Saccade-confounded image statistics explain visual crowding

Anirvan S Nandy1,3 & Bosco S Tjan1,2

Processing of shape information in human peripheral visual fields is impeded beyond what can be expected by poor spatial resolution. Visual crowding, the inability to identify objects in clutter, has been shown to be the primary factor limiting shape perception in peripheral vision. Despite the well-documented effects of crowding, its underlying causes remain poorly understood. Given that spatial attention both facilitates learning of image statistics and directs saccadic eye movements, we propose that the acquisition of image statistics in peripheral visual fields is confounded by eye-movement artifacts. Specifically, the image statistics acquired under a peripherally deployed spotlight of attention are systematically biased by saccade-induced image displacements. These erroneously represented image statistics lead to inappropriate contextual interactions in the periphery and cause crowding.

In humans, the central 2° of the visual field is extensively represented in both the retina (with the highest cone density in the fovea) and the primary visual cortex (V1), and is therefore well suited for tasks that require resolving fine visual details. Saccadic eye movements allow the visual system to bring objects of interest to the central visual field (that is, to foveate). The rest of visual space falls on the peripheral retina. Shape perception, or form vision, suffers in the visual periphery. Among the various deficits in peripheral form vision, perhaps the most disruptive ones are those attributable to visual crowding. Crowding is the inability to recognize target objects in clutter (Fig. 1a). In the periphery, surrounding objects (flankers) that are within a critical distance of the target impair target identification. This deficit cannot be explained by the lower spatial resolution in the periphery. In normally sighted individuals, crowding is less consequential because it is compensated by foveating saccades. However, crowding is detrimental for individuals who do not have a functioning fovea, as they must rely on their peripheral visual fields for everyday tasks, such as reading and object recognition. There has been persistent, but unresolved, debate about the neural underpinnings of crowding (see ref. 1 for a review) since Bouma formally described crowding four decades ago. Many theories invoke some form of pre-attentive processing in the early stages of visual processing, such as inappropriate feature integration or positional averaging, as the underlying cause of crowding. Others claim a lack of spatial resolution in the attentional mechanism itself as the primary cause.

The crowding zone, the spatial extent over which flankers affect target identification, exhibits several robust characteristics (Fig. 1b). First, the size of the crowding zone scales linearly with eccentricity. Along the radial axis (the line connecting the fovea to the target), the crowding zone extends roughly to half the target eccentricity. This is referred to as Bouma’s law. Second, flankers have an asymmetric effect on the target in that an outward flanker, that is, the one more eccentric than the target, has a greater crowding effect than an equally spaced inward (less eccentric) flanker. We refer to this as inward-outward asymmetry. Third, the crowding zone is not circular, but is instead markedly elongated along the radial axis so that radially positioned flankers produce more interference than laterally (that is, tangentially) positioned flankers. We refer to this as radial-tangential anisotropy. Any viable model of crowding must reproduce these well-defined properties of the crowding zone.

Previous studies have offered explanations for some, but not all, of these characteristics. Bouma’s law on scaling has been addressed in terms of ‘combining fields’ that are implemented by a fixed number of cortical neurons irrespective of eccentricity. The inward-outward asymmetry has been explained in terms of asymmetric cortical distances of near and far flankers that are otherwise equidistant from the target in visual space. It has also been speculated that ‘ecological factors’, including optic flow and saccadic eye movements, might underlie the inward-outward asymmetry. Currently, there is no satisfactory explanation for radial-tangential anisotropy. Past studies have chosen to incorporate anisotropy as an assumption in their models. Moreover, no existing model of crowding can simultaneously account for all three spatial characteristics of the crowding zone much less explain the possible neural underpinnings of crowding. We propose such a unified model, with all but one of the parameters constrained by anatomical and behavioral data from studies unrelated to crowding, and provide testable predictions of the model, some with pertinent clinical implications.

Statistical regularities of the visual environment are thought to be important for shaping the connectivity and the response properties of the visual cortex. Receptive field properties of neurons in V1 can be derived from the statistics of natural images. The response of a V1 neuron further depends on the context surrounding the
neuron’s classical receptive field. Such contextual interactions are mediated in part by anatomical connections extending laterally across multiple cortical columns. The patterns of these lateral connections suggest that orientation statistics of natural images have shaped their formation.

We argue that the acquisition of the orientation statistics of natural images in peripheral vision is confounded by eye movements. Specially, we propose that the same spatial attentional mechanism that directs gaze and helps to acquire relevant image statistics in central vision causes an acquisition of misrepresented image statistics in peripheral vision. Given the fundamental importance of orientation statistics in form vision, these erroneously represented image statistics would, in turn, lead to contextual interactions in the periphery that are inappropriate for form vision and cause crowding.

Attending a spatial location promotes neural responses and enhances contextual effects at the attended location. Spatial attention also mediates learning in the visual cortex. We assume that spatial attention is important for promoting the learning of image statistics at the attended location and facilitating the formation of cortical connectivity that conforms to these statistics. By image statistics, we specifically refer to the pair-wise statistics of oriented edges, which we assume are encoded in terms of the functional weights of lateral connections. In other words, if two nearby neurons tend to be correlated in their stimulus-evoked activity, then, under the spotlight of attention, they are more likely to form long-lasting lateral connections that encode this correlation.

Another important role of spatial attention is to drive saccadic eye movements. Covert shifts of spatial attention to salient objects in the periphery are typically followed by a saccadic eye movement that brings the fovea to the peripheral target (Fig. 2a). If we assume that the onset of the saccade happens before the retraction of the attentional spotlight, a critical difference between central and peripheral vision emerges: the window of temporal overlap between spatial attention and saccade-produced image displacement is present in the periphery, but not in the fovea (Fig. 2b).

Thus, the learning of orientation statistics at any particular peripheral cortical location in V1 will essentially be confounded by the saccade-produced image displacement, which is not part of the natural scene. This should cause an overestimation of repeated patterns along the direction that connects the saccade target in the periphery to the fovea (radial direction). We refer to such misrepresented statistics as saccade-confounded image statistics. These saccade-confounded statistics, if represented in lateral connections, would lead to inappropriate and radially biased contextual interactions in the periphery. The inappropriate contextual interactions would lead to crowding and form the basis of an elongated crowding zone, with the long axis pointing toward the fovea.

RESULTS

Here we present a quantitative model that implements our theory on the anisotropic processing of image statistics in peripheral vision.
Our model is based on three crucial and specific assumptions, the first two of which have been well established. The first assumption is that the acquisition of image statistics occurs primarily at attended spatial locations\textsuperscript{21,22}. It is worth noting that the feedback connections from the secondary visual cortex (V2) to V1, which may mediate top-down attention, have roughly the same anatomical spread (~6 mm in radius, independent of eccentricity) in V1 as do the lateral connections in V1 (ref. 18). Thus, our second assumption is that the physiological footprint of spatial attention, be it defined by the spatial extent of the lateral or the feedback connections, is constant in size (6 mm in radius) in V1 and is independent of eccentricity. Finally, we assume that spatial attention and any subsequent eye movement that it elicits overlap in time; that is, the eyes move before the spotlight of attention that elicited the eye movement is fully retracted.

The underlying cortical architecture of our model consists of a mosaic of cortical ‘hypercolumns’ (Fig. 3a and Online Methods) that are laterally connected. Each model hypercolumn consists of a set of filters that extract orientation information from a local region of visual space (receptive field) in a fashion that is analogous to that of orientation tuned neurons in a V1 cortical column. The receptive fields of the hypercolumns tile visual space and their sizes scale linearly with eccentricity\textsuperscript{23}. We refer to the set of hypercolumns with which a reference hypercolumn has lateral connections as the lateral interaction zone (Fig. 3a).

**Geometry of lateral interactions**

Our model assumes that the physiological footprint of the spotlight of spatial attention and the lateral interaction zone of a reference hypercolumn are approximately the same in V1. They are isotropic on the cortex and independent of eccentricity. Assuming that the lateral connections in an interaction zone are modified under the spotlight of spatial attention, a geometric analysis of the footprint of spatial attention or, equivalently, the lateral interaction zone, should reveal the maximum spatial extent of crowding.

The spatial extent of lateral interaction zones of constant cortical size scales up with eccentricity (Fig. 3b). To quantify this result, we calculated the end-to-end extent of the receptive fields along the radial axis (Fig. 3c). The coincidence with Bouma’s law is simply a result of the linear scaling of the receptive fields with eccentricity and the cortical size of the interaction zone being independent of eccentricity. The radius of the interaction zone that is required to match Bouma’s law is about six hypercolumns (Online Methods), which is consistent with the measured extent of lateral connections in V1 (ref. 18).

Furthermore, when we split the end-to-end radial extent into two parts, the distance from the receptive field center of the reference hypercolumn to the outer extremity ($d_{\text{out}}$) and to the inner extremity ($d_{\text{in}}$) of the interaction zone is plotted against the eccentricity of the reference hypercolumn. That $d_{\text{out}}$ is always greater than $d_{\text{in}}$ (for nonzero eccentricities) explains the inward-outward asymmetry.

**Saccade-confounded image statistics**

Our basic premise is that spatial attention serves a dual role of driving saccadic eye movements and facilitating the learning of orientation statistics of the visual world. We further assume a temporal overlap between the deployment of spatial attention and the subsequent saccade that it drives. To examine the nature of image statistics seen at a peripheral location under such conditions, we performed simulations of saccadic eye movements. The simulated system makes saccades to different attended locations in the periphery.

In the context of a visual scene, we measured pair-wise joint spiking statistics (mutual information; Online Methods, equation (11)) between each of the oriented filters in a reference hypercolumn and each of the oriented filters in neighboring hypercolumns in the lateral interaction zone (Fig. 4a). Such pair-wise statistics may determine the strength of lateral interactions between V1 neurons. In our simulations, spatial attention (Fig. 2b) enables the learning of these

---

**Figure 3** Spatial consequences of isotropic lateral interaction zone in V1. (a) A simple geometry of V1 is assumed with cortical hypercolumns arranged in a hexagonal mosaic. The receptive fields of the computational elements in the hypercolumns scale up linearly with eccentricity. Each hypercolumn is assumed to have lateral (long-range horizontal) connections in an isotropic neighborhood of hypercolumns on the cortex (lateral interaction zone). The radius of the neighborhood in cortical distance or, equivalently, the number of hypercolumns is independent of eccentricity. (b) The extent of the lateral interaction zone is projected to visual space for three reference hypercolumns at eccentricities 2°, 4° and 6°. The radius of the zones is six hypercolumns as suggested by several studies (Online Methods). (c) Half the end-to-end distances of the interaction zones along the radial axis (the line joining the receptive field center of the hypercolumn to the fovea) are plotted against the eccentricity of the corresponding reference hypercolumn. The dashed line is the prediction of Bouma’s law\textsuperscript{2} (Fig. 1b). (d) The radial distance from the receptive field center of a reference hypercolumn to the outer extremity ($d_{\text{out}}$) and to the inner extremity ($d_{\text{in}}$) of the interaction zone is plotted against the eccentricity of the reference hypercolumn. That $d_{\text{out}}$ is always greater than $d_{\text{in}}$ (for nonzero eccentricities) explains the inward-outward asymmetry.
pair-wise mutual information (Online Methods, equation (12)) was calculated between an oriented filter in the reference hypercolumn and each of the neighboring oriented filters in the interaction zone. Two such filters in the reference hypercolumn were selected for illustration: one of the filters (green) is oriented along the radial axis, whereas the other (blue) is oriented along the orthogonal tangential axis. (b,c) True (veridical) statistics of the simulated visual environment: pair-wise mutual information between the reference filters (green in b, blue in c) and all neighboring filters in the interaction zone, gathered under a spotlight of attention without eye movements. At each neighboring location, the oriented thick and thin bars depict the oriented filters at this location. The colors of the oriented bars depict the magnitude of the mutual information. For each hypercolumn, the oriented filter with the highest mutual information is highlighted with a thick line. (d,e) Saccade-confounded statistics: pair-wise mutual information of the same simulated visual environment gathered under the attentional spotlight during temporally overlapping eye movements. Although the veridical statistics implicate smooth continuation of contours (illustrated with overlaid dashed circles in c), the saccade-confounded ones favor repetition of co-oriented fragments (overlaid dashed lines in e). The time constant of the decay of spatial attention ($\lambda$) was 16 ms for these simulations.

Predictions of the theory
We argue that the patterns of lateral connectivity in the cortex, and therefore the nature and extent of crowding, reflect the patterns of saccadic eye movements and the statistics of the visual world. Besides providing a unified and coherent account of the existing data on crowding, our theory also predicts that changes in the pattern of saccades or in the statistics of the visual world will lead to reorganization in the patterns of lateral connectivity and to the shape of the crowding zone. Here we offer several empirically testable predictions regarding the shape of the crowding zone in situations in which either the pattern of saccades or the image statistics deviate from the normal scenario that we have considered so far.

Given that most saccades in humans have magnitudes of 15° or less,$^{31}$ our theory predicts that the radial-tangential anisotropy
Figure 5 Zones of inappropriate integration. The pooled and normalized differences (Online Methods, equation (13)) between saccade-confounded (Fig. 4d,e) and veridical (Fig. 4b,c) image statistics (mutual information) between a reference hypercolumn and neighboring hypercolumns are shown in visual space for three reference hypercolumns at 2°, 4° and 6°. The color scale shows the magnitude and sign of the deviation from the veridical statistics, indicative of inappropriate integration: shades of red indicate that the mutual information between a reference hypercolumn and an adjacent hypercolumn is higher in saccade-confounded statistics than in veridical statistics, implying over-integration; shades of blue indicate lower mutual information than the veridical, implying under-integration. Elliptical fits (dashed lines at 40% of peak normalized mutual information) indicate lower mutual information than the veridical, implying under-integration; shades of red indicate higher mutual information than the veridical, implying over-integration. The color scale shows the magnitude and sign of the deviation from the veridical statistics, indicative of inappropriate integration: shades of red indicate that the mutual information between a reference hypercolumn and an adjacent hypercolumn is higher in saccade-confounded statistics than in veridical statistics, implying over-integration; shades of blue indicate lower mutual information than the veridical, implying under-integration. Elliptical fits (dashed lines at 40% of peak normalized mutual information) indicate lower mutual information than the veridical, implying under-integration; shades of red indicate higher mutual information than the veridical, implying over-integration. The color scale shows the magnitude and sign of the deviation from the veridical statistics, indicative of inappropriate integration: shades of red indicate that the mutual information between a reference hypercolumn and an adjacent hypercolumn is higher in saccade-confounded statistics than in veridical statistics, implying over-integration; shades of blue indicate lower mutual information than the veridical, implying under-integration. Elliptical fits (dashed lines at 40% of peak normalized mutual information) indicate lower mutual information than the veridical, implying under-integration; shades of red indicate higher mutual information than the veridical, implying over-integration.

DISCUSSION

Our model explains the qualitative differences in form vision between the fovea and periphery, as exemplified by visual crowding, without having to postulate a specialized mechanism that is not shared between central and peripheral vision. We began by assuming that lateral interaction zones in V1 are isotropic and constant in size on the cortex. Specific interactions in the zones are learned under the spotlight of attention, which overlaps in time with the subsequent saccadic eye movements that it elicits. We found that this minimal set of assumptions can explain form vision deficits in peripheral vision. Specifically, we found that the scaling law and the inward-outward asymmetry of crowding are consequences of the extent of lateral connections in V1 being isotropic and independent of eccentricity, and the sizes of the receptive fields of V1 neurons increasing linearly with eccentricity. The elliptical shape of the crowding zone can be caused by distorted image statistics encoded in lateral connections between V1 hypercolumns. The distortion is attributable to the fact that spatial attention facilitates the acquisition of image statistics at the attended retinal location and that there is temporal overlap between the duration of the spatial attention at a retinal location and the saccade that it elicits. Given that saccades in normal vision are generally radial with respect to the fovea, the acquired image statistics are mostly confounded in the radial direction.

Our quantitative results illustrate an important aspect of the anomalous contextual interactions underlying crowding that has not been fully explored empirically: diminished binding of target features resulting from proximal weakening of connectivity combined with inappropriate and spurious binding of distracter features resulting from distal strengthening of connectivity in the lateral interaction zone. This dual nature of the binding deficiency explains our previous finding with classification images that crowding reduces the use of valid features while increasing the number of invalid features used by the visual system. Furthermore, the co-oriented connectivity pattern (Fig. 4d,e) suggests a texture-like processing of the peripheral field, rather than a Gestalt-like smooth contour integration process. We surmise that such a texture-like representation of the peripheral field,

would be less pronounced for eccentricities beyond 15° and should approach the aspect ratio predicted by the cortically isotropic lateral interaction zone alone (about 1.05). The crowding zone in infants should be similarly circular and defined mainly by the geometry of the lateral interaction zone, as their visual systems would not have had sufficient exposure to the biased statistics resulting from saccades. Amblyopic individuals with strong foveal crowding are also likely to have circular crowding zones at the fovea, where there should not be a directional bias in the saccade statistics, particularly in cases of anisometropic amblyopia in which there is no gaze offset between the amblyopic and the fellow eye.

Prevailing differences in image statistics between the upper and lower visual fields should result in different shapes and spatial properties of crowding zones across the horizontal meridian. We made detailed measurements of the crowding zone (Supplementary Fig. 2 and Supplementary Note) in the lower and upper visual field and found evidence that the crowding zones in the upper visual field were less elongated than those in the lower visual field (Supplementary Fig. 3 and Supplementary Table 1). This difference likely reflects the greater incidence of oriented structure that is typically present in the lower visual field compared with the upper field. These statistics would help drive the greater elongation in the lower field.

Neurons in V1 typically respond most vigorously to moving stimuli whose orientation is orthogonal to the direction of motion. Our theory would predict a greater spatial extent of crowding for flankers oriented orthogonally to the radial axis as compared to those oriented in parallel to the radial axis. We measured the spatial extent of crowding for such oriented flankers in both radial and tangential arrangements (Supplementary Fig. 4a and Supplementary Note). Data from six observers revealed that flankers oriented orthogonally to the radial axis had a greater extent of crowding irrespective of their spatial arrangement, consistent with our prediction (Supplementary Fig. 4b).

Many individuals with central vision loss as a result of age-related macular degeneration develop the use of a stable retinal locus in the periphery for fixation during form-vision tasks. This is known as the preferred retinal locus (PRL) and is typically located just outside the central scotoma. Given that the stable PRL is used for fixation, saccadic eye movements for some of these individuals are radial with respect to the PRL and not to the anatomical fovea, which is in the scotoma. Under such circumstances, the visual system is exposed to PRL-centric saccade statistics, and our theory would predict that the crowding zone measured at the PRL should no longer be elongated, as the PRL no longer experiences a radial bias in eye movements, and that the elongated axes of the crowding zones at other peripheral locations should point toward the PRL (Supplementary Fig. 5). Preliminary results measured with a scanning laser ophthalmoscope from individuals with age-related macular degeneration suggest that the zone of crowding measured at the PRL is indeed circular (S.T.L. Chung & Y. Lin, ARVO Abstr., 49:1509, 2008). Further studies are needed to determine the shape of the crowding zones at non-PRL locations and to assess the time course of the predicted reorganization.
although insufficient for accurate object identification, may serve a useful purpose such that there is no ecological reason to impose a strict temporal separation between covert spatial attention and the subsequent saccade that it elicits. Suppression of detailed form information from the vast expanse of the peripheral fields might prevent upstream object processing areas (for example, LOC) from getting overloaded, whereas the texture-like representation may aid in the detection of salient objects. Three issues raised by our theory warrant additional scrutiny: temporal overlap between attention and saccades, saccadic suppression, and the neural loci of crowding.

Temporal overlap between attention and saccades

Although attention has been a highly active area of research, the temporal dynamics concerning the extinction of attention during saccadic eye movements, as opposed to immediately before or after a saccade, has not been characterized. We assumed an exponential decay function to model the temporal overlap between attention and a saccade and chose to parametrically explore the effect of varying the time constant of the decay. Our simulation results indicate that even moderate values of overlap between attention and saccadic eye movement, as little as 4 ms, are able to produce anisotropy in lateral connection weights. Electrophysiological and psychophysical experiments are needed to confirm the parameters of the overlap.

It is possible that there could be a small, but significant, temporal overlap between spatial attention and eye movements, even at the fovea. For example, this could happen if attention is divided between the fovea and the periphery. In this case, the periphery will continue to exhibit the radial bias, whereas the bias at the fovea will essentially be isotropic. This is consistent with the finding that interaction zones in the fovea are approximately circular.

Saccadic suppression

One of the objections that could be raised to our model is that the phenomenon of saccadic suppression would prevent the retinal motion blur from affecting the plasticity of the early visual cortex. There is considerable debate about the mechanisms underlying saccadic suppression; some have argued for an extra-retinal suppressive mechanism, whereas others have argued for a visual-masking mechanism.

Although both mechanisms might contribute toward saccadic suppression, albeit unequally, there is little evidence of complete suppression in the early visual cortex. Instead there is a growing consensus that peri-saccadic stimuli are indeed processed by the early visual system and that these signals are prevented from reaching awareness at a later stage in visual processing. By attributing crowding to contextual interactions in V1, we allow crowding to be shaped by retinal motion blur, yet remain consistent with the observation that any such eye-motion induced motion is not perceived under normal circumstances.

Anisotropy and the neural loci of crowding

Area V2 has been proposed as a possible locus of crowding because the scaling of its receptive fields with eccentricity matches that of the crowding zones; however, radial-tangential anisotropy is not evident in V2 receptive fields and the theory that implicated V2 did not address the issue of anisotropy. Area V4 has also been proposed as a result of the reported anisotropy in V4 receptive field size. There is evidence that a V4 receptive field represents a convergence of information from a circular patch of V1. As illustrated in our geometric analysis of the lateral interaction zone, this anisotropy, with an average aspect ratio of 1.05, is insufficient to explain the human data (aspect ratio ≈ 2.2). This finding lends credence to our theory that crowding originates in V1 as a result of extra-classical interactions. At the same time, our theory does not preclude the possibility that crowding occurs at multiple levels in the visual system.

The V4 finding further suggests that the anisotropy in the crowding zone cannot be a result of any anisotropy in the CMF along the radial and tangential axes. With the purported anisotropy in the CMF, as was suggested in a functional magnetic resonance imaging study, a circle on the cortex will project to an ellipse in visual space, but with the major axis along the tangential direction, orthogonal to the observed crowding zone.

Conclusion

Form vision in the periphery is markedly degraded beyond its limited spatial resolution, as demonstrated by the phenomenon of crowding. We found that a small amount of temporal overlap between spatial attention and saccadic eye movements can cause the acquisition of erroneous image statistics by the neurons in the visual cortex that serve peripheral vision.

The misrepresented statistics exhibit preferences for co-orientation and repetition and are spatially elongated along the radial axis. The consequent contextual interactions would thus render object identification against a cluttered background particularly difficult in the periphery. The spatial extents of the inappropriate interactions dictated by our theory quantitatively match the observed size and scaling of the zones of visual crowding.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/natureneuroscience/.

Note: Supplementary information is available on the Nature Neuroscience website.

ACKNOWLEDGMENTS

The authors would like to thank I. Biederman, A. Disney, J. Hirsch, M. Jadi, R. Millin and J. Reynolds for helpful comments on the manuscript. This research was supported by US National Institutes of Health grant EY017707 (B.S.T.).

AUTHOR CONTRIBUTIONS

A.S.N. and B.S.T. developed the theory, designed the experiments and collected the data. A.S.N. analyzed the data and ran the model simulations. A.S.N. and B.S.T. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at http://www.nature.com/natureneuroscience/.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Levi, D.M. Crowding—an essential bottleneck for object recognition: a mini-review. Vision Res. 48, 635–654 (2008).
2. Bouma, H. Interaction effects in parafoveal letter recognition. Nature 226, 177–178 (1970).
3. Levi, D.M., Harirhan, S. & Klein, S. Suppressive and facilitatory spatial interactions in peripheral vision: peripheral crowding is neither size invariant nor simple contrast masking. J. Vis. 2, 167–177 (2002).
4. Pelli, D.G., Palomares, M.C. & Majaj, N.J. Crowding is unlike ordinary masking: distinguishing feature integration from detection. J. Vis. 4, 1136–1169 (2004).
5. Nandy, A.S. & Tjan, B.S. The nature of letter crowding as revealed by first- and second-order classification images. J. Vis. 7, 1–26 (2007).
6. Greenwood, J.A., Bex, P. & Dakin, S.C. Positional averaging explains crowding with letter-like stimuli. Proc. Natl. Acad. Sci. USA 106, 13130–13135 (2009).
7. He, S., Cavanagh, P. & Intriligator, J. Attentional resolution and the locus of visual awareness. Nature 383, 334–337 (1996).
ONLINE METHODS

Geometry of V1. We assumed a columnar architecture of the primary visual cortex (V1) with the cortical hypercolumns packed hexagonally in cortical space (Fig. 2a). The computational units in each hypercolumn have receptive fields centered at the same retinal location, but each was tuned to different orientations and spatial scales. For our purposes, we were concerned mainly with the orientation tuning of these units. We used eight broadband oriented filters to extract the corresponding orientation energy in the image patch under the receptive field (θ = [0, 0, ..., θ], radians; Fig. 4a). The average diameter of the receptive fields increase linearly with eccentricity with a slope of 0.1 (ref. 24).

We found that this geometry captures the logarithmic cortical magnification as follows. Let \( d_\theta \) be the receptive field diameter (in degrees) at eccentricity \( \theta = 0 \), \( s \) be the slope of the linear function relating receptive field diameters to eccentricity and \( \gamma \) be the proportion of receptive field diameter overlap between adjacent hypercolumns (assumed constant across all eccentricities). In a V1 hypercolumn, the eccentricity \( \phi \) of an receptive field at the cortical position \( p \), measured in units of hypercolumns (\( p = 1 \) represents the center of the fovea), is given by

\[
\phi(p) = \frac{\alpha}{\beta - 1} (\beta^{p-1} - 1)
\]  

where \( \alpha = d_\theta (1 - \gamma) \) and \( \beta = 1 + s(1 - \gamma) \). For our model, \( d_\theta = 0.01 \), \( s = 0.1 \) (ref. 24) and \( \gamma = 0.3 \) (average value in ref. 49). Away from the immediate vicinity of the center of gaze, equation (1) can be written in approximate form as

\[
\phi(p) = \frac{\alpha}{\beta - 1} \beta^p
\]  

Inverting the function, we can express \( p \) as a function of \( \phi \) as

\[
p(\phi) = \frac{\ln(\phi)}{\ln(\beta)} - K
\]  

where

\[
K = \frac{\ln(\frac{1}{1 - \gamma})}{\ln(\beta)}.
\]  

Equation (3) thus gives the logarithmic cortical magnification outside of the immediate center of gaze.

If we assume the critical spacing in the visual field for crowding to be \( \Delta \phi = \phi_b \), where \( \phi_b \) is Bouma’s constant (about 0.5), then the critical spacing in the cortex is

\[
\Delta p = p(\phi + \Delta \phi) - p(\phi) = \frac{\ln(\phi + \Delta \phi) - \ln(\phi)}{\ln(\beta)} = \frac{\ln(1 + b)}{\ln(\beta)}
\]  

where \( b = 0.5 \), we get \( \Delta p \approx 6 \). That is, the critical spacing for crowding corresponds to six hypercolumns in V1, independent of eccentricity. This is consistent with the anatomical extent of lateral (long-range horizontal) connections in V1 (ref. 18) and with the estimated extent of combining fields in V1 (ref. 10) if each hypercolumn is roughly 1 mm on the cortex.

Conversely, equation (4) shows that, if a computational unit in a particular hypercolumn has lateral connections to all computational units in neighboring hypercolumns up to an isotropic extent of a constant number of hypercolumns on the cortex, then the resulting spatial interaction in the visual field must follow Bouma’s law of linear scaling (\( \Delta \phi = \phi_b \)). We refer to the set of hypercolumns (Fig. 2a) to which a reference hypercolumn (Fig. 2a) has lateral connections as the lateral interaction zone. In our model, we set the radius of the lateral interaction zone equal to six hypercolumns.

Saccadic eye movements. For the eye-movement simulations, the saccadic velocity profile was modeled as follows. Let \( A \) be the saccade amplitude (the distance between successive fixations in degrees of visual angle), \( T \) the duration and \( v(t) \) the velocity profile of a saccade. \( v(t) \) must satisfy the following conditions:

\[
\begin{align*}
  v(0) &= 0 \\
  v(T) &= 0 \\
  v_{peak} &= v(T) = k \sqrt{A}
\end{align*}
\]  

where \( k \) is a constant. A sinusoidal velocity profile of the following form satisfies the constraints in equation (5) in the range \( 0 \leq t \leq T \)

\[
v(t) = k \sqrt{A} \sin \frac{\pi t}{T}
\]  

Given that \( A = \int_0^T v(t) dt \), we have

\[
T = \frac{\pi}{2k} \sqrt{A}
\]  

The distance traversed, \( D(\tau) \), in time \( \tau \) is therefore given by

\[
D(\tau) = \int_0^\tau v(t) dt = \frac{A}{1 - \cos 2k \tau}
\]  

The distribution of saccade amplitudes along the radial axis from the fovea was modeled as an exponential distribution with the following probability density function: \( f(x) = \lambda e^{-\lambda x} \), \( \lambda = \frac{1}{T} \) and \( \Delta \phi = \phi_b \). The distribution along the iso-eccentric axis was assumed to be uniform.

Eye-movement simulations and image statistics. Using the distribution of saccade amplitudes and the corresponding velocity profile described above, we simulated saccadic eye movements in which the visual stimulus presented to the system was a random clutter of uppercase letters (Palatino font) at various sizes and orientations. For computational tractability, we calculated the outputs of the set of eight broadband oriented filters in each hypercolumn at discrete time points in the interval \( 0 \ldots T \). Each filter measures the contrast energy along a given orientation in the image patch that is in the receptive field of the hypercolumn. Let \( r(t) \) denote the response of one of the filters at time \( t \). The cumulative response of the filter over the time course of the eye movement is

\[
r = \sum_{t=0}^{T} r(t) e^{-\frac{t}{\lambda}}
\]  

where the modulation of spatial attention during its overlap with saccadic eye movement (Fig. 2b) is modeled as an exponential decay function with a time constant \( \lambda \), a free parameter of the model. Such a characterization captures the probabilistic distribution of the overlap period. For the purpose of calculating joint image statistics, the cumulative filter response, \( r \), is first converted into a firing probability \( p \) with a saturating nonlinearity, with \( k \) as an arbitrary constant

\[
p = \tanh(kr)
\]  

Let \( \Theta_{ij} \) be a random variable associated with a filter with orientation \( \theta \) in the reference hypercolumn \( R \). \( \Theta_{ij} \) is equal to 1 if the cell has fired within a temporal window, else it is zero. For simplicity, the temporal window used in our simulations was the entire duration of a saccade. The joint probability distribution \( P(\Theta_{ij}, \Theta_{jN}) \) between the oriented filter in the reference hypercolumn and another oriented filter in a neighboring hypercolumn (N) can be calculated by accumulating and averaging the joint firing probabilities across many eye-movement traces (30,000 in our simulations). To obtain robust estimates of the joint probability distribution, we used the bootstrap procedure. For any saccade trace, the probabilities are accumulated only if both the reference and
the neighboring hypercolumn are under the spotlight of attention. Finally, the statistical dependence between $\Theta_{i,R}$ and $\Theta_{j,N}$ can be calculated in terms of the pair-wise mutual information

$$I(\Theta_{i,R}, \Theta_{j,N}) = \sum_{\Theta_{i,R} = [0,1]} \sum_{\Theta_{j,N} = [0,1]} P(\Theta_{i,R}, \Theta_{j,N}) \log_2 \frac{P(\Theta_{i,R}, \Theta_{j,N})}{P(\Theta_{i,R})P(\Theta_{j,N})} \quad (11)$$

The mutual information is zero when the two random variables are statistically independent.

For a reference hypercolumn $R$, pooled mutual information (pooled across all orientations) between $R$ and a neighboring hypercolumn $N$ is defined as

$$I_{SC}(R; N) = \sum_i \sum_j I_{SC}(\Theta_{i,R}; \Theta_{j,N})$$

$$I_{V}(R; N) = \sum_i \sum_j I_{V}(\Theta_{i,R}; \Theta_{j,N}) \quad (12)$$

where $I_{SC}$ and $I_{V}$ are the pair-wise mutual information for the saccade-confounded and veridical conditions respectively (equation (11)). We express the gross difference between the saccade-confounded and veridical statistics in term of the normalized difference between saccade-confounded and veridical mutual information

$$\Delta I(R; N) = \frac{I_{SC}(R; N) - I_{V}(R; N)}{I_{V}(R; N)} \quad (13)$$

This normalized difference when plotted in visual space for all neighboring hypercolumns maps the amplitude and spatial extent of inappropriate feature integration for a reference hypercolumn. Image features from hypercolumns with negative difference (reduced mutual information in the saccade-confounded statistics compared with the true statistics) would have weaker interactions and thus be only loosely bound to the reference features, leading to an under-integration of features. Conversely, features from hypercolumns with positive difference (excessive mutual information in the saccade-confounded statistics) would strongly influence the reference, leading to excessive and possibly erroneous feature integration.

49. Wilson, J.R. & Sherman, S.M. Receptive-field characteristics of neurons in cat striate cortex: changes with visual field eccentricity. *J. Neurophysiol.* **39**, 512–533 (1976).

50. Lebedev, S., Gelder, P.V. & Tsui, W.H. Square-root relations between main saccadic parameters. *Invest. Ophthalmol. Vis. Sci.* **37**, 2750–2758 (1996).