A Neural Model of Figure–Ground Organization

Edward Craft,[1,3] Hartmut Schütze,[3] Ernst Niebur,[2,3] and Rüdiger von der Heydt[2,3]

[1]*Department of Biophysics,* [2]*Department of Neuroscience, and* [3]*The Zanvyl Krieger Mind/Brain Institute; Johns Hopkins University; Baltimore, Maryland*

Craft E, Schütze H, Niebur E, von der Heydt R. A neural model of figure–ground organization. *J Neurophysiol* 97: 4310–4326, 2007. First published April 18, 2007; doi:10.1152/jn.00203.2007. Psychophysical studies suggest that figure–ground organization is a largely autonomous process that guides—and thus precedes—allocation of attention and object recognition. The discovery of border-ownership representation in single neurons of early visual cortex has confirmed this view. Recent theoretical studies have demonstrated that border-ownership assignment can be modeled as a process of self-organization by lateral interactions within V2 cortex. However, the mechanism proposed relies on propagation of signals through horizontal fibers, which would result in increasing delays of the border-ownership signal with increasing size of the visual stimulus, in contradiction with experimental findings. It also remains unclear how the resulting border-ownership representation would interact with attention mechanisms to guide further processing. Here we present a model of border-ownership coding based on dedicated neural circuits for contour grouping that produce border-ownership assignment and also provide handles for mechanisms of selective attention. The results are consistent with neurophysiological and psychophysical findings. The model makes predictions about the hypothetical grouping circuits and the role of feedback between cortical areas.

INTRODUCTION

Visual perception begins at the retina, with cluttered two-dimensional (2D) projections of the three-dimensional (3D) world. In the early twentieth century, Gestalt psychologists observed that we tend to organize this clutter through a process of figure–ground segregation—i.e., by identifying those regions of the retinal images that are object-related (figures) for further processing, and relegating other regions to the background (Rubin E 1921, 2001; Wertheimer 1923, 2001). Borders appear where objects occlude what is behind them and can often be identified by discontinuities in color, luminance, or texture. However, simply locating these borders is not enough to distinguish figure from ground. It is also necessary to infer from the image context which of the two regions abutting along each border corresponds to the closer, occluding surface, and is defined by—or "owns"—the border (Nakayama et al. 1989, 1995; Rubin E 1921, 2001). As the illustrations in Fig. 1 demonstrate, this assignment of border ownership critically determines how we perceive individual regions of a scene (Fig. 1, *A* and *B*) and how objects are recognized (Fig. 1, *C* and *D*).

Although there is evidence suggesting that high-level object understanding can influence figure–ground segregation (see, e.g., Peterson et al. 1991), most studies have concluded that figure–ground segregation precedes high-level processing. Recognition of objects or contour shapes depends on the assignment of border ownership (Driver and Baylis 1996; Nakayama et al. 1989). Such an effect can be observed in Fig. 1*C*, where manipulation of border ownership impedes the recognition of familiar objects. It has also been observed that, in cluttered displays, a shape that pops out in visual search when it "owns" its borders can only be found with scrutiny when the apparent depth ordering is changed so that some of its borders get assigned to an alternate foreground object (He and Nakayama 1992; Rensink and Enns 1998). Thus border-ownership assignment seems to precede the process of object recognition, and the deployment of attention (such as in a search task) seems to work on a structured representation in which borders are already assigned to the corresponding regions or objects. The explanation of these findings presents a conundrum, though: If higher-level processing of objects depends on figure–ground segregation, how can figure–ground segregation depend on understanding image context?

Background

The problem of figure–ground organization was previously addressed by several theoretical studies. In this section, we briefly review the models proposed by these studies, in the context of recent neurophysiological findings. Although some were quite effective in resolving the occlusion structure of images and successfully explained the perceptual phenomena, we argue that consideration of neural coding mechanisms, distance and speed limitations on cortical processing, and the need for an interface to central processes such as selective attention requires a different approach. In subsequent sections, we then present an alternative model architecture that satisfies these constraints.

PRINCIPLES OF CODING. Two general schemes have been proposed for figure–ground representation: region labeling and border-ownership coding. Whereas region labeling means that regions are differentiated by labeling the corresponding elements in an isomorphic surface representation (like labeling pixels in a bitmap), border-ownership coding involves the contour representation and thus orientation-selective neurons.

One class of models has used region labeling, assuming, for example, that color/brightness signals spread within a cortical sheet to fill in the regions within the boundaries given by the contour representation (Arrington 1994; Gerrits and Vendrik 1970; Grossberg 1994; Grossberg and Mingolla 1985; Roelfsema et al. 2002). Although such mechanisms may be appealing theoretically, it is not clear whether spreading of color/brightness signals in the visual cortex occurs (Roe et al. 2005;
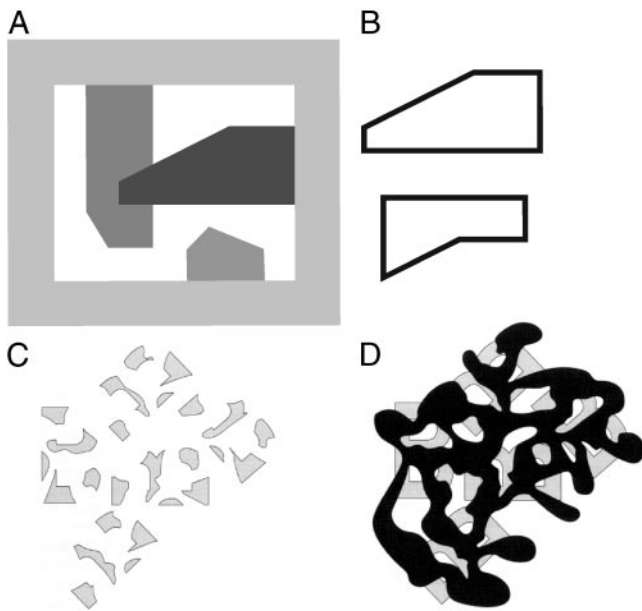
FIG. 1. Border ownership influences shape perception and object recognition. After examining the scene in *A*, observers tend to remember the top shape in *B*, but not the bottom one. Although there are regions in *A* corresponding to both shapes, one region "owns" its borders and is perceived as a definite figure, whereas the other region is lumped into the background (modified from Rubin N 2001). In *C*, the shaded regions "own" all of their borders and appear to be dissociated and meaningless. When a plausible occluding surface is introduced (*D*), the fragments lose ownership of key borders and flow together perceptually in the background. Only then can we recognize them as pieces of familiar objects (modified from Kandel et al. 2000, after Bregman 1981).

Rossi and Paradiso 1999; von der Heydt et al. 2003). Figure–ground organization has been studied in area V1, where neural responses representing a figure region were found to be enhanced compared with responses representing the ground region (Lamme 1995; Lee et al. 1998; Zipser et al. 1996). This could be interpreted as region labeling.

The present model is based on the data from recent studies of border-ownership coding. We chose to model these results because the border-ownership signals, particularly those in area V2, tend to be stronger and emerge earlier than the figure enhancement in V1, suggesting that the latter is the result of feedback from V2 or other extrastriate areas (cf. Lamme et al. 1998). Zhou et al. (2000) found that orientation-selective cortical neurons, which are usually thought to represent local contour properties such as edge contrast and orientation, are also sensitive to the global configuration of contours and code for border ownership. A neuron may respond to a contrast edge with a high firing rate if the edge is a contour of a figure on one side of the receptive field, but with a low firing rate if it is a contour of a figure on the other side. Many of these neurons combine side-of-figure selectivity with selectivity for the depth order of surfaces, as defined by stereoscopic cues (Qiu and von der Heydt 2005) or by dynamic occlusion (von der Heydt et al. 2003). This shows that side-of-figure selectivity is not just a random asymmetry in the wiring of receptive fields, but has to do with the strife for a 3D interpretation of the image. Thus the side-of-figure–related response modulation reflects figure–ground organization. The most intriguing aspect of the neurophysiological findings is the influence of the image context. Apparently, V2 neurons have some knowledge of global shape even when the shapes are much larger than their receptive fields.

In essence, border-ownership selectivity means that each contour element is represented by two pools of neurons, one for each side of ownership, whose differential activity codes for border ownership (more precisely, the degree of border-ownership assignment), whereas their common activity codes for the local contour attributes such as orientation, motion, color/luminance contrast, and so forth. It is interesting to see the assignment of borders to regions, which is relational information, encoded in the firing rate of neurons like other contour attributes. It was often assumed that coding of relations between features requires a qualitatively different mechanism such as synchronized oscillation across neurons (Singer and Gray 1995). The preceding results suggest that border ownership is represented by opponent channels, just as light and dark are represented by on- and off-center ganglion cells, and direction of motion by neurons in MT cortex. Although the evidence described earlier comes from recordings in monkey visual cortex, psychophysical experiments indicate that border-ownership–selective neurons also exist in the human visual cortex (von der Heydt et al. 2005). Taken together, these findings indicate that the visual system explicitly represents border ownership at an early cortical level following the stage of local feature representation.

Representing figure–ground relationships in terms of border ownership seems plausible, for theoretical as well as physiological reasons. The contours carry most of the information about the shape of objects. Consequently, as discussed earlier (Fig. 1), border-ownership assignment is critical for object recognition. The vast majority of neurons in V1 and V2 are edge selective and orientation tuned, and a comparison between the activity evoked by the borders and the interior of a figure shows that the border signals are five- to sixfold stronger than the surface signals (Friedman et al. 2003).

Several studies have modeled figure–ground organization in terms of border-ownership coding. Some of these assume that image context integration is achieved by lateral propagation of signals within the image representation, such as by horizontal fibers in area V2 (Baek and Sajda 2005; Kikuchi and Akashi 2001; Nishimura and Sakai 2004, 2005; Pao et al. 1999; Zhaoping 2005). As will be discussed in the next section, there are physiological constraints that limit the speed of image context integration in this type of architecture. Other models advocate feedback from higher cortical areas (Finkel and Sajda 1992; Kienker et al. 1986; Kikuchi and Fukushima 2003; Sajda and Finkel 1995; Yu et al. 2001). These models are not constrained by the limitations of feedforward and lateral connections.

Perhaps the most extensive of these models, in terms of its ability to integrate multiple cue types and bind features for visual surface construction, is that of Sajda and Finkel (1995). It uses algorithms that identify contour terminations, resolve junctions, and identify closed contours, enabling the system to "tag" complete contour segments. Border ownership, represented as a binary value, is then assigned for each segment. Although this model implements border-ownership coding, and thus parallels the physiology better than region-labeling models, questions about its neural implementation remain. For example, it is unclear how the identification of contour segments (which is a global operation) might be realized and how

the tags might be represented. The authors suggest coherent oscillation of the neurons representing the same segment as a possible mechanism. However, the functional role of such oscillations in primate visual cortex under awake conditions has been debated (Bair et al. 1994; Young et al. 1992). Our own recordings from pairs of cells ($n = 37$) failed to show significant coherent oscillations, whether both cells were activated by a common contour segment or by different, unrelated contour segments (F Qiu, H Schütze, and R von der Heydt, unpublished observations). Also, because we know now that border ownership modulates firing rates, the coherent oscillation hypothesis appears as a rather remote possibility. Finally, the gradual nature of the neural border-ownership signals (Zhou et al. 2000) argues against binary coding and the global tagging scheme. The existence of neurons with a fixed border-ownership preference calls for a revision of the basic concept.

PHYSIOLOGICAL CONSTRAINTS. The recent physiological results place important constraints on modeling. First, the extent of visual context integration in border-ownership modulation is much larger than the classical receptive field. Second, it was found that the border-ownership signal emerges with a short latency. Figure 2 illustrates the time course of border-ownership signals (and edge signals, for comparison) in the critical
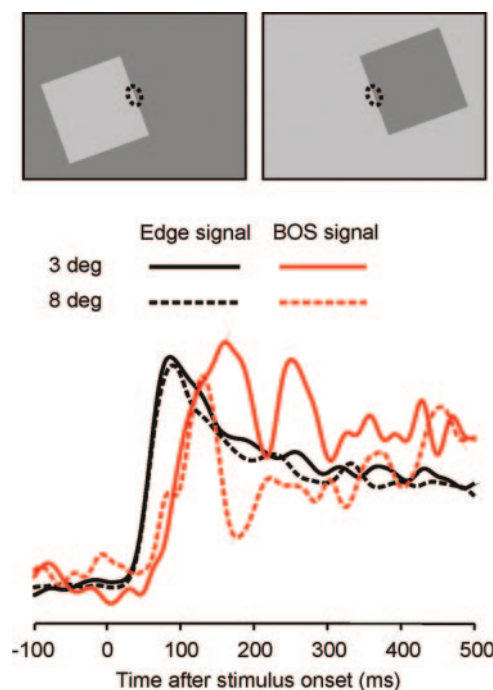

FIG. 2. Timing of border-ownership signals does not depend on the spatial extent of image context integration. Neural signals (from single-unit recordings in macaque V2) representing the center of the edge of a square are shown for square sizes of 3° (full lines) and 8° (dashed lines). Border-ownership signals (red) were defined as the difference between the responses for the 2 sides of figure location (average of smoothed normalized firing rates of 42 V2 neurons). Edge signals (defined as the mean of the 2 responses) are also shown for comparison (black). Amplitudes of border-ownership–selective (BOS) signals and edge signals are scaled differently, but the signals for the 2 figure sizes are plotted on the same scale. Note that the BOS signals for 8° figures (dashed red line) are not delayed relative to the BOS signals for 3° figures (solid red line), and that both rise at approximately the same rate, whereas models based on within-area connectivity predict significant delays, because the cortical distance from the representation of the critical context features to these neurons is considerably longer for the 8° figures than for the 3° figures. Reproduced from Sugihara et al. (2003).

tests. The stimulus configuration is shown schematically at the *top*. The receptive field (ellipse) was stimulated with a straight edge that could be the border of a square either on one side or on the other. The important point is that the entire region of visual field occupied by the two squares received identical stimulation in the two conditions. Thus any difference between the responses indicates an influence from stimulus features outside this region. The size of this region (and thus the minimum extent of spatial context that needs to be integrated for border-ownership assignment) is given by the size of the squares. In the experiment illustrated in Fig. 2, square sizes of 3 and 8° of visual angle were tested. Thus the region of identical stimulation was either 3 × 6 or 8 × 16°. The black curves show the time course of the edge signals (the average of the firing rates for the two figure locations) and the border-ownership signals (the difference between the firing rates for the two figure locations) are shown in red. It can be seen that the border-ownership signal emerges well before 100 ms, and with only a small delay after the edge signals (which are representative of V2 neurons in general). Importantly, there also appears to be no difference in latency between the border-ownership signals for large and small figures. That is, context integration over larger distances in the visual field does not take more time than context integration over smaller distances.

The integration of image context cannot be achieved by convergence in the afference because its spatial extent is much larger than the size of the classical receptive fields (8° is about tenfold the average size of the receptive fields of V2 neurons in the near-foveal region that was studied). It must involve either horizontal fibers within V2 cortex, or recurrent fibers from other extrastriate areas, or both.

A model of border ownership that aims to account for the recent neural data, as well as the perceptual phenomena, assumes that context integration occurs by horizontal fibers within V2 (Zhaoping 2005). This parsimonious model reproduces the observed border ownership data from assumptions about the lateral connectivity in V2. Briefly, it posits that neurons with nearby receptive fields are linked by excitatory and inhibitory connections depending on whether the corresponding border segments are consistent with being contours of the same figure.

The assumption made in this model (Zhaoping 2005) and others (Baek and Sajda 2005; Grossberg 1994; Kikuchi and Akashi 2001; Nishimura and Sakai 2004, 2005; Pao et al. 1999), that image context integration occurs within the area, through horizontal fibers, implies that the latency of the border-ownership signal would increase with the distance of the relevant image context from the receptive field under consideration. This is a consequence of the retinotopic representation of visual information in V2 cortex.

As pointed out before (Zhou et al. 2000), the distances in V2 cortex are considerable and the conduction through intracortical fibers is probably too slow to explain the short latencies of border-ownership signals. Conduction velocity estimates for these fibers range between 0.1 and 0.25 m/s in cat V1 (Bringuier et al. 1999; Grinvald et al. 1994) and 0.33 m/s in monkey V1 (Girard et al. 2001); we are not aware of corresponding data for V2. Note that these figures are median values; there is a range of conduction velocities of single fibers, and it has been argued that longer fibers might conduct much faster than shorter fibers (Zhaoping 2005).

To make a quantitative argument, it is necessary to consider the topography of area V2. A look at the well-known illustration of the unfolded cortical areas (Felleman and Van Essen 1991) shows that V2 is a large area whose elongated shape is quite different from that of V1. In V2, the visual field representation is split at the horizontal meridian into a ventral part and a dorsal part that are connected only by a narrow bridge. Detailed maps of V2 (Gattass et al. 1981) show that intracortical fibers would have to span considerable distances to explain the findings on border-ownership coding (Qiu and von der Heydt 2005; Zhou et al. 2000).

As an example, consider responses produced by a square figure of 8° side length. When an edge of the square is centered about the receptive field of the neuron under study, which was the condition used in the cited studies, the closest points that can provide border-ownership information are the corners on both ends of the edge, with a distance of 4° visual angle from the receptive field center. The representation of one of those corners is generally also the nearest point in cortex to the neuron where such information is available. From Fig. 9 of Gattass et al. (1981) it can be seen that two neurons with receptive fields located 2.5 and 6.5° below the center of gaze (cells 5 and 11) are separated by 21 mm [because no scale bars are provided, we estimated the scale based on brain sections of the monkeys of Zhou et al. (2000) to be 3:1]. Thus if the vertically oriented edge of the square were centered on the receptive field of cell 5, then the bottom corner would be represented at a distance of 21 mm. The representation of the top corner would be even farther away, in the ventral part of V2. For edges of horizontal orientation the situation is similar (consider cells 61 and 53): if the center of the edge is at 0.6° horizontal eccentricity, one corner would be at 4.6° eccentricity on the horizontal meridian and represented in cortex at a distance of about 27 mm, and the other corner would be represented in the opposite hemisphere. Because the maximal length of horizontal connections is only 3–4 mm (measured in V1; Table 1 in Angelucci et al. 2002), the signals would have to be relayed many times through a cascade of neurons. Moreover, activity cannot be relayed through cells that are not directly stimulated (this would contradict the notion of the classical receptive field: stimuli outside this receptive field generally do not elicit responses). Thus signals could propagate only through neurons that are excited by the given contrast borders.

Any model that attempts to explain border-ownership coding in V2 faces the problem that neurons have to communicate over those distances. If communication is assumed to occur within V2, and assuming a median conduction velocity for horizontal fibers of 0.3 m/s (Girard et al. 2001), distances of 21–27 mm would imply delays of 70–90 ms just from signal propagation, whereas the observed total delays are only about 30 ms (Fig. 2). These figures show that conduction delays are a serious problem for models based on intracortical connections, not to mention the problem of communication between points in left and right hemifields, for which intracortical connections do not exist, or between points in upper and lower quadrants of a hemifield, for which such connections may not exist. In contrast, the average length of white matter fibers connecting V2 and V4 is no more than about 20 mm (estimated from Gattass et al. 1988; Fig. 3), corresponding, at 3.5 m/s (Girard et al. 2001), to about 6-ms conduction delay, or 12 ms for the loop. Loops through V3 would be even faster. In addition, white-matter connections of comparable speeds exist within and between hemispheres (Swadlow 2000) and, presumably, the feedback connections from higher-order areas target the ventral and dorsal representations of V2 alike, without major discontinuities in the time course of signals.

In summary, considerations of the V2 topography and the timing issue indicate that the intrinsic connectivity *alone* is not sufficient to explain the findings on border-ownership coding. In the case of area V1, detailed comparisons of the spatial range of surround effects with the existing anatomical connec-
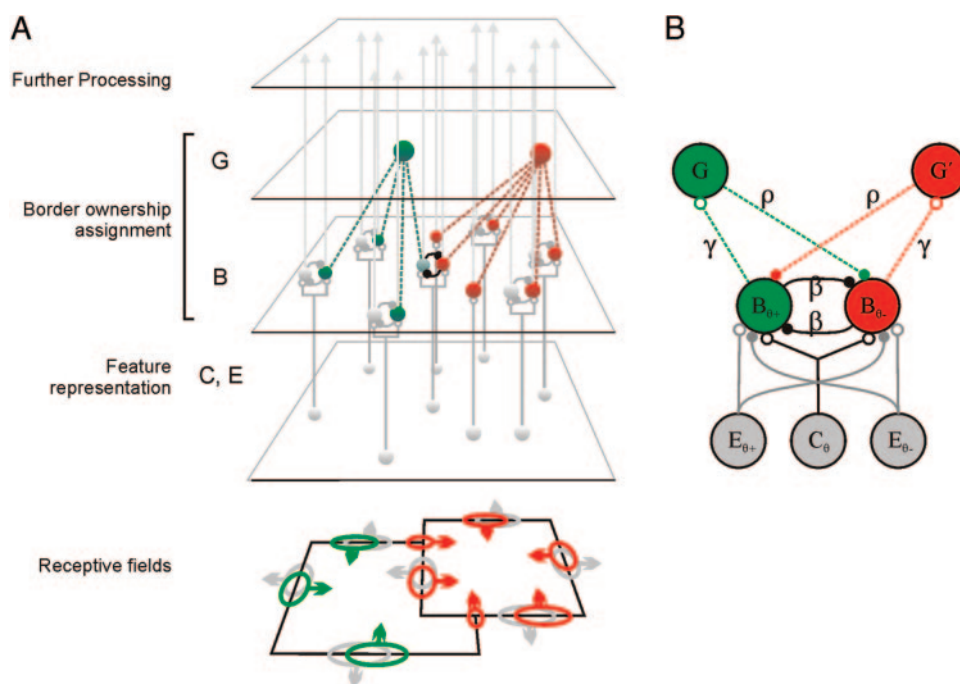

FIG. 3. Model architecture. *A*: network overview, showing border-ownership selection for a stimulus of 2 overlapping rectangles (*bottom*). Receptive fields of B cells are shown as ellipses, where attached arrows indicate their preferred side of figure. B cells with opposite arrows compete, and this competition is decided by grouping cell input (receptive fields of active cells are shown in green and red; receptive fields of suppressed cells shown in gray). Output from the B cells is then passed on to higher-order areas, not the subject of this study (but see Fig. 11 and DISCUSSION). *B*: microcircuitry of the model network. Input to the model is provided by orientation-tuned edge detectors, C_θ, and end-stopped cells, $E_{\theta\pm}$. At a particular location in the visual scene, neuron C_θ is activated by the presence of an edge with orientation θ. Each edge detector provides input to a pair of B cells, $B_{\theta+}$ and $B_{\theta-}$, which are mutually inhibitory (strength β). B cells project to grouping cells (G and G′) with strength γ and receive inhibition from those with strength ρ. For clarity, feedback from only one size of G cells (see text) has been drawn, and connections between the G cells and other pairs of B cells have been omitted. $E_{\theta\pm}$ bias edges toward one side of ownership wherever line terminations occur (e.g., at T-junctions; see METHODS).

tions have indicated that feedback connections from higher-order areas are responsible for producing the nonclassical surround effects (Angelucci and Bullier 2003; Angelucci et al. 2002). By analogy, it is plausible that image context integration for border-ownership assignment in area V2, which is a non-classical surround effect of large spatial extent, also involves fast loops through higher-order areas.

FIGURE–GROUND ORGANIZATION AND MECHANISMS OF ATTENTION. Our motivation for the present study comes also from the need for a more general point of view. Most of the existing models leave open the question of how the figure–ground assignments they compute influence, and are influenced by, higher-level processes (some exceptions are Carpenter and Grossberg 1993; Grossberg et al. 1994; Vecera and O'Reilly 1998). The models transform one retinotopic representation into another in which regions are labeled as figure and ground, or contour segments are assigned border ownership. Although the result is an improved representation, it is essentially no more than an enhanced image. To avoid referring the further processing to a homunculus that views this internal image, we have to offer at least a hypothesis of how the figure–ground representation produced by the model will interact with higher-level processes, specifically processes of selective visual attention and form recognition. As pointed out, the shapes of figure regions capture attention and are remembered, whereas the shapes of ground regions often go unnoticed (Fig. 1B). This shows that figure–ground organization plays a role in selective attention and both must be closely related.

Kienker et al. (1986) modeled figure–ground segregation in a purely top-down fashion—border ownership being determined by the location of an attentional spotlight. In contrast, most recent models treat figure–ground segregation as a purely bottom-up process, relying on local, within-area interactions at the lower stages of the visual hierarchy (e.g., Zhaoping 2005). Neither of these approaches offers a satisfactory explanation of how the whole system can function, barring the unacceptable solution of a homunculus that, in the case of the first model, determines where to shine the top-down attentional spotlight without the benefit of an organization of the sensory information or, in the case of the latter model, a homunculus that interprets the transformed versions of the retinal image created by the autonomous, bottom-up circuits. One solution is to assume iterative interaction between bottom-up signals and memory-related top-down signals in a way that converges over time (Carpenter and Grossberg 1993; Vecera and O'Reilly 1998). Inasmuch as these algorithms rely on the filling-in hypothesis and lateral signal propagation in cortex, though, the earlier concerns regarding physiological plausibility and the latency problem remain the same as for the other models.

In the present model, we propose dedicated neural circuits for perceptual grouping and figure–ground organization that also provide handles for attentional selection. Because our circuits for image context integration are separate from those representing the visual sensory information, they may include neurons in a higher-order cortical area and use recurrent white-matter projections, explaining the size invariance of the latency of the border-ownership signal. Our framework is consistent with recent neurophysiological findings (Zhou et al. 2000) and makes specific, testable predictions regarding both the physiological mechanism of border-ownership determina-

tion and the functional role of this mechanism in higher-level visual processing. Portions of this report were previously presented in abstract form (Craft et al. 2004; Schütze et al. 2003).

METHODS

Overview of network structure

Figure 3A shows the overall architecture of our model network and Fig. 3B highlights specific aspects of its connectivity. Input is provided by a stimulus map composed of oriented-edge detectors C_θ, akin to the topographic representation of a scene by complex cells in primary visual cortex (V1). Because an edge can be owned on either of its two sides by a figure, each edge detector C_θ provides input to a pair of mutually antagonistic border-ownership cells—one for each direction of ownership, $B_{\theta+}$ and $B_{\theta-}$ ("B cells"; see Fig. 5A for direction notation). The B cells inhibit each other (connections labeled β in Fig. 3B). The current version of the model uses only horizontal and vertical edges, resulting in four border-ownership channels. The B cell pairs receive additional input from end-stopped cells, $E_{\theta+}$ and $E_{\theta-}$ (see *Border-ownership cells* and Fig. 5A).

The balance of activity in each pair of B cells is influenced by feedback from a third layer in the network, consisting of contour-grouping cells ("G cells"). Residing at a higher level of the visual hierarchy, G cells have larger receptive fields than those of the B cells and sample the activity of the B cells in such a way that they respond preferentially to combinations of cocircular contour segments. The annular patterns shown in Fig. 4 illustrate the relative distribution of B cells at different radii that provide input to each G cell. A given G cell is driven by the activity of those B cells that signal ownership in the direction of the center of its annulus; the G cell then provides feedback that inhibits the B cells' signaling ownership in the opposing direction. For example, in Fig. 3B, cell $B_{\theta+}$ provides input to grouping cell G (connection labeled γ), so G inhibits $B_{\theta-}$ (connection labeled ρ). This is functionally similar to a circuit in which G applies positive feedback directly to $B_{\theta+}$, but does not require any additional mechanisms to preserve the classical receptive field property of the B cells (see RESULTS), allowing us to focus on the performance of the grouping mechanism. Note also that the inhibitory connections in Fig. 3 require inhibitory interneurons that we omitted in this schematic for the sake of clarity.

This configuration causes the G cells to bias B cells with a Gestalt-like preference for convexity and proximity of stimulus contours, as illustrated in Fig. 4B. The preference for convexity is a direct consequence of grouping cocircular contour segments (see annulus "C"). A proximity preference (see annulus "P") results because the individual synaptic weights are larger for the smaller, less-diffuse annuli. Global border ownership is determined by the net balance of these biases as they interact in the recurrent network.

FIG. 4. Grouping cell connectivity. A: spatial distribution of connections; darker pixels indicate stronger connection weights. B: annular connectivity emphasizes convexity ("C") and proximity ("P") of contours.

Grouping cell connections

Each pixel of the connection pattern images in Fig. 4*A* indicates the point of origin and strength (in grayscale coding; darker meaning stronger) of a single feedforward connection from a B cell at the perimeter of the annulus to the G cell at the center. These connection patterns are generated by convolving circles of the desired radii (*r*, Fig. 5*B*) with a normalized 2D Gaussian filter (parameter *σ*, Fig. 5*B*), resulting in the annuli shown. The annular regions become increasingly diffuse at larger radii to preserve scale invariance, and we normalize the strength of the synaptic inputs (connection weights) to give a common total weight (of unity) for all radii.

Next, we determine which border-ownership orientation channel must give rise to each of the connections along the perimeter of these annuli. The preferred direction of border ownership is defined at each point in the connection pattern image by a vector pointing toward the center of the annulus. Because our model currently implements only four orientations of border ownership, we resolve these vectors into positive and negative components along the horizontal and vertical axes of ownership. Based on these vector components, we then partition the annuli into separate sets of connections arising from each of the four border-ownership orientation channels (cocircular contour fragments $\mathbf{K}_{\theta\pm}^{(r)}$, Fig. 5*C*; see APPENDIX for details).

All G cells receive their input through a translated version of this same relative pattern of connections. In such a homogeneously connected network, 2D cross-correlation (i.e., spatial filtering) provides a natural description of the interactions between layers, using the fixed connectivity pattern as the correlation kernel. The four annulus components, $\mathbf{K}_{\theta\pm}^{(r)}$, are correlated separately with the four border-ownership channels, $\mathbf{B}_{\theta\pm}$, as described by

$$\mathbf{B}_{\theta\pm} \otimes \mathbf{K}_{\theta\pm}^{(r)} = \sum_{i}\sum_{j} B_{\theta\pm}(x+i, y+j)K_{\theta\pm}^{(r)}(i,j) \qquad (1)$$

For brevity, we have used boldface symbols **B** and **K** to represent the 2D arrays $B(x,y)$ of neuron firing rates and $K(x,y)$ of connection strengths, and "$\otimes$" denotes the 2D cross-correlation operation.

If we adopt the shorthand notation $\mathbf{S}_{\theta\pm}^{(r)} = \mathbf{B}_{\theta\pm} \otimes \mathbf{K}_{\theta\pm}^{(r)}$, we can regard the input from each kernel as representing individual curvature segments, "$S_{\theta\pm}^{(r)}(x,y)$," with center of curvature (x,y), radius of curvature r, and orientation of ownership $\theta\pm$. We would like the G cell activity to reflect a preference for cooccurrences among these segments so, rather than summing them linearly, we combine them as follows

$$[\mathbf{S}_0^{(r)}\mathbf{S}_{\pi/2}^{(r)} + \mathbf{S}_0^{(r)}\mathbf{S}_{\pi}^{(r)} + \mathbf{S}_0^{(r)}\mathbf{S}_{3\pi/2}^{(r)} + \mathbf{S}_{\pi/2}^{(r)}\mathbf{S}_{\pi}^{(r)} + \cdots]^{1/2} \qquad (2)$$

Here, unique pairwise products of the cocircular inputs at each spatial location (x,y) are summed, and the square root is applied as a compressive nonlinearity to ensure that the result stays approximately in the same dynamic range as the individual inputs. *Equation 2* represents just one way to model selectivity for cooccurrences among inputs. Our simulations have shown that other forms of input nonlinearity (e.g., logarithmic) can also be used to produce a similar selectivity.

The activity of the G cells is then governed by firing-rate equations of the form

$$\tau_G \frac{d\mathbf{G}^{(r)}}{dt} = -\mathbf{G}^{(r)} + \gamma^{(r)}[\mathbf{S}_0^{(r)}\mathbf{S}_{\pi/2}^{(r)} + \mathbf{S}_0^{(r)}\mathbf{S}_{\pi}^{(r)} + \mathbf{S}_0^{(r)}\mathbf{S}_{3\pi/2}^{(r)} + \mathbf{S}_{\pi/2}^{(r)}\mathbf{S}_{\pi}^{(r)} + \cdots]^{1/2}$$

$$(3)$$

where the term in brackets on the right-hand side is the total input described earlier, τ_G is a time constant common to all G cells, and $\gamma^{(r)}$ scales the feedforward connection weights for each G cell receptive field size. Again, the boldface symbol **G** represents a 2D array, $G(x,y)$, of neuron firing rates.

Border-ownership cells

As was illustrated in Fig. 3, every feedforward connection from a cell $B_{\theta\pm}$ to a G cell is accompanied by an inhibitory feedback connection from that same G cell to the opposing border-ownership cell, $B_{\theta\mp}$. Because pixels in the annulus images shown in Fig. 4*A* represent border-ownership cell locations that a particular G cell "observes," we can characterize the feedback connections by asking the reciprocal question: What is the pattern of all G cells that a single B cell location observes? Mathematically, this pattern can be obtained directly from the feedforward correlation kernels $\mathbf{K}_{\theta\pm}^{(r)}$ described in the previous section (see APPENDIX). As with the grouping cells, these feedback connections to the B cells are implemented using spatial filtering [$\mathbf{G}^{(r)} \otimes \mathbf{K}_{\theta\pm}^{(r)}$]. Together with a term $\beta\mathbf{B}_{\theta\pm}$ that accounts for the mutual inhibition between a pair of border-ownership cells, the
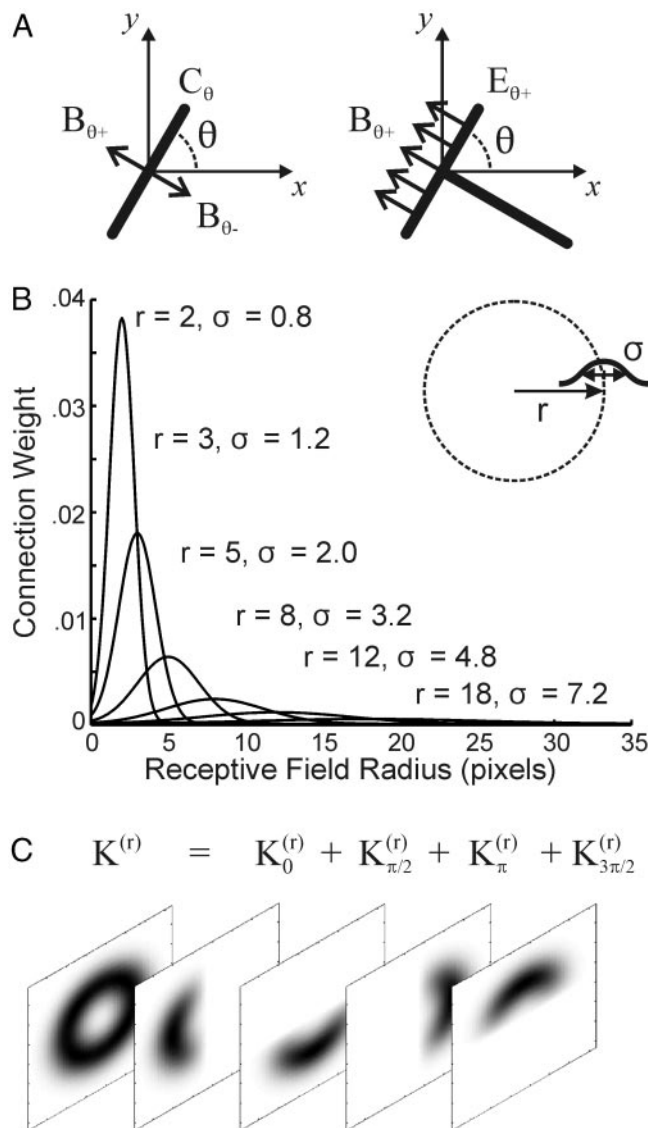

FIG. 5. Model implementation details. *A*: notation for directions of ownership and end-stopping at *θ*-oriented edges. *B*: cross-sectional profiles of receptive field annuli, showing radial variation of connection weights. Large cross-sections have small peak magnitudes because each annulus is normalized to a total weight of unity. *Inset*: illustration of relation of annulus parameters to the 2D view in Fig. 4*A*. *C*: resolution of G cell–receptive field into correlation kernels for each orientation of border ownership, as described in METHODS and by *Eqs.* A6–A10 in the APPENDIX.

sum over these cross-correlations yields the total inhibitory input (see *Eqs. 4* and *5*, below) to the border-ownership cells.

Excitatory feedforward input to $B_{\theta\pm}$ is provided by orientation-selective edge detectors, C_θ (see Fig. 5*A* for directional notation). Additional contributions from local stimulus features can also be taken into account in the input stage. For instance, cues such as binocular disparity and dynamic occlusion contribute to the perception of border ownership and have been shown to influence the neural border-ownership signals accordingly (Qiu and von der Heydt 2005; von der Heydt et al. 2003). In the present study, we model the influence of T-junctions as one example of such local cues.

T-junctions are represented in visual cortex by the activity of end-stopped cells, which are common in V1 and V2 (Heider et al. 2000; Hubel and Wiesel 1968). Because these cells respond selectively to terminations of edges and lines, they are well suited for detecting occlusion features (Heitger et al. 1992). Model end-stopped cells have been used successfully in prior computational models for representing occluding contours (Heitger and von der Heydt 1993; Heitger et al. 1998).

In our model, oriented edge segments that form the "hat" of a T-junction (encoded by complex cells, C_θ) are biased toward ownership on the side where termination of the "stem" of the junction (encoded by end-stopped cells, $E_{\theta\pm}$) suggests occlusion (see Fig. 5*A*). It is important to note that this T-junction bias does not depend on specialized detectors that classify image features. All that is needed are edge selective elements that are also sensitive to the presence of an intersecting contour on one side, similar to what was proposed in the model of illusory contours of Heitger et al. (1998). B cells whose activity is influenced by orthogonally oriented, end-stopped cells achieve this naturally.

End-stopped cells may either excite or inhibit border-ownership cells to produce this bias (Fig. 3*B*). Thus the combined input from C_θ and $E_{\theta\pm}$ effectively yields three distinct levels of *total* feedforward input: each cell $B_{\theta\pm}$ may receive a baseline input (proportional to the activity of C_θ) that conveys local edge contrast; a stronger input (e.g., proportional to twofold the activity of C_θ, as a consequence of complementary excitation from $E_{\theta\pm}$), where end-stopping indicates ownership in the B cell's preferred direction; or a weaker input (e.g., diminishing toward 0, as a result of inhibition from $E_{\theta\mp}$), where end-stopping indicates ownership in the opposite direction. For simplicity, these three levels of feedforward input were combined into a single term in the stimulus maps, as indicated in *Eqs. 4* and *5* using the mathematical shorthand $CE_{\theta\pm}$ (see APPENDIX for more details).

The activity of all complementary pairs of B cells is thus described by

$$\tau_B \frac{d\mathbf{B}_{\theta+}}{dt} = -\mathbf{B}_{\theta+} + \left\{ \mathbf{CE}_{\theta+} - \beta\mathbf{B}_{\theta-} - \sum_r \rho^{(r)}[\mathbf{G}^{(r)} \otimes \mathbf{K}_{\theta-}^{(r)}] \right\}_+ \quad (4)$$

$$\tau_B \frac{d\mathbf{B}_{\theta-}}{dt} = -\mathbf{B}_{\theta-} + \left\{ \mathbf{CE}_{\theta-} - \beta\mathbf{B}_{\theta+} - \sum_r \rho^{(r)}[\mathbf{G}^{(r)} \otimes \mathbf{K}_{\theta+}^{(r)}] \right\}_+ \quad (5)$$

where τ_B is the B cell time constant, β reflects the strength of mutual inhibition between opposing border-ownership cells, and $\rho^{(r)}$ scales the feedback from the grouping cells. The subscripted plus signs on the brackets indicate that B cell responses are rectified—i.e., if the net input to a B cell becomes negative, its firing rate will not decrease below zero.

Based on *Eqs. 3, 4*, and *5*, we implemented a model network of four orientation channels of border ownership and six sizes of G cell receptive fields (see Figs. 4*A* and 5*B*), with a resolution of one neuron per pixel in each visuotopic B cell map. To maintain uniform coverage of the B cell layers by each of the different sizes of G cell annuli, the number of G cells present at each scale (radius, r) was decreased in proportion to $1/r^2$ (see DISCUSSION). Each pixel of the B cell map corresponds to the base of an arrow in Figs. 6 and 9 representing the vectorial modulation index, as defined below. For comparison with
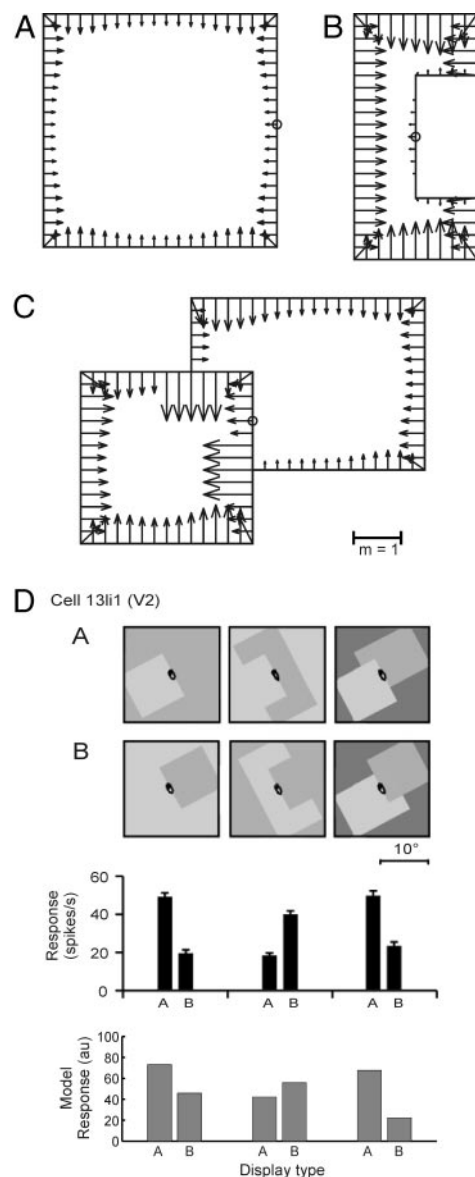

FIG. 6. Model border-ownership assignments for stimuli used in neurophysiological experiments. Responses shown for single square (*A*), C-shape (*B*), and overlapping squares (*C*). Base of each arrow represents one pixel of the input edge map, and the arrows reflect the net activity of the B cells at each location according to vectorial modulation index (see METHODS). Scale bar indicates length of arrow corresponding to a modulation index of magnitude 1, which would correspond to a 100% certain assignment of border ownership in a particular direction. *D*: comparison of model performance to neurophysiological findings for border ownership in alert, behaving macaques. *Top 2 rows*: stimulus configuration with a cell's receptive field (indicated by small oval; cross indicates center of gaze). *Third row*: responses of a B cell recorded in area V2; bars indicate average firing rate of neuron for each stimulus condition. *Bottom row*: response of a model B cell to analogous stimulus conditions (location of cells indicated by circles in *A–C*). Part of this figure was adapted from Fig. 23 of Zhou et al. (2000), with permission.

physiology, we assumed that the size of a pixel in the edge map corresponds to 0.5° of visual angle. Cross-correlations were computed using zero padding at the boundaries of the B cell maps and by expanding the G cell maps to preserve feedforward signals from boundary B cells.

Decreasing the number of G cells at larger scales weakens the contribution of these scales to the overall grouping feedback beyond the proximity preference discussed earlier. This is offset by constrain-

ing the weight parameters in *Eqs. 3, 4,* and *5* according to the expression $\gamma^{(r)} = \rho^{(r)} = \sqrt{\gamma_0} \cdot r$, where γ_0 is a fixed proportionality constant. This gives the recurrent loops between the B cells and the G cells a net weight of $\gamma_0 \cdot r^2$ (the product of the separate feedforward and feedback weights) at each scale r, which counterbalances the similarly proportioned decrease in cell numbers. β and γ_0 represent the only parameters that were used to tune the model. Our results were generated using values $\beta = 0.5$ and $\gamma_0 = 4.5$, which were chosen based on an exploration of the model's parameter space (see APPENDIX, Fig. A1). Steady-state [$d\mathbf{B}/dt = 0$ and $d\mathbf{G}^{(r)}/dt = 0$] border-ownership assignments and response time courses were then generated by numerically integrating the system of model equations with time constants $\tau_B = \tau_G = 10$ ms and a one-way conduction delay of 6 ms between the B cell and G cell layers (thus 12-ms "round-trip" for the loop from B cells to G cells and back), as estimated under PHYSIOLOGICAL CONSTRAINTS in the *Background* section. To keep our model simple and focused on the border-ownership computation, we assumed that the arrays of complex and end-stopped cell firing rates were given; that is, we did not compute them from responses of retinal cells.

Vectorial modulation index

The strength of the border-ownership signal is described by a generalization of the modulation index. We define a vector quantity by the expression

$$\vec{v}(x, y) = m_i(x, y)\hat{\imath} + m_j(x, y)\hat{\jmath} \qquad (6)$$

where $\hat{\imath}$ and $\hat{\jmath}$ are unit vectors along the horizontal and vertical image axis, respectively, and the components $m_i(x,y)$ and $m_j(x,y)$ are the usual modulation indices along their respective axes, defined as

$$m_i(x, y) = \frac{B_0(x, y) - B_\pi(x, y)}{B_\pi(x, y) + B_0(x, y)} \qquad (7)$$

$$m_j(x, y) = \frac{B_{\pi/2}(x, y) - B_{3\pi/2}(x, y)}{B_{\pi/2}(x, y) + B_{3\pi/2}(x, y)}$$

Clearly, both components in *Eq. 7* are limited to values between $+1$ and -1. For the *x*-component, for instance, a positive value of $m_i(x,y)$ signifies that the figure is to the right of position (x,y) and a negative value signifies that the figure is to the left. Its absolute value indicates the "strength" of the border-ownership signal, with zero being equivalent to ambivalence between left and right. The corresponding comments apply to the *y*-component, $m_j(x,y)$, regarding the figure's position upward or downward of (x,y). The direction of the vectorial modulation index $\vec{v}(x,y)$ defined in *Eq. 6* indicates the position of the foreground figure in the 2D image plane relative to the point (x,y). For instance, positive values in both components [$m_i(x, y) > 0$, $m_j(x,y) > 0$] indicate that the figure is located upwards and to the right of (x, y).

RESULTS

Comparison with neurophysiology

We began by testing our model with stimuli similar to those from the neurophysiological experiments detailed in Zhou et al. (2000). Figure 6, *A–C* shows edge maps for these stimuli, along with border-ownership assignments made by the model. Arrows on a border point toward the region determined by the model to own the border, and the length of the arrows indicates the "strength" of the border-ownership signal (vectorial modulation index; see METHODS, *Eq. 6*). As can be seen, the model correctly determines the direction of border ownership—i.e., it assigns the borders to those regions that are perceived as foreground—everywhere for all of these stimuli.

Figure 6*D* compares the behavior of the model B cells at a single location in each of these figures (indicated by a circle in Fig. 6, *A–C*) to neurophysiological responses. In all three cases, the direction of ownership indicated by the pair of model B cell responses is consistent with experimental findings. Notice that the magnitude of the model's side-of-figure distinction is considerably smaller along the inner arm of the C-shape than for the other two shapes (Fig. 6, *B* and *D*). Zhou et al. (2000) reported a similar trend in the proportion of neurons they observed with the correct border-ownership modulation for these three stimuli (cf. their Fig. 27), which they attributed to an incomplete use of available cues by B cells.

Zhou et al. (2000) also found that the responses of border-ownership cells for the single square were fairly insensitive to stimulus size. In our model, G cell feedback provides B cells with a broad range of contextual information about local edges. Accordingly, as Fig. 7 (*left column*) shows, our model is able to maintain its border-ownership distinction over a similar range of stimulus sizes as the neurons in area V2.

In our model, feedback is applied through inhibition to rectifying B cells. This scheme accounts naturally for experimental observations that surround stimulation has only a modulatory effect on responses to stimuli in the classical receptive field (Zhou et al. 2000; their Fig. 13). Figure 7 (*right column*) shows that despite a broad range of image context integration for border ownership, no response is elicited from either the observed cells (Fig. 7, *middle right*) or model cells (Fig. 7, *bottom right*) unless an edge is placed directly in their classical receptive field. It is important to show that the activity produced by the model is confined to the locations of the contours as defined in the stimulus. In neurophysiology, this rule corresponds to the observation of small, sharply defined minimum response fields. Figure 13, *A* and *B* of Zhou et al. (2000) illustrates this for a border-ownership–selective cell and Fig. 7 shows that the present model is consistent with this observation.

As discussed in the *Background* section, another important consideration is the timing of contextual integration. Figure 8*A* shows the time course of the model's responses for square stimuli at two different sizes. In agreement with physiological findings (Fig. 2), it can be seen that the model's border-ownership signals emerge with only a short delay following the onset of the edge responses ($\sim$20 ms in these simulations) and that this delay does not vary with the size of the square stimulus (i.e., the solid and dashed red curves emerge at the same time). Moreover, both signals rise at the same rate and reach their half-maximal height well before 100 ms, demonstrating that context integration in the reentrant circuits is rapid and independent of stimulus size.

Figure 8*B* shows that modifying the model slightly, by including inhibitory self-feedback to the B cells through local interneurons (a common gain control mechanism in cortical circuits; cf. Wilson 1999), is sufficient to account for the additional overshoot seen at the onset of the physiological border-ownership signals. This improves the agreement of the model's time courses with the data without changing its steady-state assignments and makes it even more apparent that the model's border-ownership signals are rising at the same rate—note that the initial portions of the two red curves overlap and that the border-ownership signal for the large square (red, dashed curve) reaches its peak slightly earlier than the signal
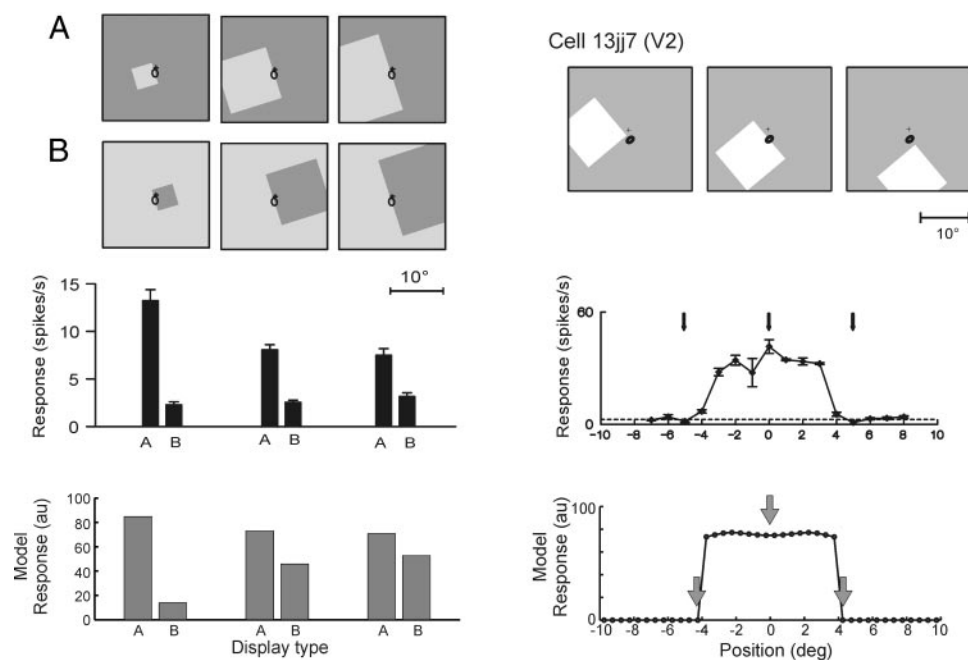
FIG. 7. Comparison of model performance to neurophysiological findings for border ownership. *Left column*: border-ownership responses of real and model cells persist over a range of sizes of the square stimulus. *Right column*: real and model cells have similar classical receptive field properties, demonstrating the lack of activity when the figure is outside the receptive field, despite the large amount of contextual modulation in the model. *Left column*, *top 2 rows*: stimulus configuration. Size of square increases >3-fold from *left* to *right*, with either right (*A*) or left (*B*) edge placed in cell's receptive field (indicated by small oval). *Third row*: responses of a B cell recorded in area V2 of an alert, behaving macaque; bars indicate average firing rate of neuron to each edge of square. *Bottom row*: response of a model B cell to analogous stimulus conditions. *Right column*, *top row*: stimulus configuration. *Left* and *right*: stimulus (white edge) is outside the classical receptive field. *Center*: stimulus is in the receptive field. *Middle row*: firing rate as function of distance between the center of the oriented edge and the center of the receptive field. Cell responds only when the stimulus is within the receptive field. Three positions in the *top row* are indicated by arrows. *Bottom row*: response of a model border-ownership cell to analogous stimulus conditions. Parts of this figure were reproduced from Figs. 5 and 13*B* of Zhou et al. (2000), with permission.

for the small square (red, solid curve). The inclusion of this gain control mechanism also introduces damped oscillations that are suggestive of those seen in the physiological border-ownership signals. We have not attempted to model the later structure of the time courses here because these are likely to reflect a mixture of additional influences (for instance, feedback from higher cortical areas) that are outside the scope of the present study.

Model results: predictions

In addition to using the stimuli that had been used in the neurophysiological recordings by Zhou et al. (2000), we tested our model with three new stimulus edge maps. Figure 9*A* shows a modified version of the C-shape, in which the contour that formed the inner arm of the "C" now appears to be owned by an occluding rectangle to the right of the contour. As is seen by the arrows plotted in the figure, the model correctly reverses its border-ownership assignment along this contour. Similarly, the stimulus in Fig. 9*B* is commonly perceived by viewers as a vertical bar occluding a horizontal bar of the same length, rather than as a vertical bar flanked by two squares. The model again produces results that are in agreement with perception, which is particularly remarkable given the strong model responses for isolated square figures shown earlier.

Although all stimuli considered thus far have been closed figures or combinations thereof, our model is not limited to this stimulus set. As Fig. 9*C* shows, the model is also able to account for more general perceptual grouping based only on

the proximity of contours. Furthermore, if we introduce cues of an alternate figure–ground relationship (Fig. 9*D*), the model's assignments reverse, in agreement with perception. No electrophysiological results are as yet available for these stimuli so our model results are genuine predictions. The model response shown in Fig. 9*C* is important because it demonstrates a general grouping ability and shows that contour closure is not required. Thus the model will produce robust border-ownership assignment even in situations where the input edge map is incomplete, as is the rule when natural images are processed. Models based on collinear facilitation cannot relate distant, isolated contours as in Fig. 9*C*, and are thus unlikely to yield border-ownership assignment in such displays (see DISCUSSION).

DISCUSSION

Our network is constructed from generic model neurons that communicate with each other using firing-rate–based mechanisms. We introduce two classes of neurons: B cells and G cells. The basic hypothesis is that at a given location and orientation, input from V1-like edge detector cells (that do not respond differentially to side-of-figure) is routed to a pair of B cells, with each member of this pair responding preferentially to a figure located on one or the other side of an edge. Groups of B cells project to G cells such that activation of a specific G cell represents the presence of a visual object at or near a given location. The two B cells forming a pair inhibit each other and their competition is biased by feedback from the G cells, resulting in border-ownership–sensitive responses. These own-
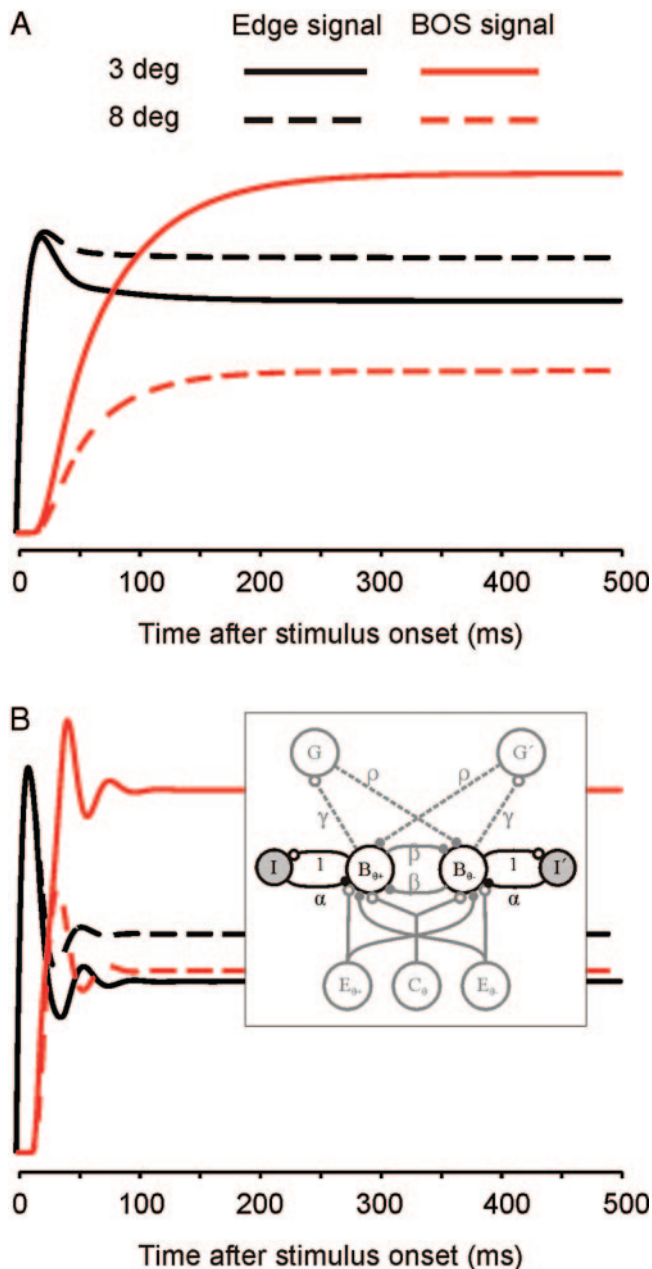
FIG. 8. Timing of model border-ownership signals does not depend on the spatial extent of image context integration. Activity of model border-ownership cells centered on the edge of a square is shown for square sizes of 3° (full lines) and 8° (dashed lines). Border-ownership signals (red) were defined as the difference between the responses of a pair of opposing model cells. Edge signals (defined as the sum of the 2 responses) are also shown for comparison (black). Amplitudes of BOS signals and edge signals are scaled differently, but the signals for the 2 figure sizes are plotted on the same scale. A: responses of the model configured as depicted in Fig. 3B and described in METHODS. Note that the BOS signals for both square sizes (red lines) emerge after the same, short (20-ms) delay after the onset of the edge responses (black lines), and that both signals rise at the same rate, reaching their half-maximal value well before 100 ms. B: responses of the model with inhibitory self-feedback applied to the B cells through local interneurons ("I"; see *inset*), as a form of gain control. This additional feature is sufficient to account for the overshoot seen at the onset of the physiological border-ownership signals. It also introduces damped oscillations that are suggestive of those in the physiological responses. Parameter values used for this simulation were $\alpha = 2$, $\tau_I = 10$ ms, and $\gamma_0 = 18$ (see METHODS); all other model parameters were identical to those of the simulations presented in the other figures and described in METHODS. For comparison to physiological data under similar stimulus conditions, see Fig. 2.

ership assignments are then passed on to subsequent stages of form analysis. As we discuss later, in addition to facilitating border-ownership assignment through contour grouping, our G cells also provide a specific, intermediate-level structure for the attentional selection of these assignments that is lacking in other models. The assumption that each physiological B cell has a counterpart with which it forms mutually inhibitory connections implements a processing strategy that exploits the physical law that an occluding contour can be "owned" in only one direction at a time. We propose that the B cells in our model reflect the average behavior of the border-ownership–selective neurons recorded in primate visual cortex (Zhou et al. 2000).

Reentrant versus intrinsic circuits

A key premise of our model is that contextual modulation of border-ownership cell responses occurs through recurrent interactions with subsequent levels of the cortical hierarchy (i.e., grouping cells), rather than within-area, lateral interactions. Our choice of mechanisms is supported by a number of physiological and functional considerations. Foremost among these is the relationship between border-ownership–modulation latencies and figure size.
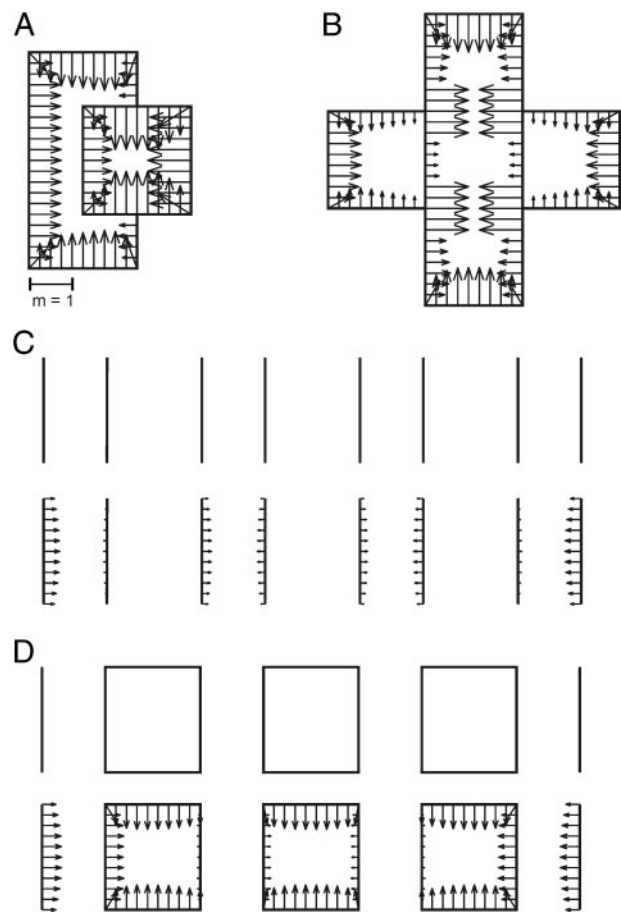

FIG. 9. Predictions made by the model for additional stimuli. Shown are border-ownership assignments made by the model for a modified C-shape (A) and an occluding bar (B). Model also predicts perceptual grouping of isolated contours by proximity (C), which reverses (D) when horizontal contours are added. Arrows as in Fig. 6.

Lateral interactions may well play a role in context integration. However, as we have argued in the *Background* section, models relying exclusively on signal propagation through horizontal fibers within V2 are unrealistic because they imply long conduction delays, in contradiction to the neurophysiological findings. This is in contrast to Zhaoping (2005), who presented a model based on within-V2 connections in which border-ownership signals emerge without excessive delays. We have estimated the delays produced by bridging the cortical distances in V2 corresponding to 4° of visual angle, the minimum required for generating a border-ownership signal in the center of the edge of a square figure of 8° size. Based on the published median conduction velocity measured for horizontal fibers in V1 (Girard et al. 2001) we estimated that axonal conduction alone (not counting synaptic transmission times) would produce delays of the border-ownership signal of 70–90 ms relative to the edge signals, whereas only 30 ms has been found (Fig. 2).

Zhaoping (2005), citing Angelucci et al. (2002), argues that it is reasonable to assume that the longest horizontal fibers in V2 bridge a distance corresponding to 3° and that the conduction for that length would take 8–10 ms. This means that, for a square of 8° size, transmitting information from a corner to the center of a side of the square (a cortical distance corresponding to 4°) would require only one relay of activity, and conduction would cause delays of little >10 ms. However, our calculations indicate that the visual angle corresponding to the longest horizontal fibers is much smaller. Angelucci et al. (2002) measured the extent of the fields that were labeled by tracer injections (in V1) and found 6–8 mm on average (see their Table 1), corresponding to a maximum length of fibers of 3–4 mm. According to our estimates of cortical distances in V2 cortex (see *Background*), a visual angle of 4° in a typical situation corresponds to 21- to 27-mm distance, which is five- to ninefold the maximum length of intracortical fibers (if one assumes, for lack of comparable measurements in V2, that this length is the same in V2 as in V1). The exact visual angular distance that can be bridged by intracortical fibers depends on the eccentricity. For example, Fig. 4 of Angelucci et al. (2002) shows that, at an eccentricity of 6.5°, the radius subtended by lateral connections in layer 4B of V1 is about 2°. Most of the data of Zhou et al. (2000) came from smaller eccentricities (median of 1.5° for V1, 2.0° for V2) where the cortical magnification factor is higher and the longest length of lateral connections corresponds to smaller visual angles. Clearly, more comprehensive studies of the neuroanatomy and topography of area V2 are needed to make a definitive argument.

Besides the conduction delays, the scale of the retinotopic representation of V2 alone shows that models of figure–ground organization that rely exclusively on intracortical circuits are implausible. Signals would have to be relayed through several neurons, requiring that every neuron in the chain produce action potentials. This contradicts the small size of the classical receptive fields. Most neurons in foveal and parafoveal V2 do not respond to stimuli outside a radius of 2° from the center of their receptive field (even in situations where neurons signal illusory contours, activity spreads over only small distances; Peterhans and von der Heydt 1989). This feature of the neurons characterizes the precision of spatial localization. To be compatible with this basic feature, models can use propagation of border-ownership signals only along chains of neurons that are also directly activated by contrast borders. This means, for example, that intracortical network models cannot produce grouping between two parallel lines, as shown in Fig. 9*C*, unless the cortical representations of the two lines are within the reach of monosynaptic connections. Taking the measurements from V1 by Angelucci et al. (2002) (for lack of measurements in V2), this would be distances of <4 mm in cortex (corresponding to about 1° visual angle at 2° of eccentricity; Gattass et al. 1981).

Grouping cells

The grouping cell architecture offers a number of advantages over mechanisms that have been proposed in previous models. By forming recurrent circuits between different cortical areas, G cells are able to integrate and disseminate contextual information rapidly, over much greater distances than purely feedforward or within-area lateral connections (Baek and Sajda 2005; Grossberg and Raizada 2000; Kikuchi and Akashi 2001; Nishimura and Sakai 2004, 2005; Pao et al. 1999; Zhaoping 2005) would allow. In addition, our G cells rely only on a single, simple pattern of connections at multiple scales to achieve a sensitivity for proximity, convexity, and smoothness/cocircularity among contour segments. As will be discussed in the following text (see *Grouping and attention*), grouping cells organize low-level feature information in a structure that is suitable for attentional selection and processing.

In our construct, grouping cells have annular receptive fields (see Fig. 4) of various sizes. These cells should respond optimally to circular disks of certain size and location, and less to any part of such a disk. The existence of cells with these properties—a key prediction of our model—is suggested by prior psychophysical evidence of special integration mechanisms for concentric circular patterns (see, e.g., Wilson et al. 1997), as well as neurophysiological demonstrations of neurons selective for concentric gratings (Gallant et al. 1996). Note, however, that the results presented in this paper can be obtained without assuming that grouping cells have complete circular integration fields. Indeed, in our computation, the individual kernels that make up the G cell annuli ($\mathbf{K}_{\theta\pm}^{(r)}$, Fig. 5*C*) can be viewed as modeling cells tuned to curved contour segments. Combining these components as shown in Fig. 5*C* then yields single cells that respond preferentially to complete circular contours. Cells that respond selectively to curved contour segments and to combinations of such segments have been shown to exist in extrastriate cortex (Brincat and Connor 2004; Pasupathy and Connor 2001).

Border-ownership coding was observed neurophysiologically for stimuli of linear dimensions ranging ≤20° (Zhou et al. 2000). Assuming a spatial resolution of 0.5° of visual angle per pixel in the model, the diameters of model G cell annuli, ranging from four to 36 pixels, then correspond to G cell connections that span from 2 to 18° in the visual scene (not including their Gaussian spread). Because receptive fields in V4 are large (Gattass et al. 1988), and the spread of the back projections to V2 (Rockland et al. 1994) would add to the range of context integration, the findings of Zhou et al. could probably be explained by our model with grouping cells in V4.

As was illustrated in Fig. 4 and discussed earlier, the annular connection patterns and heavier weighting of shorter-distance connections (i.e., annuli of smaller diameters) causes the G cell

bias to favor compact arrangements of contours over less-compact ones, as well as convex over concave patterns. The Gaussian profiles of the grouping cell annuli (see METHODS and Fig. 5B) reflect a gradual drop-off in the probability of forming connections around a particular radius and provide robustness for variability of locations of contour segments. By constraining the standard deviation parameter σ of the Gaussian profile to increase with annular radius, scale invariance is achieved.

Interestingly, because of these annular connection patterns, activity in the G cell layer conveys a skeleton-like representation of each figure [akin to the medial axis transform (Blum 1967); see Fig. 10]. Through the recurrent network, these "skeletons" are closely interrelated with the boundary information (cf. Ogniewicz and Kübler 1995). It has been suggested that such skeletal representations play a role in visual perception of form (Kovacs and Julesz 1994; Psotka 1978) as well as in neural coding (Lee et al. 1998).

In addition to the very existence of G cells, our model makes a prediction about their numerosity, relative to that of other cell types. We expect that the number of G cells is much lower than that of either edge detectors or B cells. The reason is that the resolution of these two latter classes of cells must be close to the spatial resolution of the system (at a given eccentricity). On the other hand, the spatial resolution of the grouping process can be considerably lower. As discussed in the following text, we argue that the network of grouping cells serves as a substrate for object-related visual attention. This means that the resolution of grouping cells should be comparable to the resolution with which observers are able to individuate items in a group of similar items. For example, if a row of ten equally spaced vertical bars is viewed with fixation on a point 3° above the row, it is impossible to attend to individual bars (to count them, for example) unless the spacing of the bars is ≳1° of visual angle. However, the bars can easily be resolved (seen as discrete elements) at a spacing of ≤0.1°. Experiments of this kind led to the conclusion that the spatial resolution of attention is five to 20 times lower than visual resolution for two-point

discrimination (the factor depends on eccentricity; the given range corresponds to eccentricities between 0.5 and 15°; Intriligator and Cavanagh 2001). We assume that the resolution of the annular templates of grouping cells is similar to the resolution of attention. Thus the density of the G cells with the smallest templates may be about ten times lower than the density of B cells on a linear scale. To achieve 2D coverage, the number of the smallest G cell templates might then be no >1% of the number of B cells, and the total number of G cells about 2% (because for the larger templates, the numbers required for equal coverage decrease by a factor of about 2 with each step of scale in our model).

Scope and limitations of the model

Our model's responses agree both with perception and with recordings of single-unit neuronal activity (Figs. 6 and 7). Although no response is elicited from a B cell unless an edge is presented directly in its classical receptive field, the cell's activity is modulated by portions of the visual field far outside its classical receptive field (Fig. 7). Border-ownership coding is found for both simple convex figures (Fig. 6A) and for more complex situations (Fig. 6, B and C), in that the activity level of the "correct" B cell exceeds that of its partner in all cases (Fig. 6D). Furthermore, the difference in activity between competing B cells is smaller at perceptually more "difficult" locations (such as at the inside of the C-shape, Fig. 6B), again in agreement with neural responses. Border-ownership coding in the model, as in the physiological experiments, is found at all figure sizes tested, and the activity difference within pairs of B cells decreases with figure size (Fig. 7). These signals also appear with only a short delay after stimulus presentation, and neither the length of this delay nor the rise time of the signals increases with stimulus size (Fig. 8). No neurophysiological data are available for the types of stimuli shown in Fig. 9, so the modeling results shown in these figures, although consistent with perception, are genuine predictions with respect to neuronal responses.

A limitation of the present model is that it describes the behavior of an "average" border-ownership–selective neuron, whereas the physiological results show a variety of selectivity patterns in different cells. For example, some neurons assigned border ownership consistently for isolated squares as well as overlapping figures and C-shaped figures, but other neurons failed to produce a border-ownership difference for one or the other of these conditions (Zhou et al. 2000; their Fig. 27). A similar variation between cells was found with respect to the integration of configuration cues with stereoscopic cues (Qiu and von der Heydt 2005). These variations probably reflect the sampling of signals corresponding to different stages of cue integration rather than a random variability. For this reason, we do not expect an exact quantitative agreement of the output of the present model with the responses of example neurons.

It can also be seen in Fig. 8 that our model tends to exhibit a larger difference in steady-state responses for the two square sizes than what is shown by the neural data (compare with the difference between the solid and dashed red curves in Fig. 2). We have incorporated a relatively strong proximity bias in the grouping feedback, which diminishes the strength of the border-ownership assignment for the larger figure. This suggests that the weighting of contextual feedback from different spatial
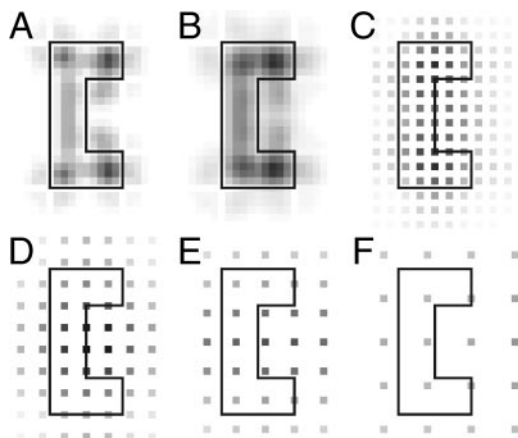

FIG. 10. Grouping cells convey skeleton-like information about figural regions over multiple scales. A–F: visuotopic maps of grouping cell activity (arranged from smallest annulus diameter, A, to largest, F), in response to C-shape stimulus. Each pixel indicates the firing rate of a G cell in grayscale, with darker pixels indicating higher firing rates, and the stimulus outline has been superimposed in each panel, for reference. For a given stimulus size, a certain subpopulation of grouping cells is optimal for representing the medial axis of the figural region (e.g., B). Activity at other scales may convey additional, useful information such as center of gravity (e.g., D).

scales in the cortex may differ from the exact drop-off in weights used here (see Fig. 5*B* and METHODS). Nevertheless, given the simplicity of the present model (except for Fig. 8*B*, only two parameters were tuned to generate all results in this study), it is important that its predictions are qualitatively correct for a broad range of results.

A model that reproduces the responses of individual neurons was proposed recently by Nishimura and Sakai (2004, 2005). They accomplish this by varying the spatial arrangement and strength of excitatory and inhibitory connections that convey contextual information to border-ownership cells in their model. It is relatively easy to see, qualitatively, how our model can also explain the different behavior of single cells, and how it could be made to fit these data. The cells presented in Zhou et al. (2000) differ from each other mainly along three lines: *1*) the relative strength of border-ownership modulation (their Fig. 15), *2*) the degree of variation of the border-ownership signal with size of figure (their Fig. 19), and *3*) the sensitivity to configuration (their Fig. 27). In our model, *1*) means that each B cell should have an individual value of the parameter ρ; *2*) means that B cells receive different strengths of feedback from the different sizes of G cell templates. Thus γ_0 should also vary with template size, in an individual fashion (presently, all sizes are weighted equally). B cells that communicate with G cells of all template sizes would assign border ownership correctly for C-shaped figures, as illustrated in our Fig. A1, but B cells communicating only with a limited range of G cell templates may assign the inner contour incorrectly, according to the nearby contours. With respect to *3*), Zhou et al. found that some B cells produce a greater border-ownership signal for one configuration than another, such as isolated squares versus pairs of overlapping squares. To fit this variation, we could assume that the relative weights of the inputs from regular complex cells (C) and from end-stopped cells (E) to the B cells vary. This would mean that G cells receive varying amounts of information about low-level cues when they sample the activity of the B cells, and that individual B cells would then receive feedback from G cells that differs in this respect. Some G cells may pick up little from E cells (by B cells), and their grouping signal would thus be determined mainly by evidence for cocircular arrangement of contours; other G cells may sample more from E cells and their signal would thus be influenced strongly by the presence of T-junctions. These G cells would produce a strong border-ownership bias on the occluding contour of overlapping figures and less for isolated figures, whereas the former would signal no bias for the occluding contour, but a strong bias for isolated squares.

To resolve situations of occlusions, we have assumed that T-junctions locally determine border-ownership modulation. Thus end-stopped cells play an important role in these situations. In other models, this result is achieved by means of the connectivity scheme (e.g., lateral collinear facilitation) without the explicit use of end-stopped cells or T-junction detectors (e.g., Zhaoping 2005). The lateral connectivity schemes (by horizontal cortical fibers) and our grouping cell scheme should not be viewed as mutually exclusive alternatives. On the contrary, it seems likely that the cortex uses both, one for integrating the global image context and the other for incorporating local cues such as T-junctions and other occlusion features.

Our model does not account for situations in which object recognition is a *prerequisite* for scene segmentation—e.g., the famous mottled image of a Dalmatian dog. In simpler cases, though, the output from our model provides a representation in which borders are assigned (more or less strongly) to one of two adjacent image regions. For the subsequent stages of shape analysis, this is a significant improvement over an unassigned edge representation. In fact, shape selectivity of inferior temporal (IT) neurons depends on border ownership (Baylis and Driver 2001; Kovacs et al. 1995). By demonstrating how border ownership can be assigned without resorting to object memory, our model can explain these findings in IT cortex.

Grouping and attention

At the same time, one of the most fundamental tasks of visual perception is the organization of raw input into a representation of separate visual objects that can be addressed efficiently by cognitive processes. G cell activity carries a condensed representation of the low-level image features associated with individual objects, emphasizing the relationships among these features. In this way, they provide "handles" that allow efficient addressing of this information by attentional-selection mechanisms.

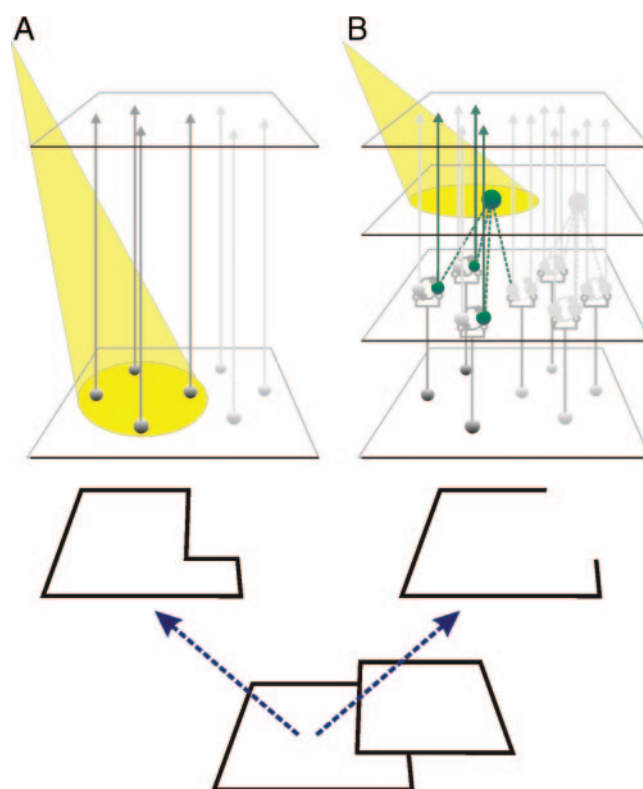As Fig. 11*A* illustrates, the traditional spatial model of

FIG. 11. Proposed role of grouping cells in visual processing. Figure illustrates the situation when attention is directed to a partially occluded object. *A*: traditional spatial selection model of attention (spotlight) addresses local features and incorrectly includes shared contours in its selection, emphasizing an L-shape. *B*: G cell network assigns border ownership to each edge in the stimulus by lumping the given features together into clusters (proto-objects). At the same time, these G cells serve as "handles" through which attention (spotlight) can selectively address the high-resolution representations of the proto-objects. In this way, only those contours that "belong" to the occluded shape are selected for further processing (ascending green arrows).

top-down attention shines a spotlight into the local feature representation of a scene. In the case when attention is directed toward a region corresponding to a partially occluded object, this mechanism would give rise to an incorrect interpretation of the object's form. By contrast, G cell–mediated selection at the proto-object level would correctly emphasize only those contours that "belong" to the occluded shape for further processing (Fig. 11B; see also Fig. 1, C and D). At the same time, the grouping cell "handles" offer an advantage for bottom-up (saliency-driven) attention. Rather than emphasizing the salience only of local features, grouping cell activity would draw attention to *regions* of the visual scene that correspond to the most salient proto-objects. This would explain why, in Fig. 1A, the foreground region attracts attention and is remembered (even though it is not a recognizable object), whereas the form of the adjacent background region goes unnoticed. The grouping cell scheme also explains how, in the case of ambiguous figures such as Rubin's famous vase/face figure (Rubin E 1921, 2001), border ownership can be flipped voluntarily between two regions by directing attention to one side or the other. We therefore conjecture that G cells, in addition to their *implicit* role of providing a bias signal to B cells and thus arbitrating between competing local interpretations of visual input, also play an *explicit* role in the representation of proto-objects for higher-level processing.

We have shown that a simple, biologically plausible recurrent network can account for basic neurophysiological and perceptual assignments of border ownership. Moreover, we suggest a specific neural mechanism that interrelates the assignment of border ownership with higher-level processing in a structured fashion, allowing these assignments to attract or be selectively addressed by object-related attention before object recognition. Future neurophysiology experiments will allow us to test our model's predictions and, at the same time, will broaden the stimulus set with which we constrain our model. In addition, it will be interesting to examine the interaction of attention effects with border-ownership coding responses. Perceptual grouping of object-related features and their attentional selection are seen in our model as complementary processes that go hand in hand from an early stage. This concept may help to solve the conundrum that high-level processing of objects requires segregation of specific regions from the visual scene while, at the same time, the segregation of these regions requires some "understanding" of their context in the image.

APPENDIX

Grouping cell connections

For a particular G cell $G^{(r)}(x_0, y_0)$, the overall pattern of connections is specified by starting with a discrete approximation to a circle of radius r centered at position (x_0, y_0) in visuotopic space

$$I^{(r)}(x, y) = \begin{cases} 1 & \text{if } |\sqrt{(x - x_0)^2 + (y - y_0)^2} - r| \leq \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (A1)$$

The small positive constant ε (0.5 in our simulations) ensures that the pixelated circle has a closed perimeter. Receptive fields with larger circumferences (larger r) include more points, corresponding to an increased number of feedforward connections. Thus G cells with larger receptive field radii receive a larger number of inputs, biasing them toward higher levels of activity than those with smaller receptive fields. To eliminate this bias, we normalize the strength of synaptic

inputs (connection weights) to give receptive fields of all radii the same total weight of unity. This is achieved by setting the weights at

$$\hat{I}^{(r)}(x, y) = \frac{I^{(r)}(x, y)}{\sum_i \sum_j I^{(r)}(i, j)} \quad (A2)$$

Finally, the sharply defined edges are smoothed by convolving $\hat{I}$ with a normalized 2D Gaussian filter

$$K^{(r)}(x, y) = \hat{I}^{(r)}(x, y) * \hat{h}(x, y) \quad (A3)$$

where * denotes convolution

$$\hat{I}^{(r)}(x, y) * \hat{h}(x, y) = \sum_i \sum_j \hat{I}^{(r)}(i, j)\hat{h}(x - i, y - j) \quad (A4)$$

and $\hat{h}(x,y)$ is a discrete, 2D normalized Gaussian (see Fig. 5B for choice of parameters)

$$\hat{h}(x, y) = \frac{\exp[-(x^2 + y^2)/2\sigma^2]}{\sum_x \sum_y \exp[-(x^2 + y^2)/2\sigma^2]} \quad (A5)$$

The radius of the smallest kernel was chosen as two pixels, in accordance with the discrete nature of the input. The radii of the sequence of larger kernels were chosen to ensure adequate coverage by requiring that the maximum of each kernel falls onto the half-height point of the next-smaller kernel. The spread (σ) of the kernels was chosen arbitrarily such that $r/\sigma = 2.5$ for all r, ensuring that the kernels decay to a very small value at their centers.

As defined in *Eq. A3*, $K^{(r)}(x, y)$ describes the spatial pattern and strengths of all feedforward connections to a given G cell. We now determine which border-ownership orientation channel gives rise to each of these connections. Direction of border ownership is defined by a set of unit vectors oriented inward-normal to (toward the center of) the annular receptive field

$$\tilde{n}(x, y) = n_x(x, y)\hat{\imath} + n_y(x, y)\hat{\jmath} \quad (A6)$$

where $\hat{\imath}$ and $\hat{\jmath}$ are horizontal and vertical unit vectors, respectively, and the vector components for a receptive field centered at (x_0, y_0) are given by

$$n_x(x, y) = \frac{-(x - x_0)}{\sqrt{(x - x_0)^2 + (y - y_0)^2}} \quad (A7)$$

$$n_y(x, y) = \frac{-(y - y_0)}{\sqrt{(x - x_0)^2 + (y - y_0)^2}}$$

For instance, a B cell (coordinates x, y) whose receptive field is exactly to the right of the center of the annulus defining the receptive field of a G cell (coordinates x_0, y_0) has $x - x_0 > 0$ and $y - y_0 = 0$ and therefore $n_x(x, y) < 0$ and $n_y(x, y) = 0$.

Scaling these vectors by the connection weights of $K^{(r)}(x,y)$ computed in *Eq. A3*, we obtain

$$\tilde{K}^{(r)}(x, y) = K^{(r)}(x, y)\bar{n}(x, y) \quad (A8)$$

$$= K^{(r)}(x, y)n_x(x, y)\hat{\imath} + K^{(r)}(x, y)n_y(x, y)\hat{\jmath}$$

Here, $K^{(r)}(x, y)n_x(x, y)$ includes only those connections arising from horizontally oriented border-ownership cells (B_0 and B_π), whereas $K^{(r)}(x, y)n_y(x, y)$ accounts for the connections from vertically oriented B cells ($B_{\pi/2}$ and $B_{3\pi/2}$).

In the final step, we segregate these components according to where the border-ownership cell is relative to the center of the receptive field. We define

$$K_0^{(r)}(x, y) = \begin{cases} |K^{(r)}(x, y)n_x(x, y)| & \text{if } x < 0 \\ 0 & \text{otherwise} \end{cases} \quad (A9)$$
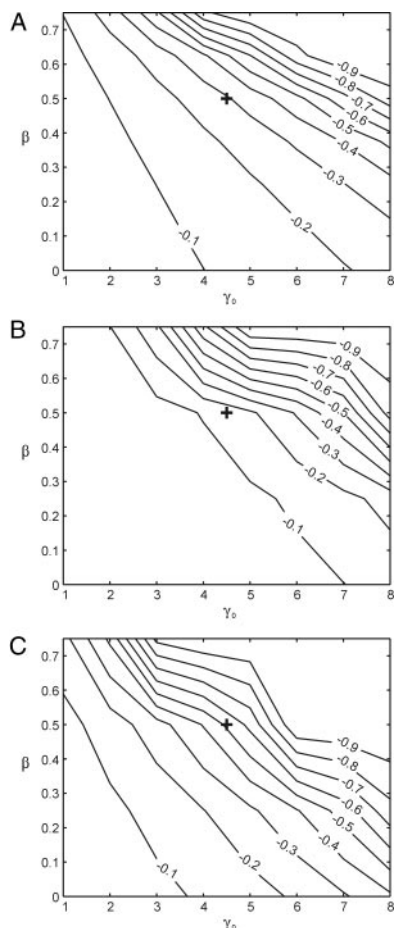
and

FIG. A1. Parameter space of the model. For each of the 3 stimuli in Fig. 6 (*A*: single square; *B*: C-shape; *C*: overlapping squares), contours indicate the effect that varying parameters γ_0 and β has on the modulation index $m_{\bar{i}}$ of the circled model cells (see *Eq. 7*; $m_{\bar{i}} < 0$ indicates the "correct" direction of border ownership). Parameter values used in the current study are indicated by a cross in each plot. Note, however, that these exact values are not essential to the proper performance of the model. Despite variations in response magnitude (see DISCUSSION), the contour plots show that the model determines the direction of border ownership correctly for all 3 stimuli over the entire range of γ_0 and β plotted. As can be seen in the *top right corner* of each plot, increasing either parameter value beyond the range shown results in saturated (i.e., $|m_{\bar{i}}| = 1$) border-ownership assignments in the correct direction.

$$K_\pi^{(r)}(x, y) = \begin{cases} |K^{(r)}(x, y)n_x(x, y)| & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases} \quad (A10)$$

and, by replacing x by y appropriately, similar definitions apply for $K_{\pi/2}^{(r)}(x, y)$ and $K_{3\pi/2}^{(r)}(x, y)$. This yields a set of channelwise patterns of feedforward connections as shown in Fig. 5C.

Border-ownership cells

Consider the feedforward connections generated by a single B cell—for example, $B_0(x_0, y_0)$. We can identify the pattern of G cells that these connections target simply by substituting a delta function, $\delta(x_0, y_0)$, for $B_0(x_0, y_0)$ in *Eq. 1*

$$\delta(x_0, y_0) \otimes K_0^{(r)}(x, y) = \sum_i \sum_j \delta(x + i, y + j) K_0^{(r)}(i, j) \quad (A11)$$

$$= K_0^{(r)}(-x, -y)$$

$$= K_\pi^{(r)}(x, y)$$

This is the pattern of G cells that must send inhibitory feedback

connections to $B_\pi(x_0, y_0)$. Similar relationships apply for the other orientation channels, so feedback from the G cells to the border-ownership layer is implemented using cross-correlations of the form $[\mathbf{G}^{(r)} \otimes \mathbf{K}_{\theta\pm}^{(r)}]$.

The feedforward input to $\mathbf{B}_{\theta\pm}$ is determined by the combined activity of edge detectors $\mathbf{C}_\theta$ and end-stopped cells $\mathbf{E}_{\theta\pm}$. For simplicity, the influence of the latter two cell types was represented as a single term in the stimulus maps, using the following shorthand notation

$$CE_{\theta+}(x, y)$$

$$= \begin{cases} 2 \times C(x, y) & \text{if T-junction indicates } \theta + \pi/2 \text{ directed ownership} \\ 0 & \text{if T-junction indicates } \theta - \pi/2 \text{ directed ownership} \\ C(x, y) & \text{if there is no T-junction} \end{cases}$$

$$(A12)$$

where $C(x, y)$ represents the local edge contrast conveyed by complex cell activity in the absence of junctions. (A similar definition applies for $CE_{\theta-}$, by exchanging each "+" with a "−" and vice versa.) According to this definition, additional excitation provided by end-stopped cells at T-junctions effectively doubles the total feedforward input provided to the like-oriented border-ownership channel. Similarly, inhibition from the end-stopped cells effectively cancels the edge input to the opposing B cell. For a pair of border-ownership cells $B_{\theta+}$ and $B_{\theta-}$, the excitatory feedforward input in *Eqs. 4* and *5* is then given by $CE_{\theta+}$ and $CE_{\theta-}$, respectively (see Fig. A1).

REFERENCES

Angelucci A, Bullier J. Reaching beyond the classical receptive field of V1 neurons: horizontal or feedback axons? *J Physiol (Paris)* 97: 141–154, 2003.

Angelucci A, Levitt J, Walton E, Hupé J, Bullier J, Lund J. Circuits for local and global signal integration in primary visual cortex. *J Neurosci* 22: 8633–8646, 2002.

Arrington K. The temporal dynamics of brightness filling-in. *Vision Res* 34: 3371–3387, 1994.

Baek K, Sajda P. Inferring figure-ground using a recurrent integrate-and-fire neural circuit. *IEEE Trans Neural Syst Rehabil Eng* 13: 125–130, 2005.

Bair W, Koch C, Newsome W, Britten K. Power spectrum analysis of bursting cells in area MT in the behaving monkey. *J Neurosci* 14: 2870–2892, 1994.

Baylis GC, Driver J. Shape-coding in IT cells generalizes over contrast and mirror reversal, but not figure-ground reversal. *Nat Neurosci* 4: 937–942, 2001.

Blum H. A transformation for extracting new descriptors of shape. In: *Models for the Perception of Speech and Visual Form*, edited by Whaten-Dunn W. Cambridge, MA: The MIT Press, 1967, p. 362–380.

Bregman AL. Asking the "what for" question in auditory perception. In: *Perceptual Organization*, edited by Kubovy M, Pomerantz JR. Hillsdale, NJ: Erlbaum, 1981.

Brincat S, Connor C. Underlying principles of visual shape selectivity in posterior inferotemporal cortex. *Nat Neurosci* 7: 880–886, 2004.

Bringuier V, Chavane F, Glaeser L, Frégnac Y. Horizontal propagation of visual activity in the synaptic integration field of area 17 neurons. *Science* 283: 695–699, 1999.

Carpenter G, Grossberg S. Normal and amnesic learning, recognition and memory by a neural model of cortico-hippocampal interactions. *Trends Neurosci* 16: 131–137, 1993.

Craft E, Schütze H, Niebur E, von der Heydt R. A physiologically inspired model of border ownership assignment [Abstract]. *J Vision* 4: 728, 2004.

Driver J, Baylis G. Edge-assignment and figure-ground segmentation in short-term visual matching. *Cognit Psychol* 31: 248–306, 1996.

Felleman DJ, Van Essen D. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1: 1–47, 1991.

Finkel L, Sajda P. Object discrimination based on depth-from-occlusion. *Neural Comput* 4: 901–921, 1992.

Friedman HS, Zhou H, von der Heydt R. The coding of uniform color figures in monkey visual cortex. *J Physiol* 548: 593–613, 2003.

Gallant J, Connor C, Rakshit S, Lewis J, Van Essen D. Neural responses to polar, hyperbolic, and Cartesian gratings in area V4 of the macaque monkey. *J Neurophysiol* 76: 2718–2739, 1996.

Gattass R, Gross C, Sandell J. Visual topography of V2 in the macaque. *J Comp Neurol* 201: 519–539, 1981.

Gattass R, Sousa A, Gross C. Visuotopic organization and extent of V3 and V4 of the macaque. *J Neurosci* 8: 1831–1845, 1988.

Gerrits H, Vendrik A. Simultaneous contrast, filling-in process and information processing in man's visual system. *Exp Brain Res* 11: 411–430, 1970.

Girard P, Hupé J, Bullier J. Feedforward and feedback connections between areas V1 and V2 of the monkey have similar rapid conduction velocities. *J Neurophysiol* 85: 1328–1331, 2001.

Grinvald A, Lieke E, Frostig R, Hildesheim R. Cortical point-spread function and long-range lateral interactions revealed by real-time optical imaging of Macaque monkey primary visual cortex. *J Neurosci* 14: 2545–2568, 1994.

Grossberg S. 3-D vision and figure-ground separation by visual cortex. *Percept Psychophys* 55: 48–120, 1994.

Grossberg S, Mingolla E. Neural dynamics of form perception: boundary completion, illusory figures, and neon color spreading. *Psychol Rev* 92: 173–211, 1985.

Grossberg S, Mingolla E, Ross W. A neural theory of attentive visual search: interactions at boundary, surface, spatial and object recognition. *Psychol Rev* 101: 470–489, 1994.

Grossberg S, Raizada RD. Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Res* 40: 1413–1432, 2000.

He ZJ, Nakayama K. Surfaces versus features in visual search. *Nature* 359: 231–233, 1992.

Heider B, Meskenaite V, Peterhans E. Anatomy and physiology of a neural mechanism defining depth order and contrast polarity at illusory contours. *Eur J Neurosci* 12: 4117–4130, 2000.

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

Heitger F, von der Heydt R. A computational model of neural contour processing: figure-ground segregation and illusory contours. In: *Proceedings of the 4th International Conference on Computer Vision*. Los Alamitos, CA: IEEE Comput. Soc. Press, 1993, p. 32–40.

Heitger F, von der Heydt R, Peterhans E, Rosenthaler L, Kübler O. Simulation of neural contour mechanisms: representing anomalous contours. *Image Vision Comput* 16: 409–423, 1998.

Hubel D, Wiesel T. Receptive fields and functional architecture of monkey striate cortex. *J Physiol* 195: 215–243, 1968.

Intriligator J, Cavanagh P. The spatial resolution of visual attention. *Cognit Psychol* 43: 171–216, 2001.

Kandel E, Schwartz J, Jessell T. *Principles of Neural Science* (4th ed.). New York: McGraw-Hill, 2000.

Kienker PK, Sejnowski TJ, Hinton GE, Schumacher LE. Separating figure from ground with a parallel network. *Perception* 15: 197–216, 1986.

Kikuchi M, Akashi Y. A model of border-ownership coding in early vision. In: *ICANN 2001*, edited by Dorffner G, Bischof H, Hornik K. Berlin: Springer-Verlag, 2001, p. 1069–1074.

Kikuchi M, Fukushima K. Assignment of figural side to contours based on symmetry, parallelism, and convexity. In: *Seventh International Conference on Knowledge-Based Intelligent Information and Engineering Systems*. Berlin: Springer-Verlag, 2003, vol. 2774, pt. 2, p. 123–130.

Kovacs G, Vogels R, Orban G. Selectivity of macaque inferior temporal neurons for partially occluded shapes. *J Neurosci* 15: 1984–1997, 1995.

Kovacs I, Julesz B. Perceptual sensitivity maps within globally defined visual shapes. *Nature* 370: 644–646, 1994.

Lamme V. The neurophysiology of figure-ground segregation in primary visual cortex. *J Neurosci* 15: 1605–1615, 1995.

Lamme V, Zipser K, Spekreijse H. Figure-ground activity in primary visual cortex is suppressed by anesthesia. *Proc Natl Acad Sci USA* 95: 3263–3268, 1998.

Lee TS, Mumford D, Romero R, Lamme VAF. The role of the primary visual cortex in higher level vision. *Vision Res* 38: 2429–2452, 1998.

Nakayama K, He ZJ, Shimojo S. Visual surface representation: a critical link between lower-level and higher-level vision. In: *Visual Cognition: An Invitation to Cognitive Science* (2nd ed.), edited by Kosslyn S, Osherson D. Cambridge, MA: The MIT Press, 1995, vol. 2, chapt. 1, p. 1–70.

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

Nishimura H, Sakai K. Determination of border-ownership based on the surround context of contrast. *Neurocomputing* 58–60: 843–848, 2004.

Nishimura H, Sakai K. The computational model for border-ownership determination consisting of surrounding suppression and facilitation in early vision. *Neurocomputing* 65: 77–83, 2005.

Ogniewicz R, Kübler O. Hierarchic Voronoi skeletons. *Pattern Recognit* 28: 343–359, 1995.

Pao HK, Geiger D, Rubin N. Measuring convexity for figure/ground separation. *Proc 7th Int Conf on Computer Vision*, Kerkyra, Greece, 1999.

Pasupathy A, Connor CE. Shape representation in area V4: position-specific tuning for boundary conformation. *J Neurophysiol* 86: 2505–2519, 2001.

Peterhans E, von der Heydt R. Mechanisms of contour perception in monkey visual cortex. II. Contours bridging gaps. *J Neurosci* 9: 1749–1763, 1989.

Peterson M, Harvey E, Weidenbacher H. Shape recognition contributions to figure-ground reversal: which route counts? *J Exp Psychol Hum Percept Perform* 17: 1075–1089, 1991.

Psotka J. Perceptual processes that may create stick figures and balance. *J Exp Psychol Hum Percept Perform* 4: 101–111, 1978.

Qiu F, von der Heydt R. Figure and ground in the visual cortex: V2 combines stereoscopic cues with Gestalt rules. *Neuron* 47: 155–166, 2005.

Rensink RA, Enns JT. Early completion of occluded objects. *Vision Res* 38: 2489–2505, 1998.

Rockland K, Saleem K, Tanaka K. Divergent feedback connections from areas V4 and TEO in the macaque. *Vis Neurosci* 11: 579–600, 1994.

Roe A, Lu H, Hung C. Cortical processing of a brightness illusion. *Proc Natl Acad Sci USA* 102: 3869–3874, 2005.

Roelfsema P, Lamme V, Spekreijse H, Bosch H. Figure-ground segregation in a recurrent network architecture. *J Cogn Neurosci* 14: 525–537, 2002.

Rossi A, Paradiso M. Neural correlates of perceived brightness in the retina, lateral geniculate nucleus, and striate cortex. *J Neurosci* 19: 6145–6156, 1999.

Rubin E. *Visuell wahrgenommene Figuren.* Copenhagen: Gyldendalske, 1921.

Rubin E. Visuell wahrgenommene Figuren. In: *Visual Perception: Essential Readings*, edited by Yantis S. Philadelphia, PA: Psychology Press, 2001, chapt. 12, p. 225–229.

Rubin N. Figure and ground in the brain. *Nat Neurosci* 4: 857–858, 2001.

Sajda P, Finkel L. Intermediate-level visual representations and the construction of surface perception. *J Cogn Neurosci* 7: 267–291, 1995.

Schütze H, Niebur E, von der Heydt R. Modeling cortical mechanisms of border ownership coding [Abstract]. *J Vision* 3: 114a, 2003.

Singer, W, Gray CM. Visual feature integration and the temporal correlation hypothesis. *Annu Rev Neurosci* 18: 555–586, 1995.

Sugihara T, Qiu FT, von der Heydt R. Border ownership coding in monkey area V2: dynamics of image context integration. *Soc Neurosci Abstr* 29: 819.12, 2003.

Swadlow HA. Information flow along neocortical axons. In: *Time and the Brain* (Conceptual Advances in Brain Research), edited by Miller R. Reading, UK: Harwood Academic, 2000, p. 131–155.

Vecera S, O'Reilly R. Figure-ground organization and object recognition processes: an interactive account. *J Exp Psychol Hum Percept Perform* 24: 441–462, 1998.

von der Heydt R, Friedman HS, Zhou H. Searching for the neural mechanisms of color filling-in. In: *Filling-in: From Perceptual Completion to Cortical Reorganization*, edited by Pessoa L, Weerd PD. Oxford, UK: Oxford Univ. Press, 2003, p. 106–127.

von der Heydt R, Macuda T, Qiu F. Border-ownership dependent tilt aftereffect. *J Opt Soc Am A Optics Image Sci Vis* 22: 2222–2229, 2005.

von der Heydt R, Qiu FT, He ZJ. Neural mechanisms in border ownership assignment: motion parallax and gestalt cues [Abstract]. *J Vision* 3: 666a, 2003.

Wertheimer M. Untersuchungen zur Lehre von der Gestalt II. *Psychol Forsch* 4: 301–350, 1923.

Wertheimer M. Drei Abhandlungen zur Gestalttheorie. In: *Visual Perception: Essential Readings*, edited by Yantis S. Philadelphia, PA: Psychology Press, 2001, chapt. 11, p. 216–224.

Wilson HR. *Spikes, Decisions, and Actions: The Dynamical Foundations of Neuroscience.* New York: Oxford Univ. Press, 1999.

Wilson HR, Wilkinson F, Asaad W. Concentric orientation summation in human form vision. *Vision Res* 37: 2325–2330, 1997.

Young M, Tanaka K, Yamane S. On oscillating neuronal responses in the visual cortex of the monkey. *J Neurophysiol* 67: 1464–1474, 1992.

Yu SX, Lee TS, Kanade T. A hierarchical Markov random field model for figure-ground segregation. In: *Energy Minimization Methods in Computer Vision and Pattern Recognition* (Lecture Notes in Computer Science), edited by Figueiredo M, Zerubia J, Jain AK. Third International EMMCVPR Workshop, September 3–5, Sophia Antipolis, France, 2001, vol. 2134, p. 118–133.

Zhaoping L. Border ownership from intracortical interactions in visual area V2. *Neuron* 47: 143–153, 2005.

Zhou H, Friedman HS, von der Heydt R. Coding of border ownership in monkey visual cortex. *J Neurosci* 20: 6594–6611, 2000.

Zipser K, Lamme VAF, Schiller PH. Contextual modulation in primary visual cortex. *J Neurosci* 16: 7376–7389, 1996.