of the occluded surface. T-junctions mark the confluence of three luminance values, and the regions on the hat and stem sides of the junction are associated with the occluding and occluded surfaces, respectively. Figure 5a focuses on the dynamics of G and R cells with receptive fields on the side of the occluded surface

(stem). Because G cells respond to convexity, they elicit high activation on the stem side (bright green circles superimposed on input). As shown above the input in the one-dimensional cross-section of the G cell activity, the strong response to the corners is concentrated in certain visuotopic locations: G cells with receptive fields far away from the T-junction on the stem side will elicit a weak response (blue). R cells with larger receptive fields integrate G cell activity in an on-surround manner. The R cell that monosynaptically receives interareal projections from both highly active G cells (bright green) will also be highly active before the

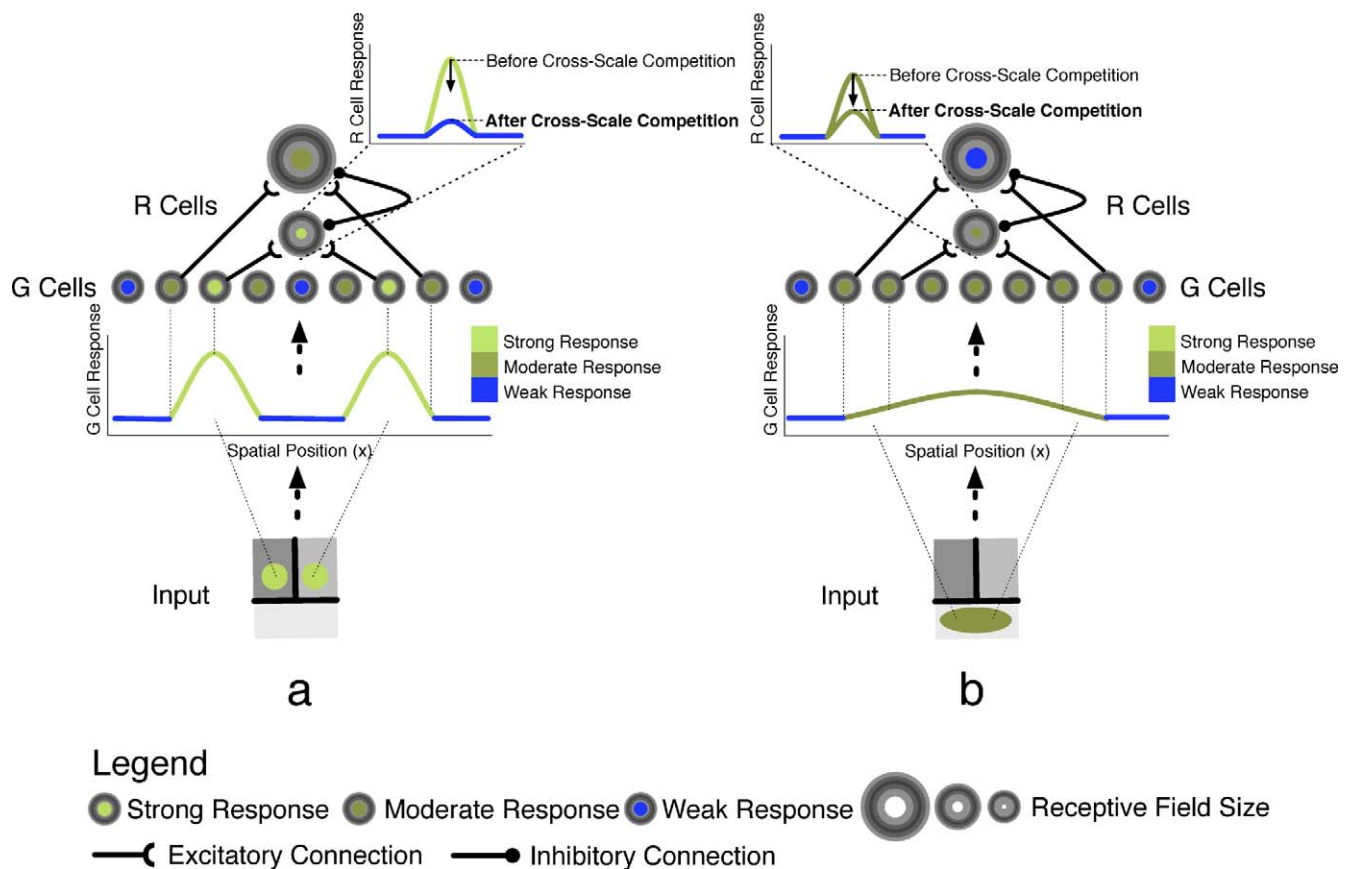


Figure 5. Model R cells produce activity peaks at the location of an occluding surface. The (a) and (b) panels focus on the dynamics of R cells with receptive fields located on the stem and hat sides of the T-junction, respectively. G cells in the (a) panel with receptive fields near the T-junction on the stem side are highly active (bright green disks). The one-dimensional cross-section of G cell activity positioned above the model input shows the expected G cell responses at different locations along the stem side. R cells perform on-surround integration of the G cell units. Before R cell competition occurs in the model, R cells that receive interareal projections from the highly (moderately) active G cells will also be highly (moderately) active. The expected R cell response before competition occurs is shown at the top. R cells with receptive fields centered on the same visuotopic location compete in shunting manner: the higher the unit activation, the more it inhibits other units. Due to the concentration of highly and moderately active R cells on the stem side, competition will be fierce and cell responses will drop precipitously following the competition. Conversely, G cells in the (b) panel with receptive fields on the hat side of the T-junction indicated by the oval overlaid on the input are only weakly active, as indicated in the one-dimensional cross-section of G cell activity shown above the input. Before R cell competition occurs in the model, R cells that receive projections from the weakly active G cells will also be weakly active. The expected R cell response before competition occurs is shown at the top. Because R cell activity is weak on the hat side, the competition is less fierce and the activity following competition is higher on this side compared to the stem side. R cells indicate via feedback projections to B cells that the hat side of the T-junction is the occluding figure. Note that at the top of the diagram only small-scale activity is shown, but the large-scale will begin with similar peaks.

competition occurs. Conversely, the R cell with a larger receptive field that misses the highly active G cells receives projections from moderately active (green) G cells and will therefore be moderately active. The R cells shown will inhibit one another proportional to their present level of activity due of their shunting interactions. Because both units are at least moderately active, the activity of both units will be strongly suppressed relative to their initial levels after the competition occurs (top of Figure 5a).

Figure 5b illustrates on the dynamics of G and R cells whose receptive fields are on the occluding surface side (hat). Because G cells prefer convexity and do not receive information about corners as they did on the occluded surface side, the G cell response will uniformly be much weaker on this side (green ellipse superimposed on the input). As shown above the input in the one-dimensional cross-section of the G cell activity, G cells still receive contrast information from the line, so the G cells whose receptive fields intersect with the line from the occluding surface side will be moderately active. The R cell that integrates the moderate G activity will also be moderately active before competition (green). Conversely, the R cell with a larger receptive field size does not integrate the moderate G cell activity and produces a weak response (blue). As in Figure 5a, R cells inhibit each other proportional to their current level of activity. Because the R cells shown are only weakly or moderately active initially, the inhibition they receive from one another is weak. After the competition occurs, their responses will decrease, but not as much as the R cells shown in Figure 5a. Hence, R cells reverse border-ownership polarity (left vs. right, up vs. down) at a T-junction and respond more strongly to the occluding surface: Two strong G cell activity peaks are annihilated by R cells on the stem side and one shallow peak on the hat side survives. The surviving peak on the occluding surface is conveyed to B cells via interareal projections to suppress the response of B cells that have preferred directions away from the occluding surface.

Model equations

Differential equations in the model specify the activity of cells or cell populations with receptive fields centered on each pixel of the input displays. Since the operations within each equation apply to all cells, we use matrix notation. For example, \mathbf{x} stands for the set of cells at every spatial location (p, q) in the input display. Convolution between a matrix \mathbf{x} and kernel \mathbf{F} , specified by the $*$ operator, is always centered at each cell position (p, q) . In all convolutions, we tile the boundary values beyond the image border as far as necessary. All ordinary differential equations were

numerically integrated using a Runge-Kutta routine. Our simulations were performed on a 2.66 Ghz 8-core Apple Mac Pro with 64 GB RAM in Wolfram Mathematica 8.

Model neurons are represented by a single-compartment voltage $V(t)$ that obeys the following shunting equation:

$$C_m \frac{dV(t)}{dt} = - \left(V(t) - E_{leak} \right) \delta_{leak} - \left(V(t) - E_{excite} \right) \delta_{excite}(t) - \left(V(t) - E_{inhib} \right) \delta_{inhib}(t). \quad (1)$$

In Equation 1, C_m specifies the membrane capacitance, δ_{leak} denotes the constant leakage rate, $\delta_{excite}(t)$ signifies the lumped excitatory inputs to the cell over time, and $\delta_{inhib}(t)$ specifies lumped inhibitory inputs to the cell over time. The terms E_{leak} , E_{excite} , and E_{inhib} refer to the cell's leak, excitatory, and inhibitory reversal potentials, respectively.

Model LGN cells

The first stage of the model consists of isotropic on-center/off-surround processing. Equation 1 can be rewritten by setting $\mathbf{x} = V$, $\alpha_{LGN} = \delta_{leak}$, $E_{leak} = 0$, $\beta_{LGN} = E_{excite}$, and $\gamma_{LGN} = -E_{inhib}$:

$$\frac{d\mathbf{x}}{dt} = -\alpha_{LGN}\mathbf{x} + (\beta_{LGN} - \mathbf{x})\mathbf{F}_{excit}^{LGN} * \mathbf{I} - (\mathbf{x} + \gamma_{LGN})\mathbf{F}_{inhib}^{LGN} * \mathbf{I}. \quad (2)$$

In Equation 2, \mathbf{x} signifies the activation of model LGN neurons with receptive fields centered at each pixel in the input display at time t , α_{LGN} represents the passive decay of the model cell, β_{LGN} specifies the saturation upper bound, γ_{LGN} is the inhibitory lower bound, and \mathbf{I} refers to the input image. Both \mathbf{F}_{excit}^{LGN} and \mathbf{F}_{inhib}^{LGN} are Gaussian kernels with $\sigma_{excit}^{LGN} = 0.25$ and $\sigma_{inhib}^{LGN} = 0.5$. In our simulations, we set $\alpha_{LGN} = 5$, $\beta_{LGN} = 1$, and $\gamma_{LGN} = 0.5$. To interpret Equation 2 as cell or population firing activity, we half-wave rectify the output (i.e., consider the nonnegative component of the response):

$$\mathbf{X} = \max(\mathbf{x}, 0). \quad (3)$$

Model complex cells

We construct complex cell units by summing the responses garnered by pairs of simple cell units that differ in orientation preference by π radians. We fix $\theta \in \{0, \pi/2\}$. Simple cell units also possess orientation preferences but are sensitive to the polarity of contrast. According to the Steering Theorem (Freeman &

Adelson, 1991) we can express a simple cell kernel of arbitrary orientation preference \mathbf{S}_θ via a weighted sum of the partial derivatives of a 2D Gaussian kernel \mathbf{F}^{V1} with respect to the x and y directions:

$$\mathbf{S}_\theta = \cos(\theta)\mathbf{S}_0 + \sin(\theta)\mathbf{S}_{\frac{\pi}{2}}. \quad (4)$$

In Equation 4, the notation \mathbf{S}_θ means the first derivative of kernel \mathbf{F}^{V1} in the direction of θ , rotated such that the kernel has a preferred orientation θ radians. Since \mathbf{S}_0 and $\mathbf{S}_{\pi/2}$ are orthogonal, they form a basis. Hence, we may interpret Equation 4 as basis elements rotated such to yield a simple cell kernel, which has a preferred orientation of θ radians. In our simulations, we chose $\theta \in \{0, \pi/2, \pi, 3\pi/2\}$. Complex cell units \mathbf{C}_θ can be obtained by summing anti-phase kernels:

$$\mathbf{C}_\theta = \mathbf{S}_\theta + \mathbf{S}_{\theta+\pi}. \quad (5)$$

Because we are simply concerned with horizontal and vertical directions of contrast, we fix $\theta \in \{0, \pi/2\}$. We set $\sigma^{V1} = 0.5$. To compute the complex cell \mathbf{Y} unit activity in model V1, we convolve the LGN unit responses with the horizontal and vertical complex cell kernels \mathbf{C}_θ , threshold by Γ , and half-wave rectify in the following additive equation:

$$\frac{d\mathbf{Y}_\theta}{dt} = -\mathbf{Y}_\theta + \max(\mathbf{C}_\theta * \mathbf{X} - \Gamma, 0). \quad (6)$$

Model border-ownership cells (B cells)

We model B cells, with border-ownership direction θ , using the following equation that integrates bottom-up input from complex cells and top-down signals from G and R units (described below):

$$\begin{aligned} \tau^B \frac{d\mathbf{B}_\theta}{dt} = & -\mathbf{B}_\theta + (\beta^B - \mathbf{B}_\theta) \mathbf{Y}_\theta \\ & -\gamma^B \mathbf{B}_\theta \left(\sum_s \zeta^s (\mathbf{G}^s * \mathbf{K}_{\theta+\pi}^s) + \sum_l \eta^l (\mathbf{R}_2^l * \mathbf{L}_{\theta+\pi}^l) \right). \end{aligned} \quad (7)$$

In Equation 7, \mathbf{G}^s and \mathbf{R}_2^l signify the grouping cell (G cell) and second stage R cell activity at scale s and l , respectively. $\mathbf{K}_{\theta+\pi}^s$ and $\mathbf{L}_{\theta+\pi}^l$ may be viewed as subunits of the G and R cell units' receptive fields. We set the scale indices s and l of $\mathbf{K}_{\theta+\pi}^s$ and $\mathbf{L}_{\theta+\pi}^l$ such that the feedback from G and R cells projects to nearby B cells that receive bottom-up input from complex cells, respectively. ζ^s and η^l serve to differentially weight feedback contributions to B cells due to scale-dependent receptive field differences. We perform a scale-dependent weighting by $\zeta^s = \eta^l = \sqrt{4.5}r$ similar to that of Craft et al. (2007), where r is the radius of the G or R cell receptive field kernel. The parameter γ^B regulates

the gain of inhibitory G and R feedback signals. In Equation 7, we set $\tau^B = 10$ ms, $\beta^B = 1$, and $\gamma^B = 100$. At every spatial position, we model B cells with four directions of border-ownership: $\theta \in \{0, \pi/2, \pi, 3\pi/2\}$.

Model grouping cells (G cells)

G cells possess radially symmetric annular or ring-like receptive fields that integrate B cell activity. We achieve feedback between B and G units in Equation 7 by convolving \mathbf{G}^s with either the left, top, right, or bottom piece of the G cell receptive field annulus $\mathbf{K}_{\theta+\pi}^s$, which selects B cells in regions of the G cell receptive field and suppresses B cells with direction of border-ownership preference π radians away from that pointing inward toward the annulus center.

We construct the G cell kernels by taking a difference-of-Gaussians (i.e., Mexican hat) between kernels \mathbf{F}_1^G and \mathbf{F}_2^G of radius r^G and standard deviations σ_1^G and σ_2^G , respectively. For the Gaussians \mathbf{F}_1^G and \mathbf{F}_2^G , we constrain the ratios $r^G/\sigma_1^G = 2$ and $r^G/\sigma_2^G = 2.22$. For computational tractability and simplicity, we select two G cell receptive field sizes within the range reported in the neurophysiological literature. We set r^{G1} and r^{G2} to 2 and 3, respectively. In order to obtain the kernel fragments \mathbf{K}_θ^s we simply extract the necessary half of the annulus. For example, to obtain \mathbf{K}_π^s , we take the left half component of the G cell kernel at scale s .

In order to model the G cell dynamics, we employ an equation similar to that of Craft et al. (2007):

$$\begin{aligned} \tau^G \frac{d\mathbf{G}^s}{dt} = & -\mathbf{G}^s \\ & + \zeta^s (1 - \mathbf{G}^s) \sqrt{\sum_{m,n} (\mathbf{B}_{\theta_m} * \mathbf{K}_{\theta_m}^s) (\mathbf{B}_{\theta_n} * \mathbf{K}_{\theta_n}^s)} \end{aligned} \quad (8)$$

In Equation 8, we multiplicatively combine all pairs of directions of ownership permutations m and n . For example, when $m = 1$ and $n = 2$, $\theta_m = 0$ and $\theta_n = \pi/2$, since $\theta \in \{0, \pi/2, \pi, 3\pi/2\}$. We fix $\tau^G = 10$ ms. Each combination may be interpreted as functional subunits of the G cell receptive field. The multiplicative subunit structure affords G cells a nonlinear response to regions with convexity and closure. Hence, G cells detect co-occurrences between pairs of edge signals with different orientations.

Model R cells

The R cell layer consists of stages of competition across scale and distance-dependent spatial competition. R cells pool over local populations of G cells that possess a common scale. We introduce a temporal

competitive neural network to model R cells in model area V4 (Grossberg, 1973). In the following R cell network, we select a faster-than-linear signal function $f(w) = w^2$ to form a winner-take-all network.

$$\begin{aligned} \tau^{R_1} \frac{dR_1^l}{dt} &= -R_1^l + (\beta^{R_1} - R_1^l) \left(f(R_1^l) + \delta^l (G^s * F_{excit}^{l,R_1}) \right) \\ &\quad - R_1^l \sum_{k \neq l} F_{inhib}^{R_1} * f(R_1^k). \end{aligned} \quad (9)$$

In our simulations, we let $\tau^{R_1} = 10$ ms and $\beta^{R_1} = 2$. By having R units that perform inter-scale competition, the network selects the spatial locations of salient figures at each scale. As with G cell spatial scales, we employ two R cell receptive field sizes. For the purposes of our simulations, we assume G and R cells each possess two scales. We assume R_1^l , which represents the set of R cells of scale l , may only receive projections from G^s , which represents the set of G cells of scale s , when $s = l$. In all simulations, we set $r^l \in \{2,3\}$. δ^l specifies a scale-dependent proportionality constant. We set in all simulations δ^1 and δ^2 to 1200 and 350, respectively. We produced the ring-shaped kernel F_{excit}^{l,R_1} by subtracting a 2D disk kernel of radius r with another of radius $r/2$. The radius three disc kernel $F_{inhib}^{R_1}$ specifies which proximal R cells with different receptive field sizes enter in the inter-scale competition.

Finally, we have a second R cell stage that performs local spatial competition to enhance the contrast in the network responses, which is useful for high spatial frequency displays:

$$\begin{aligned} \tau^{R_2} \frac{dR_2^l}{dt} &= -R_2^l + A(\beta^{R_2} - R_2^l) \left(f(R_2^l) + R_1^l \right) \\ &\quad - (R_2^l) \left(f(R_1^l * F_{inhib}^{l,R_2}) \right). \end{aligned} \quad (10)$$

In Equation 10, we set $\tau^{R_2} = 10$ ms and $\beta^{R_2} = 1$. Equation 10 represents a choice network as well with $f(w) = w^2$. A specifies the scale-specific attentional signal, which we model with a broad 2D Gaussian, spatially centered at the locus of attention. The attentional signal multiplicatively enhances the R cell signal at a particular spatial scale. Hence, it only enhances existing R cell activity. A was set to one in all simulations except for in the double-reversal transparency display (Figure 7e) to differentially weight small and large R cell peak activities. Figure 7f (left) shows the model output when $A = 1$ and Figure 7f (right) shows the model output when A is changed to amplify the activity peaks shown in Figure 7e (right). F_{inhib}^{l,R_2} represents an annular kernel with diameter 3° . R units feedback to B cells (Equation 7) to suppress those with antipreferred directions of border-ownership.

Vectorial modulation index

We use the following vector modulation index $\vec{v} = (m_x, m_y)$ to relate our B cell responses to those reported in other studies (Craft et al., 2007; Mihalas et al., 2011):

$$\begin{aligned} m_x &= \frac{B_0 - B_\pi}{B_0 + B_\pi} \\ m_y &= \frac{B_{\frac{\pi}{2}} - B_{\frac{3\pi}{2}}}{B_{\frac{\pi}{2}} + B_{\frac{3\pi}{2}}}. \end{aligned} \quad (11)$$

Due to the normalizing difference over a sum of B cell activity with antipreferred directions of border-ownership, each component lies between -1 and 1 . Negative values correspond to computed border-ownership in the leftward and downward directions for (x, y) components at each spatial location (p, q) , respectively. By contrast, positive values correspond to computed border-ownership in the rightward and upward directions for (x, y) components at spatial location (p, q) , respectively. A zero valued component indicates an indifference in border-ownership in that particular axis. Since Zhou et al. (2000) discovered border-ownership modulation emerging 10–25 ms from the onset of the visual presentation, we ran the model for 25 ms in model time.

Results

T- and L-junctions

Figure 6 shows simulation results to a T-junction, which is an important test for the model, because model B cells should indicate border-ownership toward the occluding surface, as reported by Zhou et al. (2000). Figure 6a shows the vertical and horizontal complex cell activity that projects to B cells (Figure 6b and d). Figure 6c depicts the activity of G cells with different receptive field sizes. Consider the case wherein the model consists only of B and G cells (Figures 6a–c); that is, B cells do not receive feedback from R cells. Because the G cells prefer convexity and the geometry of a T-junction includes corners of contrast on the stem (Figure 6a), the response of G cell is greatest inside the corners of the T-junction stem (Figure 6c; region enclosed by the dashed white contours). Conversely, G cells respond poorly to the T-junction hat (indicated by weak green color) because G cells may at best make one point of contact with the complex cell signal. The G cells that respond well on the left side project to B cells that respond to bottom-up contrast on the bottom left half of the vertical edge (Figure 6a) to enhance the activity of B cells with side-of-figure preferences toward

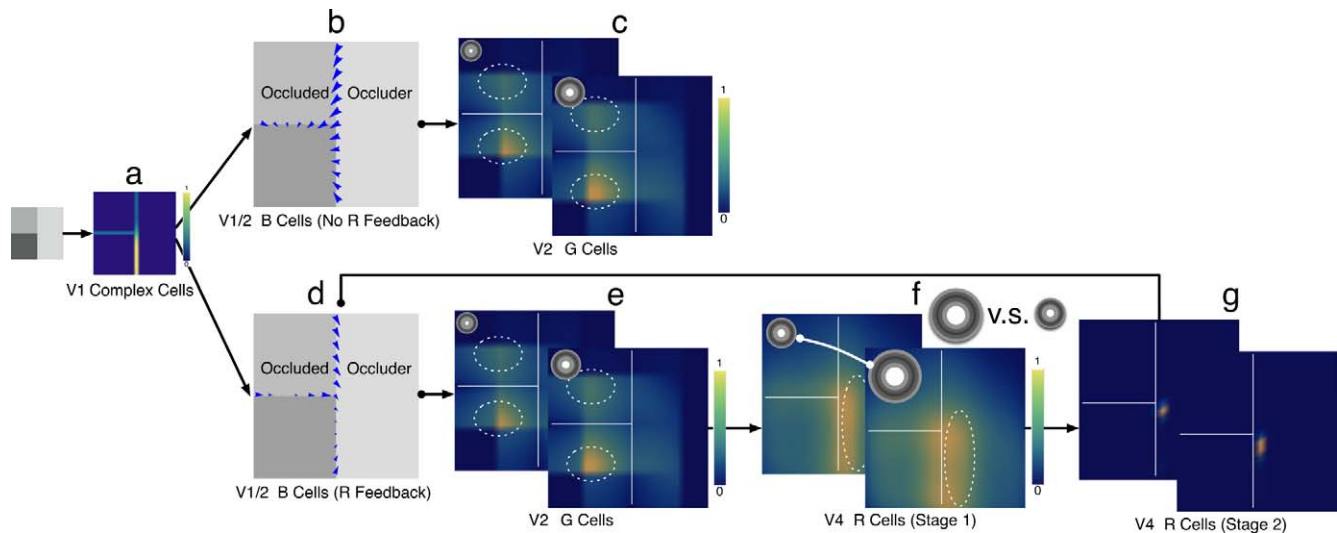


Figure 6. T-junction simulation. The V1 complex cell response to the T-junction display is shown in (a). The top (b–c) and bottom (d–g) rows show simulation results with and without V4 R cells, respectively. Without feedback from R cells, the vectorial modulation index (Vmod) of B cells points toward the stem side of the T-junction, rather than to the occluding surface (b). This is due to the high G cell response on either side of the T-junction stem, indicated by the dashed white contour (c and e). The solid white lines are drawn to indicate the contrast boundaries present in the T-junction display. R cells perform on-surround competition across scale (Stage 1) and respond with high activity on the side of the T-junction hat (the occluding surface) due to shunting inhibition in the network competition, indicated by the dashed white contour (f). A second stage of R cells performs local spatial competition to identify peaks from Stage 1 that may represent figure locations (g). With R cell feedback, the Vmod of B cells indicates the presence of a figure on the side of the T-junction hat (d), consistent with Zhou et al. (2000). The lengths of each vector component in (b) and (d) are proportional to the difference in activity between cells signaling border-ownership along that particular axis. Thirty-one B cells (19 are shown in [b] and [d] for clarity) and 22×22 G and R cells were each simulated.

the G cells' receptive field locations. As shown in Figure 6b, the B-G interaction results in a preference toward the corners on the side of the stem. However, this region is not always consistent with the usual interpretation of occlusion. The B-G microcircuit alone cannot account for the figure-ground properties associated with the T-junction or the border-ownership response observed by Zhou et al. (2000) (see figures 23 and 24). This means that with only B and G cells, G cells can respond outside of figures. Their responses are also biased by high contrast regions. Since G cells respond to any region of convexity at all spatial scales, G cells alone cannot convey the appropriate scale range within which an object may be perceived a figure.

Figures 6f–g show simulation responses of the activation of R cells. R cells with receptive fields centered at the same spatial location but with different receptive field sizes compete with each other, which results in the pattern of activity shown in Figure 6f. R cells are inhibited proportional to their present level of activation (shunting inhibition). A highly active R cell at a particular spatial scale receives strong inhibition when spatially proximal R cells with different receptive field sizes are also active. The result is a shift in the activity distribution of R cell units compared to that of G cells (compare activity peaks in Figure 6e and g): R cells that have the same receptive field location as a

highly active G cell will have low activity and vice versa. R cells that have receptive fields on the side of the T stem integrate two G cell peaks across scales and as a result encounter strong suppression. When an R cell feeds back to B cells to enhance units with side-of-figure preferences toward the location of the R cell, Figure 6d shows B cells now show a net modulation in activity toward the side of the T-junction typically associated with the occluding surface. This reversal is consistent with the B cell responses reported neurophysiologically (see Zhou et al., 2000, figures 23 and 24).

Although R cells exhibit an inversion with respect to G cell peak activity in the case of T-junctions, this does not occur for L-junctions (Figure 7a). For the sake of the present discussion, a T-junction can be thought as the combination of an L-junction and its reflection along an axis defined by the T-junction stem. As we would expect from the T-junction simulation, G cells respond most strongly inside the L-junction. Although R cells are engaged outside the L-junction, the activity is lower than that within. Without an adjacent corner, which would be present with a T-junction, the R cell on-surround shunting competition does not receive as much suppression as it would in the T-junction case. Hence, a shift in peak R cell activity does not occur, which is consistent with the tendency for the L-junction

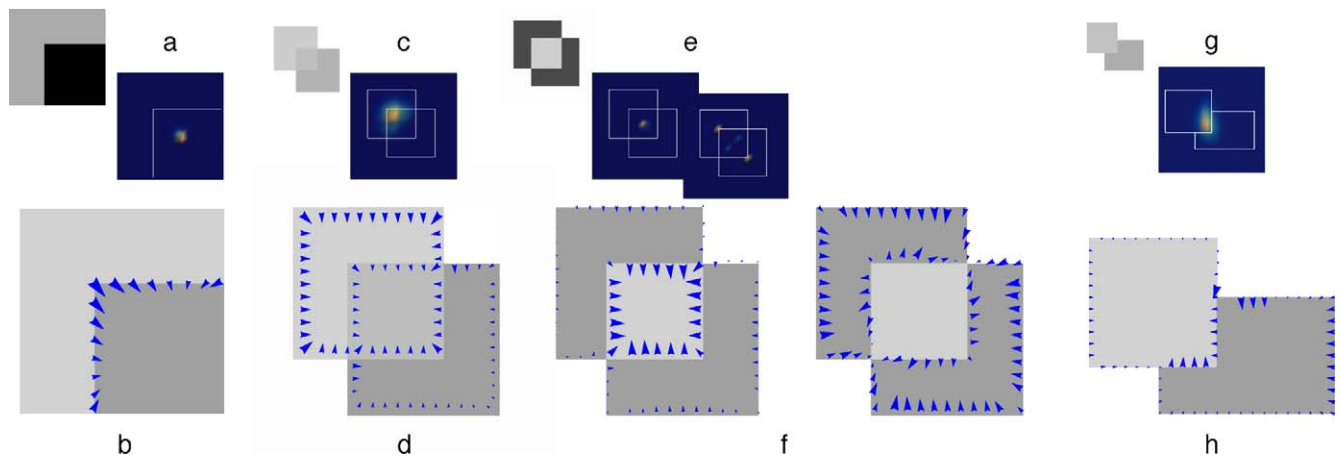


Figure 7. The top and bottom rows show the model R and B cell responses, respectively, to a number of visual displays. Unlike the T-junction simulation in which the G and R cells respond maximally to different regions of the visual input, G and R cells both respond inside the L-junction contrast-defined corner (a). The presence of two strong G cell activity peaks in the T-junction simulation on the stem side and a weak activity peak on the hat side induced the peak R cell activity to shift to the hat side. The L-junction display only results in one distinct G cell peak, which does not produce enough inhibition to move the peak R cell activity to a different location (b). The response of R and G cells to similar locations extends to simple rectangular shapes. The (c) panel shows a configuration of X-junctions (defined in the text) that produces a percept of a transparent overlay on top of a square. R cells respond highest in the center of the transparent overlay and signal to B cells that it is the occluding surface (d). When X-junctions are arranged such that they do not support the percept of transparency (e), R cells yield three activity peaks of different magnitudes: The peak due to the center of the visual display (e, left) is the strongest and the other two peaks near the corners of the two L-shapes are of equal magnitudes (e, right). The left subpanel of (f) shows the B cell response, which favors the center region. If the activity peaks produced by the large receptive field R cells (e, left) are weighted higher than those yielded by the small receptive field R cells (e, right), the B cell response favors the L-shapes. The (g) panel shows the model R cell response to a square occluding a rectangle. Due to the presence of T-junctions, R cells respond on the interior of the occluding surface and bias B cell responses (h) toward to square.

to contain a figure on the inside of its corner. The junction geometry is important because when four L-junctions are joined to form a square, the R cell peak inversion does not occur and R cells maintain the G cell peaks inside the square's region. Hence, the mere adjacency of two L-junctions does not induce a R cell peak shift—their spatial relations matter. Junctions with different spatial configurations give rise to different sets of G and R cell dynamics, which induces border-ownership signals consistent with neurophysiological findings and human figure-ground perception.

Transparency

Although the present model cannot fully explain perceptual transparency, the model correctly assigns border-ownership to the perceived figure in the case shown in Figure 7c. Figures 5c–f show the model output to a transparency display with two X-junctions. When the contrast polarity reverses once about the X-junctions, the configuration favors the percept of a lighter square serving as a transparent filter over an equally-sized darker square. Figure 7c show the peak R cell activity after spatial competition, which is concentrated close to where the top left corner of the centrally

location medium gray square intersects with the lighter luminance L-shape. Feedback from the R cells to B cells generates modulation that encloses the square that perceptually appears as a transparent overlay (Figure 7d). The average magnitude of the B cells that enclose this region is higher than that of any other closed region within the visual display, such as the center medium gray square (Figure 7d). The nontrivial response obtained around the borders of the center square shows the model can simultaneously maintain more than one figure-ground segmentation of a visual scene.

Figure 7e contains two X-junctions that exhibit a double reversal in contrast polarity, which does not perceptually support transparency. The display is perceptually segmented into two dark gray L-shapes and a centrally located light gray square. R cells yield three activity peaks of different magnitudes: the peak due to the center of the visual display (e, left, large spatial scale) is the strongest, and the other two peaks near the corners of the two L-shapes are of equal magnitudes (e, right, small spatial scale). Figure 7f (right panel) shows that the chosen parameter configuration favors the segmentation of the dark L-shapes as figure. If we give higher weight to the peak produced by the larger spatial scale R cells (e, left), which is centered

on the square, compared to the peak produced by R cells with smaller receptive fields (e, right), the B cell responses reverse in their modulation along the borders of the L-shapes and square to support the square as being the figure (Figure 7f, left panel).

Occlusion

We show in Figures 7g–h an occlusion display containing a square in front of a rectangle. As in Zhou et al. (2000), we set the visible rectangle area to match that of the square. In the absence of R cells, the B cells exhibit response modulation preferring the higher contrast areas. As predicted by our T-junction simulation (Figure 6g), the activity of R cells peak on the hat side of the T-junction, and the B cells reverse their overall direction of modulation to prefer the light gray square.

Displays with concavities

Figure 8 shows the simulation results to a C-shape display. The display is an important test for models because the shape is not convex and cells whose receptive fields are centered around the corners of the concave region demonstrate a border-ownership preference toward the C-shape instead of the background (Zhou et al., 2000). Without R cells, the B cells mostly develop preferences consistent with the convex C-shape appearing as a figure, except for some B cells whose receptive fields are centered along the concavity corner (Figure 8d, see red circles). This is because G cells respond maximally to the interior of both the C-shape (Figure 8b, see top and bottom dashed ellipses) and concave region (Figure 8b, see middle dashed ellipse). G cells with receptive fields centered along the interior of the C-shape generally elicit weak activity except for the cases indicated by the top and bottom dashed ellipses (Figure 8b). Note the similarity between the G cell activity distribution in the C-shape and T-junction cases: There are proximal regions of high G cell activity (dashed ellipses) and an extent of weak activity on the interior of the C-shape. As in the T-junction simulation (Figure 6), R cells yield a peak shift relative to the G cell peak activity distribution, in this case from the peaks indicated by ellipses to where there was weak G cell activity on the interior of the C-shape (Figure 8c). Hence, B cells that respond inconsistently with the representation of the C shape as a figure (Figure 8d, indicated by red circles) in the absence of R cells now exhibit consistent modulations (Figure 8e).

The convexity display of Peterson & Salvagio (2008) may be challenging for the model as well as others, because G units with appropriate receptive field sizes

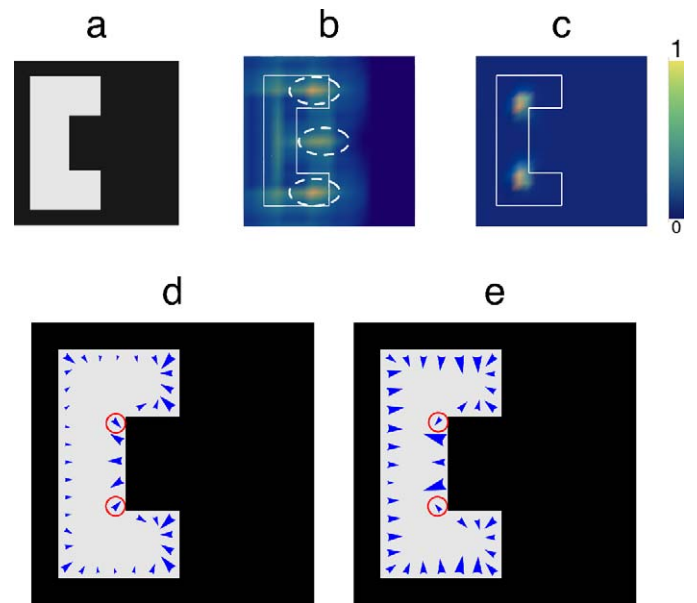


Figure 8. C-shaped display simulation. By virtue of their annular receptive field structure, G cells respond maximally to the top and bottom portions of the convex C-shape and the concave region (b, indicated by dashed white ellipses). In the absence of R cells, some B cell vectorial modulation indices show preferred direction of ownership in the direction of the concavity (circled in red), which is not consistent with the cell data, indicating the C-shape as the figure (d). The distribution of G cell activity is similar to that produced in the T-junction case (see Figure 6); there are proximal regions in which G cells elicit high activation (indicated by dashed white ellipses) and a nearby region in which there is uniformly weak G cell activity inside the vertical part of the convex C-shape (b). Competition between units with different receptive field sizes produces peak R cell activity within the C-shape (c) which biases the B cells to demonstrate vectorial modulation indices toward the C-shape interior (e).

would respond to the interior and exterior cavities of the segments just as well. A solution has been proposed in the form of a Bayesian belief propagation network to bias border-ownership cells toward a “skeleton” along the medial axis of a shape (Feldman & Singh, 2006; Froyen, Feldman, & Singh 2010). It has also been shown that radial kernels can locate the medial axis of a shape (Pizer et al., 1998). The combination of radially symmetric R cell kernels and network dynamics in our model can also bias border-ownership toward the medial axis of a shape, as shown in Figure 9. Because G cells respond to convexity, units with appropriate receptive field sizes yield higher activities in the locally convex portions (*1) in the convex region than in the locally concave portions (*2) of the concave region. In addition, the sharp protrusions (*3) from the curved vertical contour toward the medial axis of the convex region elicit weaker G cell responses than the convex protrusions into concave region (*4). When R cells with

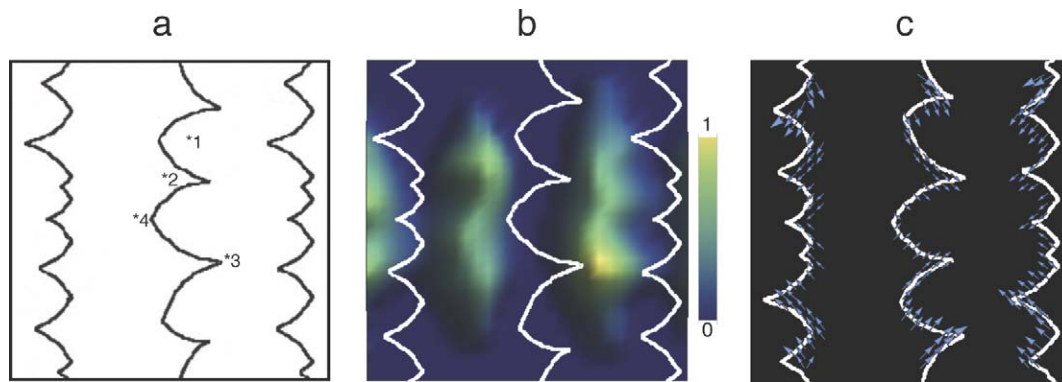


Figure 9. (a) Convexity display used in Peterson and Salvagio (2008). Human subjects were more likely to indicate the convex region (right) as figure, as compared to the concave region (left). (b) Stage 2 R cell responses. R cells respond to the convex, concave, and far-left regions, but exhibit peak activity to the convex region. We argue this reflects the human bias to indicate the convex region as figure compared to the concave region. Some R cells responded on the far left side, too, because their receptive fields are on the convex side of the farthest left contour. (c) B cells indicate directions of border ownership toward the convex regions, consistent with human psychophysical data (Peterson & Salvagio, 2008). We downsampled (10x) the convexity display shown in (a) within our simulations and the luminance was inverted in (c) for visualization.

different receptive field sizes compete and undergo normalization across scale, similar to the T-junction case, R cell activity in the locally convex portions in the convex region (*1) is low because G cell activity is high. R cells in the convex region with receptive fields near the medial axis that integrate the sharp protrusions (*3) elicit higher activation than the concave region because G cell activity is low. The R cell dynamics bias B cells toward the convex region (Figure 9b), which is consistent with the human judgments found by Peterson and Salvagio (2008). The B cell bias toward the convex region is independent of the distance between the curved vertical contours, and rather is related to the relative locations of concave and convex turns around each vertical contour. The normalized inter-scale competition in the RGB model assists in identifying the location of a proximal figure candidate and a best fit approximation to its size.

In Figure 10, we compare the ordinal relationship between B cell unit responses to square, C-shape, and square occluder displays in our model (top right) and those shown in figure 23 of Zhou et al. (2000) (bottom right). Irrespective of the shape of the figure present in the visual display and its contrast polarity, the model B cell, whose classical receptive field is indicated by the red ellipse, demonstrates border-ownership of the figure when it is to the left of the receptive field. At the spatial location indicated by the red ellipse, we take the mean activity of the most responsive B cell in time bins of 1 ms over 25 ms following the onset of the visual display. The model cells demonstrate the same relative mean firing rates as reported in neurophysiology when reflecting the figure position about the vertical axis defined by the center of the B cell's receptive field while preserving the same contrast that is present in the B

cell's receptive field (Craft et al., 2007). For example, when the C-shape appears darker than the background (Figure 10c) and the B cell has a receptive field centered on the border, indicated by the red ellipse, the cell yields less mean normalized activity compared to when the C-shape is lighter than the background and the C-shape configuration is reflected about the vertical axis (Figure 10d).

Kanizsa square

Finally, we examined the model behavior to the Kanizsa square display. In Figures 11a–b, we show the differences in B cell modulation in the presence and absence of R cells in the circuit, respectively, and inspect the modulation of border-ownership signals as inducers are progressively removed (Figures 11c–e). Without R cells, border-ownership signals favored the representation of the four individual inducers as figures (Figure 11a). The addition of R cells reversed the direction of B cell modulation for those located along the concave borders of the inducers. The vectorial modulation is toward the center of the illusory square (Figure 11b). When one inducer is removed (Figure 11c), most B cells have vectorial modulation indices that still are directed toward the illusory square. With only two inducers (Figure 11d), few B cells now exhibit vectorial modulation indices that are directed toward the illusory square, while most are directed toward the inducers. Only one inducer yields vectorial modulation indices that indicate that the inducer is the figure. The behavior of the model as inducers are removed is similar to that of the C-shape (Figure 8), since it detects the salient figure despite the presence of local concavity.

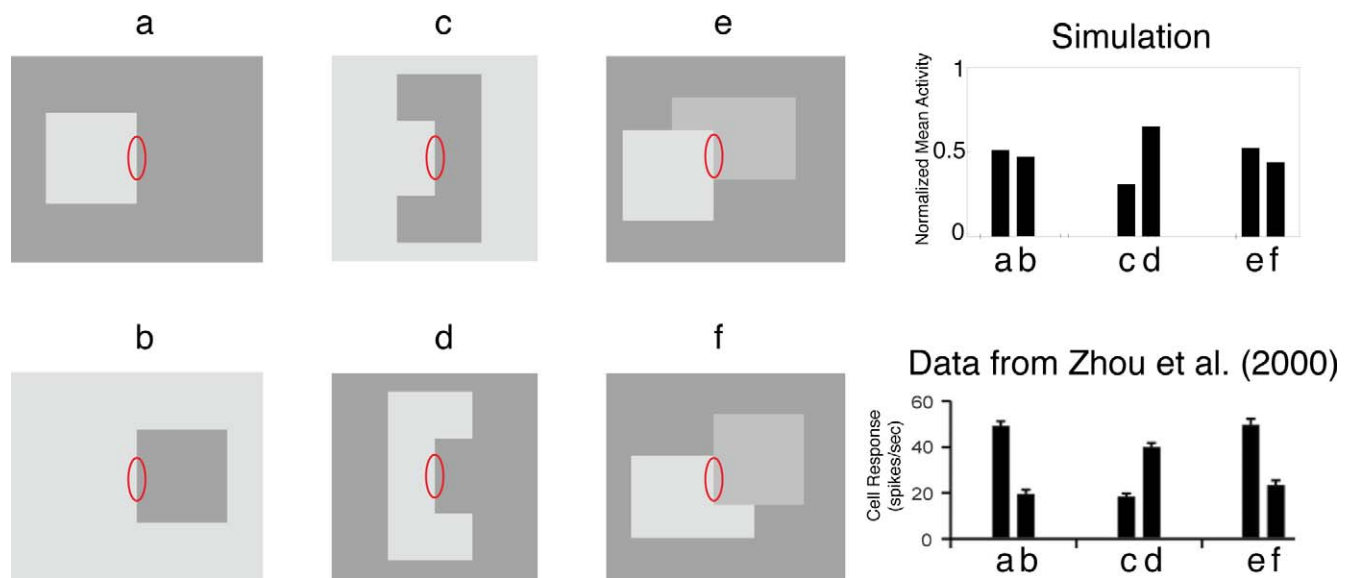


Figure 10. Irrespective of the shape of the figure present in the visual display and its contrast polarity, the model B cell demonstrates border-ownership of the figure when it is to the left of the receptive field. The model B cells (top right bar graph) exhibit the same relative mean firing rates as reported in neurophysiology (bottom right bar graph), when reflecting the figure position about the B cell’s receptive field, while maintaining the same local contrast. Consistent with the cell responses, for example, model B cells elicit greater activation when the square is positioned to the left (a) compared to the right (b) of the unit’s receptive field. Red ellipses designate the location of the B cell classical receptive field both in the simulation and in the experiments of Zhou et al. (2000). Bottom bar graph adapted from Zhou et al. (2000).

Discussion

We have shown that the RGB model network of a subset of cells identified in primate V1, V2, and V4 connected by fast interareal connections gives rise to border-ownership selectivity in a variety of visual displays in a manner that is consistent with neurophysiological and psychophysical data. The model predicts that competition between convexity-sensitive cells with on-surround receptive fields acting in a network across primate visual areas V2 and V4 can explain the

response of border-ownership cells in V2. B cells quickly provide vector information (i.e., a direction of ownership and a confidence measure in the strength of the response) that may prove useful in domains other than figure-ground segregation, such as the coordination of motor actions. For example, in situations whereby a rapid motor response is required to act on rapidly approaching objects, B cells could quickly via interareal connections provide enough information to the motor cortex without fully analyzing the visual scene. Rapid development of vector information about figures could also contribute to the stability of visual

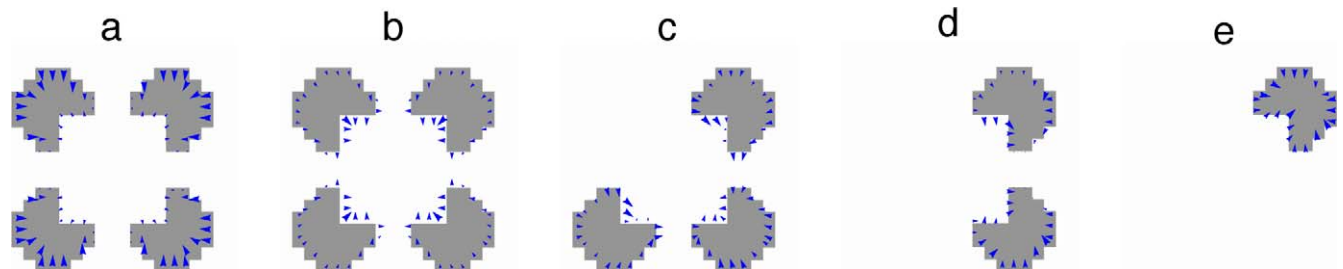


Figure 11. Kanizsa square simulation. In the absence of R cells, B cells along the pacmen inducer borders have vectorial modulation indices consistent with the percept that the inducers are figures (a). Through competition and feedback from R cells, B cells along the interior borders of pacmen reverse their vectorial modulation index directions to point toward the illusory square (b). Since progressively removing inducers diminishes the percept of the illusory square, we studied the B cell vectorial modulation indices as a function of number of inducers remaining. After removing an inducer (c), all B cells still exhibit vector directional components in the direction of the illusory square. With two inducers, 40% of B cell vectorial modulation indices have vector components toward the illusory square (d). Finally, the presence of only one inducer clearly favors it as the figure.

perception in the presence of eye movements and moving objects (Ballard & Hayhoe, 2009). Our analysis focused on the resolution of border-ownership signals in visual scenes at a glance. Evidence exists that when the interpretation of the visual scene changes, border-ownership cell activity may take some time (~100 ms) to adjust to the new information about the figures (O’Herron & von der Heydt, 2009).

Are local visual junctions special?

The occlusion of one by another is a frequent occurrence in nature. B cells in neurophysiology demonstrate sensitivity to changes in perceived occluding figure when the properties of small local junctions are altered. For example, B cells show a diminished response when a dark gray rectangle occludes a light gray rectangle (Figure 10f), compared to when the light gray rectangle is in front (Figure 10e). We tested the model’s competency in resolving border-ownership in the presence of T-junctions. Without feedback from R cells in model V4 to B cells in model V2, border-ownership signals were always directed toward the T-junction corners on the occluded surface (stem side), because of the increased convexity compared to the occluding surface (hat side; Figure 6). We showed that R cells changed the border-ownership signals to point toward the occluding surface (hat side). This reversal occurs due to the model’s on-surround competition. Qiu and von der Heydt (2007) showed that in a transparency display, border-ownership signals reversed direction within 50–100 ms of the presentation of the display. This reversal is consistent with an initial bias by G cell activity and eventual R cell modulation. It may take 50–100 ms for the competition within the intercortical network across V1, V2, and V4 to stabilize border-ownership signals in such situations. Unlike some other types of models, such as Bayesian belief propagation networks, wherein connecting model results to timing is an additional degree of freedom (free parameter), the dynamic structure of our model naturally affords a link to actual temporal units (ms). We believe that the latency that is required for border-ownership signals to develop is evidence for the dynamic coding of figure-ground segregation in the cortex.

Many have proposed that the visual system detects local luminance junctions and that they are important for determining whether objects occlude others (Finkel, 1992; Bayerl & Neumann, 2006; Craft et al., 2007; Weidenbacher & Neumann, 2009). The on-surround competitive dynamics between B, G, and R cells in model V1, V2, and V4 allow the model to determine the occluding surface or figure without explicitly detecting local luminance junctions. If the primate visual system

detects local luminance signatures such as L- and T-junctions and uses that information to determine whether an object is in front of or behind another, humans should be good at determining if a local luminance pattern marks a point of occlusion in synthetic and natural images. McDermott (2004) studied human performance on this task when subjects viewed through a small aperture synthetic junctions or junctions sampled from larger images. His analysis showed that local information alone cannot fully explain the psychophysical results. Humans erred ~11% and ~11%–27% in determining points of occlusion in the synthetic and natural junction conditions, even for the largest aperture sizes of degrees of visual angle. After considering scaling artifacts, ~25% of points that indicated the presence of occlusion could not be judged as such based on local junction information alone (McDermott, 2004). T-junctions are also not necessary to obtain a robust perception of occlusion (Shimojo, 1990; Zaidi, Spehar, & Shy, 1997; Peterson & Salvagio, 2008). There is limited neurophysiological evidence that supports the existence of cells that selectively respond to luminance junctions in lower visual areas, presumably where such an operation would need to take place (Lazareva et al., 2002). If local junctions were critical to the detection of occlusion, the scale at which they are viewed matters (Koenderink, 1984). Since V1 cells possess small receptive field sizes, any potential junction detector system would need to only consider junctions in a select number of spatial scales. Such a system would require more context outside the small V1 receptive field and very specific connectivity anatomies.

Occlusion and figure-ground interpretation in visual scenes does not always accompany the presence of T-junctions. Figure 9 shows that the RGB model also yields border-ownership results in nonrectangular displays without T-junctions, consistent with the psychophysical data that humans are more likely to indicate that convex regions are figures compared to those that are concave. The simple ring-like receptive fields of G and R cells resemble the on-surround receptive fields of cells in primate V2 and V4 and intrinsically respond to regions of convexity. The RGB model predicts that border-ownership cells would prefer the convex regions over the concave one in display shown in Figure 9.

Border-ownership recruits an interareal cortical network within the early visual system

Intra-areal fiber properties and conduction speeds impose constraints on how contextual information may propagate without feedback. Considering that the

interareal distance between V1 and V2 is only several millimeters and axonal diameters are approximately 1 μm , conduction times may be as small as 1 ms (Rockland & Virga, 1990; Nowak & Bullier, 1997). The rapid speed of interareal connections is consistent with data showing minimal V2 response latencies of roughly 10–20 ms after V1 becomes active (Bullier, 2001), and certain higher visual areas having lower response latencies than lower ones in the visual hierarchy (Hochstein & Ahissar, 2002; Paradiso, 2002). The data of Sugihara et al. (2003) do not show changes in the slope of B cell responses as a function of figure size. Although the model of Zhaoping (2005) that exclusively used intra-areal connections showed a roughly equivalent latency in the onset of border-ownership signal development, in the case of these data, the slope of the border-ownership signal strengths decreases as a function of the square display side length (see Zhaoping, 2005 Figures 6E–G). Unlike the measured cell responses, border-ownership units in Zhaoping’s model take longer to develop the same strength responses for larger figure sizes. Consistent with the data of Sugihara et al. (2003) interareal conduction speeds would not predict a difference in the slopes of border-ownership signals as a function of the figure size (Craft et al., 2007). It is unclear whether models with extensive intra-areal connections can handle the description of border-ownership responses at arbitrary scales within which humans can perform figure-ground segregation. Our model relies on fast interareal connections rather than lateral connections among retinotopically proximal units in the same area.

Border-ownership sensitivity occurs in large populations of neurons in visual areas V2 and V4, yet approximately 20% of cells in V1 also possess this property. Without feedback, it remains unclear how 20% of V1 cells can be robustly sensitive to border-ownership via intra-areal connections alone. Intra-areal interactions between neurons of different orientation and border-ownership preferences, such as those used by Zhaoping (2005), may locally contribute to border-ownership, but we suspect interareal feedback is necessary to satisfy the neural timing constraints observed in the data.

Comparison with existing models

The use of feedback between units that detect edges and those with on-surround receptive fields has existed in the literature for some time. Instead of the B/G cell nomenclature, Pizer et al. (1992) describe the theory of Cores in which these units are called “boundariness” and “medialness” detectors, respectively (Pizer et al., 1992, 1998). Boundariness and medialness detectors feature a similar connectivity anatomy by which

boundariness detectors do not directly communicate via horizontal-like connections but rather indirectly propagate information via the medialness detector. Hot spot clusters that obtain many votes from the edge detectors become known as cores and may reside on a so-called medial axis of a figure. While the theory of cores provides an algorithmic description of a system capable of identifying figures, our model proposes a biological implementation that considers known neurophysiological information about B cells.

Existing models either employ junction detectors or extensive intra-areal connectivity. Kogo, Strecha, Van Gool, and Wagemans (2010) introduced the Differentiation-Integration for Surface Completion (DISC) model to reconstruct relative surface brightness and border-ownership. Although the DISC model performs many operations often attributed to neurons, such as reconciling border-ownership, it is algorithmic instead of neurally described. In a process similar to that performed by the Retinex algorithm (Land, 1986), the DISC model first constructs a map representing luminance ratios between different regions of the input display. In parallel, the algorithm assigns border-ownership throughout the image. DISC determines border-ownership based on a priori knowledge of the junction distribution and labels T- and L-junctions using explicit rules. The DISC model simulates the illusory brightness percept in the Kanizsa square display as a function of different choices of inducers. Unlike the DISC model, the RGB model can simulate the predicted border-ownership responses yielded to the Kanizsa square display (Figure 11) in a neurally described rather than an algorithmic framework.

Zhaoping (2005) developed a neural model to describe how, in the absence of feedback from higher visual areas, cells in V2 could exhibit border-ownership sensitivity. The model uses intra-areal connectivity rules built on Gestalt grouping assumptions, such as convexity. For example, the model confers larger synaptic weights between neurons separated by a right-turn compared to a left-turn and adjacent neurons sampling like contrast or possess like border-ownership preferences signals facilitate or otherwise suppress each other. Units in the model are also assumed to possess end-stopped properties. Although the model does not incorporate an explicit T-junction detection stage, the Gestalt connectivity rules effectively afford border-ownership cells with receptive fields centered at T-junctions, a bias consistent with occlusion. Because T-junctions feature bottom-up contrast signals of orthogonal contrast orientations, border-ownership units sampling the T-junction hat and stem would compete if the units had differing border-ownership preferences. Competitive ties are broken through the use of random noise or the inhibitory/facilitatory influence of adjacent units with (in)consis-

tent border-ownership or contrast orientation preferences. Unlike the model of Zhaoping (2005), which cannot reconcile the presence of figures at multiple spatial scales with connectivity limited by known physiological fiber lengths, our model is fundamentally based on physiological data of interareal connections that do not face the same temporal constraints as intra-areal fibers and cells with a variety of receptive fields sizes.

While the Form-And-Color-And-Depth (FACADE) model does not address the dynamics of border-ownership cells in primate visual cortex, its goal is to explain how presentations of 2D displays can give rise to 3D figure-ground perception by reconstructing surface brightness at different depth planes, despite whether the surface is visible (modal) or not (amodal). FACADE theory argues that representations of both modal and amodal percepts are required to act on visual information and recognize partially occluded objects, respectively (Kelly & Grossberg, 2000). FACADE builds on the boundary contour system/feature contour system (BCS/FCS) (Cohen & Grossberg, 1984; Grossberg & Mingolla, 1985) that describes how a complementary neural representation of borders and surfaces can give rise to form and brightness in visual perception. First, the system performs on-center/off-surround processing to discount the illuminant but maintains relative contrast strengths in the borders. The BCS performs filtering to determine the borders in the visual display, regardless of contrast polarity and orientation. The FCS performs boundary-gated, nearest-neighbor diffusive filling-in jointly based on the outputs of the output BCS and on-center/off-surround processing. Through the use of recurrent competition between bipole long-range grouping units, which possess a bow-tie-like receptive field with two excitatory lobes, and hypercomplex units, which act as end-stopped complex cells, the FACADE model does not require explicit T-junction detectors. While obtaining a border-ownership readout based on the FACADE model is possible, the process is indirect. Our model specifically focuses on the dynamics of cells that produce border-ownership signals.

Free-space border-ownership

Our results of the Kanizsa square display demonstrate that the model can reconcile figure-ground segregation in the presence of illusory figures. Importantly, we show with the progressive removal of inducers, the border-ownership responses become less consistent with an illusory form and more consistent with the individual inducers serving as the figure. The DISC model of Kogo et al. (2010) uses an interesting phenomenon called *free-space border-ownership* where-

by units develop border-ownership preferences in the absence of bottom-up contrast signals. Some models (Zhaoping, 2005; Craft et al., 2007), including our own, assume border-ownership units require bottom-up stimulation to acquire side-of-figure preferences. Current neurophysiology has not reported the existence of cells with free-space border-ownership sensitivity and indicates that border-ownership cells are edge-gated (e.g., see figure 11 of Zhou et al., 2000). Nevertheless, the possible existence of free-space border-ownership cells has theoretical interest and should be investigated further in displays with illusory contours, such as the Kanizsa square. Such cells would complement our results shown in Figure 11 obtained with only edge-gated B cells.

The RGB model performs figure-ground segregation and is sensitive to occlusion without the use of junction detectors. Perception of local junctions may be an outcome of operations such as captured in our model, as suggested by the data of McDermott (2004), which show that allegedly local junctions can best be discerned only with large context in natural scenes. Evidently the visual system does not rely on specialized junction circuits to perform figure-ground segregation. Occlusion information conferred by T-junctions may be a specialized case of a more general process of inter-scale competition within a multi-area intercortical network.

Acknowledgments

This work has been supported in part by CELEST, an NSF Science of Learning Center (NSF SBE-0354378 and NSF OMA-0835976), AFOSR FA9550-12-1-0436, and the Office of Naval Research (ONR N00014-11-1-0535). The authors would like to thank Anne Martin for helpful discussions and two anonymous reviewers for their valuable feedback.

Commercial relationships: none.

Corresponding author: Arash Yazdanbakhsh.

Email: yazdan@bu.edu.

Address: Center for Computational Neuroscience and Neural Technology, Program in Cognitive and Neural Systems, Boston University, Boston, MA, USA.

References

- Adelson, E. H., & Anandan, P. (1990). *Ordinal characteristics of transparency*. Retrieved July 29, 2012 from Massachusetts Institute of Technology, Media Laboratory Vision and Modeling Group

- Web site: http://web.mit.edu/persci/people/adelson/pub_pdfs/ordinal90.pdf.
- Anderson, B. L. (1997). A theory of illusory lightness and transparency in monocular and binocular images: The role of contour junctions. *Perception*, 26(4), 419–453.
- Angelucci, A., Levitt, J., Walton, E., Hupe, J., Bullier, J., & Lund, J. (2002). Circuits for local and global signal integration in primary visual cortex. *The Journal of Neuroscience*, 22(19), 8633–8646.
- Ballard, D. H., & Hayhoe, M. M. (2009). Modelling the role of task in the control of gaze. *Visual Cognition*, 17(6–7), 1185–1204.
- Bayerl, P., & Neumann, H. (2006). Disambiguating visual motion by form-motion interaction—a computational model. *International Journal of Computer Vision*, 72(1), 27–45.
- Bullier, J. (2001). Integrated model of visual processing. Brain research. *Brain Research Reviews*, 36(2–3), 96–107.
- Bushnell, B. N., Harding, P. J., Kosai, Y., & Pasupathy, A. (2011). Partial occlusion modulates contour-based shape encoding in primate area V4. *The Journal of Neuroscience*, 31(11), 4012–4024.
- Carandini, M., & Heeger, D. J. (2011). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, 13, 51–62.
- Cohen, M. A., & Grossberg, S. (1984). Neural dynamics of brightness perception: Features, boundaries, diffusion, and resonance. *Perception & Psychophysics*, 36(5), 428–456.
- Craft, E., Schutze, H., Niebur, E., & von der Heydt, R. (2007). A neural model of figure-ground organization. *Journal of Neurophysiology*, 97(6), 4310–4326.
- Feldman, J., & Singh, M. (2006). Bayesian estimation of the shape skeleton. *Proceedings of the National Academy of Sciences of the United States of America*, 103(47), 18014–18019.
- Finkel, L. H. (1992). Object discrimination based on depth-from-occlusion. *Neural Computation*, 4(6), 901–921.
- Freeman, W. T., & Adelson, E. H. (1991). The design and use of steerable filters. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(9), 891–906.
- Froyen, V., Feldman, J., & Singh, M. (2010). A Bayesian framework for figure-ground interpretation. *Advances in Neural Information Processing Systems*, 23, 1–9.
- Gattass, R., Sousa, A. P., Mishkin, M., & Ungerleider, L. G. (1997). Cortical projections of area V2 in the macaque. *Cerebral Cortex*, 7(2), 110–129.
- Gattass, R., Gross, C. G., & Sandell, J. H. (1981). Visual topography of V2 in the macaque. *The Journal of Comparative Neurology*, 201(4), 519–539.
- Girard, P., Hupe, J., & Bullier, J. (2001). Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *Journal of Neurophysiology*, 85(3), 1328.
- Grossberg, S. (1973). Contour enhancement, short term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics*, 52(3), 213–257.
- Grossberg, S., & Mingolla, E. (1985). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception & Psychophysics*, 38(2), 141–171.
- Hegde, J., & Van Essen, D. C. (2004). Temporal dynamics of shape analysis in macaque visual area V2. *Journal of Neurophysiology*, 92(5), 3030–3042.
- Hegde, J., & Van Essen, D. C. (2006). A comparative study of shape representation in macaque visual areas V2 and V4. *Cerebral Cortex*, 17(5), 1100–1116.
- Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5), 791–804.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of Physiology*, 160, 106–154.
- Kanizsa, G., & Gerbino, W. (1976). Convexity and symmetry in figure-ground organization. In M. Henle (Ed.), *Vision and artifact* (pp. 25–32). New York: Springer.
- Kelly, F., & Grossberg, S. (2000). Neural dynamics of 3-D surface perception: Figure-ground separation and lightness perception. *Attention, Perception, & Psychophysics*, 62(8), 1596–1618.
- Koenderink, J. J. (1984). What does the occluding contour tell us about solid shape? *Perception*, 13(3), 321–330.
- Kogo, N., Strecha, C., Van Gool, L., & Wagemans, J. (2010). Surface construction by a 2-d differentiation-integration process: A neurocomputational model for perceived border ownership, depth, and lightness in Kanizsa figures. *Psychological Review*, 117, 406–439.
- Land, E. H. (1986). Recent advances in Retinex theory. *Vision Research*, 26(1), 7–21.
- Lazareva, N. A., Shevelev, I. A., Novikova, R. V.,

- Tikhomirov, A. S., Sharaev, G. A., & Tsutskiridze, D. Y. (2002). The disinhibitory zone of the striate neuron receptive field and its sensitivity to cross-like figures. *Neuroscience and Behavioral Physiology*, 32(6), 595–602.
- McDermott, J. (2004). Psychophysics with junctions in real images. *Perception*, 33(9), 1101–1127.
- Mihalas, S., Dong, Y., von der Heydt, R., & Niebur, E. (2011). Mechanisms of perceptual organization provide auto-zoom and auto-localization for attention to objects. *Proceedings of the National Academy of Sciences of the United States of America*, 108(18), 7583–7588.
- Nieder, A., & Miller, E. K. (2004). A parieto-frontal network for visual numerical information in the monkey. *Proceedings of the National Academy of Sciences of the United States of America*, 101(19), 7457.
- Nowak, L. G., & Bullier, J. (1997). The timing of information transfer in the visual system. In K. L.S. Rockland, J. H. Kaas, & A. Peters (Eds.), *Cerebral cortex: Vol. 12. Extrastriate cortex in primates* (pp. 205–241). New York: Springer.
- Nowak, L. G., Munk, M. H., Girard, P., & Bullier, J. (1995). Visual latencies in areas V1 and V2 of the macaque monkey. *Visual Neuroscience*, 12(2), 371–384.
- O'Herron, P., & von der Heydt, R. (2009). Short-term memory for figure-ground organization in the visual cortex. *Neuron*, 61(5), 801–809.
- Paradiso, M. A. (2002). Perceptual and neuronal correspondence in primary visual cortex. *Current Opinion in Neurobiology*, 12(2), 155–161.
- Pasupathy, A., & Connor, C. E. (1999). Responses to contour features in macaque area V4. *Journal of Neurophysiology*, 82(5), 2490–2502.
- Peterson, M. A., & Salvagio, E. (2008). Inhibitory competition in figure-ground perception: Context and convexity. *Journal of Vision*, 8(16):4, 1–13, <http://www.journalofvision.org/content/8/16/4>, doi:10.1167/8.16.4. [PubMed] [Article]
- Pizer, S. M., Burbeck, C. A., Coggins, J. M., Fritsch, D. S., & Morse, B. S. (1994). Object shape before boundary shape: Scale-space medial axes. *Journal of Mathematical Imaging and Vision*, 4(3), 303–313.
- Pizer, S. M., Eberly, D., Morse, B. S., & Fritsch, D. S. (1998). Zoom-invariant vision of figural shape: The mathematics of cores. *Computer Vision and Image Understanding*, 69(1), 55–71.
- Qiu, F. T., & von der Heydt, D. (2007). Neural representation of transparent overlay. *Nature Neuroscience*, 10(3), 283–284.
- Rockland, K. S., & Virga, A. (1990). Organization of individual cortical axons projecting from area V1 (area 17) to V2 (area 18) in the macaque monkey. *Visual Neuroscience*, 4(1), 11–28.
- Shepard, R. N. (1990). Mind sights. *Original visual illusions, ambiguities, and other anomalies, with a commentary on the play of mind in perception and art* (Vol. 76). New York: W.H. Freeman & Co.
- Shimojo, S. N. K. (1990). Amodal representation of occluded surfaces: role of invisible stimuli in apparent motion correspondence. *Perception*, 19(3), 285–299.
- Sincich, L. C., & Horton, J. C. (2005). The circuitry of V1 and V2: Integration of color, form, and motion. *Annual Review of Neuroscience*, 28, 303–326.
- Sugihara, T., Qiu, F. T., & von der Heydt, R. (2003). Border ownership coding in monkey area V2: Dynamics of image context integration. *Society for Neuroscience Abstracts*, 29, 819.12.
- von der Heydt, R., Qiu, F. T., & He, Z. J. (2003). Neural mechanisms in border ownership assignment: Motion parallax and gestalt cues. Presented at the 3rd Annual meeting of the Vision Sciences Society, Sarasota, FL, May 2003.
- von der Heydt, R., Zhou, H., & Friedman, H. (2000). Representation of stereoscopic edges in monkey visual cortex. *Vision Research*, 40(15), 1955–1967.
- Weidenbacher, U., & Neumann, H. (2009). Extraction of surface-related features in a recurrent model of V1-V2 interactions. *Plos One*, 4(6), e5909.
- Zaidi, Q., Spehar, B., & Shy, M. (1997). Induced effects of backgrounds and foregrounds in three-dimensional configurations: The role of T-junctions. *Perception*, 26(4), 395–408.
- Zhang, N. R., & von der Heydt, R. (2010). Analysis of the context integration mechanisms underlying figure-ground organization in the visual cortex. *The Journal of Neuroscience*, 30(19), 6482–6496.
- Zhaoping, L. (2005). Border ownership from intracortical interactions in visual area V2. *Neuron*, 47(1), 143–153.
- Zhou, H., Friedman, H., & von der Heydt, R. (2000). Coding of border ownership in monkey visual cortex. *The Journal of Neuroscience*, 20(17), 6594–6611.