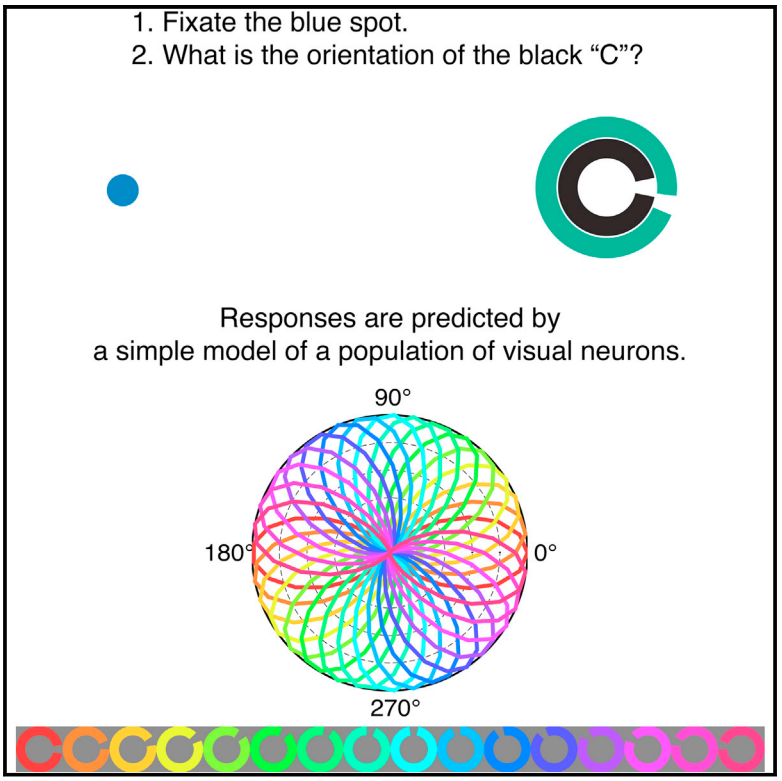
Report

Current Biology

A Unifying Model of Orientation Crowding in Peripheral Vision

Graphical Abstract



Authors

William J. Harrison, Peter J. Bex

Correspondence

willjharri@gmail.com

In Brief

Harrison and Bex studied visual "crowding," when objects that are too close together become perceptually indistinguishable; their results suggest that a biologically inspired model of a population of neurons can account for this breakdown of visual awareness in peripheral vision.

Highlights

- An object surrounded by distractors is unrecognizable in peripheral vision
- We provide a novel method to systematically study such "crowded" perception
- A simple population code model produces the diverse errors made by human observers
- A single mechanism may thus suffice to explain multiple classes of crowding

Harrison & Bex, 2015, Current Biology *25*, 3213–3219
December 21, 2015
http://dx.doi.org/10.1016/j.cub.2015.10.052

Current Biology
Report

A Unifying Model of Orientation Crowding in Peripheral Vision

William J. Harrison[1,2,3,*] and Peter J. Bex[1]
[1]Department of Psychology, Northeastern University, Boston, MA 02115, USA
[2]Department of Psychology, University of Cambridge, Cambridge CB2 3EB, UK
[3]Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072, Australia
*Correspondence: willjharri@gmail.com
http://dx.doi.org/10.1016/j.cub.2015.10.052

SUMMARY

Peripheral vision is fundamentally limited not by the visibility of features, but by the spacing between them [1]. When too close together, visual features can become "crowded" and perceptually indistinguishable. Crowding interferes with basic tasks such as letter and face identification and thus informs our understanding of object recognition breakdown in peripheral vision [2]. Multiple proposals have attempted to explain crowding [3], and each is supported by compelling psychophysical and neuroimaging data [4–6] that are incompatible with competing proposals. In general, perceptual failures have variously been attributed to the averaging of nearby visual signals [7–10], confusion between target and distractor elements [11, 12], and a limited resolution of visual spatial attention [13]. Here we introduce a psychophysical paradigm that allows systematic study of crowded perception within the orientation domain, and we present a unifying computational model of crowding phenomena that reconciles conflicting explanations. Our results show that our single measure produces a variety of perceptual errors that are reported across the crowding literature. Critically, a simple model of the responses of populations of orientation-selective visual neurons accurately predicts all perceptual errors. We thus provide a unifying mechanistic explanation for orientation crowding in peripheral vision. Our simple model accounts for several perceptual phenomena produced by crowding of orientation and raises the possibility that multiple classes of object recognition failures in peripheral vision can be accounted for by a single mechanism.

RESULTS AND DISCUSSION

Recent modeling work suggests that the perceptual experience of crowding can be synthesized by V2 neurons pooling across regions of visual space [8] and that the shape of crowding zones may be attributable to saccade-distorted image statistics in V1 [14] (see also [15]). However, these models do not attempt to account for the empirical finding that, within a pooling region in which features such as orientation signals are degraded, visual detail nonetheless affects observers' perceptual reports [13, 16]. Conversely, to our knowledge, only one approach has modeled crowding as a limited capacity to select from highly detailed representations [17], but this model does not account for perceptual averaging of orientation [9]. Here we take a reductionist approach and systematically quantify orientation discrimination using a novel paradigm and then produce the same results with our population model.

Perceptual Task and Behavioral Data

Consider the perceptual task shown in Figure 1A. An observer is required to report the orientation of the defining feature of an object [19, 20], the gap of a Landolt C (see the Supplemental Experimental Procedures). This target stimulus could be oriented any of the 360° of rotation, offering twice the variability of a Gabor element due to phase insensitivity in peripheral vision [21]. The target is presented in peripheral vision at 10° eccentricity, and, on some trials, is presented in isolation ("unflanked" condition). An observer's performance can be described by an estimate of the distribution of report errors, which we fit with a circular Von Mises function (Figure 1B; Equation S1). The SD of this function can be taken as the observer's "perceptual error," where higher values mean the observer is less precise (Equation S2).

We tested how crowding affected observers' perceptual error by surrounding the target with various flankers. Shown in Figure 1C are false-colored examples of a "one-gap flanker" condition, in which each colored ring represents a potential flanker drawn to scale. Such trials were interleaved with unflanked trials and "no-gap flanker" trials, in which flanks were solid rings. For three human observers, perceptual error across conditions is plotted on the ordinate in Figure 1D, with flanking ring size plotted on the abscissa, expressed as a proportion of the target eccentricity. Consistent with classical crowding reports, performance improved approximately linearly as the radius of the flanking ring increased relative to the target ring. In a control experiment, we rule out the possibility that systematic eye movements drove these results by reducing the stimulus duration from 500 ms to 130 ms (Figure S2). Note that we found significant crowding in both flanker conditions; as discussed in more detail below, crowding in the no-gap flanker condition cannot be explained by the confusions that occur between target and flankers in substitution and attention resolution accounts of crowding [18].

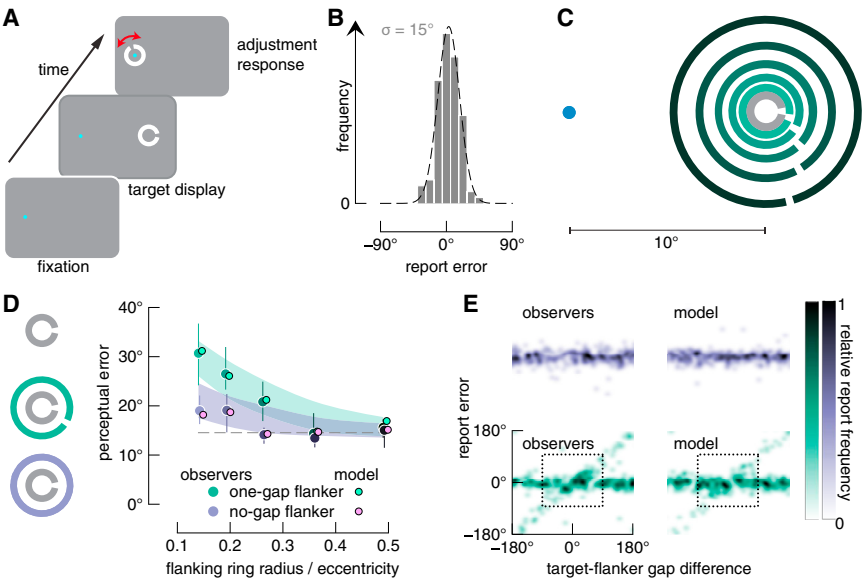


Figure 1. Experimental Design and Experiment 1 Results

(A) A typical "unflanked" trial. Observers fixated the blue spot and, after the presentation of the peripheral target, rotated a foveal Landolt C to report the perceived orientation of the target.

(B) Example error distribution. Observer's report errors conform to a circular normal distribution, the SD of which is used to define their perceptual error.

(C) False-colored examples of different "one-gap flanker" conditions. When observers fixate on the blue spot, the orientation of the central Landolt C is made difficult to report because of the flanks. Stimuli are drawn to the scale used in the model and experiments.

(D) Human and model data from experiment 1. Data show mean perceptual error across flanker sizes, with conditions colored according to the stimuli at left. Human data and model data are shown as per the inset legend. The dashed line shows unflanked performance. Error bars indicate 95% bootstrapped confidence intervals for the observed data, while shaded regions show the full range of model simulations.

(E) Distributions of report errors generated by observers and model. Data are from for the one-gap flanker condition (bottom) and no-gap flanker condition (top), with the smallest target-flanker separation. For the no-gap flanker condition, we assigned each trial to a random position on the x axis to show clearly the spread of report errors. Data from the one-gap flanker condition reveal patterns of inhomogeneous response error variability across target-flanker gap orientation differences. The dotted box shows a region within which response variability cannot easily be attributed to averaging or substitution (see Figure S1B and [18]) but is nonetheless reproduced by our model.

See also Figures S1 and S2.

We can approximate the spatial area over which a flanker interferes with orientation discrimination by taking the minimum flanker radius at which flanked performance equaled unflanked performance. The spatial extent of crowding is typically expressed as a proportion of the target's eccentricity, ϕ, and is referred to as Bouma's constant [22]. We used a hinged-line to estimate Bouma's constant (Equations S3 and S4), and found it to be 0.27 $\phi \pm 0.11$ ϕ (median ± SEM) in the no-gap flanker condition compared with 0.42 $\phi \pm 0.06$ ϕ in the one-gap flanker condition. Across previous studies, Bouma's constant ranges from 0.1–0.6 ϕ, a range that has been attributed to different crowding fields for stimuli of differing complexity [2, 23]. As we will show in the next section, our model suggests that a single, fixed weighting function may underlie the variability in Bouma's constant without invoking multiple mechanisms for different stimuli. We further note that our data show other widely reported hallmarks of crowding [1, 24] such as the anisotropic shape of the interference zone (see Figure S1A). We discuss the trial-by-trial data in the following sections to allow comparison with the data simulated by the model.

A Population Response to Crowded Orientation Signals

Our approach is based on the well-documented neurophysiology of orientation selectivity in primary visual cortex [25]. For the unflanked target, we use a population of orientation-tuned filters that encodes the stimulus to predict perceptual reports by treating this population code as the probability distribution of all possible reports. The target is represented by a bank of 32 orientation-tuned filters whose responses can be described by a circular-normal Von Mises distribution (Equation S1). This population response embodies the posterior probabilities for Landolt-C orientations at the target position. The bandwidth of this posterior distribution for an unflanked target (Figure 1B; Equation S5), rather than the tuning of orientation selective filters, determines an observer's orientation acuity (gray data in Figure 2C). For our observers, the SD of report errors, and correspondingly the model's half bandwidth, was 15°, which is higher than previous measures [26], most likely due to the eccentricity of the stimulus in our experiment.

Rather than using formal decoders (e.g., [27]), we predict perceptual reports by drawing directly from this distribution of posterior probabilities of target orientations. In Figure S3, we provide an overview of how the population code could be encoded by V1 or V2 complex filters: responses of filters convolved with the image are first pooled to calculate contrast energy across the stimulus, and then changes in contrast energy at all orientations relative to the center of the stimulus are computed. In Figure 2, we summarize a processing stage that differentially weights orientation signals according to their distance from the stimulus center. The population response thus reflects probability functions based on biologically motivated filters (see also [28, 29]). With this single quantification, we compute the probability of any given perceptual report under various viewing conditions.

Our model predicts the types of errors observers make under crowded conditions because the crowded stimulus produces an ambiguous population response. Consider the example stimuli shown in Figure 2A. We assume that observers monitor the responses of detectors across an area of space centered on the target (Figure S3). Any features registered by these detectors will influence the orientation filter used to calculate the

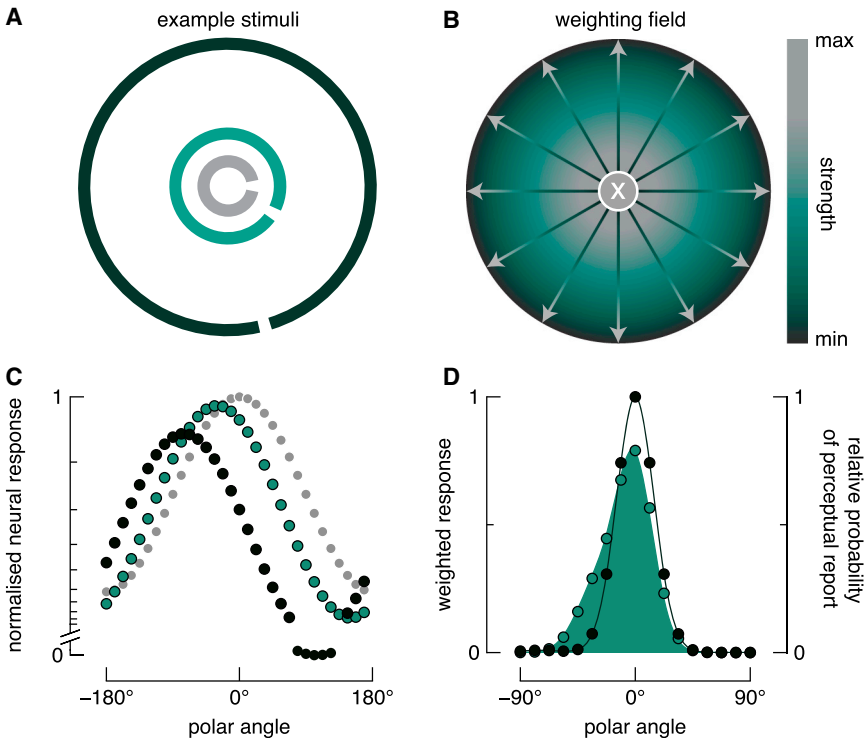


Figure 2. Population Response Model of Crowding

(A) Example stimuli for model analysis. We predict perceptual reports from each condition from the weighted and summed responses of a bank of 32 orientation-selective filters. Here, the target has an orientation of 0° in polar coordinates, and the smaller and larger flankers have orientations of −30° and −75°, respectively.

(B) Weighting field used to estimate population response to flankers. The strength of an orientation signal is modulated according to the signal's distance from the center of the stimulus. See the Results and Discussion and Supplemental Experimental Procedures.

(C) Weighted filter responses to each gap depicted in (A). 0° polar angle refers to "east" (i.e., the depicted target orientation), with orientation increasing anti-clockwise. The response to the unflanked target, shown in gray, has a circular-normal tuning function, the bandwidth of which is set to the bandwidth of the measured error distribution shown in Figure 1B. The flanker responses are tuned to each flanker gap's orientation, but with a strength that is weighted according to the gap's distance from the target center (see B). All data are normalized to the target, which is set to have a peak response of 1.

(D) Summed population responses. The weighted responses to each target-flanker combination are summed to create probability distributions of each perceptual report for each condition. Data are normalized such that a relative probability of 1 indicates the most probable report in the unflanked condition.

See also Figure S3.

orientation of the target gap. However, as depicted in Figure 2B, the farther a feature is from the center of the area over which features are detected, the less it will contribute to the orientation calculation. Therefore, to estimate the population response to a flanker, we simply down-weight the response to an unflanked target by the flanker's distance from the target center (Equation S6). Thus, whereas other models propose the flanker probabilistically either causes crowding or doesn't [9, 30], here the flanker influences the population response with a strength that diminishes with distance from the target, akin to the properties of neural response fields [31].

Figure 2C shows population responses to each feature depicted in Figure 2A, computed independently of one another. We use the weighted sum of the target response and flanker response to compute the probability of each perceptual report. Figure 2D shows these weighted responses for the two example flanker conditions in Figure 2A. Also shown for these conditions by the open and filled distributions, respectively, is the relative probability of reporting each orientation. Perceptual reports are predicted on a trial-by-trial basis by randomly drawing from these interpolated probability distributions. When the flanker feature is far from the target feature (dark-green data, unfilled distribution), the distribution is narrow and symmetric around the target orientation (0°). In this case, the most likely report of the model is near the target orientation, which predicts perceptual reports to be similar to the unflanked condition. However, when the flanker is closer to the target, the weighted response probability distribution becomes broader and skewed, which predicts an increase in the variability of perceptual reports and a skew toward the orientation of the flanker feature (−30° in this case).

Predicted and Observed Data

Presented in Figure 1D are the model simulations compared with performance from our observers. It is clear that the model accurately predicts the average perceptual error for each flanker condition and for each target-flanker separation. This is also true of the range of data produced by 1,000 model simulations (shaded regions in Figure 1D).

A key advantage of our behavioral task and model is that response variability is captured continuously (Figures 1B and 2D). In the bottom-left panel of Figure 1E, we show the distribution of report errors for the one-gap flanker condition as a function of the orientation difference between target and flanker. These raw data reveal that systematic report errors, not random guesses, contribute to the measured increase of perceptual error. As shown in Figure S1B, report error is systematically skewed away from zero when the target-flanker gap difference is relatively small (i.e., less than approximately ±90°). These data cannot distinguish between an averaging or substitution account of crowding, because both accounts predict the same data (see Figure S1B; see also [18]). Beyond this central region,

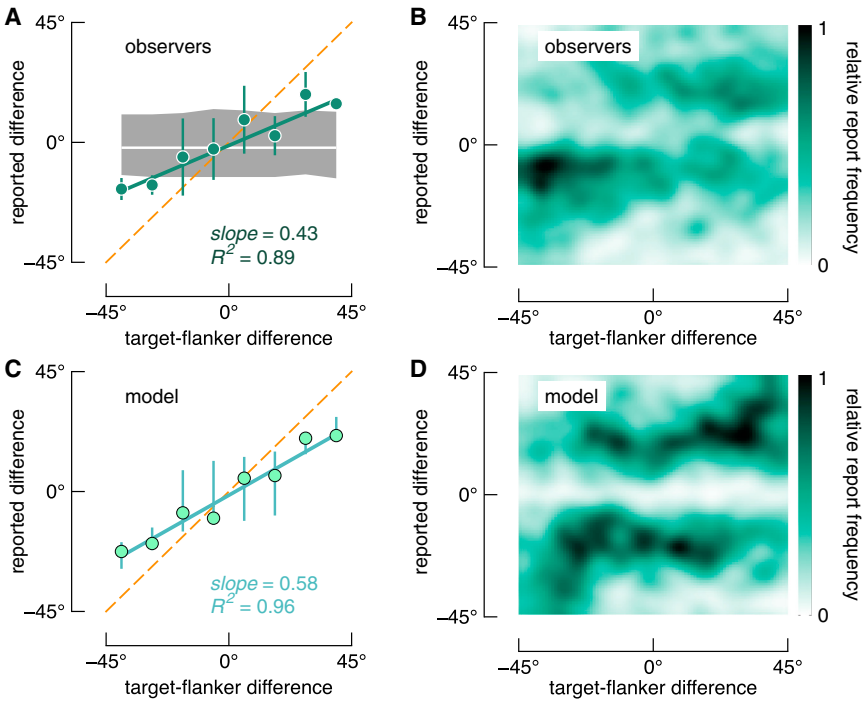


Figure 3. Detailed but Inaccurate Perceptual Reports of Crowded Stimuli

(A) Averaged reports from experiment 2. Green data show the median reported difference between target gap and flanker gap orientations as a function of the actual difference, binned in 11.25° intervals. The dark solid line is the best fitting linear function, weighted according to the inverse of the SEs shown ($F(6) = 56.9$, $R = 0.95$, $p = 0.0003$). The white line shows expected performance under the null hypothesis as estimated using permutation analyses, and the gray region shows the 95% confidence bounds for this estimate (see the main text and Figure S4 for details). The orange dashed line has a slope of 1, representative of accurate performance.

(B) Heatmap of the raw data. The absence of responses centered on 0° reveals that observers rarely reported the target and flanker as having the same orientation.

(C and D) Data from our model that incorporates the change in orientation weightings following an observer's first decision, plotted as in (A) and (B). Error bars show SEs.

See also Figure S4.

the majority of report errors clusters around zero. It is clear, however, that some errors are distributed along the diagonal line, which corresponds to the orientation of the flanker. These report errors result in increased perceptual error (Figure 1D). Importantly, these raw data suggest that observers were able to discriminate the target from the flanker but reported the orientation of the incorrect feature. It has previously been argued that such errors are the result of substitutions among target and flanker features or by the observers' limited attentional resolution [12, 13, 32]. In either case, their ability to select reliably the target item was worse than their ability to resolve the visual detail of target and flanker orientations. A probabilistic substitution model may account for this pattern of errors [18]. However, rather than invoking a substitution account to explain these data and a separate mechanism to account for the results from the no-gap flanker condition described above, our model accounts for these perceptual errors from the responses of well-known orientation selective mechanisms (Figures 2 and S3). Indeed, this same mechanism accounts for perceptual error as well as the heterogeneous patterns of reports across all conditions (Figures 1D and 1E). None of the existing crowding models alone is able to explain the full distribution of our data. We discuss limitations on the generalizability of our model in the General Discussion.

Observers Are Inaccurate but Can Nonetheless Report Detail

Of critical importance to theories of visual awareness is the amount of reportable object detail made available by the visual system. Attentional selection and substitution theories propose that more detail is encoded than can be accessed accurately [12, 33], whereas averaging theories propose that detail is irretrievably lost [22, 34]. We tested these possibilities in a second experiment by asking observers to report both the target orientation *and* the flanker orientation. Observers completed a one-gap flanker condition with the orientation of the flanker constrained such that it differed from the target uniformly between ±45°, which covers the range where reports are strongly influenced by the flanker (Figure S1B). Not only does this experiment provide a systematic quantification of the detail of visual representations in peripheral vision, but it also provides a novel and sensitive test of models attempting to explain the interactions between target and flankers.

Shown as green data in Figure 3A is the average difference between the reported target orientation and flanker orientation as a function of the actual difference. For comparison, we simulated patterns of expected report errors based on averaging, attentional selection, and substitution models (Figure S4). Although the distributions of expected reports are markedly different for each model, they all predict that the *average* reported difference between target and flanker orientations is zero (dashed lines in Figure S4). We can thus simulate expected performance under the null hypothesis that the observed error fits any of the prior models by permuting the observed data to achieve a distribution of zero differences (see the Supplemental Experimental Procedures). The white line in Figure 3A shows the null data; the gray shaded region indicates the 95% confidence intervals. The observed data systematically and significantly deviate from this prediction but were well fit by a linear function (solid green line). Therefore, observers could report the relative orientation of the target feature with respect to the flanker feature, and so our data are incompatible with all previous crowding models. However, observers were systematically inaccurate: they underestimated the target-flanker orientation difference, as indicated by the shallow slope of the data relative to the equality line (dashed

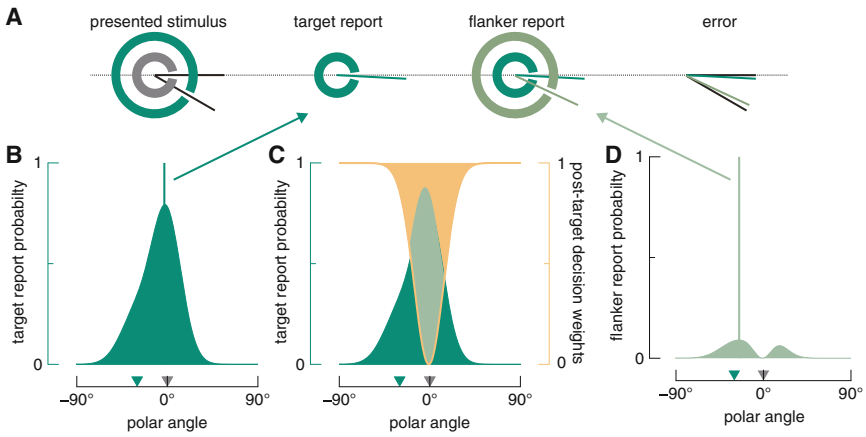


Figure 4. A Model of Detailed but Inaccurate Perceptual Reports

(A) Our model shows how an observer is able to report the correct relative orientations of a target and flanker, but with systematic report errors.

(B–D) Black and green triangles on the x axes are the orientations of the target feature and flanker feature, respectively, of the presented stimulus shown in (A). The most probable target report is predicted from the model's population response (B; see Figure 2). Note that this distribution is normalized according to the relative probability of an unflanked target report. (C) is the same as (A), with the post-target decision weights plotted on the right axis. The point wise product of these distributions yields (D), the probability distribution of flanker reports. The most probable flanker report tends toward the actual flanker orientation.

orange line). Although we cannot reject an averaging model that claims to hold only for similar target-flanker orientations, we will show in the next section that our single model is parsimonious because it reproduces the full range of data.

Modeling Detailed but Inaccurate Reports

Visual inspection of the distribution of report errors reveals that observers systematically avoided reporting the same orientation for both the target and flanker (Figure 3B), similar to a response bias away from canonical values [9]. Based on this observation, we added a decision stage to our population response model, in which the observer's target orientation decision affects the flanker orientation decision.

As described in the preceding sections, an observer first reports the target orientation by sampling from a probability distribution that combines target and flanker orientations (Figures 4A and 4B). The flanker report is also informed by this population response: we simply include in the model the observer's avoidance of their own target report by down-weighting a narrow distribution of orientations centered on the reported target orientation (Figure 4C; Equations S7 and S8). This down-weighting could represent a report bias, perceptual repulsion, overt strategy, or something else. We can approximate the extent of these changed weights using the observers' perceptual error as quantified in the unflanked condition (Figure 1B), which assumes observers have some implicit knowledge about their likely target report error. By taking the point wise product of the population response and the post-target decision weights, we obtain a post-decision probability distribution from which the observer's report of the orientation of the flanker is sampled (Figure 4D). In Figure 4D, the most probable flanker report is close to the flanker's actual orientation. As shown in Figure 4A, the model correctly predicts that the observer underestimates the difference between target and flanker features, but typically reports the correct relative orientations (see also Figure 3C). The model also reproduces the patterns of report errors made by observers (Figure 3D). It is also apparent in Figure 3B that observers' reports are not perfectly symmetrical: observers tended to report more negative (i.e., clockwise) differences than positive (i.e., anti-clockwise) differences overall. We speculate that this may reflect a response bias when they were uncertain. Note, however, that this asymmetry is not present in our modeled reports (Figure 3D) because we did not attempt to model the negative response bias with an arbitrary parameter.

General Discussion

We show that a population response model can predict the complex pattern of observers' reports in a difficult perceptual task involving the discrimination of an oriented target embedded in a cluttered scene. In a first experiment, our model predicted observers' perceptual reports of the target surrounded by different flankers across a range of spatial configurations (Figure 1D). In a second experiment, incorporating into our model an observer's decision allowed us to account for how the observer can report multiple details of a scene while being inaccurate about those details (Figures 3 and 4). Our data are derived from a specific stimulus at a single eccentricity but provide in-principle evidence that a single mechanism can account for the variety of perceptual reports produced by an orientation discrimination task in peripheral vision.

Researchers have previously attributed perceptual errors to averaging [9], substitution [18], and limits of attentional resolution [13]. Our first experiment showed robust crowding effects when the flank contained no gap and thus no substitutable features. This finding challenges all crowding accounts that are based on confusions between target and flanks [18, 35]. Instead, our model accounts for both averaging and substitution qualitative reports within a common, biologically plausible framework (Figure S3 and [34]). Our model also captures the perceptual effects taken to support an attentional account of crowding. Indeed, our paradigm is well suited to study the effects of attention on peripheral vision. The influence of classical top-down manipulations, such as providing advance positional or temporal information [36, 37], can be investigated by measuring their effect on perceptual accuracy and precision for single and multiple perceptual reports. Our approach can easily incorporate changes in orientation tuning brought about by spatial attention [38] and can thus provide a quantitative tool to study attention's controversial role in visual crowding.

A consideration of the generalizability of our model is important. Our aim was to provide evidence that a single model can, in principle, account for a variety of categorically unique errors made when discriminating an indivisible target element. It is important to note, however, that in this endeavor we used unique

stimuli, rather than tilted lines, letters, or Gabors that are more common in the crowding literature. Our model could be extended to predict and describe crowding of as many stimulus dimensions as can be expressed by population response distributions (Figure S3). For example, our model could provide insight into the crowding of feature conjunctions [39], such as the binding of color, position, and orientation. Indeed, we have extended the present model to account for changes in crowding over time [40].

Two emergent properties of orientation feature integration zones arose in our data. First, a flanker's influence over the target response increases as its orientation approaches that of the target: perceptual reports follow the target-flanker average when elements have similar orientations; flankers with dissimilar orientations resulted in apparent substitution responses. Because our aim was to reconcile conflicting accounts of crowding and not to localize the neural area in which it occurs (cf. [8, 14]), we built into our model certain spatial characteristics of the integration zone (see the Supplemental Experimental Procedures). Nonetheless, our model explains computationally why averaging is stronger when target and flanker are similar than when they are dissimilar (see Figure S1B). This similarity effect has been reported previously for orientation [35, 41] and other stimulus properties [23, 42, 43] but is simply asserted by averaging models. Second, following Bouma's constant, our flankers became ineffectual at a distance that approximates the estimated size of receptive fields in area V2 in primate brains, according to a recent meta-analysis [8]. This suggests that the orientation crowding in our tasks occurs at or later than area V2 and may thus help to explain why orientation adaptation aftereffects, thought to occur in V1, can occur without invoking a high-level cognitive filter, like attentional selection [13].

Conclusions

Several stereotypical patterns of perceptual errors characterize object recognition in peripheral vision. Although multiple mechanisms have previously been invoked to account for these data, we show that a single model can reproduce the main classes of errors for orientation crowding: perceptual averaging, substitutions, and apparent attentional failures. Our results show that the loss of object discriminability in visual clutter may not be an all-or-none process; visual detail is not lost, per se, but instead perceptual reports conform very well to continuous probabilistic distributions.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and four figures and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2015.10.052.

AUTHOR CONTRIBUTIONS

W.J.H. conceived the psychophysical paradigm and computational model, conducted the experiments and analyses. Both authors designed the experiments, coded different elements of the model, and wrote the manuscript.

ACKNOWLEDGMENTS

We thank Julie Golomb for help with fitting circular distributions to the behavioral data, Guido Maiello for help with the hinged-line fits, Kate Storrs for many helpful discussions about population codes, and three reviewers who provided insightful feedback. This work was supported by NIH grant R01EY021553 (P.J.B.) and a National Health and Medical Research Council of Australia CJ Martin Fellowship (APP1091257; W.J.H.). The Northeastern University IRB approved procedures, and all observers consented to participation.

Received: July 23, 2015
Revised: October 21, 2015
Accepted: October 26, 2015
Published: November 25, 2015

REFERENCES

1. Bouma, H. (1970). Interaction effects in parafoveal letter recognition. Nature *226*, 177–178.
2. Pelli, D.G., and Tillman, K.A. (2008). The uncrowded window of object recognition. Nat. Neurosci. *11*, 1129–1135.
3. Levi, D.M. (2008). Crowding–an essential bottleneck for object recognition: a mini-review. Vision Res. *48*, 635–654.
4. Anderson, E.J., Dakin, S.C., Schwarzkopf, D.S., Rees, G., and Greenwood, J.A. (2012). The neural correlates of crowding-induced changes in appearance. Curr. Biol. *22*, 1199–1206.
5. Kwon, M., Bao, P., Millin, R., and Tjan, B.S. (2014). Radial-tangential anisotropy of crowding in the early visual areas. J. Neurophysiol. *112*, 2413–2422.
6. Chen, J., He, Y., Zhu, Z., Zhou, T., Peng, Y., Zhang, X., and Fang, F. (2014). Attention-dependent early cortical suppression contributes to crowding. J. Neurosci. *34*, 10465–10474.
7. Parkes, L., Lund, J., Angelucci, A., Solomon, J.A., and Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. Nat. Neurosci. *4*, 739–744.
8. Freeman, J., and Simoncelli, E.P. (2011). Metamers of the ventral stream. Nat. Neurosci. *14*, 1195–1201.
9. Greenwood, J.A., Bex, P.J., and Dakin, S.C. (2009). Positional averaging explains crowding with letter-like stimuli. Proc. Natl. Acad. Sci. USA *106*, 13130–13135.
10. Balas, B., Nakano, L., and Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. J. Vis. *9*, 1–18.
11. Chastain, G. (1982). Confusability and interference between members of parafoveal letter pairs. Percept. Psychophys. *32*, 576–580.
12. Strasburger, H., and Malania, M. (2013). Source confusion is a major cause of crowding. J. Vis. *13*, 24.
13. He, S., Cavanagh, P., and Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. Nature *383*, 334–337.
14. Nandy, A.S., and Tjan, B.S. (2012). Saccade-confounded image statistics explain visual crowding. Nat. Neurosci. *15*, 463–469, S1–S2.
15. Harrison, W.J., Mattingley, J.B., and Remington, R.W. (2013). Eye movement targets are released from visual crowding. J. Neurosci. *33*, 2927–2933.
16. Yeh, S.-L., He, S., and Cavanagh, P. (2012). Semantic priming from crowded words. Psychol. Sci. *23*, 608–616.
17. Chaney, W., Fischer, J., and Whitney, D. (2014). The hierarchical sparse selection model of visual crowding. Front. Integr. Nuerosci. *8*, 73.
18. Ester, E.F., Klee, D., and Awh, E. (2014). Visual crowding cannot be wholly explained by feature pooling. J. Exp. Psychol. Hum. Percept. Perform. *40*, 1022–1033.
19. Pelli, D.G., Burns, C.W., Farell, B., and Moore-Page, D.C. (2006). Feature detection and letter identification. Vision Res. *46*, 4646–4674.
20. Pelli, D.G., Farell, B., and Moore, D.C. (2003). The remarkable inefficiency of word recognition. Nature *423*, 752–756.
21. Morrone, M.C., Burr, D.C., and Spinelli, D. (1989). Discrimination of spatial phase in central and peripheral vision. Vision Res. *29*, 433–445.

22. Pelli, D.G. (2008). Crowding: a cortical constraint on object recognition. Curr. Opin. Neurobiol. *18*, 445–451.
23. Bernard, J.-B., and Chung, S.T.L. (2011). The dependence of crowding on flanker complexity and target-flanker similarity. J. Vis. *11*, 11.
24. Toet, A., and Levi, D.M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. Vision Res. *32*, 1349–1357.
25. Hubel, D.H., and Wiesel, T.N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. J. Physiol. *160*, 106–154.
26. Andrews, D.P. (1967). Perception of contour orientation in the central fovea. I: short lines. Vision Res. *7*, 975–997.
27. Ma, W.J., and Pouget, A. (2009). Population codes: theoretic aspects. In Encyclopedia of Neuroscience (Elsevier), pp. 749–755.
28. Jazayeri, M., and Movshon, J.A. (2006). Optimal representation of sensory information by neural populations. Nat. Neurosci. *9*, 690–696.
29. Matthey, L., Bays, P.M., and Dayan, P. (2015). A probabilistic palimpsest model of visual short-term memory. PLoS Comput. Biol. *11*, e1004003.
30. Dakin, S.C., Cass, J., Greenwood, J.A., and Bex, P.J. (2010). Probabilistic, positional averaging predicts object-level crowding effects with letter-like stimuli. J. Vis. *10*, 14.
31. Ringach, D.L. (2002). Spatial structure and symmetry of simple-cell receptive fields in macaque primary visual cortex. J. Neurophysiol. *88*, 455–463.
32. Petrov, Y., and Meleshkevich, O. (2011). Locus of spatial attention determines inward-outward anisotropy in crowding. J. Vis. *11*, 11.
33. Intriligator, J., and Cavanagh, P. (2001). The spatial resolution of visual attention. Cognit. Psychol. *43*, 171–216.
34. van den Berg, R., Roerdink, J.B.T.M., and Cornelissen, F.W. (2010). A neurophysiologically plausible population code model for feature integration explains visual crowding. PLoS Comput. Biol. *6*, e1000646.
35. Ester, E.F., Zilber, E., and Serences, J.T. (2015). Substitution and pooling in visual crowding induced by similar and dissimilar distractors. J. Vis. *15*, 4.
36. Posner, M.I., Snyder, C.R., and Davidson, B.J. (1980). Attention and the detection of signals. J. Exp. Psychol. *109*, 160–174.
37. Harrison, W.J., and Bex, P.J. (2014). Integrating retinotopic features in spatiotopic coordinates. J. Neurosci. *34*, 7351–7360.
38. Lee, D.K., Itti, L., Koch, C., and Braun, J. (1999). Attention activates winner-take-all competition among visual filters. Nat. Neurosci. *2*, 375–381.
39. Greenwood, J.A., Bex, P.J., and Dakin, S.C. (2012). Crowding follows the binding of relative position and orientation. J. Vis. *12*, 12.
40. Harrison, W.J., and Bex, P.J. (2015). A population response model of spatial crowding over time. 38th European Conference on Visual Perception.
41. Andriessen, J.J., and Bouma, H. (1976). Eccentric vision: adverse interactions between line segments. Vision Res. *16*, 71–78.
42. Kooi, F.L., Toet, A., Tripathy, S.P., and Levi, D.M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. Spat. Vis. *8*, 255–279.
43. Harrison, W.J., Retell, J.D., Remington, R.W., and Mattingley, J.B. (2013). Visual crowding at a distance during predictive remapping. Curr. Biol. *23*, 793–798.

Current Biology

Supplemental Information

A Unifying Model of Orientation Crowding in Peripheral Vision

William J. Harrison and Peter J. Bex

SUPPLEMENTAL FIGURES

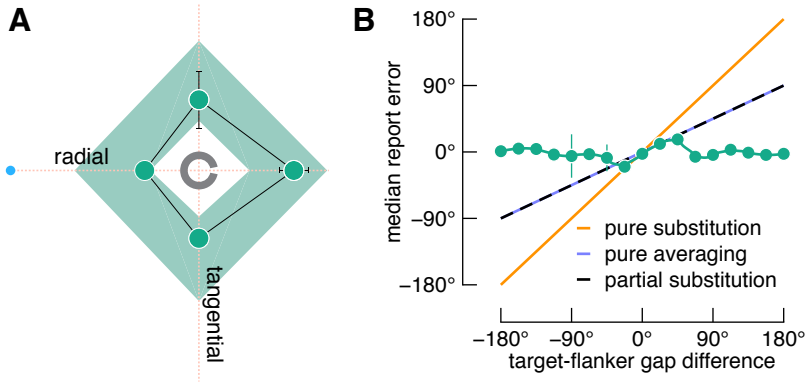


Figure S1. Related to Figure 1. A) The zone within which a one-gap flanker feature interferes with perception of a target feature, drawn to scale. Observers' fixation point is shown as a blue spot, with critical spacing zones shown as green data with white outlines. These data show that our measure produces the characteristic radial-tangential anisotropy reported previously [S1] (measurements described in Supplemental Experimental Procedures). Black error bars show standard error, which in some cases is smaller than the point size. Shaded green regions show the 95% confidence intervals produced by our model. Although it is clear that the model captures the data, the modeled interference zones are not anisotropic. Anisotropy could be incorporated by, for example, scaling the flanker weights according to their distance in visual cortex [S2] or by assuming radially-skewing of image statistics by saccadic eye movements [S3]. B) Average report errors. Green data show the median report error per target-flanker gap orientation difference (in bins of 22.5°) for the smallest target-flanker separation from the one-gap condition in Experiment 1. Error bars show one standard error, which is smaller than the point size in most cases. These averaged data reveal that perceptual reports are least accurate at relatively small target-flanker gap differences, and that previous models make similar predictions about these errors. The orange diagonal (slope = 1) indicates where reports would fall were observers to consistently report the flanker orientation ("pure substitution"). The magenta and black dashed lines show, respectively, expected performance were observers to report the target-flanker average ("pure averaging"), and expected performance were they to report the target on half the trials and the flanker on the remainder of trials ("partial substitution"). Despite being generated by different models, these predictions overlap. The data fall directly on this region of ambiguity, and so the data do not exclusively support either of these previous models. This issue is discussed further by Ester et al [S4].

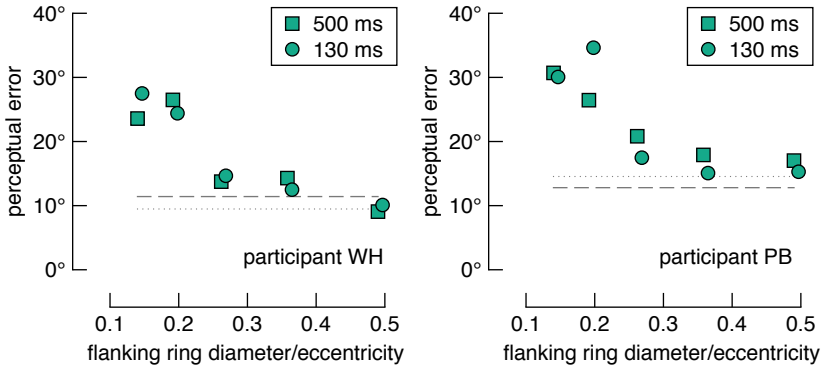


Figure S2. Related to Figure 1. Results from a control experiment (circles) compared with results from Experiment 1 (squares). To rule out the possibility that systematic eye movements influenced the main results (Fig. 1D), the two authors repeated 1200 trials of the one-gap flanker condition from Experiment 1 but with a target duration of 130 ms. The shorter target duration precludes the contamination of eye movements; the details of the experiment were the same as Experiment 1 in all other regards. Squares and circles show perceptual error across all target-flanker separations for Experiment 1 (500 ms) and the control experiment (130 ms), respectively. The 500 ms and 130 ms unflanked conditions are shown as dotted and dashed lines, respectively. For both observers, perceptual error from the control experiment matched very well the original results with longer target durations. Thus, eye movements are unlikely to have contributed to the main results. Moreover, these results demonstrate our novel method is stable across target durations, and that the measured levels of crowding are stable over the year after which the control data were collected.

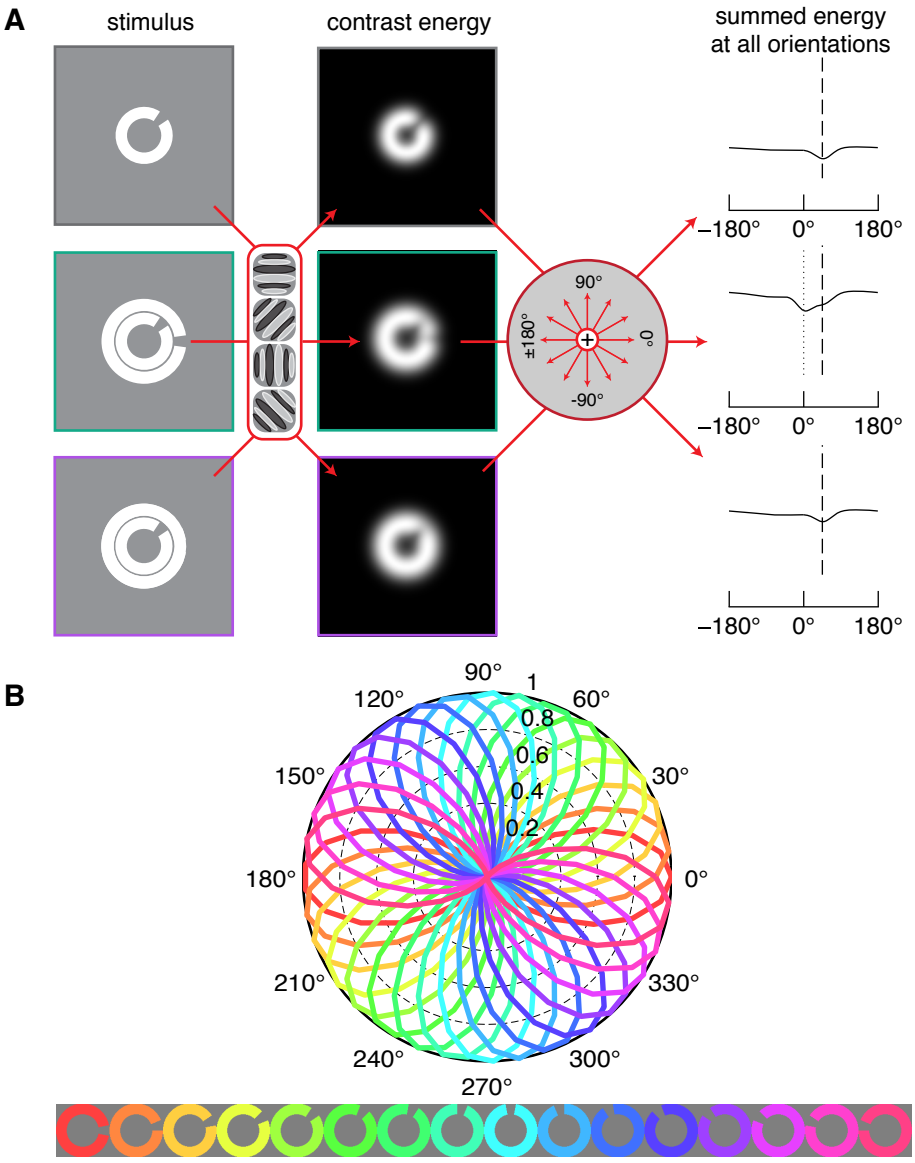


Figure S3. Related to Figure 2. Biologically motivated filter responses. A) Model frontend. In the left column from top to bottom are stimulus examples of the unflanked, one-gap flanker, and no-gap flanker conditions, respectively. By convolving quadrature phase Gabor complex filters of varying orientations (shown schematically in the red rectangle), we compute contrast energy, as shown in the middle column. Note the differences in contrast energy around the target gap position (45° polar orientation) across conditions. Gap orientation can be computed by summing the contrast energy along radii at each orientation (see red circle). In the right column, we show the resulting summed energy at each orientation. In the unflanked condition (top), a local minimum corresponds to the target orientation, indicated by the dashed line. The same minimum exists in the one-gap flanker condition (middle), but the 0°-oriented flanker creates a second larger minimum. For the no-gap flanker condition in the bottom row, the absolute contrast energy is greater than the unflanked energy for all orientations, resulting in a reduced minimum and thus lower signal-to-noise ratio. The model described in the main text represents a later stage process in which the variability in contrast energy at all orientations is weighted according to the distance from the center of the stimulus (Fig. 2). B) Normalized responses of oriented filters underlying the responses at each location in (A). Each color in the top plot corresponds to the signal strength at all orientations in response to the same colored Landolt C in the legend at bottom. Signals are pooled from filters positioned within a circular region centered on the target.

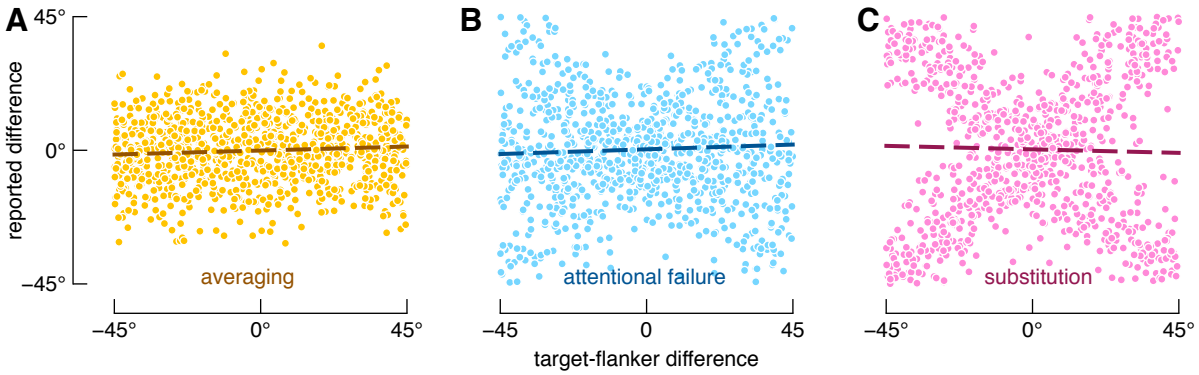


Figure S4. Related to Figure 3. Three unique predictions for Experiment 2 based on simulations using previous models of crowding. A) Simulations from an averaging model. B) Simulations from an attentional model. C) Simulations from a substitution model. Despite predicting unique patterns of raw data, all models predict that the average difference between the reported target feature and flanker feature should be zero degrees (dashed lines). Despite these predictions, the observed data were non-zero and thus are not captured by these models of crowding (see Fig. 3A). Our population response model, however, accounts for these data very well (see Figs. 3 and 4). Note that if we were to include repulsion [S5] in these models to incorporate the absence of similar target-flanker reports, the averages (dashed lines) would remain unchanged. See Supplemental Experimental Procedures for simulation details.

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Psychophysical testing

Three experienced psychophysical observers, two authors and one naïve, participated in both experiments. Observers had typical or best-corrected visual acuity. Although no eye tracking was used, all observers have had extensive experience in experiments requiring steady fixation. The Northeastern University IRB approved procedures, and all observers consented to participation.

Observers sat in a dimly lit room with their head in a chinrest 58cm from a CRT (1280 x 1024 resolution; 85 Hz). Stimuli were programmed with the Psychophysics Toolbox Version 3 [S6, S7] in MATLAB (MathWorks). All stimuli were white and presented on a uniform grey background. False-colored examples of the stimuli are shown in Figures 1 and 2. Targets had a diameter of 2°, a line width of 0.4°, and were always presented at 10° directly to the right of a fixation spot positioned in the center of the display. Note that our stimuli are modified versions of a Landolt C: the bounds of the gap section of the target are not parallel, and the proportions of the flanking stimuli vary from standard proportions. The gap width, measured at the midpoint of the line width, was 0.4°. In the one-gap flanking condition, the gap size remained constant across flanker diameters. For all flanking conditions, the line width remained constant. In the first experiment, the target was presented in one of the three conditions shown in Figure 1D. For the flanker conditions, the distance from the target's outer edge to the flanker's outer edge was randomly selected from one of five possible distances: 0.4°, 0.92°, 1.62°, 2.58°, or 3.9°.

The target and flankers were presented for 500 ms in the first experiment, and 60 ms in the second experiment. Immediately following the target-flanker offset, the adjustment stimulus was presented at fixation (Fig. 1A). Observers used the arrow keys to rotate the adjustment stimulus so that its orientation matched that of the target, and then pressed the space bar to confirm their report. In the second experiment, observers were first presented with one Landolt C to adjust the apparent target orientation, and, after confirming its orientation, an additional Landolt C was presented with the same target-flanker separation as presented in peripheral vision. Observers then adjusted its orientation to match that of the flanker. Following confirmation of the flanker orientation, they were given a chance to change their reports by pressing an arrow key, or to continue to the next trial by pressing the space bar. The following trial began immediately after confirming the perceptual report. All conditions were randomly intermingled within a testing session using MATLAB's rng() functions.

For the first experiment, observers completed at least 5 testing sessions each, giving at least 1100 trials in total per observer. Each session took approximately 10 minutes to complete. For the second experiment, each observer completed 4 sessions, giving at least 350 trials per person. Because observers had to make two reports per trial in the second experiment, each of these sessions took approximately 15 minutes.

Quantifying performance

We used maximum-likelihood estimation to fit Von Mises circular distributions to observers' report errors for each condition. The Von Mises distribution is defined as:

$$p(x) = \frac{e^{kcos(x-\mu)}}{2\pi I_0(k)} \qquad \text{(S1)}$$

where p(x) is the probability of report error x, μ is the mean report error, k is the distribution concentration, and μ and x are in radians. I_0 is a Bessel function of order 0. Perceptual error is defined as the standard deviation of responses, σ, converted to degrees after calculating:

$$\sigma = \sqrt{1/_k} \qquad \text{(S2).}$$

95% confidence intervals shown in Figure 2B were calculated via standard bootstrapping procedures [S8], resampling with replacement 2000 times. This non-parametric analysis is appropriate for small sample sizes.

Estimating Bouma's constant

We estimated Bouma's constant, the distance over which features are integrated as a proportion of target eccentricity, by fitting a hinged-line to the individual observer data of the first experiment. The hinged-line is defined as:

$$\sigma_{est} = \begin{cases} -ax + (s - ad), & x \leq d \\ s, & x > d \end{cases} \qquad \text{(S3)}$$

where σ_{est} is estimated perceptual error in degrees, ***a*** specifies the rate of change of perceptual error with changing target-flanker separation, s is unflanked performance set to 14.54° in our experiments, and, importantly, d is the target-flanker separation required for crowding performance to reach unflanked performance. Bouma's constant, b, is then:

$$b = \frac{d}{\phi} \qquad \text{(S4)}$$

where ϕ is the target eccentricity, 10°. To compare this value with the size of receptive fields in monkeys [S9], we multiply b by two to give the diameter of the area over which features interact

with each other in our experiments.

To find the shape of the integration zone (Fig. S1A), we first categorized each one-gap flanker trial according to the flanker-feature orientation relative to the center of the target. The flanker feature was considered to be above, below, left or right, of the target when it's polar angle deviated by less than 45° from 90°, 270°, 180°, or 0°, respectively. We then calculated d as described above for each flanker position.

Perceptual model

The internal representation of the orientation of an unflanked target can be described by the circular distribution described in Equation S1 assessed at 32 equally spaced orientations, each position of which is representative of an orientation-tuned filter. When integrated to one, the distribution is a probability density function that describes the probability of any given report. In the first experiment, we measured observers' average perceptual error for an unflanked target to be 14.54°. This value, σ in radians, was converted to the distribution's concentration parameter, k, by re-arranging Equation S2 as follows:

$$k = \frac{1}{\sigma^2} \qquad \text{(S5)}.$$

From k, we know the internal representation of the unflanked target for any given orientation (grey data in Fig. 2C). Because orientation filters are spatially dependent [S10, S11], we reduced the model's response to the flanker feature according to its distance from the target center using a symmetrical two-dimensional Gaussian centered on the target, with the form:

$$w_i = f(\omega \times d_i \mid \mu, sd) \quad \text{(S6)}$$

where w_i is the weight of the response to flanker orientation on trial i, and $f(d_i)$ is a normal probability density function (pdf) assessed at the distance of the flanker feature from the target center on trial i. The mean and standard deviation of the pdf were 0° and 2.5°, respectively. The flanker distance on each trial was weighted by ω, which was found to produce the best fits when set to a value of 2. The target feature was similarly weighted, where d_i was set to a distance of 1° to represent the radius of the target (i.e. the distance of the target feature from the receptive field center), and ω was set to 1. The weighted flanker response of each filter was summed with their response to the target feature to give the response distributions shown in Figure 2D.

Because the no-gap flanker produces crowding but has no explicitly oriented feature, we assumed that the solid ring causes noise in the population response at all orientations. The

output of the model frontend shown in Figure S3 supports this idea. To predict the report probabilities in the no-gap flanker condition in which there is no unique flanker feature (see Fig. 1D), we therefore set each filter's response to its maximum possible response to any orientation, divided by the total number of filters. The response was then weighted according to its distance from the target feature as described above and in Equation S6. Note that the generalizability of the no-gap flanker model is unclear without further testing. We can nonetheless conclude from the observed crowding in the no-gap flanker condition that a substitution or attentional selection account alone cannot explain crowding in our experiments.

We ran 1000 model simulations (see Fig. 1D for the range of model predictions). Because the model has no free parameters, variability in the predictions arose only from the probabilistic report selection as described above.

Decision model

Following Experiment 2, we modeled the change in population response weights following the target report as a Gaussian distribution with a standard deviation equal to perceptual error, σ in degrees, a mean equal to the most probable target orientation given the presented stimulus, and normalized to have a peak of 1. However, because perceptual reports are predicted by randomly drawing from the population response (Fig. 2), the observer's reported target orientation does not necessarily correspond to the peak of the probability distribution in our model. We thus assume that on a given trial the observer's target report corresponds to the most probable target orientation. The pattern of raw data in Fig. 3B justifies this assumption: observers rarely reported the flanker orientation as the same as the target, suggesting some confidence in their first report corresponding to the target orientation. We therefore estimate the difference between the most probable report given the stimulus and the actual report:

$$\beta_j = x_j - \theta_j \quad \text{(S7)}$$

where x_j, and θ_j are, respectively, the reported target orientation and the most probable target orientation, and β_j is their difference, for trial j. The flanker report on this trial, τ_j, is therefore predicted by adding this difference, β_j , to the random sample:

$$\tau_j = \varepsilon_j + \beta_j \quad \text{(S8)}$$

where ε_j is an orientation randomly sampled from the distribution resulting from the point wise product of the population response distribution and post-target weights on trial j (Fig. 4D).

Averaging, attentional selection, and substitution model simulations

We compared performance in the second experiment to three simple models in which 1) features are averaged into a unitary percept (Fig. S4A), 2) attentional resolution is too coarse to individuate features (Fig. S4B), and 3) observers sometimes substitute target and flanker features (Fig. S4C). For all models, we first randomly selected 1000 hypothetical trials in which the target-flanker orientations differed as per Experiment 2. Each feature was detected independently by a bank of filters, as described above, with a perceptual resolution of the unflanked condition in Experiment 1. For each trial in the averaging model, we first took the estimated target report as the average of a random sample from the target response distribution and a random sample from the flanker response distribution. We repeated this process for the estimated flanker report. For the attentional model, we would expect that all features are represented by their unique response distributions, but that the hypothetical observer has difficulty selecting from which distribution to report [S12]. Therefore, we simulated target and flanker reports under this model by twice selecting randomly, with replacement, from the target or flanker distributions. Under the substitution model, we simulated target and flanker reports by twice selecting randomly *without* replacement from the target or flanker distributions. For each model, the difference between target and flanker estimates gives a datum shown in Figure S4. As shown by the dashed lines in Figure S4, all models yield an average reported difference of zero. We can therefore use a permutation analysis [S8] to simulate expected performance under the null hypothesis (that performance is not different from these previous models) by taking reported differences between target and flanker features, and randomly re-assigning each report to a different target-flanker configuration. Repeating this process 2000 times creates surrogate averages of perceptual reports that are uncorrelated with the presented stimuli (white line, Fig. 3A). The 95% confidence intervals of these permuted averages are the 2.5^{th} and 97.5^{th} quantiles of the surrogate data. Note that, if we were to include a reference repulsion as per previous models of crowding [S5], the averages would be unchanged.

SUPPLEMENTAL REFERENCES

S1. Toet, A., and Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. Vision Research *32*, 1349–1357.

S2. van den Berg, R., Roerdink, J. B. T. M., and Cornelissen, F. W. (2010). A neurophysiologically plausible population code model for feature integration explains visual crowding. PLoS Computational Biology *6*, e1000646.

S3. Nandy, A. S., and Tjan, B. S. (2012). Saccade-confounded image statistics explain visual crowding. Nature Neuroscience *15*, 463–469.

S4. Ester, E. F., Klee, D., and Awh, E. (2014). Visual crowding cannot be wholly explained by feature pooling. Journal of Experimental Psychology: Human Perception and Performance *40*, 1022–1033.

S5. Greenwood, J. A., Bex, P. J., and Dakin, S. C. (2009). Positional averaging explains crowding with letter-like stimuli. Proceedings of the National Academy of Sciences *106*, 13130–13135.

S6. Brainard, D. H. (1997). The Psychophysics Toolbox. Spatial Vision *10*, 433–436.

S7. Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. Spatial Vision *10*, 437–442.

S8. Efron, B., and Tibshirani, R. (1993). An Introduction to the Bootstrap (New York: Chapman and Hall).

S9. Freeman, J., and Simoncelli, E. P. (2011). Metamers of the ventral stream. Nature Neuroscience *14*, 1195–1201.

S10. Hubel, D. H., and Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. The Journal of Physiology *160*, 106–154.

S11. Bredfeldt, C. E., and Ringach, D. L. (2002). Dynamics of spatial frequency tuning in macaque V1. The Journal of Neuroscience *22*, 1976–1984.

S12. He, S., Cavanagh, P., and Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. Nature *383*, 334–337.