Saccade-confounded image statistics explain visual crowding

Anirvan S. Nandy[1,*] and **Bosco S. Tjan**[1,2]

[1]Department of Psychology, University of Southern California

[2]Neuroscience Graduate Program, University of Southern California

SUMMARY

Processing of shape information in human peripheral visual fields is impeded beyond what can be expected by poorer 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 are poorly understood. Since 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 is systematically biased by saccade-induced image displacements. These erroneously represented image statistics lead to inappropriate contextual interactions in the periphery and cause crowding.

INTRODUCTION

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 (i.e., 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 from 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 patients who do not have a functioning fovea, since such individuals must rely on their peripheral visual fields for everyday tasks such as reading and object recognition.

*Current address: Systems & Computational Neurobiology Laboratories, The Salk Institute for Biological Studies

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[2]. Many theories invoke some form of pre-attentive processing in the early stages of visual processing—inappropriate feature integration[3–5], positional averaging[6]—as the underlying cause of crowding. Others claim a lack of spatial resolution in the attentional mechanism itself as the primary cause[7].

The crowding zone—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[2]. This is referred to as *Bouma's law*. Second, flankers have an asymmetric effect on the target in that an outward flanker, i.e., the one more eccentric than the target, has a greater crowding effect than an equally spaced inward (less eccentric) flanker[8]. We will refer to this as the *inward-outward asymmetry*. Third, the crowding zone is not circular but is markedly elongated along the radial axis so that radially positioned flankers produce more interference than laterally (i.e. tangentially) positioned flankers[9]. We will refer to this as the *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[10]. 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[11]. It has also been speculated that "ecological factors" including optic flow and saccadic eye movements might underlie the inward-outward asymmetry[12]. Currently, there is no satisfactory explanation for the radial-tangential anisotropy. Past studies have chosen to incorporate anisotropy as an assumption in their models[13]. 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. We provide testable predictions of the model, some with pertinent clinical implications.

THEORY

Statistical regularities of the visual environment are thought to play a key role in shaping the connectivity and the response properties of the visual cortex[14]. Receptive field properties of neurons in V1 can be derived from the statistics of natural images[15,16]. The response of a V1 neuron further depends on the context surrounding the neuron's classical receptive field[17]. Such contextual interactions are mediated in part by anatomical connections extending laterally across multiple cortical columns[18]. The patterns of these lateral connections suggest that orientation statistics of natural images have shaped their formation[19].

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 acquire relevant image statistics in central vision, causes an acquisition of misrepresented image statistics in peripheral vision. Given the fundamental importance of orientation statistics to form vision[20], these erroneously represented image statistics, in turn, would lead to contextual interactions in the periphery that are inappropriate for form vision and cause crowding. We elaborate next on how two well-understood functions of spatial attention can in combination lead to the acquisition of such mis-represented image statistics.

Attending a spatial location promotes neural responses and enhances contextual effects at the attended location[21]. Spatial attention also mediates learning in the visual cortex[22]. We assume that one key role of spatial attention is to promote the learning of image statistics at the attended location and to facilitate the formation of cortical connectivity that conforms to these statistics. By "image statistics", we specifically mean 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 key 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[23] (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 only in the periphery, 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 will refer to such mis-represented statistics as *saccade-confounded image statistics*. These saccade-confounded statistics, if represented in the lateral connections in the periphery, would lead to inappropriate and radially biased contextual interactions. 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

In the following sections 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[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 within V1[18]. Thus we assume 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; i.e., 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; 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 analogous to that of orientation tuned neurons in a V1 cortical column. The receptive fields (RFs) of the hypercolumns tile visual space and scale linearly with eccentricity[24]. We will refer to the set of hypercolumns with which a reference hypercolumn has lateral connections, as the *lateral interaction zone* (Fig. 3a).

We will elaborate our results in two steps: first by considering the geometry of lateral interactions, and then by computing image statistics acquired under the influence of saccades.

Geometry of lateral interactions

As stated above, 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 within 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 RFs along the radial axis (Fig. 3c). The coincidence with Bouma's Law is simply due to the linear scaling of the RFs 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 6 hypercolumns (Online Methods), which is in good agreement with the measured extent of lateral connections in V1[18].

Further, if we split the end-to-end radial extent into two parts—the distance from the RF center of the reference hypercolumn to the outer and to the inner extremity of the radial extent—the asymmetry is readily apparent (Fig. 3d). If we consider distances from the RF of the reference hypercolumn, the RFs at the outer extremity are farther away *in visual space* than the RFs at the inner extremity, but the corresponding hypercolumns are equidistant from the reference hypercolumn *on the cortex*[11].

As has been proposed separately in previous studies[10,11], we have verified that the simple assumption of constant-sized lateral interaction zones in V1, explains both the properties of scaling (Bouma's Law) and the inward-outward asymmetry of the crowding zone. However, to explain the radial-tangential anisotropy of the crowding zone and to account for crowding, we need to know the strength of the lateral interactions.

Saccade-confounded image statistics

As stated earlier, 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 within 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 pair-wise joint statistics within the interaction zone at the attended location. The time constant, λ, which represents the expected temporal overlap between spatial attention and a saccade, is the only undetermined parameter of our simulated system (others have been set by anatomical and behavioral data from published studies unrelated to visual crowding).

The true ("veridical") statistics of the simulated visual environment consist of co-circular patterns consistent with second-order orientation statistics of natural images[17,20] and with patterns that have been proposed in mathematical models[25] (Fig. 4b–c). Such statistics, if encoded in the weights of the lateral connections, would facilitate grouping of related features into continuous contours. In contrast, saccade-confounded statistics (Fig. 4d–e, 30,000 simulated saccades, $\lambda = 16$ ms) deviate from true statistics in two major aspects: (a) there is a preference for co-orientation[18,26] and (b) the spatial extent of the mismatch between the veridical and the confounded statistics has a strong radial bias irrespective of the reference orientation. Essentially, as a result of the saccade, a given representation of a visual pattern under the attentional spotlight repeats itself across hypercolumns whose receptive fields lie along the eye-movement trajectory.

Contextual interactions, if dominated by co-orientation as in the saccade-confounded statistics, should bias either the integration of features of similar orientation into a texture field or the estimation of orientations to be the average of the surround. This provides a basis for the finding that crowding appears to be a process of averaging[27,28]. It also offers a rationale for the recent proposals that crowding is caused by a peripheral visual system that encodes texture as opposed to form information[29,30] and justifies the future development of these theories to include anisotropic texture processing.

To provide an estimate of the spatial extent and strength of the putative inappropriate feature interactions in the peripheral field due to the use of non-veridical image statistics, we computed the deviation of saccade-confounded statistics from veridical statistics (Fig. 5; Online Methods, Equation 13). The overall region of inappropriate integration is elongated along the radial axis and has the shape and anisotropic extent of the psychophysically measured zone of crowding (Fig. 1b), with an aspect ratio between 1.54 and 2.48, depending on eccentricity. Our calculations further reveal a zone of under-integration in the proximal neighborhood of the reference hypercolumn and a zone of over-integration is in the distal neighborhood. This suggests that the process of inappropriate integration, thought to be the underlying cause of crowding, is two-fold: features from the target object are weakly bound while features from the clutter surrounding the target are excessively bound. The qualitative shape of the zones (radial elongation, proximal under-integration, distal over-integration) is preserved across moderate values of the parameter λ (4 or 8 ms; Supplementary Fig. 1). However, the simulations suggest that a larger time constant (about 16 ms as shown in Fig. 5) is necessary to match the spatial extent dictated by Bouma's law. We note that the learning process in our simulations is idealized and noise free. In a real neural system, the process of attention-gated learning will essentially be subjected to noise and reflect the frequency of exposure to the statistics.

Predictions of the theory

We have argued that the patterns of lateral connectivity in the cortex, and hence 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 would lead to reorganization in the patterns of lateral connectivity and hence to the shape of the crowding zone. Here we offer several empirically testable predictions regarding the shape of the crowding zone in situations where either the pattern of saccades or the image statistics deviate from the normal scenario that we have been considering so far.

Since most saccades in humans have magnitudes of 15° or less[31], our theory predicts that the radial-tangential anisotropy 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, since their visual systems would not have had sufficient exposure to the biased statistics due to saccades. Amblyopic individuals with strong foveal crowding[32] 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 where there is no gaze offset between the amblyopic and the fellow eye.

Prevailing differences in image statistics between the upper and lower visual fields[33] 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 text for experiment design) in the lower and upper visual field and found evidence that the

crowding zones in the upper visual field are less elongated as compared to those in the lower visual field (Supplementary Fig. 3, Supplementary Table 1). This difference likely reflects the greater incidence of oriented structure typically present in the lower visual field compared to 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 text for experiment design). Data from six observers show that flankers oriented orthogonally to the radial axis have a greater extent of crowding irrespective of their spatial arrangement, in agreement with our prediction (Supplementary Fig. 4b).

Many patients with central vision loss due to age-related macular degeneration (AMD) develop the use of a stable retinal location in the periphery for fixations during form-vision tasks. This is known as the preferred retinal locus (PRL) and is typically located just outside the central scotoma. Since the stable PRL is used for fixations, saccadic eye movements for some of these patients are now radial with respect to the PRL[34] and not to the anatomical fovea, which is within the scotoma. Under such circumstances, the visual system is exposed to PRL-centric saccade statistics, and our theory would predict that (a) the crowding zone measured at the PRL should no longer be elongated since the PRL no longer experiences a radial bias in eye movements and (b) the elongated axes of the crowding zones at other peripheral locations should point toward the PRL (Supplementary Fig. 5). Preliminary results from AMD patients measured with a scanning laser ophthalmoscope 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.

DISCUSSION

We have explained 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 within the zones are learned under the spotlight of attention, which overlaps in time with the subsequent saccadic eye movements it elicits. We have shown that this minimal set of assumptions can explain form vision deficits in peripheral vision. Specifically, we have shown that the scaling law and the inward-outward asymmetry of crowding are consequences of: (a) the extent of lateral connections in V1 being isotropic and independent of eccentricity and (b) 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: (c) the fact that spatial attention facilitates the acquisition of image statistics at the attended retinal location and (d) that there is temporal overlap between

the duration of the spatial attention at a retinal location and the saccade it elicits. Since 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[35] due to proximal weakening of connectivity combined with inappropriate and spurious binding of distracter features due to 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 at the same increasing the number of invalid features used by the visual system[5]. Further, the co-oriented connectivity pattern (Fig. 4d–e) suggests a texture-like processing of the peripheral field[29], rather than a Gestalt-like smooth contour integration process. We surmise that such a texture-like representation of the peripheral field, 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 it elicits. Suppression of detailed form information from the vast expanse of the peripheral fields might prevent upstream object processing areas (e.g., LOC) from getting overloaded, while at the same time the texture-like representation may aid in the detection of salient objects[36].

Three issues raised by our theory warrant additional scrutiny: (a) temporal overlap between attention and saccades, (b) saccadic suppression, and (c) 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 show 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, while the "bias" at the fovea will essentially be isotropic. This is consistent with the finding that interaction zones in the fovea are approximately circular[9].

Saccadic suppression

One of the objections that could be raised against 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 in the literature about the mechanisms

underlying saccadic suppression; some have argued for an extra-retinal suppressive mechanism[37] while others have argued for a visual masking mechanism[38]. While both mechanisms might contribute toward saccadic suppression, albeit unequally[39], there is little evidence of complete suppression in the early visual cortex[40]. Instead there is growing consensus that peri-saccadic stimuli are indeed processed by the early visual system[41] and that these signals are prevented from reaching awareness at a later stage in visual processing[42]. 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-movement induced motion is not perceived under normal circumstances[43].

Anisotropy and the neural loci of crowding

Area V2 has been suggested as a possible locus of crowding because the scaling of its receptive fields with eccentricity matches that of the crowding zones[30]; however, radial-tangential anisotropy is not evident in V2 receptive fields[44] and the theory that implicated V2 did not address the issue of anisotropy. Area V4 has been suggested[1] due to the reported anisotropy in V4 receptive field size[45]. There is recent evidence that a V4 receptive field represents a convergence of information from a circular patch of V1[46]. The observed asymmetry and anisotropy in a V4 receptive field is completely determined by the transformation of visual space according to the cortical magnification factor (CMF) of V1. As illustrated in our geometric analysis of the lateral interaction zone (Fig. 3b), 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 due to extra-classical interactions. At the same time, our theory does not preclude the possibility that crowding occurs at multiple levels in the visual system[47].

The V4 finding[46] further suggests that the anisotropy in the crowding zone cannot be due to any anisotropy in the CMF along the radial and tangential axes[10], as suggested in an fMRI study[48]. With the purported anisotropy in the CMF, 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. In this study, we have shown 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. These 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.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

REFERENCES

1. Levi DM. Crowding–an essential bottleneck for object recognition: a mini-review. Vision Res. 2008; 48:635–654. [PubMed: 18226828]

2. Bouma H. Interaction effects in parafoveal letter recognition. Nature. 1970; 226:177–178. [PubMed: 5437004]

3. Levi DM, Hariharan S, Klein S. Suppressive and facilitatory spatial interactions in peripheral vision: peripheral crowding is neither size invariant nor simple contrast masking. JOV. 2002; 2:167–177.

4. Pelli DG, Palomares MC, Majaj NJ. Crowding is unlike ordinary masking: distinguishing feature integration from detection. JOV. 2004; 4:1136–1169.

5. Nandy AS, Tjan BS. The nature of letter crowding as revealed by first- and second-order classification images. JOV. 2007; 7:5.1–5.26.

6. Greenwood J, Bex P, Dakin SC. Positional averaging explains crowding with letter-like stimuli. Proc Natl Acad Sci USA. 2009; 106:13130–13135. [PubMed: 19617570]

7. He S, Cavanagh P, Intriligator J. Attentional resolution and the locus of visual awareness. Nature. 1996; 383:334–337. [PubMed: 8848045]

8. Petrov Y, Popple AV, McKee SP. Crowding and surround suppression: not to be confused. JOV. 2007; 7:12.1–12.9.

9. Toet A, Levi DM. The two-dimensional shape of spatial interaction zones in the parafovea. Vision Res. 1992; 32:1349–1357. [PubMed: 1455707]

10. Pelli DG. Crowding: a cortical constraint on object recognition. Current Opinion in Neurobiology. 2008; 18:445–451. [PubMed: 18835355]

11. Motter BC, Simoni DA. The roles of cortical image separation and size in active visual search performance. JOV. 2007; 7:6.1–6.15.

12. Petrov Y, Popple AV. Crowding is directed to the fovea and preserves only feature contrast. JOV. 2007; 7:8.1–8.9.

13. van den Berg R, Roerdink JBTM, Cornelissen FW. A neurophysiologically plausible population code model for feature integration explains visual crowding. PLoS Comput Biol. 2010; 6:e1000646. [PubMed: 20098499]

14. Geisler WS. Visual perception and the statistical properties of natural scenes. Annual review of psychology. 2008; 59:167–192.

15. Karklin Y, Lewicki MS. Emergence of complex cell properties by learning to generalize in natural scenes. Nature. 2009; 457:83–86. [PubMed: 19020501]

16. Olshausen BA, Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. Nature. 1996; 381:607–609. [PubMed: 8637596]

17. Kapadia MK, Ito M, Gilbert CD, Westheimer G. Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in V1 of alert monkeys. Neuron. 1995; 15:843–856. [PubMed: 7576633]

18. Stettler DD, Das A, Bennett J, Gilbert CD. Lateral connectivity and contextual interactions in macaque primary visual cortex. Neuron. 2002; 36:739–750. [PubMed: 12441061]

19. Sigman M, Cecchi GA, Gilbert CD, Magnasco MO. On a common circle: natural scenes and Gestalt rules. Proc Natl Acad Sci USA. 2001; 98:1935–1940. [PubMed: 11172054]

20. Geisler WS, Perry JS, Super BJ, Gallogly DP. Edge co-occurrence in natural images predicts contour grouping performance. Vision Res. 2001; 41:711–724. [PubMed: 11248261]

21. Ito M, Gilbert CD. Attention modulates contextual influences in the primary visual cortex of alert monkeys. Neuron. 1999; 22:593–604. [PubMed: 10197538]

22. Gilbert CD, Ito M, Kapadia MK, Westheimer G. Interactions between attention, context and learning in primary visual cortex. Vision Res. 2000; 40:1217–1226. [PubMed: 10788637]

23. Deubel H, Schneider WX. Saccade target selection and object recognition: evidence for a common attentional mechanism. Vision Res. 1996; 36:1827–1837. [PubMed: 8759451]

24. Motter BC. Crowding and object integration within the receptive field of V4 neurons. JOV. 2002; 2:274–274.

25. Ben-Shahar O, Zucker S. Geometrical computations explain projection patterns of long-range horizontal connections in visual cortex. Neural computation. 2004; 16:445–476. [PubMed: 15006089]

26. Bosking WH, Zhang Y, Schofield B, Fitzpatrick D. Orientation selectivity and the arrangement of horizontal connections in tree shrew striate cortex. J Neurosci. 1997; 17:2112–2127. [PubMed: 9045738]

27. Greenwood JA, Bex PJ, Dakin SC. Crowding changes appearance. Curr Biol. 2010; 20:496–501. [PubMed: 20206527]

28. Parkes L, Lund J, Angelucci A, Solomon JA, Morgan M. Compulsory averaging of crowded orientation signals in human vision. Nat Neurosci. 2001; 4:739–744. [PubMed: 11426231]

29. Levi DM, Carney T. Crowding in Peripheral Vision: Why Bigger Is Better. Curr Biol. 2009; 19:1988–1993. [PubMed: 19853450]

30. Freeman J, Simoncelli EP. Metamers of the ventral stream. Nat Neurosci. 2011; 14:1195–1201. [PubMed: 21841776]

31. Bahill AT, Adler D, Stark L. Most naturally occurring human saccades have magnitudes of 15 degrees or less. Invest Ophthalmol Vis Sci. 1975; 14:68–69.

32. Levi DM, Song S, Pelli DG. Amblyopic reading is crowded. JOV. 2007; 7:21.1–21.17.

33. Bruce NDB, Tsotsos JK. A statistical basis for visual field anisotropies. Neurocomputing. 2006; 69:1301–1304.

34. White JM, Bedell HE. The oculomotor reference in humans with bilateral macular disease. Invest Ophthalmol Vis Sci. 1990; 31:1149–1161. [PubMed: 2354915]

35. Neri P, Levi DM. Spatial resolution for feature binding is impaired in peripheral and amblyopic vision. J Neurophysiol. 2006; 96:142–153. [PubMed: 16421195]

36. Itti L, Koch C. Computational modelling of visual attention. Nat Rev Neurosci. 2001; 2:194–203. [PubMed: 11256080]

37. Diamond MR, Ross J, Morrone MC. Extraretinal control of saccadic suppression. J Neurosci. 2000; 20:3449–3455. [PubMed: 10777808]

38. Campbell FW, Wurtz RH. Saccadic omission: why we do not see a grey-out during a saccadic eye movement. Vision Res. 1978; 18:1297–1303. [PubMed: 726271]

39. Wurtz RH. Neuronal mechanisms of visual stability. Vision Res. 2008; 48:2070–2089. [PubMed: 18513781]

40. García-Pérez MA, Peli E. Intrasaccadic perception. J Neurosci. 2001; 21:7313–7322. [PubMed: 11549741]

41. Pisapia ND, Kaunitz L, Melcher D. Backward masking and unmasking across saccadic eye movements. Curr Biol. 2010; 20:613–617. [PubMed: 20346671]

42. Watson T, Krekelberg B. The Relationship between Saccadic Suppression and Perceptual Stability. Curr Biol. 2009; 19:1040–1043. [PubMed: 19481454]

43. Castet E, Jeanjean S, Masson GS. Motion perception of saccade-induced retinal translation. Proc Natl Acad Sci USA. 2002; 99:15159–15163. [PubMed: 12417765]

44. Gattass R, Gross CG, Sandell JH. Visual topography of V2 in the macaque. J Comp Neurol. 1981; 201:519–539. [PubMed: 7287933]

45. Piñon MC, Gattass R, Sousa AP. Area V4 in Cebus monkey: extent and visuotopic organization. Cereb Cortex. 1998; 8:685–701. [PubMed: 9863696]

46. Motter BC. Central V4 receptive fields are scaled by the V1 cortical magnification and correspond to a constant-sized sampling of the V1 surface. J Neurosci. 2009; 29:5749–5757. [PubMed: 19420243]

47. Whitney D, Levi DM. Visual crowding: a fundamental limit on conscious perception and object recognition. Trends Cogn Sci. 2011; 15:160–168. [PubMed: 21420894]

48. Larsson J, Heeger DJ. Two retinotopic visual areas in human lateral occipital cortex. J Neurosci. 2006; 26:13128–13142. [PubMed: 17182764]

49. Wilson JR, Sherman SM. Receptive-field characteristics of neurons in cat striate cortex: Changes with visual field eccentricity. J Neurophysiol. 1976; 39:512–533. [PubMed: 948006]

50. Lebedev S, Gelder PV, Tsui WH. Square-root relations between main saccadic parameters. Invest Ophthalmol Vis Sci. 1996; 37:2750–2758. [PubMed: 8977491]
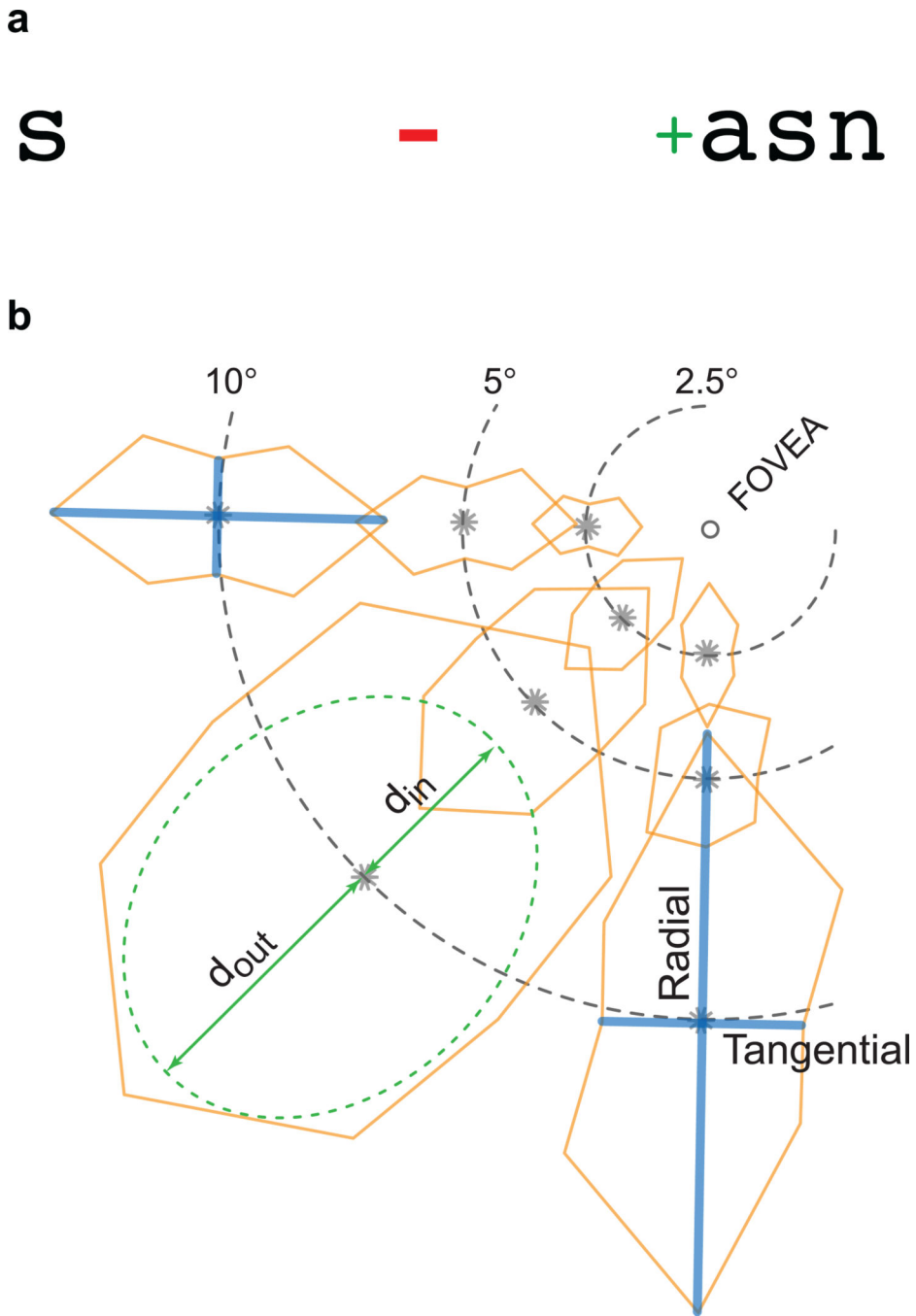
a

b

Figure 1. Characteristics of crowding in peripheral vision

(**a**) Demonstration of crowding: fixating on the red '−' it should be easy to identify the letter s on the left; the equidistant s on the right, which is flanked (crowded) by other letters, is much harder to identify. When fixation is shifted to the green '+', the formerly-crowded s becomes easier to identify. (**b**) The extent of crowding ("crowding zone", orange polygons) can be estimated by measuring target-identification performance at peripheral locations (demarcated by *) with flankers placed at various relative positions around the target. The estimated zones have three robust signatures: they scale up linearly with eccentricity of the

target (Bouma's Law); they are markedly elongated along the axis connecting the target to the fovea (radial axis); when tested with a single flanker (green dotted contour), as opposed to a pair (orange solid contours), flankers that are more eccentric than the target are more effective in crowding the target than are flankers that are less eccentric. (**a**: adapted from REF [10]; **b**: adapted from REF [9])
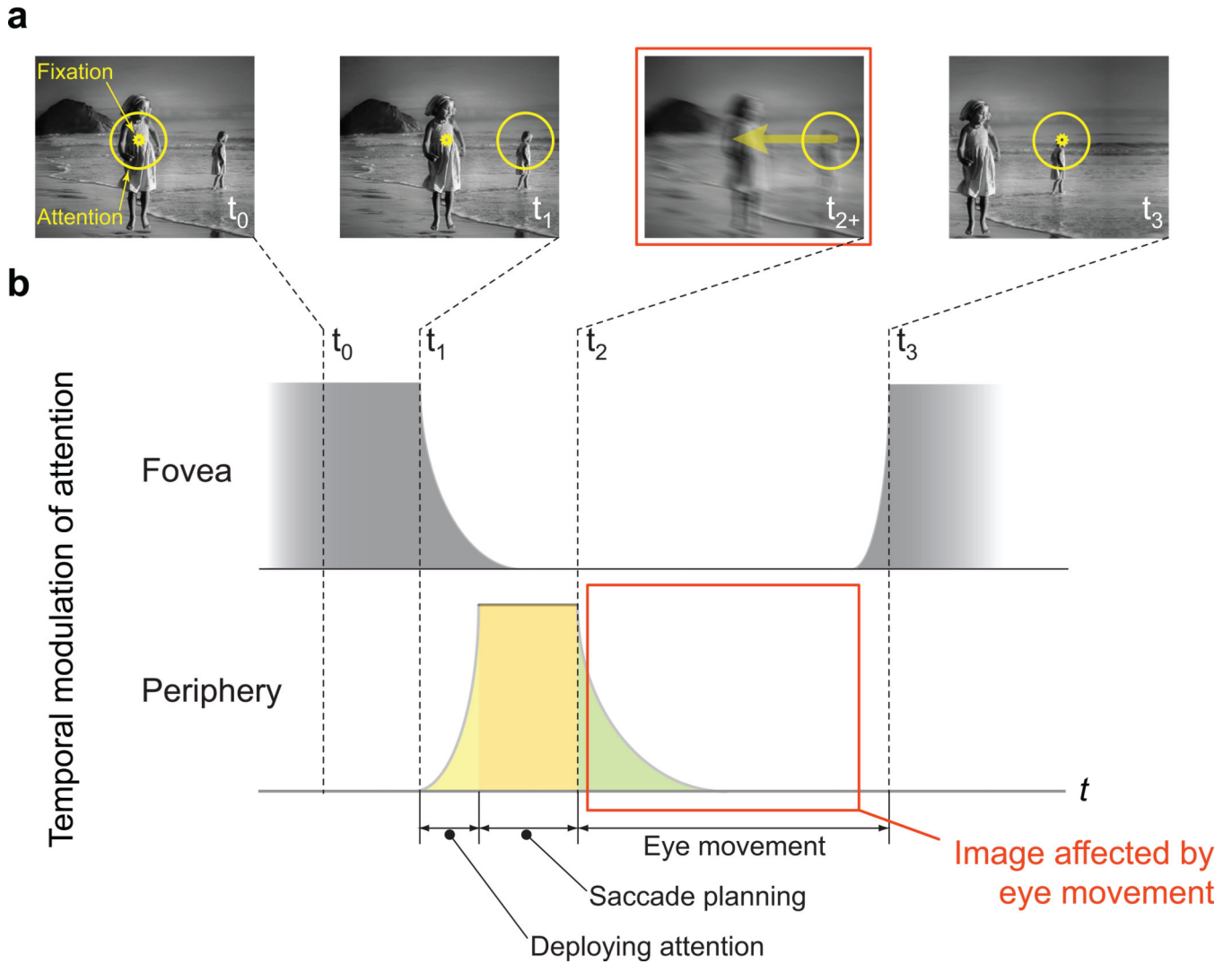
Figure 2. The interaction of spatial attention and saccades
(**a**) Typical sequence of fixation, covert deployment of attention to a salient object in the periphery and subsequent saccade to the attended spot. All illustrations are in retinal coordinates. The arrow (third panel) shows the direction of image displacement during the saccade. (**b**) Schematic of temporal modulation of spatial attention at the fovea and at the peripheral retinal location where covert attention was deployed. Because the saccade is elicited by covert attention, temporal overlap between attention and eye movement is most likely to occur when attention is at the covertly attended location in the periphery. We assume that at this peripheral location, during the time interval from the start of the execution of the saccade (t_2) till the next fixation (t_3) (red box), attention and eye movement overlap with a non-zero probability. The image displacement due to the saccade confounds the image statistics acquired during the overlap.
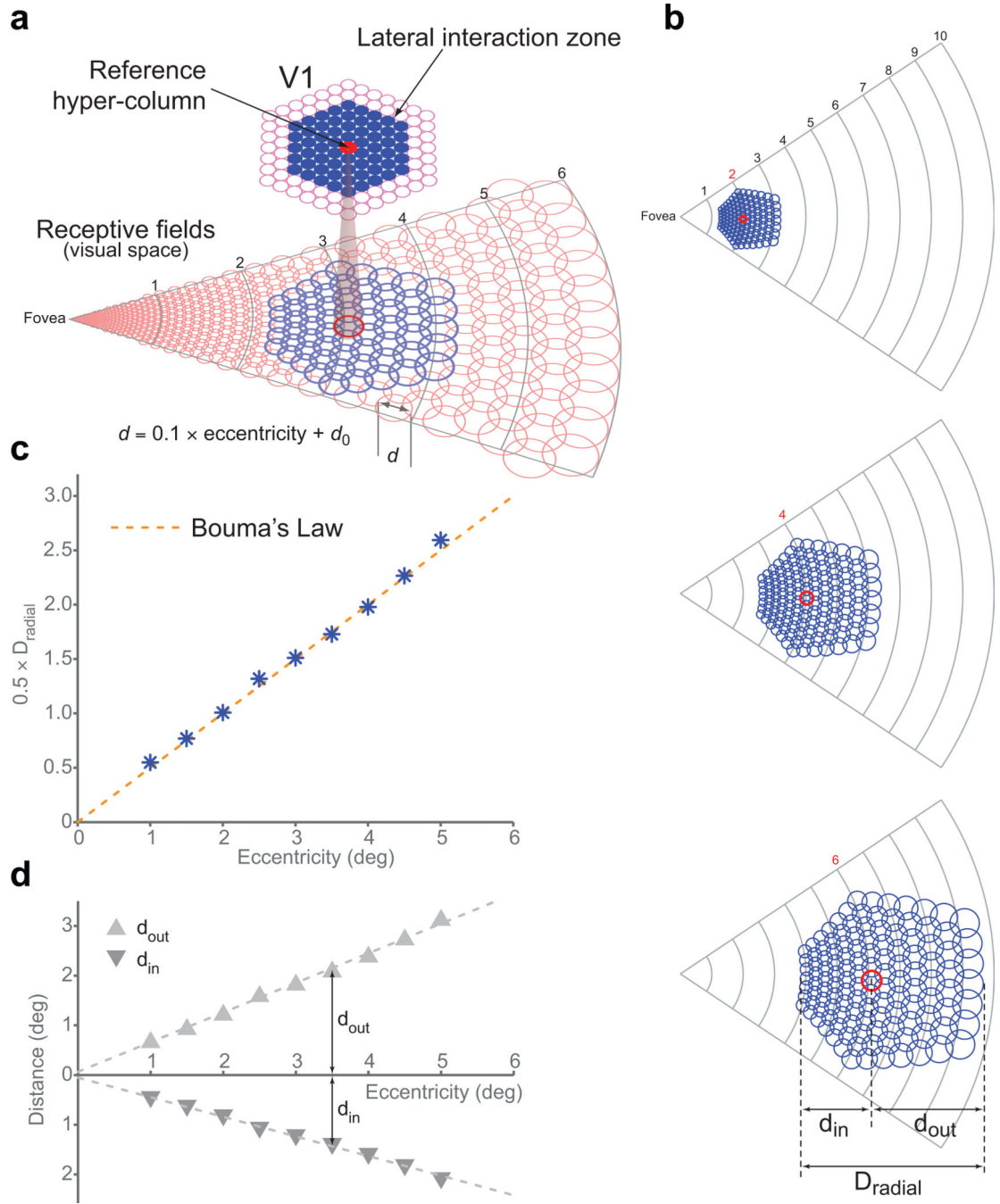
Figure 3. Spatial consequences of isotropic lateral interaction zone in V1
(**a**) A simple geometry of V1 is assumed with cortical hypercolumns arranged in a
hexagonal mosaic. The receptive fields of the computational elements within the
hypercolumns scale up linearly with eccentricity. Each hypercolumn is assumed to have
lateral (long-range horizontal) connections within 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 6 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 dotted line is the prediction of Bouma's law (Fig. 1b; Bouma, 1970). (**d**) The radial distance from the receptive field center of a reference hypercolumn to the outer extremity (d_{out}) and to the inner extremity (d_{in}) of the interaction zone is plotted against the eccentricity of the reference hypercolumn. That d_{out} is always greater than d_{in} (for non-zero eccentricities), explains the inward-outward asymmetry.
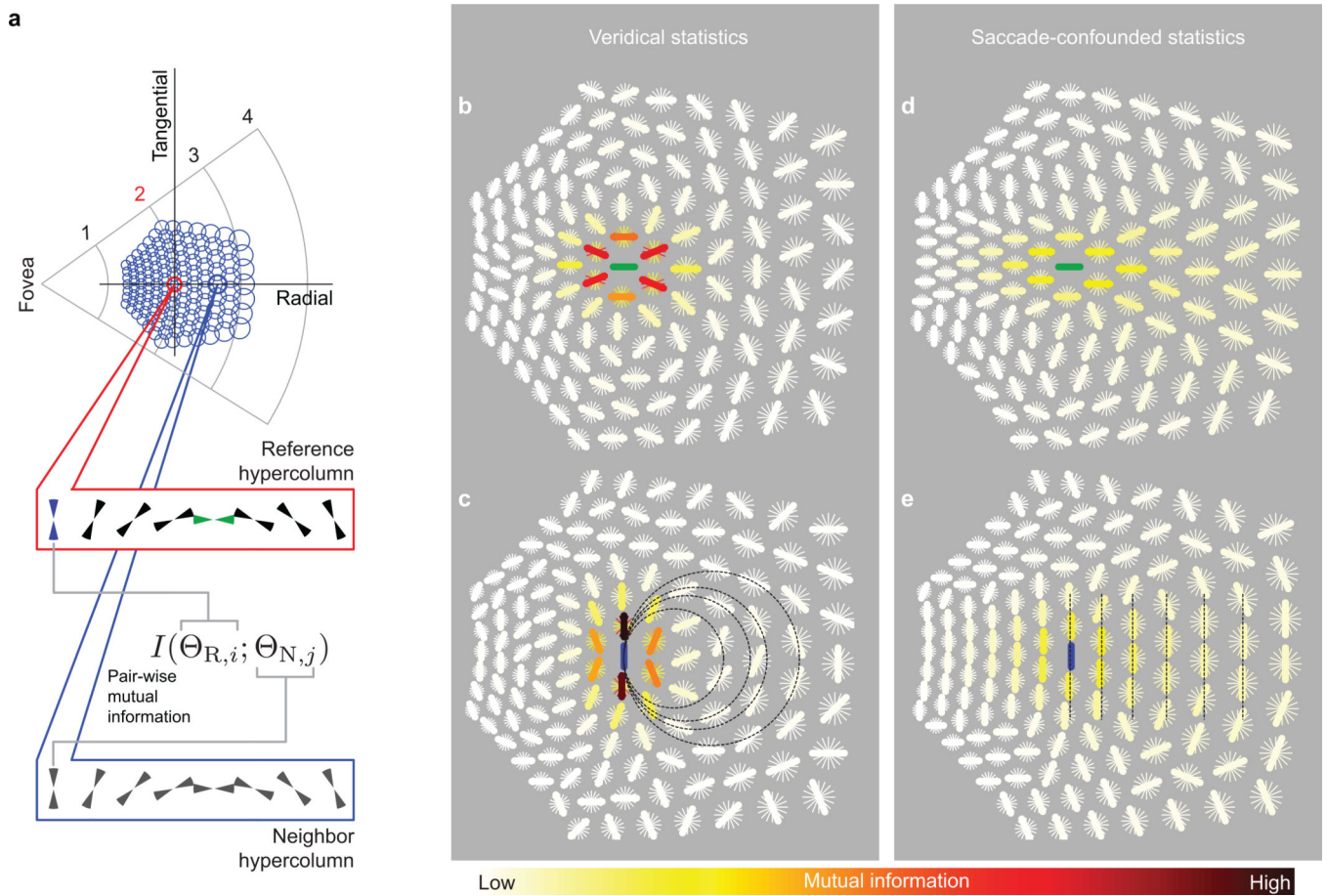
Figure 4. Pair-wise image statistics

(**a**) Each model hypercolumn (blue circles) consists of a set of 8 oriented filters that extract orientation information from a local patch of a visual scene. The highlighted red circle shows a reference hypercolumn whose receptive field is centered at 2° in the peripheral field. The ensemble of circles depict the extent of the lateral interaction zone of this reference. Pair-wise mutual information (Online Method, Equation 12) was calculated between an oriented filter in the reference hypercolumn and each of the neighboring oriented filters within 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, while 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 within 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 statistics of the same simulated visual environment gathered under the attentional spotlight during temporally overlapping eye movements. While the veridical statistics implicate smooth continuation of contours (illustrated with overlaid dotted circles in **c**), the saccade-confounded ones favor repetition

of co-oriented fragments (overlaid dotted lines in **e**). The time constant of the decay of spatial attention (λ) was 16 ms for these simulations.
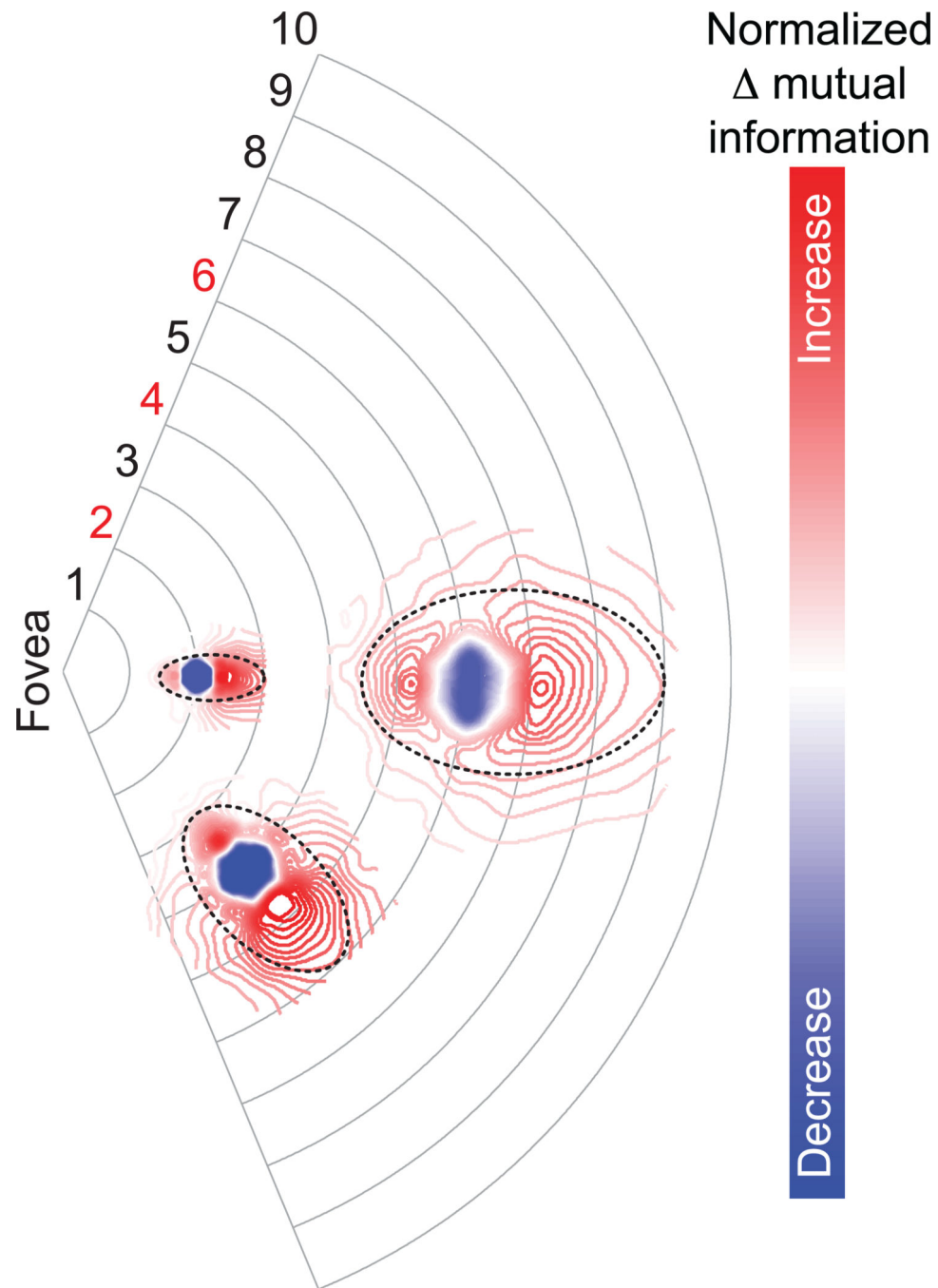
Figure 5. Zones of inappropriate integration

The pooled and normalized difference (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 is 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 (dotted lines at 40% of peak normalized difference) illustrate the elongated shape of the spatial extent of inappropriate integration. The time constant of the decay of spatial attention (λ) was set at 16 ms.