

Neural Dynamics of Grouping and Segmentation Explain Properties of Visual Crowding

Gregory Francis

Purdue University and École Polytechnique
Fédérale de Lausanne

Mauro Manassi

University of California Berkeley

Michael H. Herzog

École Polytechnique Fédérale de Lausanne

Investigations of visual crowding, where a target is difficult to identify because of flanking elements, has largely used a theoretical perspective based on local interactions where flanking elements pool with or substitute for properties of the target. This successful theoretical approach has motivated a wide variety of empirical investigations to identify mechanisms that cause crowding, and it has suggested practical applications to mitigate crowding effects. However, this theoretical approach has been unable to account for a parallel set of findings that crowding is influenced by long-range perceptual grouping effects. When the target and flankers are perceived as part of separate visual groups, crowding tends to be quite weak. Here, we describe how theoretical mechanisms for grouping and segmentation in cortical neural circuits can account for a wide variety of these long-range grouping effects. Building on previous work, we explain how crowding occurs in the model and explain how grouping in the model involves connected boundary signals that represent a key aspect of visual information. We then introduce new circuits that allow nonspecific top-down selection signals to flow along connected boundaries or within a surface contained by boundaries and thereby induce a segmentation that can separate the visual information corresponding to the flankers from the visual information corresponding to the target. When such segmentation occurs, crowding is shown to be weak. We compare the model's behavior to 5 sets of experimental findings on visual crowding and show that the model does a good job explaining the key empirical findings.

Keywords: vernier, crowding, neural network, grouping, segmentation

In visual crowding, perception of a target deteriorates when nearby elements are presented (Flom, Heath, & Takahashi, 1963; Levi, 2008; Pelli, Palomares, & Majaj, 2004; Strasburger, Rentschler, & Jüttner, 2011; Whitney & Levi, 2011). In addition to

being a phenomenon that enables scientists to study the properties of perception and recognition, crowding has practical implications for topics such as reading (Legge, 2007; Pelli et al., 2007), dyslexia (Doron, Manassi, Herzog, & Ahissar, 2015; Gori & Facoetti, 2015), and schizophrenia (Roinishvili et al., 2015). Thus, a model of crowding has potential benefits for both basic and applied sciences. In line with feedforward hierarchical models of object recognition (DiCarlo, Zoccolan, & Rust, 2012; Hubel & Wiesel, 1962; Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007; Thorpe, Delorme, & Van Rullen, 2001), crowding is often explained by pooling or substitution models. In pooling models, neurons with large receptive fields in higher visual areas pool information from lower level neurons with smaller receptive fields (Freeman, Chakravarthi, & Pelli, 2012; Freeman & Simoncelli, 2011; Greenwood, Bex, & Dakin, 2010; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Van Den Berg, Roerdink, & Cornelissen, 2010; Wilkinson, Wilson, & Elleberg, 1997). Because of such pooling, target and flanker signals are combined, thereby impairing target identification. In substitution models, crowding is thought to occur because features of the flankers (or flankers in their entirety) are confused with the target (Ester, Klee, & Awth, 2014; Ester, Zilber, & Serences, 2015; Huckauf & Heller, 2002; Krumhansl & Thomas, 1977; Strasburger, Harvey, & Rentschler, 1991; Zhang, Zhang, Liu, & Yu, 2012), which leads to target misidentification.

This article was published Online First April 24, 2017.

Gregory Francis, Department of Psychological Sciences, Purdue University, and Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne; Mauro Manassi, Department of Psychology, University of California Berkeley; Michael H. Herzog, Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne.

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement n° 604102 (HBP). Mauro Manassi was supported in part by the Swiss National Science Foundation (SNF) fellowship P2ELP3 158876 and the grant “What crowds in crowding?” by the SNF.

All authors contributed in a significant way to the manuscript and have read and approved the final manuscript for submission. Some of these results were presented at the 2016 annual meeting of the Vision Sciences Society.

Correspondence concerning this article should be addressed to Gregory Francis, Department of Psychological Sciences, Purdue University, 703 Third Street, West Lafayette, IN 47907-2004. E-mail: gfrancis@purdue.edu

These standard perspectives on crowding share three common assumptions. The first assumption is that when observers are asked to identify the target, the flankers are essentially treated as noise in the visual system and, hence, increasing flanker strength (by size or number) leads to a predicted increase in crowding strength (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Wilkinson, Wilson, & Elleberg, 1997). The second assumption is that the effect of the flankers on the target reflects local interactions. This assumption is based on the empirical observation that crowding occurs only within a spatial window with the size of half the target eccentricity (Bouma's law); and flankers presented outside this window are presumed to not affect crowding strength (Bouma, 1970; Pelli, 2008; Pelli et al., 2004; Pelli & Tillman, 2008). The third assumption is that crowding occurs only between similar features; for example when the target and flankers have the same color (Kooi, Toet, Tripathy, & Levi, 1994; Pöder, 2007), orientation (Andriessen & Bouma, 1976), shape (Kooi et al., 1994; Nazir, 1992), spatial frequency (Chung, Levi, & Legge, 2001), or global configuration (e.g., faces; Farzin, Rivera, & Whitney, 2009; Louie, Bressler, & Whitney, 2007).

Following the theoretical perspective based on these assumptions, much crowding research has focused on low level properties such as contrast or spacing and has typically presented only single flankers next to the target (e.g., Levi, Hariharan, & Klein, 2002; Levi, Klein, & Hariharan, 2002; Pelli et al., 2004; Strasburger et al., 1991; Toet & Levi, 1992). The data accumulated from these studies were commonly interpreted within a hierarchical framework, where the flanker influences identification of the target by interference with a template matching process (Levi, Klein, & Carney, 2000; Mareschal, Morgan, & Solomon, 2008). Some alternative models interpret crowding effects at the population level (Chaney, Fischer, & Whitney, 2014; Harrison & Bex, 2015; Van den Berg, et al., 2010), but these calculations, although using very different mechanisms, largely reflect the same kind of computational processes. Encouragingly, the effect of low level properties on crowding are readily explained by the standard views and models.

However, some empirical findings in the last decade have shown that the assumptions of the standard perspective do not always hold true and thereby cast doubts about the completeness of standard views of crowding and their corresponding models. In many of these experiments the observer performed a vernier judgment task, by reporting the offset direction of two slightly offset vertical lines. The threshold offset magnitude indicates the sensitivity of the observer to the target vernier, and crowding produces larger thresholds. The findings from these studies are summarized as five key properties of crowding.

1. Levi and Carney (2009); Malania, Herzog, and Westheimer (2007); and Manassi, Sayim, and Herzog (2012) showed that "bigger can be better." Crowding was strong when the vernier was flanked by lines with the same length (e.g., see Figure 1, Condition 1), as vernier offset thresholds strongly increased relative to a vernier in isolation. However, crowding was weak when the target was placed with short (e.g., see Figure 1, Condition 2) or long (e.g., see Figure 1, Condition 3) flankers. Hence, increasing flanker size (i.e., flanker energy) does not always increase crowding. Additional studies found sim-

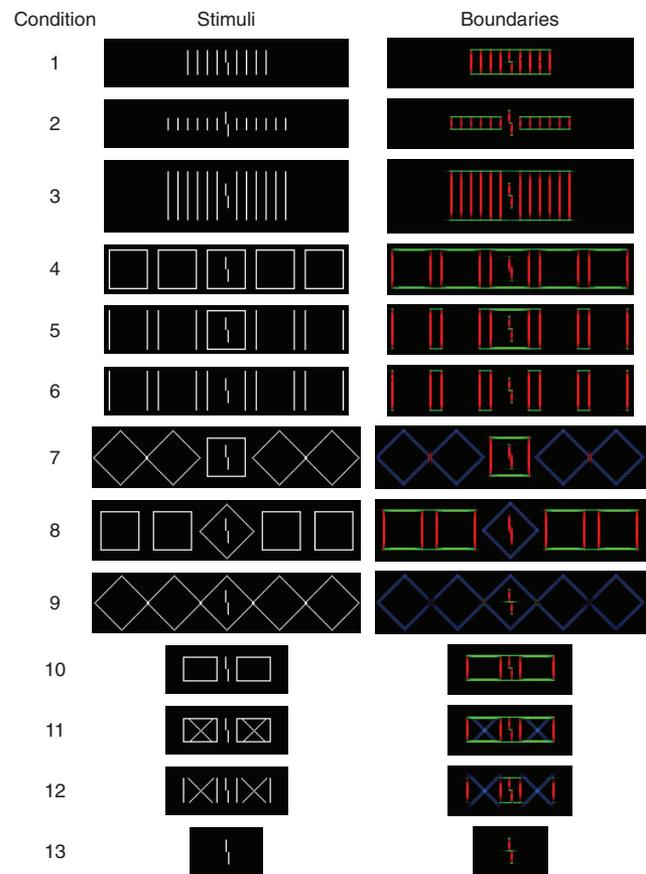


Figure 1. The types of stimuli used in crowding experiments to investigate effects of perceptual organization, and the pattern of boundaries they generate. Each boundary codes the local orientation of an edge (green for horizontal, red for vertical, blue for either oblique). The intensity of the color at each point represents a frequency count of the number of model spikes from bipole neurons during a 50-ms window. Notice the generation of "illusory" horizontal boundaries that connect disparate stimulus elements. The boundaries correspond to perceptual grouping, and a set of connected boundaries forms a group that guides a segmentation process. See the online article for the color version of this figure.

ilar results with Gabor (Saarela, Sayim, Westheimer, & Herzog, 2009; Yeotikar, Khuu, Asper, & Suttle, 2011) and letter stimuli (Saarela, Westheimer, & Herzog, 2010).

2. When increasing the number of equal-length flankers from two to 16, crowding strength remained nearly constant. In contrast, crowding strength decreased when increasing the number of smaller or longer flanking lines (Malania et al., 2007; Manassi et al., 2012; for similar results see Banks, Larson, & Prinzmetal, 1979; Pöder, 2006; Wolford & Chambers, 1983). Again, stronger flankers do not necessarily produce stronger crowding.
3. Manassi et al. (2012) and Sayim et al. (2010) showed that crowding strength depends on the global configuration of the stimulus. Using stimuli that had identical local char-

acteristics (e.g., a vertical line on either side of a target vernier), they found that target identification was substantially better if the vertical line flankers were part of a flanking rectangle (e.g., see Figure 1, Condition 10). Hence, flankers do not seem to be treated as mere noise by the visual system, rather their configural properties strongly matter. In these studies, it seemed to be critical that the flanking stimuli formed a coherent Gestalt to reduce crowding effects.

4. Using stimuli that demonstrated an “uncrowding” effect, Manassi, Sayim, and Herzog (2013) showed that elements outside Bouma’s window can strongly influence crowding strength. For example, when a vernier was embedded in an outline square that is within Bouma’s window, vernier offset discrimination deteriorated: a classic crowding effect. However, crowding decreased when additional flanking squares were presented (e.g., see Figure 1, Condition 4). Importantly, crowding decreased even when the additional flanking squares were presented outside Bouma’s window. Hence, some crowding effects are the result of global interactions across large regions of the visual field (for other global effects see Harrison, Retell, et al., 2013; Harrison, Mattingley, & Remington, 2013; Manassi, Hermens, Francis, & Herzog, 2015; Manassi et al., 2012; Rosen & Pelli, 2015; Sayim, Manassi, & Herzog, 2014; Vickery, Shim, Chakravarthi, Jiang, & Luedeman, 2009).
5. Further investigations into uncrowding revealed that high level similarity strongly influences crowding effects (Manassi, Lonchampt, Clarke, & Herzog, 2016; Manassi et al., 2013). For example, the uncrowding effects in 4) nearly vanished when removing the horizontal lines of flanking squares (e.g., see Figure 1, Condition 6). Hence, uncrowding occurred with square-square interactions but not with line-line interactions. In addition, strong crowding occurred if the target was surrounded by a square and the other flankers were diamonds (e.g., see Figure 1, Condition 7). Rotating those diamonds to be squares produced uncrowding (e.g., see Figure 1, Condition 4). The same results were found in the inverse conditions, that is, with a central flanking diamond and flanking diamonds/squares (e.g., see Figure 1, Conditions 8 and 9). Hence, target-flanker interactions are not only feature specific and are more complicated than previously thought.

These five sets of findings suggest that, contrary to standard perspectives, crowding is strongly influenced by the perceptual organization of the entire visual stimulus. A summary of the earlier findings is that crowding is strong when the target groups with the flankers; and crowding is weak when the target does not group with the flankers (for reviews see Herzog & Manassi, 2015 and Herzog, Sayim, Chicherov, & Manassi, 2015). More generally, crowding strength has been shown to depend on several grouping cues, including target-flanker similarity on various levels (low level: Malania et al., 2007; Manassi et al., 2012; Yeotikar et al., 2011; high level: Farzin et al., 2009; Ikeda, Watanabe, & Ca-

vanagh, 2013; Kimchi & Pirkner, 2015; Louie et al., 2007; Manassi et al., 2013;), good Gestalt (Manassi et al., 2012; Sayim et al., 2010), regularity (Manassi et al., 2012; Rosen & Pelli, 2015; Saarela et al., 2010), contour integration (Chakravarthi & Pelli, 2011; Livne & Sagi, 2007, 2010) and pattern completion (Manassi et al., 2015; Hermens, Scharnowski, & Herzog, 2009). Consistent with the grouping proposal, subjective ratings about the distinctiveness of the target relative to the flankers—a measure of grouping—show good correlations with crowding strength (Malania et al., 2007; Manassi et al., 2012; Saarela et al., 2009; Wolford & Chambers, 1983). If perceptual grouping plays an important role in crowding, then the standard interpretations of crowding (which largely do not consider grouping effects) will be unable to fully account for the properties of crowding or may invoke improper mechanisms to account for psychophysical data (see Herzog & Manassi, 2015; Herzog et al., 2015). Indeed, the standard interpretations cannot explain how figural interactions between multiple flanking squares (Manassi et al., 2013) can influence crowding strength; nor can quantitative models that instantiate these standard interpretations explain psychophysical data that explore the effects of perceptual grouping on crowding. More generally, despite vigorous effort, there is currently no model that can account for the all of the earlier effects of perceptual grouping on crowding (see also Agaoglu & Chung, 2017; Clarke et al., 2014; Harrison & Bex, 2016; Manassi et al., 2015; Pachai, Doerig, & Herzog, 2016).

A promising approach, but one that is not considered here, supposes that crowding effects are the result of how visual information is represented among summary statistics of a visual scene (e.g., Balas, Nakano, & Rosenholtz, 2009; Keshvari & Rosenholtz, 2016). Unlike the standard interpretations of crowding, a model based on summary statistics considers many disparate elements of a scene, and so in principle might be able to account for the earlier effects. However, there are currently no model simulations demonstrating that these kinds of models can account for these data sets.

In this article we describe a neural network model of visual perception that uses feedback to generate perceptual groupings among responses to visual stimuli. A decision mechanism based on template matching implements the classic ideas of crowding, where the flankers reduce the sensitivity of the template comparison mechanism. In a way that fundamentally differs from the classic ideas of crowding, we show that the model’s groupings can support a segmentation process whereby an observer uses top-down signals to alter the cortical representation of visual information in a way that sometimes enables isolation of the signals that correspond to the target. Such isolation can free the target identification process from the crowding effects of the flankers. We show through computer simulations that the model’s behavior is quite similar to that of human observers for the same kinds of stimuli. In particular, it accounts for the five effects summarized in the preceding text.

Model Description

Figure 2 schematizes the stages of the model that are based on previous work (Cao & Grossberg, 2005; Raizada & Grossberg, 2001), which is hypothesized to exist in areas V1 and V2 of visual cortex. Because the model is an extension of previous work, not every model property plays a fundamental role in explaining

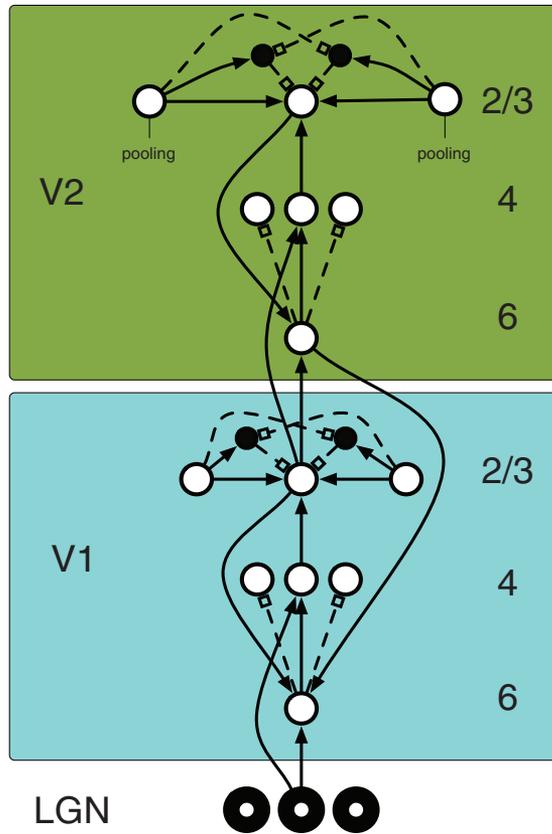


Figure 2. A schematic of the model circuits for processing oriented boundary information from a visual scene. Circles indicate model neurons, with the central column of circles indicating neurons at a single retinotopic position. Solid lines indicate excitation, and dashed lines indicate inhibition. The central top circle indicates a bipole neuron, which plays an important role in perceptual grouping. See the online article for the color version of this figure.

crowding effects. The present discussion will emphasize the model properties (old and new) that are necessary to understand grouping, segmentation, and crowding.

Grouping

The circuits schematized in Figure 2 correspond to previous versions of the model, known as LAMINART because it explains how visual information processing is performed using neural computations among the laminar layers of visual cortex (Raizada & Grossberg, 2001). Previous work has used this model to explain properties of perceptual grouping, illusory contours, and texture discrimination, among others (Bhatt, Carpenter, & Grossberg, 2007; Grossberg, 2014) by demonstrating how the visual cortex represents oriented boundaries, which generalize contrast edges. The computations in V1 are similar to those in V2, but at a smaller spatial scale, so the present discussion will focus only on the V2 computations. The schematized V2 neurons represent oriented boundaries (edges) at specific retinotopic locations. Input from V1 complex cells arrive in Layer 6 and Layer 4, which feed into a circuit for lateral spatial grouping at Layer 2/3. A “bipole” neuron

in Layer 2/3 gathers input from neurons with similar orientation preferences that are arranged along the direction of the preferred orientation (Grossberg & Mingolla, 1985a); this pooling process is similar to the “association field” used in other models (e.g., Field, Hayes, & Hess, 1993). At each retinotopic position, a bipole neuron receives excitatory signals from pooling neurons that gather signals from opposite sides of the neuron’s retinotopic position. By themselves, such excitatory signals would lead to a run-away spreading of oriented boundaries across visual space. To prevent this problem, the pooling neurons simultaneously send an excitatory projection to an interneuron associated with the bipole neuron and an inhibitory projection to another associated interneuron (the black-filled circles in Figure 2). When active, either interneuron inhibits the bipole neuron. The net effect of this circuit is that when both pooling neurons send similarly strong signals, the excitatory and inhibitory inputs to each interneuron are nearly balanced and so the interneurons do not inhibit the bipole neuron. In contrast, if one pooling neuron sends a much stronger signal than the other pooling neuron, then one of the interneurons receives strong excitation and weak inhibition. This interneuron will then inhibit the bipole neuron. The net effect is that the circuit allows interpolation between boundary signals of appropriate orientations and relative positions, but prevents extrapolation of boundaries from an isolated boundary signal (Grossberg & Mingolla, 1985a, 1985b; von der Heydt, Peterhans, & Baumgartner, 1984). As described subsequently, this property of boundary processing plays an important role in perceptual grouping of visual stimuli.

To demonstrate the network’s boundary grouping, Figure 1 shows a variety of stimuli that have been used in crowding experiments and the pattern of responses from oriented boundaries generated by the circuits in Figure 2. The images on the right show a retinotopic representation of oriented boundary signals with a colored pixel at each position where a bipole neuron sensitive to an oriented boundary at that location is active. The intensity of the color at each position indicates the number of spikes generated by the model neuron over a 50-ms period. Different colors are used to code different boundary orientations (red for vertical, green for horizontal, and blue for either 45-degree oblique). The majority of boundaries simply reflect the bottom-up luminance signals from the stimulus, but close inspection reveals horizontal (green) boundaries that connect boundaries generated by physically separate stimuli. For example, in the top image, the line ends of the flankers and target are connected by horizontal boundaries. These illusory contours are the result of the V2 pooling circuits in Figure 2.

Different boundary patterns appear for other types of stimuli. For example, the stimulus in the second row of Figure 1 uses flankers that are half the size of the target. The horizontal boundaries generated among the flankers on the left side of the target fully connect the left side flankers but do not connect with the target or the right side flankers. As described in detail subsequently, a set of connected boundaries corresponds to a perceptual group. Thus, here the model representation indicates that there are three distinct boundary groups: the left flankers, the right flankers, and the target by itself. One sees a somewhat similar pattern for the third row stimulus (long flankers), although the two sets of flankers on either side become connected and thereby form one group that is distinct from the target by itself. Similarly, the sets of squares and diamonds for the stimuli in rows 4 through 9 of Figure 1 form

boundary groups that are disconnected from the target, but the boundaries for the flanking elements in these conditions group together in different ways. The boundaries of flanking elements in rows 10 through 12 group together and with the target's boundaries. The bottom stimulus is the target vernier by itself, which is itself a fully connected group of boundaries, despite the small offset of the line elements.

Template Matching and Crowding

The observer's task in every crowding experiment considered here is to judge the offset direction of a target vernier. Such a task could conceivably be done with a variety of mechanisms, but we opted for a template matching process that contrasts signals indicating leftward and rightward shifted verniers. As schematized in Figure 3 the templates are fairly large (which offers advantages if the template is not centered directly on the target vernier and the offset is large enough). Figure 3A demonstrates how templates for a leftward target and a rightward target are positioned relative to the boundaries generated by an isolated target.

Each template sums the neural responses within its coverage area to produce, V_{left} and V_{right} . The evidence that the vernier is shifted to the right is then the contrast of these summed values:

$$E_{\text{right}} = \frac{V_{\text{right}} - V_{\text{left}}}{100 + V_{\text{right}} + V_{\text{left}}} \quad (1)$$

where the value 100 avoids division by zero and scales the evidence magnitude without changing relative order. Larger E_{right} values correspond to more evidence that the vernier is shifted to the right and correspond to a smaller offset threshold in behavioral experiments. For a vernier by itself, $V_{\text{right}} \approx 100$ (due to the

dynamical nature of the model the value changes over time) and $V_{\text{left}} = 0$, so $E_{\text{right}} \approx 0.5$.

Consistent with classic perspectives, crowding occurs when flanking elements contribute to the template calculation. Figure 3B shows that signals from the equal sized flankers contribute to the right shift template. The flankers also contribute to the left shift template, which changes the ratio for the evidence calculation. For the case shown in Figure 3B, $V_{\text{right}} \approx 270$ and $V_{\text{left}} \approx 180$, so $E_{\text{right}} \approx 0.16$. This value is relatively small compared with the 0.5 produced for an isolated vernier, and this decrease in E_{right} corresponds to crowding that is introduced by the presence of the flankers.

Because the target was always the same, we used one fixed set of templates for all model simulations. Although mostly satisfactory for the present purposes, we are not wedded to the details of this template comparison process. Similar results would be produced by supposing that oriented detectors are used to discriminate between a leftward or rightward shifted vernier (e.g., Klein & Levi, 1985; Wilson, 1986) or by pooling across a population distribution of detectors (Harrison & Bex, 2015).

Although our theoretical perspective is that grouping determines crowding effects, boundary grouping by itself does not reduce crowding in the template calculations. For example, Figure 3C shows the template superimposed on the boundaries generated by a target with long flankers. The boundaries corresponding to the target do not connect with the boundaries corresponding to the flanker, but the flanker signals do contribute to the template calculations and would correspond to strong crowding (Manassi et al., 2015). We argue in the next section that a segmentation process is needed to capitalize on the boundary groupings in a way that separates the signals of the target and flankers.

Segmentation for Boundary Groups

The main model development presented here is a segmentation process that utilizes boundary groupings to alter the representation of visual information and thereby modify crowding effects. This segmentation process allows nonspecific, observer-guided, signals to select and segment elements of a visual scene. At a conceptual level, this selection process functions similar to the hypothesized attentional gate that is common to many theories of visual perception (e.g., Reeves & Sperling, 1986). Figure 4 schematizes the circuits that implement the segmentation computations. The box labeled "Segmentation Layer 0" corresponds to the V2 boundary processing circuit in Figure 2. The other segmentation layer boxes indicate copies of the V2 circuits that are used to represent different parts of a visual scene. In the context of the vernier discrimination task, the purpose of the segmentation process is to shift boundaries of the flankers to different segmentation layers and thereby isolate the boundaries that correspond to the target vernier within one of the segmentation layers. With such isolation, the template matching process can best discriminate between targets with different properties (e.g., vernier shift direction).

Implementing a segmentation process with a plausible neural circuit requires a complex pattern of excitation, inhibition, and disinhibition between multiple stages. The arrows between the boxes in Figure 4 indicate retinotopic projections such that a neuron at a given position and with a given orientation preference in, say V1, projects to a corresponding neuron in V2 with an

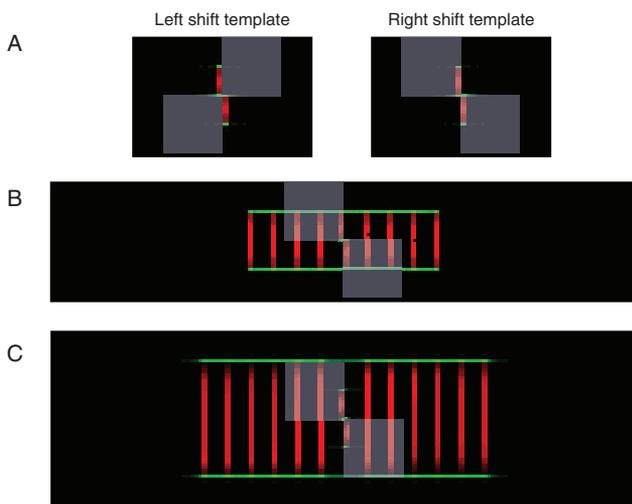


Figure 3. A schematization of the templates (purple rectangles) used to judge whether a target vernier is shifted to the left or the right. Signals within the template are summed. Panel A shows the two templates superimposed on the boundaries generated by an isolated vernier. Panel B shows the right shift template superimposed on the boundaries generated by a vernier with equal sized flanking lines. Panel C shows the right shift template superimposed on the boundaries generated by a vernier with long flanking lines. See the online article for the color version of this figure.

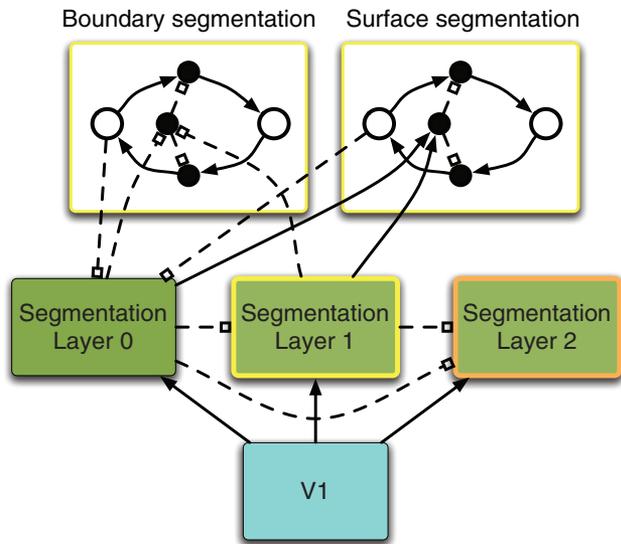


Figure 4. A schematic of the model circuits for boundary and surface segmentation. The different segmentation layers correspond to multiple copies of the V2 boundary processing circuits. See the online article for the color version of this figure.

equivalent position and orientation preference. As Figure 4 indicates, signals from V1 project to each of the segmentation layers. (For the current simulations, we suppose that there are three such layers.) However, this redundancy is checked by strong inhibition from Segmentation Layer 0 to the other two segmentation layers. Thus, the V1 boundary signals are initially only represented at Segmentation Layer 0, where they undergo the boundary grouping processes described earlier and demonstrated in Figure 1.

These boundaries then interact with nonspecific, observer-guided, selection signals at the boundary and surface segmentation stages (only the circuits for Segmentation Layer 1 are schematized in Figure 4, Segmentation Layer 2 has similar circuits). Consider the boundary segmentation circuit first. Here a retinotopic representation of the visual scene can receive selection signals that spread to neighbors. As schematized in the circuit in Figure 4 for a pair of neurons at retinotopic locations corresponding to horizontal neighbors (open circles), the spreading of the selection signals pass through interneurons (solid circles) before reaching their target neighbor. By default, these interneurons are strongly inhibited by a tonically active interneuron (the middle black circle in the boundary segmentation circuit). This tonic inhibition means that, by default, a selection signal at one location cannot spread to a neighbor because the necessary interneurons are inhibited. However, boundary signals from any of the segmentation layers at the same retinotopic location inhibit the tonically active interneuron, and this input disinhibits the interneurons that thereby spread the selection signal to neighboring neurons. Thus, when a boundary signal is present at a given location, a selection signal at that location can spread to its neighbor. Functionally, the boundary segmentation circuit enables the selection signal to flow along positions where oriented boundaries are active. Thus, a selection signal can flow along a set of connected boundaries, but will not jump across a spatial gap in boundaries. Intuitively, the segmen-

tation circuits instantiate the common idea that attention effects flow across groups of objects along connected contours (Grossberg & Raizada, 2000; Raizada & Grossberg, 2003; Roelfsema, Lamme, & Spekreijse, 1998).

To shift boundary signals, active selection signals at the boundary segmentation stage for Segmentation Layer 1 inhibit the boundaries at Segmentation Layer 0 at the same retinotopic position. The inhibition of the boundaries at Segmentation Layer 0 leads to disinhibition of the V1 to Segmentation Layer 1 signals that were initially suppressed. Thus, the boundary segmentation process suppresses selected boundaries at Segmentation Layer 0 and shifts those boundaries to be represented in Segmentation Layer 1.

Essentially the same boundary segmentation circuits exist for Segmentation Layer 2, so grouped (connected) boundaries in Segmentation Layer 0 or Segmentation Layer 1 can be selected, suppressed, and shifted to Segmentation Layer 2. As shown subsequently, an observer may have to carefully strategize how to place selection signals to maximize performance for a given task with given stimuli.

Figure 5 shows the dynamics of how the segmentation circuits select boundaries corresponding to separate groups in the scene and encodes those groups with distinct representations in the different segmentation layers. The stimulus is the target vernier with five half-length flankers on each side. At 150 ms after stimulus onset, Segmentation Layer 0 has detected and grouped the boundaries that correspond to the stimulus. These boundaries largely inhibit corresponding boundaries that would otherwise appear in the other segmentation layers.

The colored transparent disks on the stimulus in Figure 5 indicate the location of two separate selection signals; focus on the yellow (left side) disk first. For this stimulus, the yellow selection signal overlaps with boundaries generated by the flanking elements to the left of the target vernier. Because the boundaries for the left side flanking elements group together, the selection signal at the boundary segmentation stage spreads across the entire group of boundaries for the left side flankers. The selection signals from the boundary segmentation stage then inhibit the corresponding boundaries in Segmentation Layer 0, which disinhibits the V1 inputs to Segmentation Layer 1 and allows the grouped boundaries for the flankers to the left of the target to be represented in Segmentation Layer 1. The development of this representation can be seen by moving down (increasing time) in Figure 5. The functional behavior is that the selection signal spreads across a boundary group and opens an attentional gate to shift the grouped boundaries to its corresponding segmentation layer. Essentially the same story applies for the (orange) selection signal focused on the flankers to the right of the target vernier, except the boundaries are shifted to Segmentation Layer 2.

Importantly, once the flanker boundary groups have been removed from Segmentation Layer 0, the target vernier is represented by itself. Thus, when the templates in Figure 3 are used to discriminate the vernier offset, they do so without the influence of signals from the flanking elements, and vernier discrimination is nearly as good ($E_{\text{right}} \approx 0.45$) as for a physically isolated vernier. Figure 6A shows the time course of evidence for a right-shifted vernier for the simulation in Figure 5. The evidence for the signals in Layers 1 and 2 has only small values (because flanker signals contribute to both the left-shifted and right-shifted templates),

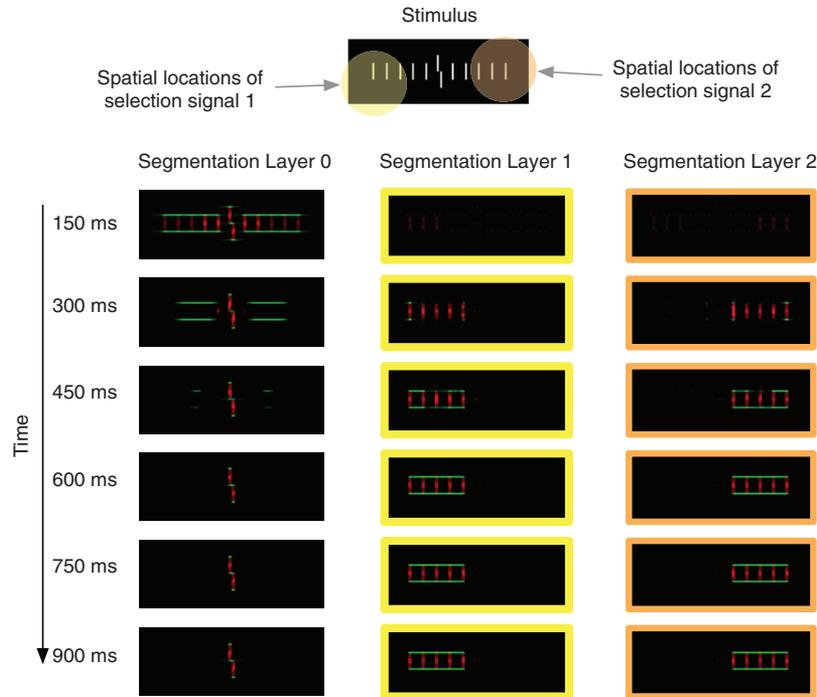


Figure 5. The dynamics of the boundary segmentation process for a target vernier with five half-length flanking lines on either side. The colored disks on the stimulus indicate selection signals that spread across connected boundaries and shift boundary input from Segmentation Layer 0 to the other segmentation layers. By 600 ms after stimulus onset, the boundaries of the target vernier becomes isolated in Segmentation Layer 0, which enables fine discrimination of the vernier offset. See the online article for the color version of this figure.

whereas the evidence for the signals in Layer 1 rise to around one half. Although it takes approximately 350 ms for the segmentation process to finish, a reliable judgment about vernier direction could be made before the segmentation has stabilized.

Note that the selection signals do not have to be precisely placed to have the desired effect for the stimuli in Figure 5. As long as a selection signal overlaps with some of the flanker-generated boundaries, the entire (connected) set of boundaries will be selected and shifted to the corresponding segmentation layer. In this sense, the selection signals can be nonspecific. It is not necessary for the selection signal to precisely pick out the boundaries that correspond to the flankers, the boundary grouping mechanisms in area V2 provide the desired selective precision.

Figure 7 shows a second example of successful segmentation when the flankers consist of a surrounding square and two additional squares on either side. As in Figure 1 (fourth row), the boundaries of the five squares group together while the boundaries of the target vernier are separate. Selection signals to the left and right of the target spread across the grouped boundaries and shift them to a different segmentation layer, in this case to Segmentation Layer 2. This shift leaves the boundaries of the target essentially isolated, which means the template discrimination proceeds with little crowding effect from the flankers ($E_{\text{right}} \approx 0.46$).

In crowding experiments, the goal of the observer is to identify the direction of the vernier offset. To accomplish that goal in the framework of the model, the observer attempts to place the selection signals in such a way as to isolate the signals generated by the

target vernier at one of the segmentation layers. For the current model simulations, we assume that the size of the selection signal is fixed (at the size indicated in Figure 5) and that there is some jitter in the placement of the selection signals so that the observer cannot perfectly place the signals. This imprecision in placement means that sometimes the segmentation process will fail to isolate the target; and two notable example failures are given in Figures 8 and 9.

In Figure 8 the target is flanked by a long line on each side. The observer tries to segment out the flanking lines and thereby leave the target signals by themselves in Segmentation Layer 0. However, because the flanking lines are so close to the target, the selection signals for the flanking lines are misplaced and cover part of the target as well as the flankers. Once the selection signals spread across all connected boundaries, the flankers and the target are both shifted to Segmentation Layer 2. The result is that the templates applied to Segmentation Layer 2 demonstrate strong crowding due to the presence of the flanking lines ($E_{\text{right}} \approx 0.16$). The simulation summarized in Figure 8 demonstrates that it is not sufficient for visual elements to be separately grouped (as defined by their boundaries being disconnected). To avoid crowding, such grouping effects must enable distinct segmentations of the flankers and the target. On some simulated trials, the segmentation process works properly for this stimulus because the signals overlap only the boundaries of the flankers (e.g., at the top and bottom of the flankers, which extend beyond the target vernier).

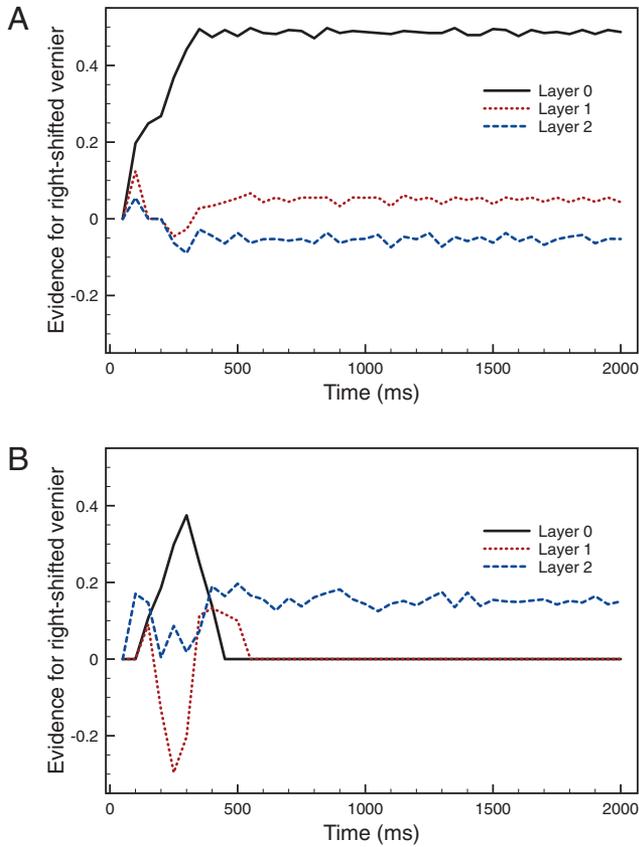


Figure 6. The time course of evidence for a right-shifted vernier at the different segmentation layers of the model. (A) When the flanking targets are half the length of the vernier target, segmentation layer 0 quickly shows the strongest evidence. (B) When the flanking targets are the same length as the vernier target, Layers 1 and 2 are inhibited by the segmentation shift of signals to Layer 2; but Layer 2 shifts both the target and the flanker signals so the evidence value is relatively weak (crowding occurs). See the online article for the color version of this figure.

In **Figure 9** the target is flanked by five equally sized lines on each side. As indicated in **Figure 1** (row 1), the target and equal length flankers form a single group of connected boundaries. The observer might try to segment out the flanking elements, but the selection signals spread across connected boundaries and shift them to a new segmentation layer. For the case shown in **Figure 9**, most of the boundaries are shifted to Segmentation Layer 2, but a few boundaries get shifted to Segmentation Layer 1. When the discrimination templates are applied at Segmentation Layer 2, strong crowding will occur because the flanking elements contribute to the template calculations ($E_{\text{right}} \approx 0.16$). The simulation summarized in **Figure 9** demonstrates that crowding occurs when the target boundaries group with the flanker boundaries, because such grouping prohibits the boundary segmentation mechanisms from isolating the target signals. **Figure 6B** shows how the evidence for a right-shifted vernier varies across the three segmentation layers as a function of time for the simulation summarized in **Figure 9**.

Segmentation for Surface Regions

Other attributes of the visual scene can also guide segmentation. Here we describe a situation where segmentation is guided by spreading selection signals across a closed surface rather than across connected boundaries. This mechanism is similar to proposed attention spreading within an object's surface (e.g., **Richard, Lee, & Vecera, 2008; Zhao, Kong, & Wang, 2013**) and the concept of an attentional shroud (**Foley, Grossberg, & Mingolla, 2012; Tyler & Kontsevich, 1995**). The circuits responsible for such guidance are very similar to those used to spread the selection signals across a group of connected boundaries, and **Figure 4** schematizes the circuits for Segmentation Layer 1. Just as for boundary segmentation, surface segmentation involves spreading a selection signal to neighboring neurons in retinotopic space. Whereas for the boundary segmentation circuit the tonic activity of the inhibitory interneuron prevents the spread of a selection signal, the inhibitory interneuron for the surface segmentation circuit is not tonically active. Thus, by default, surface selection signals easily spread (through the appropriate interneurons) to spatial neighbors; but boundary signals from V2 at the same position excite the inhibitory interneuron and thereby prevent selection signals from spatially crossing an active boundary. In this way, selection signals can become trapped within a closed set of connected boundaries. An additional "pruning" signal inhibits selection signals that fall outside a closed set of connected boundaries (this pruning process is not schematized in **Figure 4**). Just as for the boundary segmentation circuit, selection signals within the surface segmentation circuit for Segmentation Layer 1 inhibit boundaries at corresponding retinotopic locations in Segmentation Layer 0, and this inhibition shifts the boundaries from Segmentation Layer 0 to Segmentation Layer 1. A similar circuit allows a second selection signal to shift boundaries to Segmentation Layer 2.

Figure 10 summarizes a simulation where the target vernier is flanked by two rectangles. As row 10 of **Figure 1** indicates, the boundaries of the flanking rectangles group with the boundaries of the target vernier, and therefore attempting to use the boundary segmentation process would select boundaries for both the target and flankers. However, the surface segmentation process can spread selection signals (colored disks superimposed on the stimulus) across the interior regions of the flanking rectangles, and these signals selectively inhibit the boundaries at Segmentation Layer 0 that correspond to the flankers and thereby disinhibit the V1 inputs at the other segmentation layers. The net effect is that the boundaries of the flankers are shifted to Segmentation Layers 1 and 2. A side effect of the shift is that the grouping between the target vernier boundaries and the flanker boundaries is broken (as boundary grouping can only occur within a given segmentation layer). Ultimately, the surface segmentation process isolates the boundaries of the target so that the flankers cause only modest crowding ($E_{\text{right}} \approx 0.49$).

Just as for segmentation by boundary groups, the surface segmentation process can fail when the selection signals are misplaced. **Figure 11** shows an example failure where the right side selection signal barely misses the right side flanking rectangle, so the boundaries of the right side rectangle remain with the boundaries of the target vernier in Segmentation Layer 0 and thereby produce some crowding ($E_{\text{right}} \approx 0.25$). An observer might try to insure that the selection signals cover the flanking rectangles by

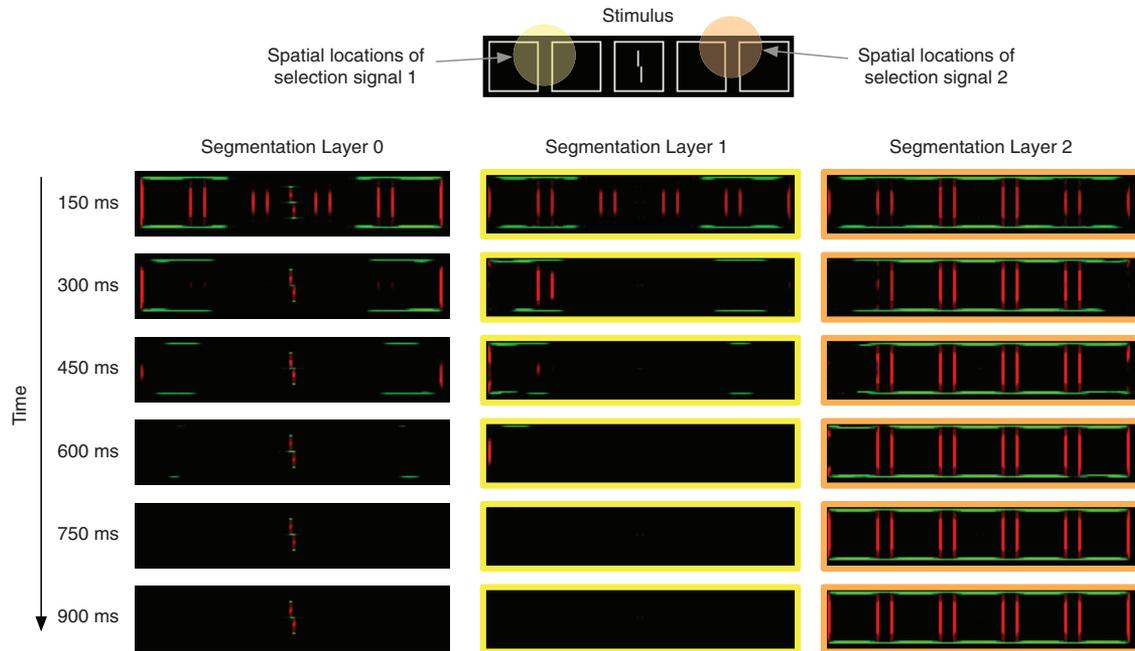


Figure 7. The dynamics of the boundary segmentation process for a target vernier with a surrounding and four flanking squares. The colored disks on the stimulus indicate selection signals that spread across connected boundaries and shift boundary input from Segmentation Layer 0 to the other segmentation layers. In this way, the target vernier becomes isolated in Segmentation Layer 0, which enables fine discrimination of the vernier offset. See the online article for the color version of this figure.

aiming for the middle of each rectangle, but then there is a risk that the selection signal accidentally overlaps the target vernier and pulls the target vernier boundaries into one of the segmentation layers (as in Figure 8), which also produces strong crowding effects.

As for the boundary segmentation process, the surface selection signals are nonspecific. The signals need only overlap the desired surface and the structure of the boundaries constrains the spread of the signals to that surface. In this way, the selection circuits take advantage of the visual information processing provided by the V2 boundary circuits.

Comparisons With Experimental Data

When an observer participates in a crowding experiment, we hypothesize that he or she tries different segmentation strategies in an effort to isolate the target vernier and thereby reduce crowding effects. This exploration process involves identifying whether boundary group segmentations or surface segmentations are effective and learning where to place the selection signals so as to capture the flanker boundaries without also accidentally grabbing the target vernier boundaries. In the simulations presented subsequently, we assumed that observers used the best possible strategy to guide the segmentation process in a way that reduces crowding effects.

Each stimulus condition was simulated 20 times with random variations added to the placement of the selection signals. For each stimulus, pilot simulations explored the best pixel coordinates, (p_1, p_2) , to place the center of each selection signal in order to maximize discrimination of the target vernier. On a given trial, the actual

placement of the center of a selection signal was $(p_1 + \varepsilon_1, p_2 + \varepsilon_2)$, where the ε terms were randomly drawn from a normal distribution with mean zero and standard deviation of 10. Segmentation signals were then generated in a circle with a radius of 20 pixels around $(p_1 + \varepsilon_1, p_2 + \varepsilon_2)$. These selection signals were initiated 100 ms after stimulus onset. None of these choices seem critical for the simulations we report, but future work should more carefully identify the appropriate size and shape of the segmentation signal area, the precision with which the signals can be placed, and the time at which they appear. Future work might also fruitfully explore whether these parameters are manipulable by the observer.

Each stimulus was presented for two simulated seconds, which allowed us to fully explore the neural dynamics of the model (crowding effects are generally unaffected by stimulus duration [e.g., Wallace, Chiu, Nandy, & Tjan, 2013]). Every 50 ms, the simulation program summed neural action potentials from the cells that fed into the templates and computed an evidence value as described in Equation 1. The average E_{right} value across the full two seconds of stimulus presentation was then taken as evidence for the vernier being shifted to the right (in the simulations all targets were shifted to the right). These calculations were computed independently for each segmentation layer, and the final evidence value was the maximum E_{right} across all three segmentation layers. We call this value E_{right}^* and refer to it as the “model evidence.”

Every neuron was modeled as an integrate-and-fire neuron and every synapse was modeled as a static synapse in the NEST software program (Gewaltig & Diesmann, 2007). The simulations were run on the Radon compute cluster at Purdue University.

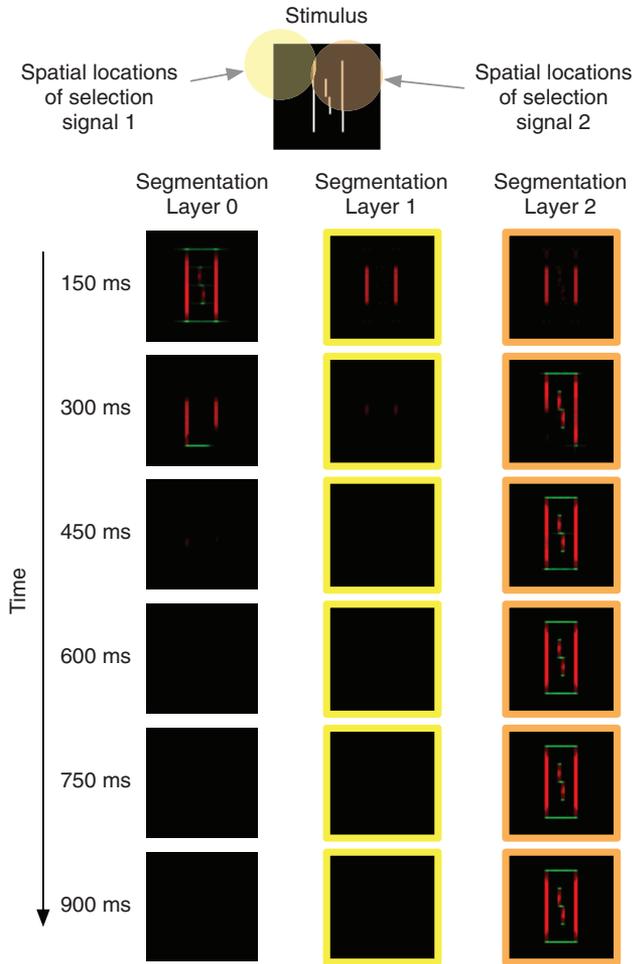


Figure 8. The dynamics of a failed boundary segmentation process for a target vernier with two flanking long lines. The colored disks on the stimulus indicate selection signals that spread across connected boundaries and shift boundary input from Segmentation Layer 0 to the other segmentation layers. Because of imprecise placement of the selection signals, the flankers and target are all shifted to Segmentation Layer 2, so the template matching process experiences crowding. See the online article for the color version of this figure.

Radon consists of 45 HP Moonshot compute nodes with 32 GB RAM. Each of the 42 stimulus conditions were simulated with a single node, which has 8 cores that can be used in parallel by the NEST simulation. Because the Radon system is for general academic use, the simulations were bundled into 15 parallel batches of two to four stimulus conditions such that conditions in a batch ran serially on a given compute node. Running the full set of simulations took almost 2 days from start to finish on this system.

To ease the computational requirements, the simulated image plane was adjusted to the size of the flanking stimuli, and the size of the network model was correspondingly adjusted. For example, the smallest simulation was for the vernier presented by itself. This simulation was 40×60 pixels and the network consisted of nearly 1 million neurons (including simulated input and recording devices) and 5.5 million synapses. In NEST, the model network has to be reconstructed for every simulation and this smallest simula-

tion took nearly 30 min to define (mostly to set up the synapses between neurons). Once established, a 2-s trial (with an additional one second intertrial interval) took approximately 6 min. The full 20 trials for this condition took around 2 hours.

The simulated neural network was much larger for other stimulus conditions. The largest, for the seven flanking squares condition (Figure 16B), used an image plane of 280×50 pixels. The corresponding neural network consisted of approximately 5.7 million neurons (including simulated input and recording devices) and nearly 33 million synapses. It took nearly 5 hr to define a network of this size in NEST (again, most of the time was spent defining the synapses), and once the network was defined each trial took approximately 40 min to simulate. Simulating the full set of 20 trials took approximately 14 hr.

The NEST code (written in Python) to reproduce all the findings reported in this article is available at the Open Science Framework (https://osf.io/4fhxs/?view_only=a1d4cee6f6514b3da9e8bf40c132b085). Random variation in the ϵ terms will mean that a new run of the code will produce slightly different results than what are reported here, but the main effects should be robust.

In addition to the location, size, and accuracy of the segmentation signals, the parameters of the model are the synaptic weights between various neurons. Not including synapses defined to emulate orientation filters for model V1 simple cells (defined by oriented Gabor functions), there were 38 synapse types defined in the model. Many of these synapse types simply passed information from one cortical area to another (e.g., the same synapse type was used to connect V1 Layer 2/3 complex cells to V2 Layer 4 and V2 Layer 6 complex cells), and the main constraint on the weight magnitude was to ensure that the signal was strong enough to evoke a sufficient response at the target layer. For other synapse types, the sign (excitatory or inhibitory) and magnitude of a weight was chosen to produce functional model behaviors such as boundary grouping, spreading of selection signals along connection boundaries, and shifting of boundaries from one segmentation layer to another. The relevant constraints were generally to insure that inhibitory signals were sufficiently strong to prevent activity in appropriate layers. For example, as schematized in Figure 2, neurons in Segmentation Layer 0 inhibit neurons in Segmentation Layer 1, and the strength of these inhibitory signals needed to be strong enough to overwhelm excitatory signals from V1 that would otherwise establish a boundary representation in Segmentation Layer 1. Likewise, inhibitory feedback from the Boundary segmentation layer to Segmentation Layer 0 needs to be strong enough to overwhelm the excitatory V1 to Segmentation Layer 0 signals. Similar requirements constrained the synapse weights in circuits for boundary grouping and segmentation signal spreading. Usually, there was a broad range of weight values that would produce similar network behaviors; no effort was made to adjust synaptic weights to match empirical data. Changes to the synaptic weights would surely affect the overall magnitude of crowding, but they are unlikely to alter the main effects of grouping and segmentation unless those functional model properties are lost.

In the following sections, we describe how the model behavior compares with measurements from psychophysical studies of crowding. The psychophysical studies report two related performance measures. Many studies simply report the mean vernier offset threshold that is needed for observers to be 75% correct. Larger threshold values correspond to stronger crowding. We

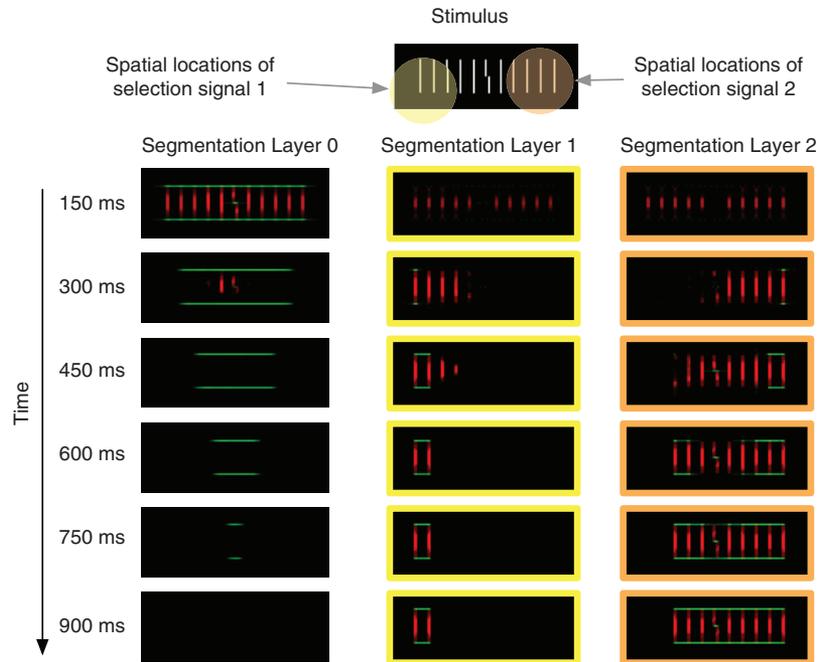


Figure 9. The dynamics of a failed boundary segmentation process for a target vernier with five flanking equal length lines on each side. The colored disks on the stimulus indicate selection signals that spread across connected boundaries and shift boundary input from Segmentation Layer 0 to the other segmentation layers. Most the boundaries get shifted to Segmentation Layer 2, where the joint representation of the target and flanker boundaries leads to strong crowding. See the online article for the color version of this figure.

propose that larger model evidence values, E_{right}^* , should correspond to smaller vernier thresholds. Rather than risk overfitting the empirical data with a quantitative mapping between model evidence and thresholds, we simply plot E_{right}^* on a reverse scale when considering empirical findings that report thresholds.

A second measure from the psychophysical studies is a ratio that indicates the “threshold elevation” relative to an isolated vernier. A threshold elevation value of one indicates that the measured threshold is the same as the threshold for an isolated vernier (no crowding). Larger threshold elevation values indicate stronger crowding. Because larger model evidence is related to smaller thresholds, we computed “evidence elevation” by taking the ratio of evidence for an isolated vernier divided by the evidence for another condition. Because crowding will produce smaller model evidence values, larger evidence elevation values correspond to stronger crowding.

Before turning to a detailed discussion of empirical data and the model’s behavior, we want to clarify the model’s explanatory scope. We do not anticipate a perfect fit between the model’s behavior and human performance. Discrepancies are expected for many reasons, including (a) the model is obviously incomplete (e.g., it lacks a retina and does not take into account differences between the fovea and periphery), (b) crowding may involve many different mechanisms that are not fully understood or included in the model, (c) empirical measures of crowding effects are often quite variable because of differences in stimuli, tasks, stimulus placement, and random sampling of observers. Where possible, we do report variability in the model’s behavior across trials as error bars (standard error of the mean). However, it is important to

recognize that these error bars are generally not directly related to the error bars reported in corresponding empirical studies. The latter describe variability in mean performance across observers, while the error bars from the simulations describe variability for a simulation with a fixed set of parameters (e.g., a single observer).

Our goal is not to fully explain crowding effects, but to demonstrate that the model’s mechanisms for perceptual grouping and segmentation enable it to account for the empirical findings that are related to perceptual organization. We anticipate that there is much room for model development to account for many properties of crowding, but that grouping and segmentation play an important role in many situations. The following sections discuss examples of the five effects described in the introduction.

Size of Flankers

Crowding effects are greatly reduced when the length of the flanking lines are shorter or longer than the target vernier (Malania et al., 2007; Manassi et al., 2012). Figure 12A shows representative results from Manassi et al. (2012). The horizontal dashed line indicates the offset threshold for a vernier in isolation (no flankers). The y-axis plots “threshold elevation,” the ratio of the offset thresholds for the various flanking conditions against the threshold for the target by itself. Larger threshold elevations indicate that the target vernier offset had to be larger. When the flanking elements are the same height as the target vernier, the threshold is large, thereby indicating crowding. When the flanking elements are a different size than the target, the threshold is modestly or hardly increased relative to the no flankers condition. These results are

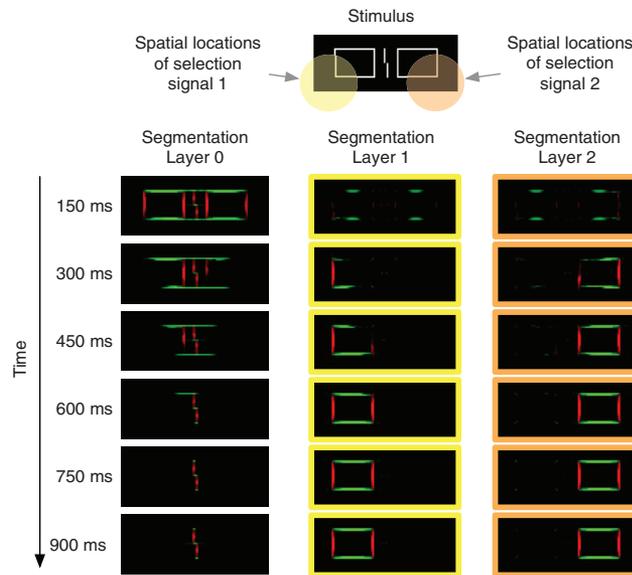


Figure 10. The dynamics of the surface segmentation process for a target vernier with flanking rectangles. The colored disks on the stimulus indicate selection signals that spread across the closed surfaces defined by the rectangles and shift boundary inputs from Segmentation Layer 0 to the other segmentation layers. In this way, the target vernier becomes isolated in Segmentation Layer 0, which should enable fine discrimination of the vernier offset. See the online article for the color version of this figure.

challenging for theories that hypothesize crowding is strictly due to a compulsory pooling mechanism. Although shorter flanking lines might understandably produce less interference in a pooling model, it is difficult to explain how longer flanking lines lead to reduced interference (Clarke et al., 2014). Despite this challenge, Figure 12B shows that the model qualitatively matches the empirical data.

As described in Figure 5, the model's grouping mechanisms connect the boundaries of short flankers but, because of the spatial layout of the stimuli, the boundaries of the target vernier are not part of these groups. These different groups then guide the boundary segmentation mechanisms and result in little crowding, relative to an isolated vernier. Essentially the same result occurs when the short flankers are replaced by long flankers. The spatial layout of the visual elements keeps the target boundaries separate from the boundary groups of the flankers. In contrast, Figure 9 shows that when the flankers have the same length as the target, all the boundaries connect in a single group. As a result of these boundary connections, the segmentation process cannot isolate the representation of the target and strong crowding occurs.

It is worth noting that although the qualitative pattern of model evidence elevation is similar to the empirical threshold elevation, there are notable quantitative differences. Empirical thresholds in the equal length condition are nearly six times as large as for an isolated vernier, while there is only a tripling of model evidence in the equal length condition. A qualitative pattern match is the best we can hope for when comparing the model to the empirical data. In part, this restriction is because the empirical data comes from quite diverse settings (sometimes data is gathered in the fovea and sometimes in the periphery and with some variations in

stimulus luminance and display equipment). Without denying that there is a need to model such variations, our intention at the moment is to explain the robust patterns in the data, and the model predicts that these robust patterns will hold for a wide variety of settings.

Number of Flanking Lines

As further evidence for an important role of perceptual grouping in crowding, Malania et al. (2007) and Manassi et al. (2012) reported that varying the number of flanking lines had effects that depended on the length of the flanking lines. Representative empirical data is presented in the left column of Figure 13 from Manassi et al. (2012; where data were presented at 9 degrees of eccentricity) and Figure 14 from Malania et al. (2007; where data were presented foveally). In every plot the x -axis varies the number of flanking lines, while the y -axis plots either threshold elevation (see Figure 13) or the threshold value (see Figure 14).

The key finding in Manassi et al. (2012), Figure 13, is that the long and short flankers produce weaker crowding (lower threshold elevation) as the number of flankers increases. In contrast, for flankers equal in length to the target, crowding stays relatively constant even as the number of flankers increases. The right column of Figure 13 shows that the model largely reproduces these characteristics. These properties follow from the model because with more short and long flankers it becomes easier to place

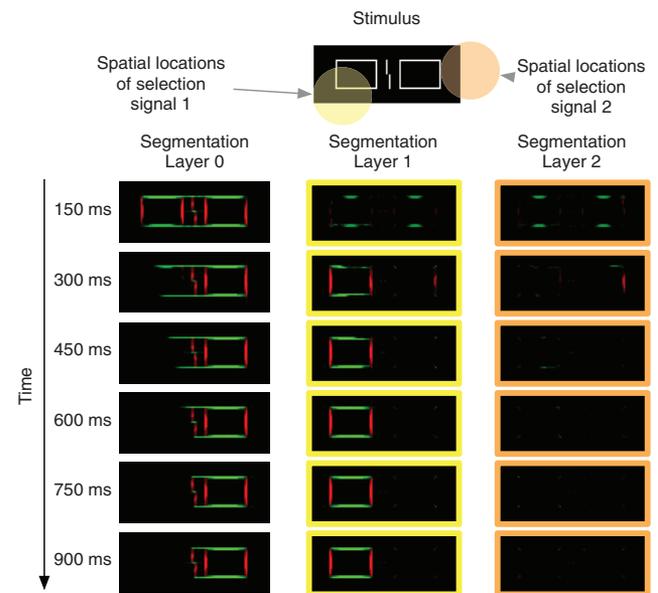


Figure 11. The dynamics of a failed surface segmentation process for a target vernier with flanking rectangles. The colored disks on the stimulus indicate selection signals that spread across the closed surfaces defined by the rectangles and shift boundary inputs from Segmentation Layer 0 to the other segmentation layers. The left side (yellow) selection signal successfully transfers the boundaries of the left flanking rectangle to Segmentation Layer 1, but the right side (orange) selection signal does not overlap the flanking rectangle, so the boundaries of the rectangle remain with the target vernier. The result is that the target boundaries are not isolated from the right side flanking rectangle, and some crowding occurs. See the online article for the color version of this figure.

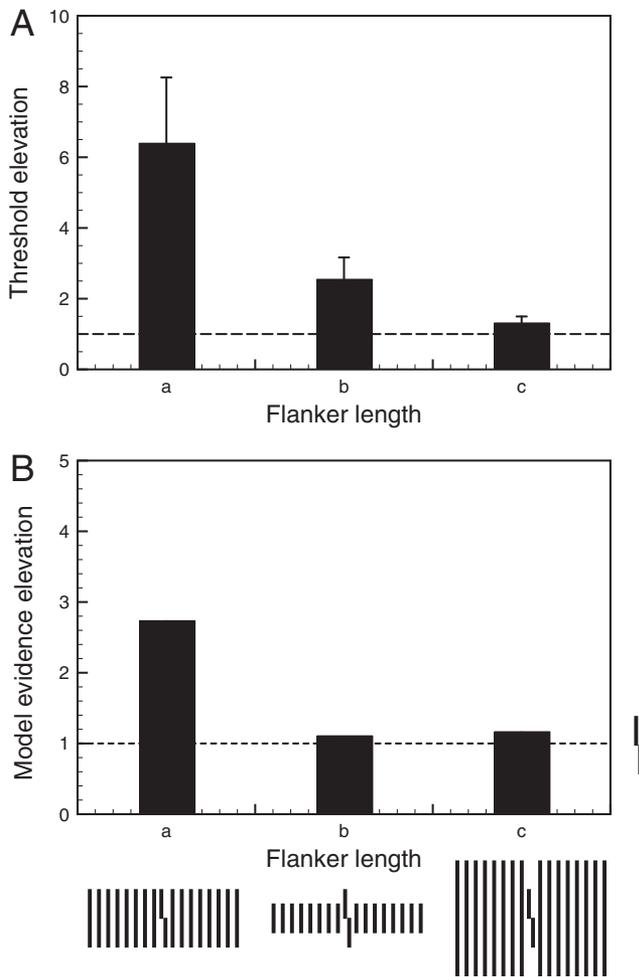


Figure 12. Crowding is strongest when the flankers are the same size as the target and weaker for shorter and longer flanking lines. Panel A shows empirical data from Manassi et al. (2013). Panel B shows model behavior for corresponding stimuli. The hatched line shows normalized performance for the vernier alone condition, where the elevation is 1.0. Please note the different y-axis scalings.

boundary selection signals so that they capture only the boundaries of the flankers and not the boundaries of the target. For flankers equal in length to the target, the boundaries of the flankers group with the boundaries of the target, so the segmentation process cannot isolate the target boundaries (see Figure 9). Thus, varying the number of equal length flankers hardly changes crowding in the model.

An exception is for the smallest number of flankers. In the model, going from two to four flankers leads to a bigger relative increase than what is seen in the experimental data of Manassi et al. (2012). More generally, although the model captures many of the basic trends of the empirical data, there are several quantitative discrepancies. Our goal is not to account for all of these discrepancies, in part because the empirical data itself is rather variable, and because it is the qualitative pattern that suggests an influence of perceptual grouping. Figure 14A shows findings from Malania et al. (2007) on the effect of the number of short flankers. Al-

though like Figure 13B it shows a general decrease in crowding as more flankers are added, the shape of the curve is rather different and so is the overall magnitude of crowding. Similar data variability is present in Figure 14B, which shows an increase in crowding for equal length flankers, whereas Figure 13A shows essentially no change. As it turns out the model does a reasonably good job of matching some of the quantitative aspects of the data in Figure 14A, but we caution readers to not take the good fits too seriously just as we caution readers to not take the quantitative discrepancies in Figure 13 as too damning. Rather, our conclusion is that the model seems to do a reasonable job of capturing many of the robust effects that are demonstrated in the various empirical data.

Figure 14B highlights one robust effect that the model does not capture. The empirical data indicate that short flankers produce stronger crowding than long flankers (the same effect can be seen by comparing the scales of the y axes in Figures 13A and 13C). In contrast, the model predicts that the long flankers produce stronger crowding than the short flankers. This model behavior occurs because of the properties of the discrimination template filters, which are large enough to pool inputs from the extended parts of the long flankers. As more empirical data clarifies the nature of these effects, it might be worthwhile to consider other discrimination templates (or other mechanisms) to account for this model discrepancy.

Closure

Figure 15A shows empirical data from Manassi et al. (2012) with flankers that created closed contours (stimuli b and d) or not (stimuli a and c). They found strong crowding from unclosed contour flankers but very modest crowding for closed contour flankers. These differences held even though the vertical flanker lines immediately adjacent to the target are the same in all conditions, and the differences held even when the flankers were matched in terms of overall amount of contour (e.g., stimuli b and c). In the model, a flanker with closed contours supports the surface segmentation process, as in Figure 10, which shifts the contours abutting a surface into a different segmentation layer. Because the target is not part of the flanker-defined surfaces, it remains at Segmentation Layer 0 and thereby is largely uncrowded by the flankers.

In contrast, for the flankers without closed contours, the surface segmentation process cannot capture the boundaries of the flanking elements. Moreover, the boundary segmentation process is ineffective because the boundaries of the target group with the boundaries from the innermost vertical lines of the flankers (see Figure 1, row 12). Such grouping means that the boundary segmentation process will shift the boundaries of both the flankers and the target together (or not at all) and that there will be crowding.

Uncrowding

Manassi et al. (2013) suggested that the immediate flankers around a target, which generally produce crowding, could themselves be crowded by additional flankers and thereby free the target; a process they called *uncrowding*. Figure 16 shows data and stimuli for this result. A single square around a target vernier

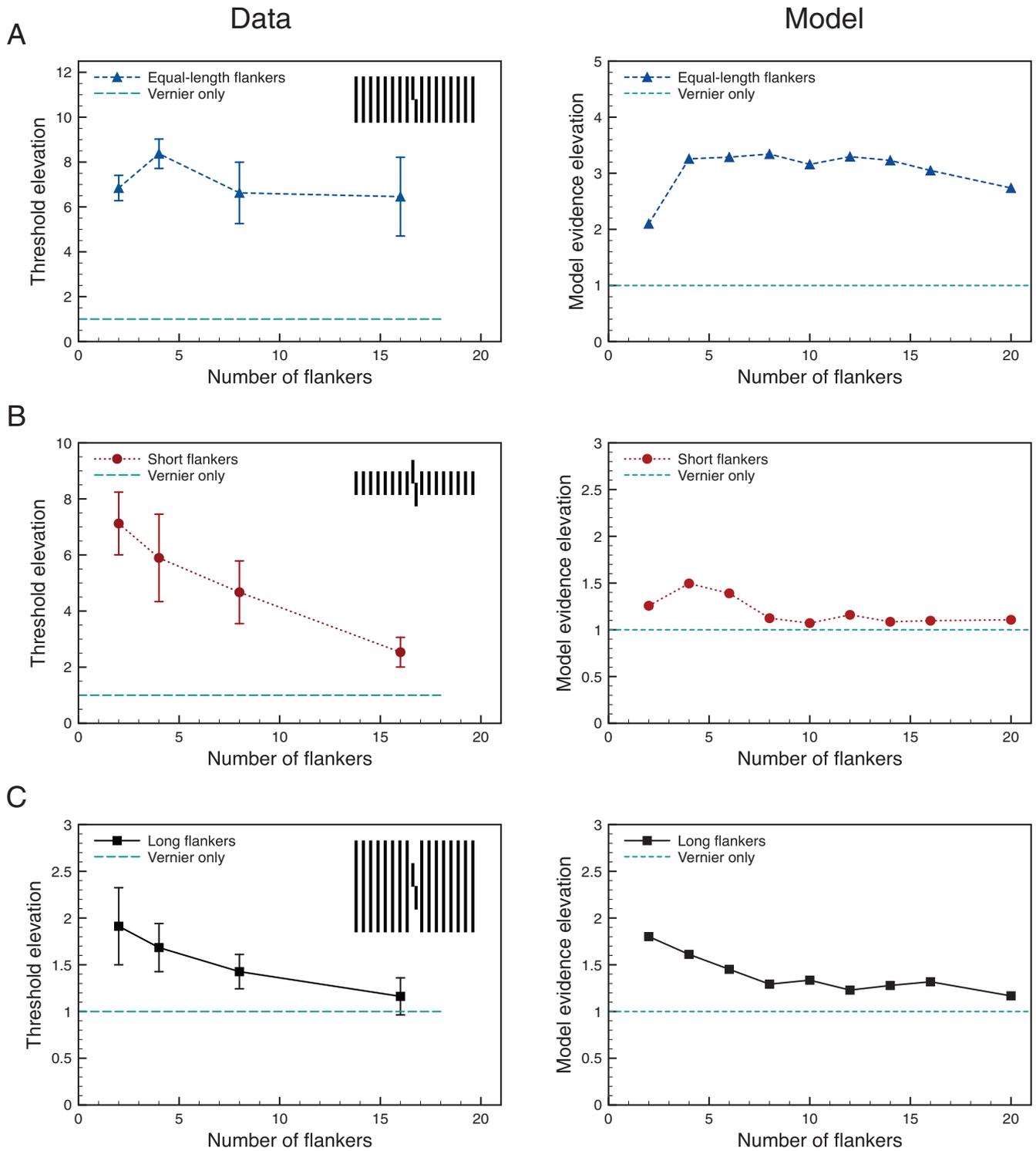


Figure 13. Crowding as a function of the number of flankers for different types of flankers when the stimuli were presented in the periphery. The left column shows empirical data from Manassi et al. (2012), whereas the right column shows the model's behavior for corresponding stimuli. Note that the plots have different scales on the y-axis to better demonstrate the effect of the number of flankers. See the online article for the color version of this figure.

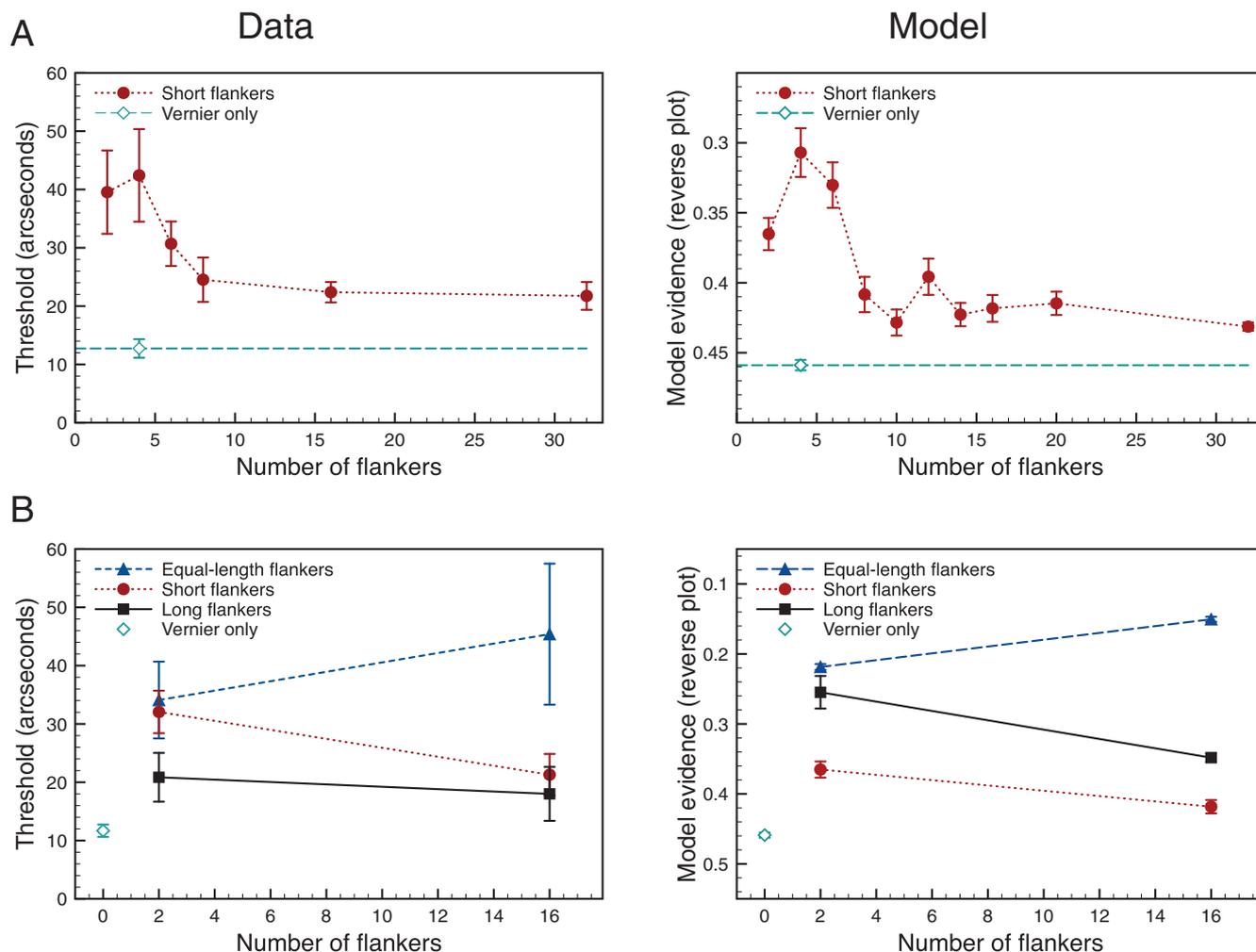


Figure 14. Crowding as a function of the number of flankers for different types of flankers when the stimuli were presented in the fovea. The left column shows empirical data from Malania et al. (2007), whereas the right column shows the model's behavior for corresponding stimuli. See the online article for the color version of this figure.

increased the vernier threshold relative to a vernier in isolation. The threshold is smaller when the same surrounding square is flanked by two or more squares.

Rather than hypothesizing inhibition between distinct representations of different stimuli, the model explains the uncrowding effect as the result of the boundary segmentation process. Although the boundaries of the target vernier do not group with the surrounding square (see row 4 of Figure 1), a single surrounding square is difficult to segment from the target because the two sets of boundaries are in close spatial proximity (similar to Figure 8). When additional flanking squares are added, they generate boundaries that group with the square surrounding the target but do not group with the target's boundaries. Because the selection signals spread along grouped boundaries, the observer can place the selection signal far from the target and thereby shift the entire set of flanker boundaries to a different segmentation layer while leaving the target's boundaries unselected. This process is demonstrated in Figure 7 for four flank-

ing squares with one central square. Importantly, unlike classic crowding models that suppose crowding effects occur only within Bouma's window (Pelli et al., 2004; Pelli & Tillman, 2008), we provide a segmentation mechanism through which crowding can be decreased by flankers across large parts of the visual field.

Uncrowding and Similarity

Manassi et al. (2013) further explored the properties of uncrowding by varying the shapes of the central and flanking elements. Figure 17A shows empirical data for one set of comparisons. Stimulus conditions a, b, and c are the same as in Figure 16, and show strong crowding for a single surround square but uncrowding when additional flanking squares are added to each side. Condition d removed the horizontal components from the additional flanking squares and thereby reintroduced a crowding effect. Condition e removed the horizontal components of the central

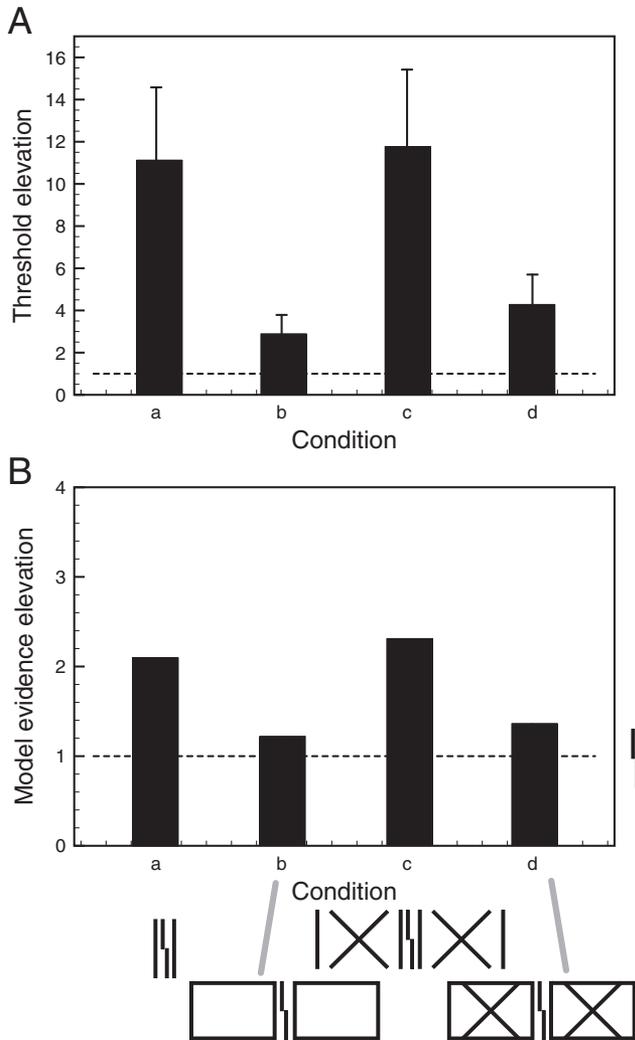


Figure 15. Crowding as a function of flanking stimuli that formed closed surfaces or not. (A) Empirical data from Manassi et al. (2012; see also Sayim et al., 2010) found that crowding was much reduced for flanking stimuli that produced closed surfaces. (B) For the same kind of stimuli, the model shows the same effect of closure due to the ability of the surface segmentation process to shift flankers with closure to a different segmentation layer than the target.

square as well and, again, showed a crowding effect, which implies that the uncrowding effects are not due to local interactions of the vertical lines but involve long range grouping mechanisms.

Figure 1 (rows 4 through 6) shows how the model boundaries group for similar stimuli (one difference is that the stimuli for the empirical study had three flanking squares on each side; to ease computation only two flanking squares per side were used). With two flanking squares on each side of a central square, all the boundaries of the squares group together, and are thereby easily segmented and shifted to a segmentation layer distinct from the boundaries of the target: uncrowding. When the flanking squares lose their horizontal parts, the boundary grouping substantially changes. The boundaries of the two innermost flanking lines group with the boundaries of the central square, but the boundaries for

the other flanking elements form smaller groups from pairs of vertical lines. The central group of boundaries thereby makes for a moderately small target for placement of selection signals, and the selection signals often either miss the flanker boundaries or inadvertently select the boundaries of the target. Thus, this stimulus typically produces crowding.

Row 6 in Figure 1 shows the boundary groupings generated when the horizontal parts of all squares are removed. Pairs of vertical lines form boundary groups, and similar to the previous case, it is difficult to place selection signals accurately enough to segment out the flanker boundaries without also grabbing the boundaries of the target; so this condition produces crowding effects. Figure 17B shows the model evidence values for the different stimulus conditions. Although there are quantitative dif-

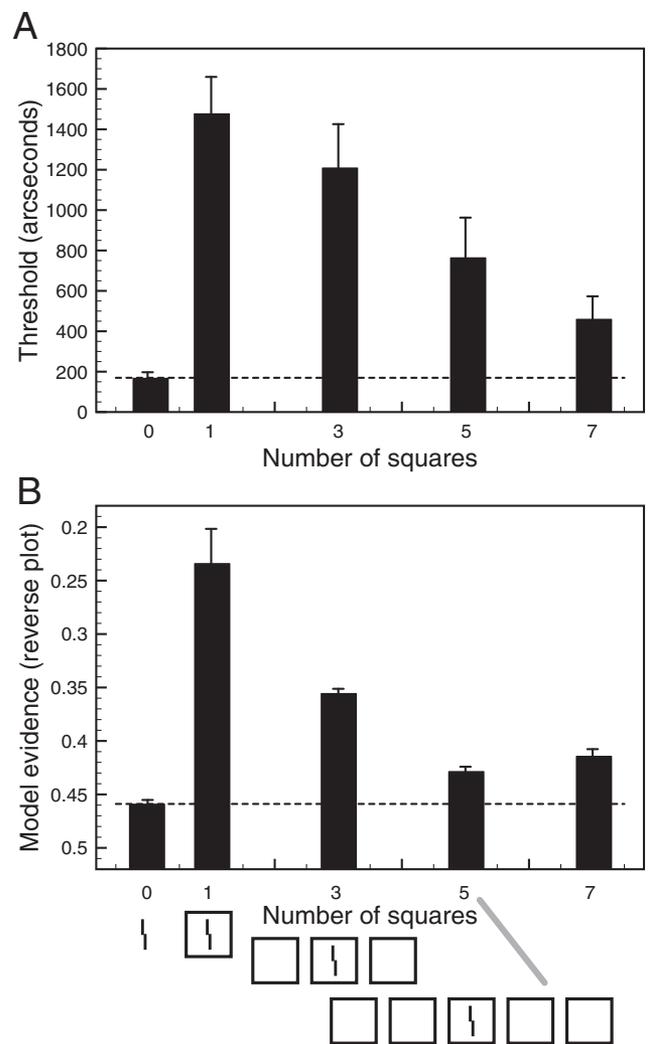


Figure 16. Conditions that produce crowding and uncrowding. (A) Empirical data from Manassi et al. (2013) found adding flanking squares around a target led to lower thresholds (uncrowding). (B) For the same kind of stimuli, the model shows a similar effect because additional flankers make it easier for the segmentation process to select the flanker boundaries and shift them to a unique segmentation layer.

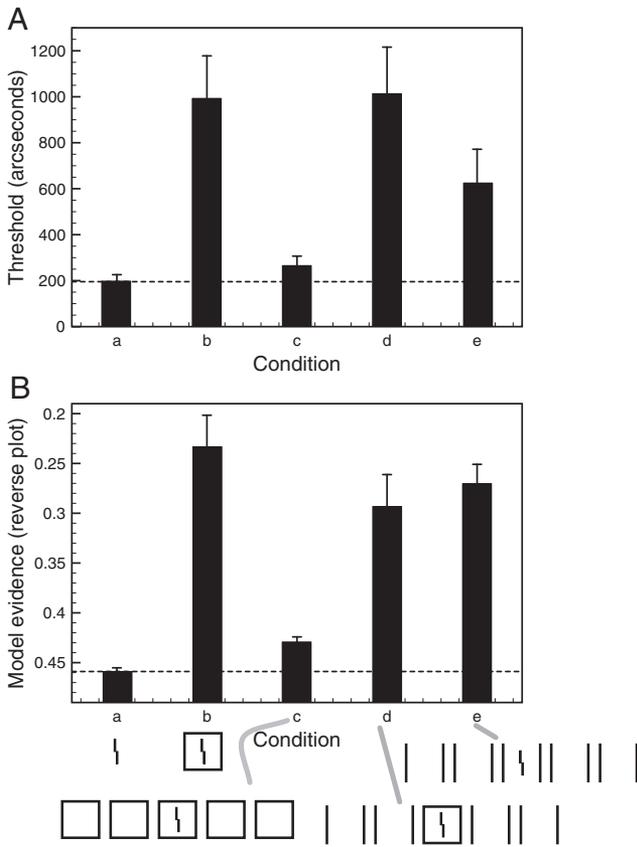


Figure 17. Stimuli that explore uncrowding and similarity. (A) Empirical data from Manassi et al. (2013) found that adding squares reduced crowding effects, but removing the horizontal lines from the flanking squares reintroduced crowding effects. (B) For the same kind of stimuli, the model shows similar crowding effects because the extra flanking squares produce grouping effects that are absent when the horizontal lines are removed.

ferences with the empirical data, the overall pattern for the data and the model is similar across different stimulus conditions.

Similar to the data in Figures 13 and 14, a precise quantitative fit from a model with one set of parameters is not conceivable given the variability in the empirical data. For example, the single surround square condition produces a threshold of around 1400 arc seconds in Figure 16A but a threshold of around 1,000 arc seconds in Figure 17A. The differences presumably reflect variation due to random sampling of observers, and they cannot be accounted for by a model with a fixed set of parameters.

In an additional study, Manassi et al. (2013) rotated the central or flanking squares by 45 degrees so that there could be various combinations of squares and diamonds. Figure 18A shows the data and a schematic of the stimulus conditions. The overall conclusion was that a square or a diamond by itself around the target vernier produced crowding, but that additional flanker elements of the same type (e.g., a central diamond among flanking diamonds) reduced the strength of crowding. When the central element and the flankers were of different types, crowding was strong.

Figure 1 (rows 4 and 7 through 9) shows how the model boundaries group for similar stimuli, and the strength of crowding

is easily predicted from the boundary groupings. When the central element is a square and the flankers are diamonds, the boundaries from the square are separate from the boundaries of the diamonds. This separation occurs because the grouping process interpolates between similarly oriented boundaries in a direction consistent with their orientation. The horizontal boundaries of the central square do not group with the diagonal boundaries of the diamonds because they have inconsistent orientations. Boundaries from the flanking diamonds do group together because their large horizontal size (due to the rotation) makes the left and right diamond points touch. Because the boundaries of the central square do not group with the boundaries of the diamond flankers, a selection signal that only covers a flanking diamond will not also segment out the

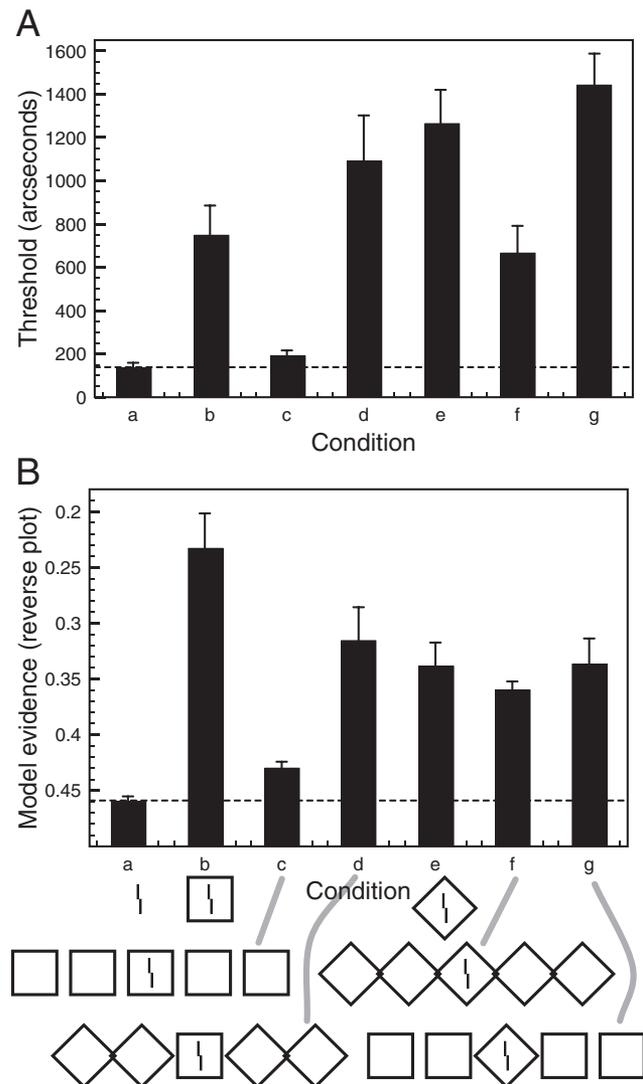


Figure 18. Stimuli that explore uncrowding and similarity. (A) Empirical data from Manassi et al. (2013) found crowding is strongest if the central and flanking elements are different shapes. (B) For the same kind of stimuli, the model shows similar crowding effects because the different shapes do not group together and thereby prevent the segmentation process from isolating the target vernier boundaries.

boundaries of the central square, and the target vernier will remain crowded. Alternatively, a selection signal that attempts to directly cover the central square will risk simultaneously selecting the boundaries of the target vernier, which will likewise produce crowding.

The situation is similar when the central element is a diamond and the flankers are squares. The flanking squares group together (due to collinearity of the horizontal lines in the squares). However, the oblique boundaries of the central diamond do not group with the horizontal boundaries of the squares. A selection signal on the flankers does not separate the boundaries of the central diamond from the target, and a selection signal on the boundaries of the diamond risks simultaneously selecting the boundaries of the target. In either case, the model produces crowding of the target.

When the central element and flanking elements are all diamonds, the points overlap, so all boundaries are connected (and thus grouped). A selection signal covering any of the diamonds will thereby spread across the entire set of boundaries and shift them to a different segmentation layer than the target boundaries (which do not group with the diamonds). There is some crowding in this condition because it takes time for the selection signals to capture the boundaries of the central diamond and shift them away from the target boundaries.

Conclusions

The cortical neural network model uses perceptual grouping and a novel segmentation process to account for challenging properties of visual crowding. In particular, simulations demonstrate that the model properly accounts for effects of flanker length, the number of flanker lines, Gestalt effects, uncrowding effects, and similarity effects. These properties of crowding are challenging (or impossible) for many other models because they assume that crowding occurs as a result of mechanisms located in a feedforward path of visual processing. For example, basic pooling models propose that crowding occurs when higher level neurons, for example, V2 neurons with larger receptive fields, pool signals from lower level neurons, for example, V1 neurons with smaller receptive fields. Pooling is inevitable for object recognition since, for example, a higher level neuron sensitive to squares needs to pool signals from lower level neurons coding for the lines making up the square. From this perspective, crowding occurs because of the unavoidable bottlenecks of object recognition.

Our model fundamentally differs from these approaches, and these differences allow the model to account for the crowding effects discussed here. First, our model comprises both bottom-up and top-down processing that enables an observer to perceptually group disparate elements of a scene, which promotes a segmentation driven change in the visual representation of those elements. Second, and more importantly, spatial resolution is at no stage per se limited by low level or high level bottlenecks (Herzog, Thunell & Ögmen, 2016). Although the top-down signals are imprecise (low-resolution), their effects are directed by the (high-resolution) grouping process. So the model circuits provide an example of how top-down and bottom-up interactions occur to promote visual processing of information relevant to the observer. The resulting model behavior is robust, in the sense that top-down signals do not have to be precisely placed but are sensitive to small stimulus changes that alter perceptual grouping. We propose that such

interfaces between top-down and bottom-up signals are a general issue to be addressed in cortical circuits because similar kinds of crowding effects also occur in domains such as audition (Oberfeld & Stahn, 2012) and haptics (Overvliet & Sayim, 2015).

It is important to note that the current model simulations are not uniformly successful. Regarding effects of flanker length, the model incorrectly predicts that long flankers produce stronger crowding than short flankers (the data show the opposite relationship). We believe this discrepancy can partly be accounted for by changing the height of the template so that it includes less input from the long flankers, even though those longer flankers are easier for the segmentation process to select. More generally, the model does not produce quite as strong a crowding effect as reported by the empirical data. It is easy to scale the model crowding effect by modifying the constant in the numerator of Equation 1, but our goal is to present the overall effects of perceptual grouping and segmentation rather than to precisely fit the empirical measurements. Indeed, there is enough variability in the empirical measurements that a precise fit to the data would likely indicate model overfitting.

There are conceptual similarities between the present work and models proposed by Jehee, Roelfsema, Deco, Murre, and Lamme (2007) and Foley et al. (2012). These earlier investigations also hypothesized that crowding effects were related to perceptual organization (grouping) and that top-down cues helped to isolate the cortical signals corresponding to the target. However, those studies only simulated a small set of crowding effects, and showed a “proof of concept” for their ideas rather than provide a systematic investigation into the relationships between grouping, segmentation, and crowding. Although our mechanisms and details differ from Jehee et al. (2007) and Foley et al. (2012), there are common ideas that might allow for valuable cross-talk between these models in future investigations. There are also similarities between the model presented here and the suggestion that crowding occurs when attention cannot access fine grained information about an element presented in clutter because the resolution of attention is limited (He, Cavanagh, & Intriligator, 1996), although the hypothesized mechanisms are quite different.

Our model complements explanations of other crowding effects (e.g., Levi, Hariharan, & Klein, 2002; Levi, Klein, & Hariharan, 2002; Pelli, 2008; Pelli et al., 2004), which typically do not consider the potential impact of perceptual organization and often use stimuli where perceptual organization hardly plays a role. For many crowding experiments (especially those involving letters), the boundaries generated by the flanker and target stimuli always group together, so segmentation by boundary grouping is not possible. Such a situation would be equivalent to the simulation described in Figure 9, where the flanking lines are the same size as the target and everything groups together. With such groupings, the segmentation process cannot separate the target from the flankers, and crowding occurs depending on the interference of the flankers on the target identification task. On the other hand, the presence of boundary grouping between the target and flankers does not necessarily imply strong crowding; for example, if the flankers are a different color or define a surface that is separate from the target (see Figures 10 and 15) then other segmentation mechanisms can isolate the target representation. Nevertheless, the model presented here has a fundamentally different representation of visual information than the summary statistics approaches of

Balas et al. (2009) and Keshvari and Rosenholtz (2016). It would be interesting to see if the summary statistics approach can account for the crowding effects reported here and to then identify empirical tests that can distinguish between the model types.

More generally, the model suggests that it is not boundary grouping but segmentation that regulates whether crowding occurs. Oftentimes these two processes go together, such that boundary grouping between the target and flankers makes segmentation impossible (e.g., see Figure 9), whereas boundary grouping among the flankers but not the target makes segmentation possible (e.g., see Figures 5 and 7). In some cases, as in Figure 8, the flanker and target boundaries do not group together, but they are in such close spatial proximity that it is difficult to select just the flanker boundaries without also selecting the target boundaries; thus strong crowding effects can occur without boundary grouping. It is important to note that the model makes a distinction between the mechanisms for boundary grouping and segmentation, but it is likely that human observers are largely unaware of these distinct processes. In a task that asks observers to judge whether flanker and target elements group together, we suspect that observers simply examine whether they can segment them. If such segmentation is easily done, then the elements may be judged to be in different groups. If such segmentation is not easily done, then observers will report that the elements are part of a common perceptual group. Thus, we propose that a perceptual group consists of visual stimuli whose cortical representations cannot be easily segmented. This perspective is consistent with earlier claims (Herzog & Manassi, 2015; Herzog et al., 2015), that grouping (within the model, a failure of segmentation) between the target and flanker boundaries is necessary for crowding to occur.

We believe that low-level models of crowding that focus solely on spatial separation or feature similarity are fundamentally incomplete because they do not consider the role of grouping and segmentation on crowding. Some properties of crowding, such as that bigger flankers produce weaker crowding, are difficult to explain by low-level crowding mechanisms. As we have shown, these seemingly paradoxical properties become understandable when considering the role of perceptual grouping and segmentation. We have restricted our focus on one visual task (vernier offset) because it allows for easy comparison across many different types of flankers, but the very same mechanisms should apply to other situations where the target and flankers differ in size, intensity, or shape that would alter perceptual grouping effects (e.g., crowding of letters as in Banks et al., 1979, Experiment 3). An extension of the segmentation mechanism to include color might likewise explain why crowding is reduced when the flankers differ in color from the target (e.g., Kooi et al., 1994; Rosen & Pelli, 2015).

Our claim that low-level models of crowding are incomplete does not mean that we believe them to be wrong in all details. For example, we can easily imagine that the current model could include characteristics of low-level models to explain Bouma's law (that crowding occurs only within a spatial window with the size of half the target eccentricity) by altering the template resolution, the features that contribute to the template calculation, or the resolution of visual space in the periphery. With such an interpretation, we anticipate that Bouma's law would be notable when grouping and segmentation effects are nearly nonexistent. Alternatively, Bouma's law might be understood as a special case

of grouping by proximity, and such properties might be incorporated into a future version of the current model.

An important property of the cortical model is that it is not a model of crowding per se, but is a model of visual perception that proposes an explanation for a variety of crowding effects. None of the model properties exist purely to account for the characteristics of crowding; rather they exist to enable the model to functionally process and represent visual information. As such, some properties of crowding can now be connected to a diverse range of other aspects of visual perception. For example, the boundary groupings that support the segmentation process that underlie crowding effects also play important roles in the model's explanations of illusory contours, neon color spreading, visual persistence, meta-contrast masking, figure-ground distinctions, texture segmentation, visual search, afterimages, scene fading, contour erasure, and many more (Bhatt et al., 2007; Francis, 1997, 2015; Francis, Grossberg, & Mingolla, 1994; Francis & Kim, 2012; Francis & Rothmayer, 2003; Grossberg, 1994; Grossberg & Mingolla, 1985a, 1985b; Grossberg, Mingolla, & Ross, 1994). A single model that simultaneously produces all of these effects does not yet exist (or, at least, cannot be simulated because the necessary computational resources would be enormous). Nevertheless, the functional behaviors of the various versions of the model appear to be consistent, so we anticipate that future experimental and theoretical work will link properties of crowding with other visual phenomena to gain a better understanding of how many different properties of visual perception interrelate. Such an understanding may help identify new methods to investigate visual processing and identify ways to mitigate crowding effects when they impair human performance. Because crowding affects a wide variety of cognitive behaviors, the model offers a new way to understand general properties of object recognition and deficits in visual processing.

References

- Agaoglu, M. N., & Chung, S. (2016). Can (should) theories of crowding be unified? *Journal of Vision*, *16*, 10.
- Andriessen, J., & Bouma, H. (1976). Eccentric vision: Adverse interactions between line segments. *Vision Research*, *16*, 71–78. [http://dx.doi.org/10.1016/0042-6989\(76\)90078-X](http://dx.doi.org/10.1016/0042-6989(76)90078-X)
- Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision*, *9*(12–13), 1–18. <http://dx.doi.org/10.1167/9.12.13>
- Banks, W. P., Larson, D. W., & Prinzmetal, W. (1979). Asymmetry of visual interference. *Perception & Psychophysics*, *25*, 447–456.
- Bhatt, R., Carpenter, G., & Grossberg, S. (2007). Texture segregation by visual cortex: Perceptual grouping, attention, and learning. *Vision Research*, *47*, 3173–3211.
- Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*, 177–178.
- Cao, Y., & Grossberg, S. (2005). A laminar cortical model of stereopsis and 3D surface perception: Closure and da Vinci stereopsis. *Spatial Vision*, *18*, 515–578.
- Chakravarthi, R., & Pelli, D. G. (2011). The same binding in contour integration and crowding. *Journal of Vision*, *11*, 1–12. <http://dx.doi.org/10.1167/11.8.10>
- Chaney, W., Fischer, J., & Whitney, D. (2014). The hierarchical sparse selection model of visual crowding. *Frontiers in Integrative Neuroscience*, *8*. <http://dx.doi.org/10.3389/fnint.2014.00073>
- Chung, S. T. L., Levi, D. M., & Legge, G. E. (2001). Spatial-frequency and contrast properties of crowding. *Vision Research*, *41*, 1833–1850.

- Clarke, A. M., Herzog, M. H., & Francis, G. (2014). Visual crowding illustrates the inadequacy of local vs. global and feedforward vs. feedback distinctions in modeling visual perception. *Frontiers in Psychology: Perception Science*, 5. <http://dx.doi.org/10.3389/fpsyg.2014.01193>
- DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, 73, 415–434. <http://dx.doi.org/10.1016/j.neuron.2012.01.010>
- Doron, A., Manassi, M., Herzog, M. H., & Ahissar, M. (2015). Intact crowding and temporal masking in dyslexia. *Journal of Vision*, 15, 1–13. <http://dx.doi.org/10.1167/15.14.13>
- Ester, E. F., Klee, D., & Awh, E. (2014). Visual crowding cannot be wholly explained by feature pooling. *Journal of Experimental Psychology: Human Perception and Performance*, 40, 1022–1033. <http://dx.doi.org/10.1037/a0035377>
- Ester, E. F., Zilber, E., & Serences, J. T. (2015). Substitution and pooling in visual crowding induced by similar and dissimilar distractors. *Journal of Vision*, 15, 1–12. <http://dx.doi.org/10.1167/15.1.4>
- Farzin, F., Rivera, S. M., & Whitney, D. (2009). Holistic crowding of Mooney faces. *Journal of Vision*, 9, 1–15.
- Field, D. J., Hayes, A., & Hess, R. F. (1993). Contour integration by the human visual system: Evidence for a local “association field.” *Vision Research*, 33, 173–193.
- Flom, M. C., Heath, G. G., & Takahashi, E. (1963). Contour interaction and visual resolution: Contralateral effects. *Science*, 142, 979–980.
- Foley, N. C., Grossberg, S., & Mingolla, E. (2012). Neural dynamics of object-based multifocal visual spatial attention and priming: Object cueing, useful-field-of-view, and crowding. *Cognitive Psychology*, 65, 77–117. <http://dx.doi.org/10.1016/j.cogpsych.2012.02.001>
- Francis, G. (1997). Cortical dynamics of lateral inhibition: Metacontrast masking. *Psychological Review*, 104, 572–594.
- Francis, G. (2015). Contour erasure and filling-in: Old simulations account for most new observations. *i-Perception*, 6, 116–126. <http://dx.doi.org/10.1068/i0684>
- Francis, G., Grossberg, S., & Mingolla, E. (1994). Cortical dynamics of feature binding and reset: Control of visual persistence. *Vision Research*, 34, 1089–1104.
- Francis, G., & Kim, J. (2012). Simulations of induced visual scene fading with boundary offset and filling-in. *Vision Research*, 62, 181–191.
- Francis, G., & Rothmayer, M. (2003). Interactions of afterimages for orientation and color: Experimental data and model simulations. *Perception & Psychophysics*, 65, 508–522.
- Freeman, J., Chakravarthi, R., & Pelli, D. G. (2012). Substitution and pooling in crowding. *Attention, Perception & Psychophysics*, 74, 379–396. <http://dx.doi.org/10.3758/s13414-011-0229-0>
- Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience*, 14, 1195–1201. <http://dx.doi.org/10.1038/nn.2889>
- Gewaltig, M.-O., & Diesmann, M. (2007). NEST (Neural Simulation Tool). *Scholarpedia*, 2, 1430.
- Gori, S., & Facoetti, A. (2015). How the visual aspects can be crucial in reading acquisition? The intriguing case of crowding and developmental dyslexia. *Journal of Vision*, 15, 1–20. <http://dx.doi.org/10.1167/15.1.8>
- Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2010). Crowding changes appearance. *Current Biology*, 20, 496–501.
- Grossberg, S. (1994). 3-D vision and figure-ground separation by visual cortex. *Perception & Psychophysics*, 55, 48–120.
- Grossberg, S. (2014). How visual illusions illuminate complementary brain processes: Illusory depth from brightness and apparent motion of illusory contours. *Frontiers in Human Neuroscience*, 8. <http://dx.doi.org/10.3389/fnhum.2014.00854>
- Grossberg, S., & Mingolla, E. (1985a). Neural dynamics of perceptual grouping: Textures, boundaries, and emergent segmentations. *Perception & Psychophysics*, 38, 141–171.
- Grossberg, S., & Mingolla, E. (1985b). Neural dynamics of form perception: Boundary completion, illusory figures, and neon color spreading. *Psychological Review*, 92, 173–211.
- Grossberg, S., Mingolla, E., & Ross, W. D. (1994). A neural theory of attentive visual search: Interactions of boundary, surface, spatial, and object representations. *Psychological Review*, 101, 470–489.
- Grossberg, S., & Raizada, R. D. S. (2000). Contrast-sensitive perceptual grouping and object-based attention in the laminar circuits of primary visual cortex. *Vision Research*, 40, 1413–1432.
- Harrison, W. J., & Bex, P. J. (2014). Integrating retinotopic features in spatiotopic coordinates. *Journal of Neuroscience*, 34, 7351–7360. <http://dx.doi.org/10.1523/JNEUROSCI.5252-13.2014>
- Harrison, W. J., & Bex, P. J. (2015). A unifying model of orientation crowding in peripheral vision. *Current Biology*, 25, 3213–3219. <http://doi.org/10.1016/j.cub.2015.10.052>
- Harrison, W. J., & Bex, P. J. (2016). Reply to Pachai et al. *Current Biology*, 26, R353–R354.
- Harrison, W. J., Mattingley, J. B., & Remington, R. W. (2013). Eye movement targets are released from visual crowding. *Journal of Neuroscience*, 33, 2927–2933. <http://dx.doi.org/10.1523/JNEUROSCI.4172-12.2013>
- Harrison, W. J., Retell, J. D., Remington, R. W., & Mattingley, J. B. (2013). Visual crowding at a distance during predictive remapping. *Current Biology*, 23, 793–798. <http://dx.doi.org/10.1016/j.cub.2013.03.050>
- He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and locus of visual awareness. *Nature*, 383, 334–337.
- Hermens, F., Scharnowski, F., & Herzog, M. H. (2009). Spatial grouping determines temporal integration. *Journal of Experimental Psychology: Human Perception and Performance*, 35, 595–610.
- Herzog, M. H., & Manassi, M. (2015). Uncorking the bottleneck of crowding: A fresh look at object recognition. *Current Opinion in Behavioral Sciences*, 1, 86–93.
- Herzog, M. H., Sayim, B., Chicherov, V., & Manassi, M. (2015). Crowding, grouping, and object recognition: A matter of appearance. *Journal of Vision*, 15, 5. <http://dx.doi.org/10.1167/15.6.5>
- Herzog, M. H., Thunell, E., & Ögmen, H. (2016). Putting low-level vision into global context: Why vision cannot be reduced to basic circuits. *Vision Research*, 126, 9–18.
- Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex. *Journal of Physiology*, 160, 106–154.
- Huckauf, A., & Heller, D. (2002). Spatial selection in peripheral letter recognition: In search of boundary conditions. *Acta Psychologica*, 111, 101–123.
- Ikeda, H., Watanabe, K., & Cavanagh, P. (2013). Crowding of biological motion stimuli. *Journal of Vision*, 13, 1–6. <http://171.67.113.220/content/13/4/20.short>
- Jehee, J. F. M., Roelfsema, P. R., Deco, G., Murre, J. M. J., & Lamme, V. A. F. (2007). Interactions between higher and lower visual areas improve shape selectivity of higher level neurons—Explaining crowding phenomena. *Brain Research*, 1157, 167–176. <http://dx.doi.org/10.1016/j.brainres.2007.03.090>
- Keshvari, S., & Rosenholtz, R. (2016). Pooling of continuous features provides a unifying account of crowding. *Journal of Vision*, 16, 39. <http://dx.doi.org/10.1167/16.3.39>
- Kimchi, R., & Pirkner, Y. (2015). Multiple level crowding: Crowding at the object parts level and at the object configural level. *Perception*, 44, 1275–1292. <http://dx.doi.org/10.1177/0301006615594970>
- Klein, S. A., & Levi, D. M. (1985). Hyperacuity thresholds of 1 second: Theoretical predictions and empirical validation. *Journal of the Optical Society of America A*, 2, 1170–1190.

- Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision*, 8, 255–279.
- Krumhansl, C., & Thomas, E. (1977). Effect of level of confusability on reporting letters from briefly presented visual displays. *Perception & Psychophysics*, 21, 269–279. <http://dx.doi.org/10.3758/BF03214239>
- Legge, G. E. (2007). *Psychophysics of reading in normal and low vision*. Mahwah, NJ: Lawrence Erlbaum.
- Levi, D. M. (2008). Crowding's an essential bottleneck for object recognition: A mini-review. *Vision Research*, 48, 635–654. <http://dx.doi.org/10.1016/j.visres.2007.12.009>
- Levi, D. M., & Carney, T. (2009). Crowding in peripheral vision: Why bigger is better. *Current Biology*, 19, 1988–1993.
- Levi, D. M., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of Vision*, 2, 167–177. <http://dx.doi.org/10.1167/2.2.3>
- Levi, D. M., Klein, S. A., & Carney, T. (2000). Unmasking the mechanisms for vernier acuity: Evidence for a template model for vernier acuity. *Vision Research*, 40, 951–972.
- Levi, D. M., Klein, S. A., & Hariharan, S. (2002). Suppressive and facilitatory spatial interactions in foveal vision: Foveal crowding is simple contrast masking. *Journal of Vision*, 2, 140–166. <http://dx.doi.org/10.1167/2.2.2>
- Livne, T., & Sagi, D. (2007). Configuration influence on crowding. *Journal of Vision*, 7, 1–12.
- Livne, T., & Sagi, D. (2010). How do flankers' relations affect crowding? *Journal of Vision*, 10, 1–14. <http://dx.doi.org/10.1167/10.3.1>
- Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding: Selective interference between configurational representations of faces in crowded scenes. *Journal of Vision*, 7, 24. <http://dx.doi.org/10.1167/7.2.24>
- Malania, M., Herzog, M. H., & Westheimer, G. (2007). Grouping of contextual elements that affect vernier thresholds. *Journal of Vision*, 7, 1. <http://dx.doi.org/10.1167/7.2.1>
- Manassi, M., Hermens, F., Francis, G., & Herzog, M. H. (2015). Release of crowding by pattern completion. *Journal of Vision*, 15, 16. <http://dx.doi.org/10.1167/15.8.16>
- Manassi, M., Lonchamp, S., Clarke, A., & Herzog, M. H. (2016). What crowding can tell us about object representations. *Journal of Vision*, 16, 35. <http://dx.doi.org/10.1167/16.3.35>
- Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision*, 12, 13. <http://dx.doi.org/10.1167/12.10.13>
- Manassi, M., Sayim, B., & Herzog, M. H. (2013). When crowding of crowding leads to uncrowding. *Journal of Vision*, 13, 110. <http://dx.doi.org/10.1167/13.13.10>
- Mareschal, I., Morgan, M. J., & Solomon, J. A. (2008). Contextual effects on decision templates for parafoveal orientation identification. *Vision Research*, 48, 2689–2695. <http://dx.doi.org/10.1016/j.visres.2008.08.020>
- Nazir, T. A. (1992). Effects of lateral masking and spatial precueing on gap-resolution in central and peripheral vision. *Vision Research*, 32, 771–777.
- Oberfeld, D., & Stahn, P. (2012). Sequential grouping modulates the effect of non-simultaneous masking on auditory intensity resolution. *PLoS ONE*, 7(10), e48054.
- Overvliet, K. E., & Sayim, B. (2015). Perceptual grouping determines haptic contextual modulation. *Vision Research*, 126, 52–58. <http://dx.doi.org/10.1016/j.visres.2015.04.016>
- Pachai, M. V., Doerig, A. C., & Herzog, M. H. (2016). How to best unify crowding? *Current Biology*, 26(9), R352–R353.
- Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4, 739–744.
- Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology*, 18, 445–451. <http://dx.doi.org/10.1016/j.conb.2008.09.008>
- Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision*, 4, 1136–1169.
- Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, 11, 1129–1135.
- Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision*, 7, 1–36.
- Pöder, E. (2006). Crowding, feature integration, and two kinds of attention. *Journal of Vision*, 6, 163–169. <http://dx.doi.org/10.1167/6.2.7>
- Pöder, E. (2007). Effect of colour pop-out on the recognition of letters in crowding conditions. *Psychological Research*, 71, 641–645. <http://dx.doi.org/10.1007/s00426-006-0053-7>
- Raizada, R., & Grossberg, S. (2001). Context-sensitive bindings by the laminar circuits of V1 and V2: A unified model of perceptual grouping, attention, and orientation contrast. *Visual Cognition*, 8, 431–466.
- Raizada, R., & Grossberg, S. (2003). Towards a theory of the laminar architecture of cerebral cortex: Computational clues from the visual system. *Cerebral Cortex*, 13, 100–113.
- Reeves, A., & Sperling, G. (1986). Attention gating in short-term visual memory. *Psychological Review*, 93, 180–206.
- Richard, A. M., Lee, H., & Vecera, S. P. (2008). Attentional spreading in object-based attention. *Journal of Experimental Psychology: Human Perception and Performance*, 34, 842–853.
- Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2, 1019–1025.
- Roelfsema, P. R., Lamme, V. A. F., & Spekreijse, H. (1998). Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, 395, 376–381.
- Roinishvili, M., Cappe, C., Shaqiri, A., Brand, A., Chkonia, E., & Herzog, M. H. (2015). Crowding, Grouping and Gain Control in Schizophrenia. *Psychiatry Research*, 226, 441–445.
- Rosen, S., & Pelli, D. (2015). Crowding by a repeating pattern. *Journal of Vision*, 15, 10. <http://dx.doi.org/10.1167/15.6.10>
- Saarela, T. P., Sayim, B., Westheimer, G., & Herzog, M. H. (2009). Global stimulus configuration modulates crowding. *Journal of Vision*, 9, 5. <http://dx.doi.org/10.1167/9.2.5>
- Saarela, T. P., Westheimer, G., & Herzog, M. H. (2010). The effect of spacing regularity on visual crowding. *Journal of Vision*, 10, 17. <http://dx.doi.org/10.1167/10.10.17>
- Sayim, B., Manassi, M., & Herzog, M. (2014). How color, regularity, and good gestalt determine backward masking. *Journal of Vision*, 14, 8. <http://dx.doi.org/10.1167/14.7.8>
- Sayim, B., Westheimer, G., & Herzog, M. H. (2010). Gestalt factors modulate basic spatial vision. *Psychological Science*, 21, 641–644.
- Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, USA*, 104, 6424–6429.
- Strasburger, H., Harvey, L. O., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics*, 49, 495–508.
- Strasburger, H., Rentschler, I., & Jüttner, M. (2011). Peripheral vision and pattern recognition: A review. *Journal of Vision*, 11, 13. <http://dx.doi.org/10.1167/11.5.13>
- Thorpe, S., Delorme, A., & Van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks*, 14, 715–725.
- Toet, A., & Levi, D. M. (1992, July). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research*, 32, 1349–1357.

- Tyler, C. W., & Kontsevich, L. L. (1995). Mechanisms of stereoscopic processing: Stereoattention and surface perception in depth reconstruction. *Perception, 24*, 127–153.
- Van Den Berg, R., Roerdink, J. B., & Cornelissen, F. W. (2010). A neurophysiologically plausible population code model for feature integration explains visual crowding. *PLoS Computational Biology, 6*(1), e1000646.
- Vickery, T. J., Shim, W. M., Chakravarthi, R., Jiang, Y. V., & Luedeman, R. (2009). Supercrowding: Weakly masking a target expands the range of crowding. *Journal of Vision, 9*, 1–15. <http://dx.doi.org/10.1167/9.2.12>
- von der Heydt, R., Peterhans, E., & Baumgartner, G. (1984). Illusory contours and cortical neuron responses. *Science, 224*, 1260–1262.
- Wallace, J. M., Chiu, M. K., Nandy, A. S., & Tjan, B. S. (2013). Crowding during restricted and free viewing. *Vision Research, 84*, 50–59.
- Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Science, 15*, 160–168.
- Wilkinson, F., Wilson, H. R., & Ellemborg, D. (1997). Lateral interactions in peripherally viewed texture arrays. *Journal of the Optical Society of America, 14*, 2057–2068.
- Wilson, H. R. (1986). Responses of spatial mechanisms can explain hyperacuity. *Vision Research, 26*, 453–469.
- Wolford, G., & Chambers, L. (1983). Lateral masking as a function of spacing. *Perception & Psychophysics, 33*, 129–138.
- Yeotikar, N. S., Khoo, S. K., Asper, L. J., & Suttle, C. M. (2011). Configuration specificity of crowding in peripheral vision. *Vision Research, 51*, 1239–1248.
- Zhang, J.-Y., Zhang, G.-L., Liu, L., & Yu, C. (2012). Whole report uncovers correctly identified but incorrectly placed target information under visual crowding. *Journal of Vision, 12*, 5. <http://dx.doi.org/10.1167/12.7.5>
- Zhao, J., Kong, F., & Wang, Y. (2013). Attentional spreading in object-based attention: The roles of target-object integration and target presentation time. *Attention, Perception & Psychophysics, 75*, 876–887. <http://dx.doi.org/10.3758/s13414-013-0445-x>

Received April 5, 2016

Revision received January 11, 2017

Accepted February 26, 2017 ■

ORDER FORM

Start my 2017 subscription to **Psychological Review**®
ISSN: 0033-295X

___ \$105.00 **APA MEMBER/AFFILIATE** _____

___ \$250.00 **INDIVIDUAL NONMEMBER** _____

___ \$1,042.00 **INSTITUTION** _____

Sales Tax: 5.75% in DC and 6% in MD and PA _____

TOTAL AMOUNT DUE \$ _____

Subscription orders must be prepaid. Subscriptions are on a calendar year basis only. Allow 4-6 weeks for delivery of the first issue. Call for international subscription rates.



AMERICAN
PSYCHOLOGICAL
ASSOCIATION

SEND THIS ORDER FORM TO
American Psychological Association
Subscriptions
750 First Street, NE
Washington, DC 20002-4242

Call **800-374-2721** or 202-336-5600
Fax **202-336-5568** : TDD/TTY **202-336-6123**
For subscription information,
e-mail: subscriptions@apa.org

Check enclosed (make payable to APA)

Charge my: Visa MasterCard American Express

Cardholder Name _____

Card No. _____ Exp. Date _____

Signature (Required for Charge)

Billing Address

Street _____

City _____ State _____ Zip _____

Daytime Phone _____

E-mail _____

Mail To

Name _____

Address _____

City _____ State _____ Zip _____

APA Member # _____

REVA17