Mechanisms of perceptual organization provide auto-focus and auto-localization for attention to objects

Stefan Mihalas, Yi Dong, Rüdiger von der Heydt, and Ernst Niebur1

Krieger Mind/Brain Institute and Department of Neuroscience, The Johns Hopkins University, Baltimore, MD 21218

Edited* by Ranulfo Romo, Universidad Nacional Autónoma de México, Mexico City, D.F., Mexico, and approved March 23, 2011 (received for review October 1, 2010)

Visual attention is often understood as a modulatory field acting at early stages of processing, but the mechanisms that direct and fit the field to the attended object are not known. We show that a purely spatial attention field propagating downward in the neuronal network responsible for perceptual organization will be reshaped, repositioned, and sharpened to match the object’s shape and scale. Key features of the model are grouping neurons integrating local features into coherent tentative objects, excitatory feedback to the same local feature neurons that caused grouping neuron activation, and inhibition between incompatible interpretations both at the local feature level and at the object representation level.

Attentional mechanisms select from the onslaught of visual input the information that is behaviorally most relevant. When we search for a face in a crowd or reach for a pencil on the table, the visual system must segregate and select information according to objects. Given the virtually infinite combinations of size, shape, and detailed position that characterize a visual object, how can we attend to any possible object? How can attention, which is controlled by regions in the frontal and parietal cortex (1, 2), access detailed visual features that are not available in those regions? Psychophysical (3–7) and neurophysiological (8–10) studies indicate that attentional selection operates on a structured representation of the visual input that is organized in terms of perceptual objects. We here propose that the neural circuitry that generates this perceptual organization also serves to shape the broad attentional feedback and to adapt it to individual perceptual objects. Attentional modulation travels downward in the network used for perceptual organization and specifically affects the local features of the objects. The internal attention field, which is coarse and purely spatial, is reshaped, repositioned, and sharpened to fit the present objects. This mechanism helps to understand why object structure interacts with attention even when the objects are irrelevant for the task (3, 11).

Figure–ground segregation, the integration of visual features into objects and the segmentation between objects and background, is a fundamental component of perceptual organization. One of its crucial elements, border ownership, is represented in the response properties of single neurons in early cortical areas, most prominently in secondary visual cortex (area V2) (12). A parsimonious explanation of how these neurons can respond highly specifically to stimuli far outside their receptive fields proposes the existence of “grouping cells” (G cells) that organize visual input into proto-objects, without going all of the way to complete object recognition (13). We propose here that the same neural mechanisms that establish this figure–ground organization automatically focus attention onto a perceptual object. Specifically, we show that attentional input to V2 is spatially broad and not tuned for object scale is repositioned to the object location (auto-localization), shaped to match the object contours, and tuned for the scale of the object (auto-zoom).

Following a recent approach by Reynolds and Heeger (14) that accounts for a large body of experimental observations, we model attention as a field that modulates neuronal activity at early stages of visual cortex. Different from their approach, which focuses on the selection of spatial locations and elementary visual features, we here consider attention to objects and contours. Therefore, in our model, attentional modulation does not act directly on the earliest stages of processing (where neurons represent locations and features) but, instead, the attentional input modulates activity in neuronal populations that mediate perceptual organization. The attentional field propagates backward to the neurons representing local features. We show that abstract object representations can thus access fine visual details that are represented only in the lower-level areas. Back propagation of attentional modulation has been used in the selective tuning model (15), but without a mechanism of perceptual organization. The mechanism we propose works with generic, zero-threshold linear neurons and the connection patterns are plausibly related to the statistics of natural visual scenes (16, 17). These patterns include reciprocal connections between cells representing local edges (B cells) and cells that group edges into proto-objects, G cells. The latter have convex, annulus-shaped receptive fields (shown in Fig. S1) and their feedback biases the activity of B cells, which thereby gain their border ownership selectivity (13). The connection patterns presented in Fig. 1 are discussed in Results with details given in SI Materials and Methods and parameters in Table S1.

Results

Multiplicative Attentional Modulation from Additive Attentional Input. Attention modulates the response of cortical neurons to visual stimuli but attention has little effect on the baseline firing rate of neurons in visual cortex (none in V1 (18), either no (10) or a small (19) increase in V2, and a small increase (19) or decrease (20) in V4). Previous studies assumed that the attentional field has functionally such a quasi-multiplicative effect on its targets (14), but did not model the underlying neuronal mechanisms. In Materials and Methods we show how a quasi-multiplicative effect can be obtained with simple additive input.

Attentional Modulation of Neural Contour Signals. We quantify the influence of attention by subtracting from the neural response to an attended visual scene the response to an identical, unattended scene. Fig. 2 shows the simplest situation, an entirely empty visual scene, with no sensory stimulation. As in all situations discussed in Fig. 2, a broad attentional modulation is directed to the Lower Left quadrant of the scene (shading). The difference between the neural response in this quadrant and in the unattended Upper Right quadrant (which is equally devoid of sensory stimulation) is shown in Fig. 2C. Even though edge (E) cells (second column) and border ownership (B) cells (third column) receive top-down input from the network (Fig. 1), excitatory and inhibitory influences cancel (Materials and Methods) and attention does not modulate their firing rates. This result holds true wherever the attended object is located in the visual scene.

Author contributions: S.M., R.v.d.H., and E.N. designed research; S.M. and Y.D. performed research; S.M. analyzed data; and S.M., R.v.d.H., and E.N. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

1To whom correspondence should be addressed. E-mail: niebur@jhu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1014655108/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1014655108
Fig. 1. Structure of the model network. Each circle stands for a population of neurons with similar receptive fields. Edges and other local features of a figure (black parallelogram) activate edge cells (E) that project to border ownership cells (B) that have the same preferred orientation and position, but opposite side-of-figure preferences (in this example B left on the side of the figure and B right on the right). B cells have reciprocal excitatory feedback connections with grouping cells (G) that integrate global contour information at multiple scales (only one scale is shown). The grouping cells bias border ownership cell activity according to the location of the figure. Because a border can be owned by only one figure, opposing border ownership cells compete directly via IE cells and indirectly via grouping cells. The B cells excite inhibitory border ownership cells (IB; again with the indexes L and R) of the same preferred side of the figure, which inhibit G cells in all directions except the preferred one. Thus, grouping cells that are activated by inconsistent figures mutually inhibit each other via B and IB cells and unstructured input creates only small activation of G cells. A weaker, broader, and less specific inhibitory interaction has been introduced in an extension of the model to explain the results presented in Fig. 5. This inhibition is realized by direct G to IB connections, depicted as dashed lines. Top–down attention is modeled as a broad and purely spatial input to the grouping cells (top). Blue and red connections are inhibitory and excitatory, respectively. High-contrast symbols indicate cells and connections activated by the shown figure. Blue, red, and orange connections are inhibitory, feedforward excitatory, and feedback excitatory, respectively.

Throughout our model: Attentional modulation affects the activity of stimulated but not of nonstimulated neurons (Fig. S2). This quasi-multiplicative effect is due to the connectivity patterns and does not require complex neural properties (our model neurons are simple linear threshold units).

Fig. 2B shows a visual scene in which two identical stimuli (squares) are present. The attention effect is shown in Fig. 2D. Both E and B cells (columns 2 and 3) show substantial modulation, in agreement with neurophysiological findings (10). There is also an attentional enhancement in all grouping cell layers (columns 4–6), most prominently at the scale closest to that of the stimulus (column 5). Column 7 shows how G-cell responses are sharpened and scaled to the stimulus size.

Fig. 2E–H shows other stimulus configurations; only the lower quadrant of the visual scene is shown with the stimulus (white lines) in the first column, together with the attentional input (superposed Gaussian shading), which is identical in all cases. As in Fig. 2C and D, the following rows show the difference of responses to attended and unattended stimuli. In Fig. 2E, the stimulus is the same square as in Fig. 2D but displaced in space. The attentional input is unchanged and therefore no longer centered on this stimulus. Nevertheless, E and B cells are appropriately modulated by attention (columns 2 and 3) and the attention field is likewise repositioned at the G-cell level (columns 5 and 7). Fig. 2F and G shows that the attentional modulation is also tuned to the object scale: Even though the stimulus is smaller (Fig. 2F) or larger (Fig. 2G) than that used in Fig. 2B, attention modulates the activities of all neuronal populations appropriately. Fig. S3 shows the activity of inhibitory populations, and Fig. S4 shows the behavior of a model with direct between-object inhibition.

The attentional mechanism does not depend on grouping cells that are tailored to specific sensory stimuli. G cells are genetically activated by a large variety of stimuli whose only common factor is that they are located in approximately the same region of the visual field. For instance, Fig. 2H shows that top–down attention can be directed to an isolated line (which has no equivalent in the grouping cell layer). Rather than creating a small focus of strong activation in the G-cell layer, as is the case for stimuli that correspond to the spatial scale of grouping cells (Fig. 2D–G), the line activates grouping cells of all sizes and on both of its sides. We also note an object-superiority effect: Activity of B cells is higher when an edge is part of an object than when the same edge is shown in isolation (compare Fig. 2D and H, column 3).

Attention needs to work in complex, cluttered environments where space alone is not sufficient to segregate foreground from background. Qiu et al. (10) studied responses in area V2 when two overlapping objects were presented and attention was directed either to the foreground or to the background object (Fig. 3). In the absence of attention, the figure–ground organization effect can be observed (Fig. 3, row 1). Of particular interest is the border ownership signal (difference between the firing rates of B cells of opposite directions) for the edge separating the foreground and the background context and both objects (including T-like junctions) assign this edge to the foreground figure (see arrows in last column), in agreement with physiological results (12). This assignment is even stronger if attention is on the foreground figure (Fig. 3, row 2). However, if attention is directed toward the background figure, the border ownership modulation of the edge between the objects is nearly abolished (Fig. 3, row 3).

This result is not only in agreement with physiological evidence (10), it is also functionally desirable. When the background object is attended, the pattern-matching mechanisms downstream should receive only edge information from this object. In the occluded area, much information is available and no information should therefore be sent to downstream processing areas; including edge information (from the foreground object) would lead them astray. Importantly, this result cannot be achieved by purely spatial attention because the occluding edge is closer to the center of the background object (the spotlight of attention) than to the foreground object. The ability to attend to objects in clutter is perhaps the most important function of the grouping mechanisms. Further tests of our model with figures embedded in noise show that the same broadly focused top–down attention can bring out figures from background noise (Fig. S5).

We performed a global sensitivity analysis (21) to determine the dependence of several attentional modulation indexes, border ownership modulation, and their interaction on several model parameters. Border ownership modulation was sensitive only to the strength of the excitatory feedback circuits (lateral connections between B cells and feedback from G to B cells). Attentional modulation of the B cells was sensitive also to the strength of the attention input. Attentional modulation of the nonpreferred B cells was sensitive to the strength of the inhibition caused by inhibitory edge (IE) cells as well. These dependencies allowed a sequential tuning of the parameters.

Quantitatively matching the observed border ownership modulation in the absence of attention constrains the strength of the positive feedback loop created by the mutual excitation between the cells coding for a proto-object and the cells coding for its local features. Subsequently, matching the experimentally observed strength of attentional modulation of B cells (10) constrains the strength of the attention input. We find that the maximal value of the attentional input to grouping cells needs to be 7% of the direct excitatory input that a grouping cell receives from a figure of its optimal size. To match the weak modulation of the opposite B cells, inhibition from IE cells needs to be just strong enough to prevent excitation by attention in the absence of edges.

The pattern and strength of the lateral inhibition are key to obtaining sharpening of the attention field and quasi-multiplicative
attention modulation. The strength of reciprocal connections from neurons representing local features and neurons integrating them and inhibition between inconsistent proto-object representations are key to repositioning and scaling.

**Effects of Perceptual Organization on Reaction Times.** Perceptual organization influences reaction times even in tasks for which grouping of features and their organization into perceptual objects are irrelevant. Egly, Driver, and Rafal (3) observed a significant reaction time difference in a detection task between two positions equidistant from the focus of attention when one of the positions was on the same perceptual object (outlined rectangle) as the focus of attention. Because our model predicts a reshaping of the attention field caused by figures, we used it to predict neuronal firing rate responses in the Egly et al. task. We assumed that higher neuronal activation lowers reaction times and we found that the...
observed activities of the edge cells (target > cued > uncued) are, indeed, consistent with the observed reaction times (uncued > cued > target; Fig. 4). Importantly, these results are a genuine prediction: All parameter values were identical to those obtained from fitting the firing rates of the border ownership neurons; they were not retuned for predicting the Egly et al. data.

Kimchi et al. (11) studied stimulus-driven capture of attention by “objecthood.” Several t-shapes of different orientations were presented and the task was to identify the color of one of them, which was precued to be the target. The elements could either be arranged randomly or be grouped such that four of them comprised the corners of a square, a perceptual object (Fig. 5A–C). It was irrelevant for the task whether the target shape was part of an object. Reaction times were fastest when the target was part of an object, slowest when the target was outside an object, and intermediate when no object was present. Using the same parameters as before, our model produced an activity enhancement of edge cells when the target was part of the object, consistent with the reaction time benefit. We did not, however, obtain a suppression when the target was outside the object (Fig. 5D). This empirically observed suppression was not reproduced by the model due to the absence of long-range inhibition between different G cells, which was not required by the constraints of the previous simulations. Competition between G cells occurred only in the presence of a common edge (via IE cells, Fig. 1). However, it is known that attention also involves competition between separate objects (e.g., refs 22–24). Accordingly, while keeping all other parameters constant, we added direct connections between G and inhibitory B (IB) cells (dashed lines in Fig. 1) that follow the same pattern as the excitatory G to B connections but with doubled radius. Through a weak but spatially broad activation of inhibitory border ownership cells (Fig. S4 D–J, column 6), grouping cells now compete with each other (Fig. S4D). This competition results in decreased activity for isolated edges (Fig. S3) and weak, broad suppression around objects. With this addition, edge modulation is weaker for items near objects (Fig. S5E) and our simulated results agree with all results observed by Kimchi et al. (11).

Discussion

In our model, attentional top–down influence is assumed to be simple: a broad Gaussian. However, the attentional modulation of the grouping cells, which pass down the attention signal to the early stages of the visual system, is no longer simple: It is focused on perceptual objects and tuned to their scale. The attentional modulation in the grouping cell layer is similar to a spotlight of attention that is shaped and positioned by bottom–up signals. This refinement of the attentional modulation results in rather even enhancement of contour signals irrespective of the size and specific position of the attended object. The effective attentional spotlight is thus not imposed from central structures; these structures are unlikely to possess information about specific properties of the local features of objects. Instead, it is formed dynamically through the interplay of central structures (frontal and parietal cortex) that provide executive guidance of a general

![Fig. 3.](https://example.com/fig3.png)

**Fig. 3.** Attention directed to one of two overlapping figures. First column: Sensory input and attentional modulation, same format as Fig. 2. Columns 2–4: Activity of grouping cells of different scales. Column 5: Vector field of border ownership cell activity where each border ownership cell is assigned a vector in the direction of the preferred figure side with a magnitude equal to its firing rate. At each location the vectors for all border ownership cells at this location are summed. In the absence of attention (Top row) or when attention is directed to the foreground (Middle row), the edge between the figures is correctly assigned to the foreground figure. If attention is directed toward the background figure (Bottom row), the border ownership signal of this edge is strongly reduced, to values consistent with experimental observations (10).

![Fig. 4.](https://example.com/fig4.png)

**Fig. 4.** Attention modulation of edge cells in a simulation of the Egly et al. (3) experiments. The outlines of two rectangles are presented. Top–down attention is directed toward the Upper Left corner. Attentional input is applied to grouping cells as a two-dimensional Gaussian (SD shown by circle). The target can be present at three positions relative to the cued location: (A) valid position, (B) invalid position but the target is in the cued rectangle, and (C) invalid position and the target is in the uncued rectangle. The distance between target and cue is the same in the invalid cases. The valid position produces the largest activity in the edge cells (A). Because in our model the attentional field is reshaped to match the figure, the activity of edge cells is higher if the target is in the cued rectangle (B) than if the target is as far away from the cue, but in the uncued rectangle (C). The depth of shading of every pixel represents the activity of the edge cells of preferred orientation at that location; see grayscale bar. (D) Activity of the edge cells representing the target for the valid, invalid cued, and invalid uncued conditions. The error bars represent the SD of the edge cells’ activity. All parameters were identical to those obtained from fitting the responses of border ownership neurons. Relative neuronal activation is consistent with observed reaction times.
Attentional modulation of edge cells in a simulation of the Kimchi et al. (11) experiments. An array of 3 × 3 L-shapes is presented and top–down attention (represented by a circle) is drawn to one of them (the target) by providing a cue (solid dot) and a direction relative to the cue (in all three examples it is downward). The task, identification of the color of the target, is independent of its orientation or of whether it is part of an object. The target can be part of an object (A), not part of an object when no object is present anywhere (B), or not part of an object that is present elsewhere (C). Shading of a pixel represents the activity of edge cells at the location of the pixel. (D) Mean edge modulation of the tips of the target (bars: SDs) for 200 simulations of random L-shape orientations in each category. The model and parameters are identical to those obtained from fitting the border ownership data. Consistent with the observed reaction times, being part of an object enhances edge activation. However, presence or absence of an object elsewhere did not make a consistent difference. (E) The same as D but after a weak and spatially broad inhibition between grouping cells is introduced (dashed lines in Fig. 1). The inhibition results in a spatially broad and nonspecific activity decrease of all cells outside an object. The smaller activity of the edge cells outside an object compared with the case in which no object is present is consistent with the observed reaction times (11). The additional inhibitory mechanism does not qualitatively change any of the previous results (Fig. S4).

Materials and Methods

Network Description. The network is described as a system of ordinary differential equations, with the dynamics of each neuron,

\[ \tau \dot{f}(t) = -f + \left[ \sum W_{ij} \right] \]

where \( f \) represents the neuron’s activity level and \( \tau \) its time constant, chosen as \( \tau = 10^{-2} \) s for all neurons, \( W \) is the neuron’s inputs, and \( [\ldots] \) means rectification. The stationary output of each neuron is thus linear in its inputs, except for a rectifying threshold that is set at 0.

The sensory input to the network is an edge map, similar to the model described by ref. 13. Perceptual organization over large spatial scales is established through specialized grouping cells that integrate over space and provide context information to border ownership selective cells (B cells, Fig. 1). A second input to the model is the top–down attention input, which is assumed to act on grouping cells. The attention input is purely spatial and, on the basis of psychophysical findings (26), it is considered to have lower spatial resolution than the sensory input. Excluding a set of simulations presented in Fig. S2, attentional input is considered to have a SD eight times the size of a border ownership cell receptive field (the median receptive field size for border ownership cells is 0.7° at 2° eccentricity) (12). The strength of the
At the level of V2, E cells activate two border ownership (B)-cell populations with opposing border ownership preferences. Parallel to its preferred orientation, a border ownership cell activates IE cells and B cells of the same preferred orientation. IE cells inhibit all neighboring B and IE cells nonspecifically. All these connections are assumed to have a Gaussian distribution pattern, with a SD twice the receptive field size. Orthogonal to the preferred orientation, these connections create a “Mexican hat” profile. The strength of the nonspecific lateral inhibition was tuned to match the measured attention modulation of B neurons. The strength of the specific lateral excitation was tuned to match the border ownership modulation in the absence of attention (10).

B cells activate grouping cells of multiple scales whose integration fields are on the border ownership cells’ preferred figure side. The connection patterns are presented in Fig. S1 and similar to those used in previous models of perceptual organization (13). The G cells are sensitive to circulatory arrangement of edges (cf. ref. 16), which combines features of the Gestalt laws of good continuation, convexity of contour, and compact shape (27, 28). The inputs to G cells of different preferred sizes are scaled such that a straight line produces equal excitation in G cells of different scales. Border ownership cells also excite IB cells that inhibit grouping cells whose activation is inconsistent with a figure edge in the receptive field of the border ownership cell. Grouping cells have reciprocal connections with border ownership cells, but with a different scaling: Each border ownership cell is equally affected by grouping cells of different spatial scales. The grouping cells also connect with the same pattern and with equal strength to all of the edge cells that coactivate with the border ownership cells that they receive input from. The feedback from each G cell is reciprocal, thus activating the same neurons that led to its activation and also the neurons that are coactivated with it. The scaling of these connections was chosen such that images of different scales are processed in a similar manner. The strength of the excitatory feedback loop from border ownership cells to grouping cells and back was tuned to match the border ownership modulation in the absence of attention (10).

In an extension of the model (see discussion of Fig. 5), the feedback from the grouping cells additionally activates the inhibitory border ownership neurons, with a connection pattern similar to the feedback to excitatory border ownership neurons and the same connection strengths but twice their spatial spread.

To constrain the model parameters we used three dimensionless quantities that were determined from experimental data (10): the border ownership modulation index when attention is directed away from the object at the receptive field and the attention modulation indices for the border ownership cells in preferred and antipreferred conditions (see ref. 10 for definitions). We performed a global sensitivity analysis on five synaptic weight constants that are not scaling parameters (Table S2). The strength of the excitatory feedback loop between B and G cells and the strength of the excitatory lateral connections between B cells were chosen such that the correct border ownership modulation was obtained. After setting these parameters, the attention input was tuned such that the experimentally observed value of the attention modulation of the preferred border ownership cells is reproduced. The attention modulation of the nonpreferred border ownership cell was used to tune the strength of the lateral inhibition produced by IE cells. The obtained value also reproduces the observed minimal effect of attention in the absence of figures (Fig. S2). From the available data, we could not determine how specific the lateral inhibition produced by IE cells is, and we assumed this inhibition to be nonspecific. See SI Materials and Methods for a complete specification of the network.

Effects of Local Circuity. We consider first attentional modulation in the absence of edge inputs (Fig. 2A and C) and we always assume that top–down attentional influence at the G-cell level is spatially broad (Fig. 1). The grouping cells provide excitatory input to a large population of border ownership (B) and IE neurons. In the absence of sensory input, the direct excitatory input from grouping cells to a B cell is compensated by the sum of all inhibitory input from the IE cells; attention therefore does not modify B-cell activities that do not receive sensory input. We note that if top–down attentional influence at the G-cell level is highly spatially localized, attention can produce a small activity increase at the B-cell level (Fig. S2), consistent with experimental observations in which the attentional focus was spatially sharp (19).

If a sensory stimulus, e.g., an oriented line, is presented in the absence of attention (Fig. S3B, Upper Right quadrant) it provides, via edge cells, bottom-up input to those B cells that are selective for its orientation. Along the axis orthogonal to the preferred orientation of the cells, due to the Mexican hat pattern of excitation/inhibition, a localized sensory stimulus produces a localized activation of both B and IE cells, with the surrounding cells being inhibited below their threshold. Along the axis parallel to the preferred orientation, the signals related to the sensory stimulus propagate, and their analysis is discussed in Results.

When a stimulus is attended, both sensory and attentional inputs are present (Fig. S3C). Lower Left quadrant, with the activity difference between the attended and the unattended stimulus shown in Fig. S3B). B cells far from the sensory input receive balanced excitation and inhibition from the attention input, as described above, and thus do not change their firing rates. At the location of the stimulus, because nearby IE cells are inhibited below their threshold by the stimulus, B cells receive excitation from the attention input that is no longer balanced by inhibition. Attentional modulation thus affects the activity of stimulated neurons but not that of nonstimulated neurons.

ACKNOWLEDGMENTS. We thank Anne Martin for a critical reading of the manuscript. This work was supported by National Institutes of Health Grants 5R01EY016281-02 and 01EY02966 and Office of Naval Research Grant N000141010278.
**Supporting Information**

Mihalas et al. 10.1073/pnas.1014655108

**SI Materials and Methods**

One pixel represents the typical size of the receptive field for border ownership cells (which at eccentricity $2^\circ$ is ~0.7') (1). The simulated visual field is assumed to be homogeneous. To avoid unbalanced inputs near the boundaries of the visual field we use periodic boundary conditions. Each neuronal unit in the model represents multiple neurons with overlapping receptive fields and similar tuning. It is referred to as a "neuron" in the following and it is represented by an ordinary differential equation, Eq. 1.

For all examples the system is simulated for 0.5 s and an equilibrium is reached within a few tens or ~100 ms. Shown in the results is the activity after 0.5 s. A typical simulation of a visual field of $64 \times 64$ units consists of 54,528 coupled ordinary differential equations that are solved using an explicit Runge–Kutta formula in Matlab (MathWorks).

Each neuron is characterized by its spatial location, its type (edge cell, grouping cell, etc.), and one additional property. The E and IE cells are indexed by the angle of their preferred orientation: horizontal, $o = 0$; vertical, $o = \pi/2$. The B and IB cells are indexed by the angle of their preferred side of figure: right, $o = 0$; up, $o = \pi/2$; left, $o = \pi$; down, $o = 3\pi/2$, which is orthogonal to the angle of their preferred orientation. The grouping cells are indexed by their preferred radii (2, 4, or 8). Grouping cells of radius 2 are present on every pixel, those of radius 4 are present every other pixel, and those of radius 8 are present every fourth pixel. The network receives three types of inputs. A binary edge map activates E cells (with value 1 if an edge of that orientation is present in the image and −1 otherwise). Local T-junction detectors provide input of a total strength 1/3 to border ownership and IE cells; these are activated only by overlapping figures. Finally, an attentional field stimulates grouping cells, with maximal strength of 0.07 (Table S1).

The edge cells E excite border ownership cells of the same preferred orientation $o$, at the same position $(x, y)$

$$W_{E|x_1,y_1,B|x_2,y_2}^{o_1} = \text{eto}_b \delta_{x_1,x_2} \delta_{y_1,y_2} \sin^2(o_1 - o_2), \quad [S1]$$

where the weight of the E to border ownership connections, eto$_b$, is a scaling parameter, and its value is set to 1. Changing it does not produce different activation patterns in the network but rather scales up all activities.

The connections from border ownership to IE are Gaussians along the preferred orientation,

$$W_{B|x_1,y_1,IE|x_2,y_2}^{o_1} = N_{btoie} \delta_{x_1,x_2} \exp\left(\frac{(x_1 - x_2)^2}{2 \text{btoie}^2}\right) \quad \text{if } o_1 \in \{\pi/2, 3\pi/2\}$$

$$W_{B|x_1,y_1,IE|x_2,y_2}^{o_1} = N_{btoie} \delta_{x_1,x_2} \exp\left(\frac{(y_1 - y_2)^2}{2 \text{btoie}^2}\right) \quad \text{if } o_1 \in \{0, \pi\}, \quad [S2]$$

where the normalization coefficient $N_{btoie}$ is determined from

$$\text{btoie} = \sum_{i,j} W_{B|x_1,y_1,IE|x_2,y_2}^{o_1} \quad [S3]$$

The weight of the border ownership to IE connections btoie is set to 2. The SD of the lateral connections, btoie, was chosen to be twice the size of the border ownership cells’ receptive field size, which is a conservative estimate. The distribution is truncated, at eight times the size of the receptive field. For simplicity, all lateral connections in V2 are assumed to have the same SD.

IE cells inhibit all border ownership and IE cells in their neighborhood. The strength of the inhibition of cells of the same preferred orientation and the strength of the inhibition of cells of orthogonal orientations were varied independently in the sensitivity analysis; however, the data are fitted well when the two are equal. We cannot draw the conclusion that the inhibition is independent of orientation because the results are largely independent of the strength of inhibition of orthogonal cells. For simplicity, we assume that whereas the excitatory connections are specific to cells of a particular tuning, and the inhibitory ones are nonspecific,

$$W_{IE|x_1,y_1,B|x_2,y_2}^{o_1} = N_{ietoie} \cos^2(o_1 - o_2) \exp\left(-\frac{(x_1 - x_2)^2}{2 \text{btoie}^2}\right)$$

$$W_{IE|x_1,y_1,B|x_2,y_2}^{o_1} = N_{ietoie} \sin^2(o_1 - o_2) \exp\left(-\frac{(y_1 - y_2)^2}{2 \text{btoie}^2}\right)$$

$$W_{IE|x_1,y_1,B|x_2,y_2}^{o_1} = N_{ietoie} \cos^2(o_1 - o_2) \exp\left(-\frac{(x_1 - x_2)^2}{2 \text{btoie}^2}\right)$$

$$W_{IE|x_1,y_1,B|x_2,y_2}^{o_1} = N_{ietoie} \sin^2(o_1 - o_2) \exp\left(-\frac{(y_1 - y_2)^2}{2 \text{btoie}^2}\right), \quad [S4]$$

where the normalization coefficients $N_{ietoie}$ and $N_{ietoie}$ are determined from

$$\text{ietoie} = \sum_{i,j} W_{IE|x_1,y_1,B|x_2,y_2}^{o_1} \quad [S5]$$

Following the assumption that the inhibition is independent of the preferred orientation of the target cell, ietoie = ietoie and ietoie = ietoie. One of the parameters in the B to IE negative feedback system is a scaling parameter, and we chose ietoie = −1. The strength of the inhibitory connections IE to border ownership, ietoie, is an important parameter in the model. Its value was chosen to be −2, which is just strong enough for the
inhibition of the border ownership cells to be complete in the case of a uniform stimulation field. In our model, attention in the absence of edges is a broad field, and experimental observations show little effect of attention alone on the firing rates of border ownership cells. This parameter also influences the value of the attention modulation of the nonpreferred border ownership cells, and the value that was a priori chosen reproduces well the observed experimental value.

B cells connect to other border ownership cells of the same preferred side of figure. These connections allow the passing of border ownership signal along a line,

\[ \text{WB}^{\delta_{11}}_{b1,b2} = N_{\text{btob}} \delta_{1,1} \delta_{1,2} \exp \left( \frac{(x_1-x_2)^2}{2 \text{btobsd}} \right) \text{ if } o_1, o_2 \in \{\pi/2, 3\pi/2\} \]

where the normalization coefficient \( N_{\text{btob}} \) is determined from

\[ \text{btobw} = \sum_{ij=-8}^{8} \text{WB}^{\delta_{11}}_{x+i,y+j}. \]  

The strength of the border ownership to grouping connections is smaller than the integral of the inhibitory connections for a line of cells along the preferred direction and that the integral of the excitatory connections is smaller than the integral of the inhibitory ones. The SD \( \text{btobsd} = 2 \) is the same as for all other lateral connections in \( V_2 \).

B cells also connect to other border ownership cells of orthogonal preferred orientation. These connections allow the passing of border ownership information along a corner,

\[ \text{WB}^{\delta_{11}}_{b1,b2} = \text{btobw} \left( \text{rot}_{1} (C_f(x_2-x_1, y_2-y_1)) \delta_{o_1, o_2 - \pi/2} + c \text{rot} \delta_{o_1, o_2} (C_f(x_2-x_1, y_2-y_1)) \delta_{o_1, o_2 - \pi/2} + \text{rot} \delta_{o_1, o_2} (C_f(x_2-x_1, y_2-y_1)) \delta_{o_1, o_2 + \pi/2} + c \text{rot} \delta_{o_1, o_2} (C_f(x_2-x_1, y_2-y_1)) \delta_{o_1, o_2 + \pi/2} \right) \]  

with the matrices

\[ C_f(x,y) = \begin{cases} N_{\text{btob}} \exp \left( \frac{-x^2 + y^2}{2 \text{btobsd}^2} \right) & \text{if } x \geq 0 \text{ and } y < 0 \\ 0 & \text{otherwise} \end{cases} \]

\[ C_f(x,y) = \begin{cases} N_{\text{btob}} \exp \left( \frac{-x^2 + y^2}{2 \text{btobsd}^2} \right) & \text{if } x \geq 0 \text{ and } y > 0 \\ 0 & \text{otherwise} \end{cases} \]

and the normalization coefficient \( N_{\text{btob}} \) is obtained from

\[ 1 = \sum_{ij=-8}^{8} C_f(i,j). \]

The values of \( \text{btobw} = 2/3 \) and \( c = 2/3 \) were chosen to be the same as \( \text{btobw} \).

B cells connect to inhibitory border ownership cells of the same preferred side of figure. These connections allow a balanced excitation/inhibition input at the level of the grouping cells,

\[ \text{WB}^{\delta_{11}}_{b1/2,b22} = N_{\text{btob}} \delta_{o_1, o_2} \delta_{1,2} \exp \left( \frac{(x_1-x_2)^2}{2 \text{btobsd}^2} \right) \text{ if } o_1, o_2 \in \{0, \pi\} \]

\[ \text{IB}^{\delta_{11}}_{b1/2,b22} = N_{\text{btob}} \delta_{o_1, o_2} \delta_{1,2} \exp \left( \frac{(y_1-y_2)^2}{2 \text{btobsd}^2} \right) \text{ if } o_1, o_2 \in \{0, \pi\}. \]  

where the normalization coefficient \( N_{\text{btob}} \) is determined from

\[ \text{btobw} = \sum_{i,j=-8}^{8} \text{WB}^{\delta_{11}}_{x+i,y+j}. \]  

The connection strength \( \text{btobw} \) is a scaling parameter and its strength is considered to be 1. As for all of the lateral connections, the SD \( \text{btobsd} \) is considered to be 2.

The connections from border ownership cells to grouping cells consist of 2D Gaussians arranged in a circular fashion,
The normalization coefficient obtained from
\[
ibtog_w = \sum_{i=-\infty}^{\infty} \frac{2}{2\pi} WIB_{x,y;i}^{*} G_{y,y} G_{x,y}^{*} \tag{S17}
\]
as well as inhibition from orthogonal orientations,
\[
W_{o}IB_{x,y;i}^{*} G_{x,y}^{*} = N_{ibtogro} \exp \left( \frac{(x_1 - x_2 + r \cos(o))^2}{2(btogr)^2} - \frac{(y_1 - y_2)^2}{2(btogr)^2} \right) \quad \text{if } o \in \{0, \pi\}
\]
\[
W_{o}IB_{x,y;i}^{*} G_{x,y}^{*} = N_{ibtogro} \exp \left( \frac{(x_1 - x_2 + r \cos(o))^2}{2(btogr)^2} - \frac{(y_1 - y_2)^2}{2(btogr)^2} \right) \quad \text{if } o \in \{\pi/2, 3\pi/2\}, \tag{S18}
\]
with the normalization coefficient obtained from
\[
ibtog_w = \sum_{i=-\infty}^{\infty} \frac{2}{2\pi} WIB_{x,y;i}^{*} G_{y,y} G_{x,y}^{*}. \tag{S19}
\]

It is assumed that the strength of the inhibition from the non-preferred side of figure ibtogw is equal to that of orthogonal preferred sides of figures ibtogen, and each of them is half the value of the excitatory strength btoew. Similar to the lateral connections in V2, this feedback loop has stronger but fewer and more specific excitatory connections, resulting in robust activity for specific inputs and more total inhibitory connection weights, resulting in little activity caused by nonspecific inputs.

The feedback from the grouping cells also follows the spatial pattern of the B to grouping connections. The feedback to E, B, and IE is identical in spatial distribution,
\[
WG_{x,y;i}^{*} B_{x,y;i} = N_{gtoembr} \exp \left( \frac{(x_1 - x_2 + r \cos(o))^2}{2(btoembr)^2} - \frac{(y_1 - y_2)^2}{2(btoembr)^2} \right) \quad \text{if } o \in \{0, \pi\}
\]
\[
WG_{x,y;i}^{*} B_{x,y;i} = N_{gtoembr} \exp \left( \frac{(x_1 - x_2 + r \cos(o))^2}{2(btoembr)^2} - \frac{(y_1 - y_2)^2}{2(btoembr)^2} \right) \quad \text{if } o \in \{\pi/2, 3\pi/2\}. \tag{S20}
\]

The normalization coefficient is chosen such that the feedback on a line is the same from all spatial scales of grouping cells. Because the number of grouping cells on a line is inversely proportional to their scale, the line integral of the weight is considered proportional to the radius of the G cell, relative to a reference radius $gtobr_w$.
\[
gtobr_w = \frac{r}{gtobr} \sum_{i=-\infty}^{\infty} \frac{2}{2\pi} WG_{x,y;i}^{*} B_{x,y;i}^{*} \tag{S21}
\]

The connection strength to the IE cells residing in V2 is assumed to be the same as the connection strength to the B cells. The value of $gtobw = 1/3$ is critical for the model, and the border ownership modulation index depends critically on it. The experimental value of 0.22 for the border ownership modulation index of a square eight times the size of the receptive field is observed when the total weight of the positive feedback $gtobw + gtoew$ is close to 1. The border ownership modulation varies only modestly with the size of the figure.

To reproduce the observed attention modulation of the non-preferred side of the figure for B cells, the weight of the feedback to edge cells needs to be double that to the border ownership cells $gtoew = 2/3$.

To fit the observed reaction time cost when irrelevant objects are outside the focus of attention (Fig. 5E), a feedback from G to IB cells is introduced. The direct feedback from G to IB cells has the same pattern and scaled weight as the feedback to B cells, but twice its radius and SDs:
\[
WG_{x,y;i}^{*} IB_{x,y;i} = N_{gtoembr} \exp \left( \frac{(x_1 - x_2 + r \cos(o))^2}{2(btoembr)^2} - \frac{(y_1 - y_2)^2}{2(btoembr)^2} \right) \quad \text{if } o \in \{0, \pi\}
\]
\[
WG_{x,y;i}^{*} IB_{x,y;i} = N_{gtoembr} \exp \left( \frac{(x_1 - x_2 + r \cos(o))^2}{2(btoembr)^2} - \frac{(y_1 - y_2)^2}{2(btoembr)^2} \right) \quad \text{if } o \in \{\pi/2, 3\pi/2\}. \tag{S22}
\]

To test the influence of background clutter we compared the edge signal enhancement and grouping cell activation between 100 simulations of randomized small line elements and 100 simulations of the same small line elements on top of which is added a figure that maximally activates a grouping cell of radius 8 (Fig. S5). The small edge elements had the length of three E cells receptive fields and were randomly distributed on a $12 \times 12$ lattice. In a first set of simulations, each lattice element had 0.25 probability of a horizontal edge element and an independent 0.25 probability of a vertical edge element (Fig. S5 A and B). In a second set of simulations, each lattice element had probability was increased to 0.5 (Fig. S5 C and D). The input from the line elements to each edge cell is 1. Due to the presence of other line elements, the feedback from the grouping cells enhances the activity of edge cells. This enhancement is larger if the edge is part of a figure than part of the random background (an unpaired $t$ test for the maximal edge cell activation in 100 simulations with figure present versus figure absent has $p = 6.7 \times 10^{-40}$ if the background clutter is 0.25 and $p = 0.0041$ if the background clutter is 0.5). The presence of a figure results in a greatly enhanced activity of the grouping cells of preferred scale (for $Gr_8 p = 5.9 \times 10^{-76}$ if the background clutter is 0.25 and $p = 3.5 \times 10^{-90}$ if the background clutter is 0.5). The activity of grouping cells of smaller scale is modestly decreased; this change is significant only for the lowest scale (Gr2) in the low clutter condition ($p = 0.01$). In the case of high clutter, whereas for a set of 100 simulations the differences in edge and grouping cell activities are different if a figure is present or not, the SDs are higher than the differences in activity. In the case of high clutter, if an attention signal is directed toward the possible location of the figure, the differences in activities become similar to those in the lower clutter conditions. With attention, in high clutter condition, at the population level, there are significant differences between maximal edge ($p = 1.2 \times 10^{-22}$) and grouping of radius 8 ($p = 1.1 \times 10^{-18}$) activities if a figure is present or not.

We performed a sensitivity analysis of the simulation of non-overlapping squares to independent variations in five synaptic weights that were not fixed from a priori reasons:

i) The weights of the feedback connections $gtoew = gtobw = 2/3$ were varied simultaneously.

ii) The weight of the excitatory lateral connection both along a line ($btoew = btobw$) and along a corner ($btoew = btobw$) were varied simultaneously.

iii) The strength of the attention field coming at the level of grouping cells, Gatt was varied.
iv) The weights of the lateral inhibition of the same orientation for inhibitory $\text{ieto}_{w}$ and excitatory $\text{ieto}_{w} = 2 \text{ieto}_{w}$ and were varied simultaneously.

v) The weights of the lateral inhibition of different orientations for inhibitory $\text{ieto}_{w}$, and excitatory $\text{ieto}_{w} = 2 \text{ieto}_{w}$ and were varied simultaneously.

The values of these parameters were randomly chosen between 9/10 and 10/9 times the value presented in Table S1. One thousand simulations were run. We used high-dimensional model representation for global sensitivity analysis, using 10 steps of the correlation method for variance reduction and a threshold of 0.01. Sobol indexes are reported in Table S2.

As the output of the model we used three dimensionless indexes that were compared with measured data from Qiu et al. (2): the border ownership modulation index in the absence of attention and the attention modulation index for border ownership cells of the preferred and the antipREFERRED side of the figure. The strength of the excitatory lateral connections between border ownership cells strongly influenced all three indexes. The strength of the feedback connection from grouping cells to earlier levels strongly influenced the attention modulation of the border ownership cells of the preferred side of the figure and, to a lesser extent, the border ownership modulation index. The strength of the attention input affects the attention modulation of the border ownership cells of the preferred side of the figure only. The strength of lateral inhibition between cells of the same preferred orientations affects the attention modulation of the border ownership cells of the nonpreferred side of the figure whereas the strength of lateral inhibition between cells of different orientations did not directly affect any of the experimentally measured indexes.


Fig. S1. Connection pattern of grouping cells of radius 2 (A), radius 4 (B), and radius 8 (C) to border ownership cells.
Fig. S2. In the absence of visual input, attention has only small effects on the activity of border ownership cells. (A) The difference in activity between attended and nonattended conditions for border ownership neurons whose preferred side of figure points toward the center of attention. (B) Vector field plot showing the overall activity of the border ownership cells in the presence of attention. The scale of the vectors is changed relative to Fig. 3; whereas an arrow of length 1 there represents an activity difference in opposing border ownership cells of 0.5, an arrow of this length here represents an activity difference of 0.1. (C and D) The effects of attention are no longer null if the strength of the nonspecific inhibition at the level of V2 is reduced to 0.9. Few other parameters influence the effect of attention in the absence of edge in the visual stimulus. (E and F) The effects of attention are also no longer null if the attention input is spatially sharp. The same total attention input is spread to fewer grouping neurons (2D Gaussian with a SD 2 rather than the SD 8 that was used for A and B), allowing excitation and inhibition to become locally unbalanced.
Fig. S3. Additional population responses showing the effect of attention directed to objects (Fig. S2). (A and C) Identical to Fig. 2 A and B. (B) The same, but the stimulus is two isolated lines. (D–I) Activity differences for several types of stimuli. Rows are organized in the same order as in Fig. 2. Columns 1, 2, 4, and 7–10 are identical to rows 1–7 of Fig. 2 in the main text and not described here. Column 3: Inhibitory edge (IE) cells of preferred orientation. Column 5: Border ownership (B) cells of preferred orientation and nonpreferred side of figure. Column 6: Inhibitory border ownership (IB) cells of preferred orientation and side of figure. In the absence of visual input (D), grouping cells of all scales are equally modulated by attention and the modulation perfectly matches the spatial spread of the attention input (E). In the presence of a figure, the attentional field is sharpened to mainly affect excitatory border ownership cells that receive an edge (E) input and have a preferred side of figure consistent with the figure presented. The results quantitatively match the attentional modulation of the responses of border ownership selective neurons (2). If the attentional field is not centered on a figure (F) attentional modulation is repositioned to match the figure’s position. For figures of different sizes (G and H), the attentional field is reshaped to maximally affect the grouping cells of their respective scale. (I) Attentional modulation of edge cells is present for objects which lead to activation of a large number of grouping cells, like an isolated line.
Fig. S4. Auto-localization and sharpening of the attention field in the presence of inhibition (via the IB pathway) between grouping cells. See Fig. S3 for details.
Fig. S5. Effects of figures on a cluttered background. (A) Example of input, edge, and grouping cell activity for a randomized set of horizontal and vertical small line elements that have 0.25 probability of being present at each location in a 12 × 12 lattice. (B) The same as A but a figure is added by activating all line elements in a square (lower left quadrant) that maximally activates one grouping cell of radius 8. (C and D) The same as A and B, but each line element has a 0.5 probability of being present. (E and F) The same as C and D, but an attention signal is added, centered on the figure (Gaussian; gray circle depicts its position and SD). In all rows, the first column represents an example input. The second column shows the enhancement of edge cells activity caused by the feedback from the grouping cells. The third column shows the mean and SD of the maximal edge signal for 100 simulations of each condition. Columns 4, 5, and 6 show examples of activation maps for grouping cells of radius 2, 4, and 8, respectively. The same intensity scale applies to columns 4, 5, and 6. The last column shows the mean and SDs of the maximal activity of grouping cells for the 100 simulations of each condition. Activity in both edge cells and grouping cells of appropriate scale is increased by the presence of a figure for background clutter of 0.25. This is the case to a smaller (but still significant) extent for background clutter of 0.5. Directing attention toward the possible location of the figure elevates the differences in both edge and grouping cell activations that then become similar to those of the low clutter condition.
Table S1. Parameter values

<table>
<thead>
<tr>
<th>Connection</th>
<th>Parameter</th>
<th>Value</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>lE</td>
<td>1*</td>
<td>Input to edge cells</td>
</tr>
<tr>
<td></td>
<td>IT</td>
<td>1/3</td>
<td>Input produced by T junction</td>
</tr>
<tr>
<td>Attention</td>
<td>Gatttw</td>
<td>0.07</td>
<td>Maximal input</td>
</tr>
<tr>
<td>Input</td>
<td>Gatttd</td>
<td>8</td>
<td>SD</td>
</tr>
<tr>
<td>E to B</td>
<td>etobw</td>
<td>1*</td>
<td>Total strength</td>
</tr>
<tr>
<td>B to IE</td>
<td>toiew</td>
<td>2</td>
<td>Total strength</td>
</tr>
<tr>
<td></td>
<td>toieid</td>
<td>2</td>
<td>SD</td>
</tr>
<tr>
<td>IE to B</td>
<td>ietobw</td>
<td>− 2</td>
<td>Total strength</td>
</tr>
<tr>
<td>IE to IE</td>
<td>ietoiew</td>
<td>− 1*</td>
<td>Total strength</td>
</tr>
<tr>
<td>B to IB</td>
<td>btoibw</td>
<td>1*</td>
<td>Total strength</td>
</tr>
<tr>
<td>B to B</td>
<td>btobw</td>
<td>2/3</td>
<td>Total strength</td>
</tr>
<tr>
<td>B to G</td>
<td>btogw</td>
<td>0.25*</td>
<td>Line integral of weights</td>
</tr>
<tr>
<td>IB to G</td>
<td>ibtogw</td>
<td>−0.125</td>
<td>Line integral of weights</td>
</tr>
<tr>
<td>IB to IB</td>
<td>ibtoibw</td>
<td>0</td>
<td>Total weight</td>
</tr>
<tr>
<td>G to B</td>
<td>gtoibw</td>
<td>1/3</td>
<td>Line integral of weights</td>
</tr>
<tr>
<td>G to IE</td>
<td>gtoiew</td>
<td>1/3</td>
<td>Line integral of weights</td>
</tr>
<tr>
<td>G to IE</td>
<td>gtoiwd</td>
<td>4</td>
<td>Reference radius</td>
</tr>
<tr>
<td>G to E</td>
<td>gtoew</td>
<td>2/3</td>
<td>Line integral of weights</td>
</tr>
</tbody>
</table>

*The scaling parameters with values that do not influence the behavior of the network. See text.

Table S2. Sensitivity analysis Sobol indexes for the sensitivity of the border ownership modulation index for unattended figures (BOmod), the attention modulation index for border ownership cells of preferred side of figure (AttPref), and the attention modulation index for border ownership cells of nonpreferred side of figure (AttNonpref)

<table>
<thead>
<tr>
<th>BOmod</th>
<th>AttPref</th>
<th>AttNonpref</th>
</tr>
</thead>
<tbody>
<tr>
<td>gtoibw</td>
<td>btoew</td>
<td>Gatt</td>
</tr>
<tr>
<td>0.14</td>
<td>0.88</td>
<td>0</td>
</tr>
<tr>
<td>0.37</td>
<td>0.42</td>
<td>0.18</td>
</tr>
<tr>
<td>0.04</td>
<td>0.85</td>
<td>0</td>
</tr>
</tbody>
</table>

No significant interactions were observed.