Current Biology
Magazine

Correspondence

How best to unify crowding?

Matthew V. Pachai, Adrien C. Doerig, and Michael H. Herzog

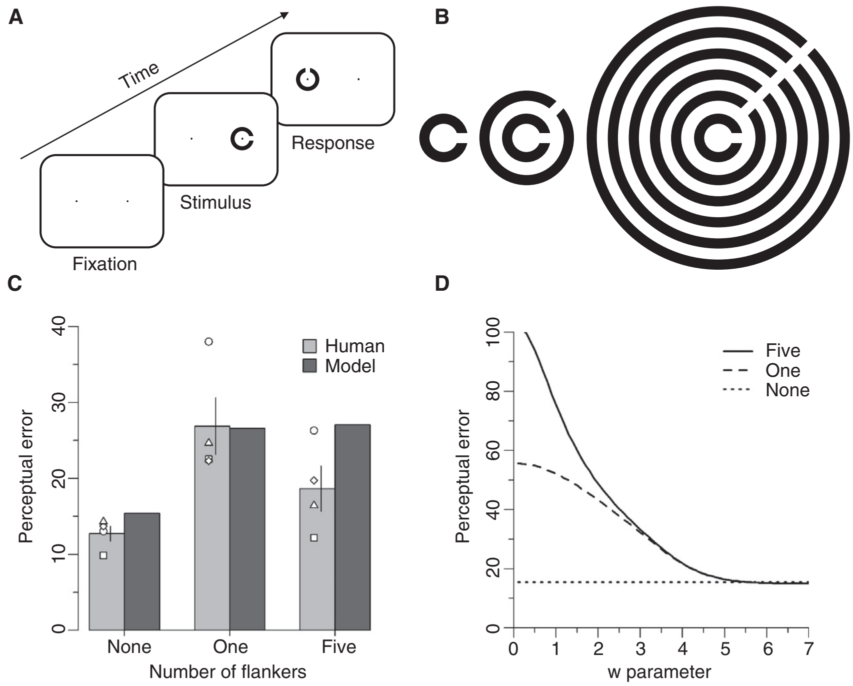

Figure 1. Experimental design and results.
(A) An example of a trial with no flankers. Observers fixated on the leftmost point, after which the stimulus appeared 10° in the visual periphery for 500 ms, followed by a response screen. Four observers (two authors, two naïve) completed the experiment in a dimly-lit room at a viewing distance of 85 cm from a 24” LCD monitor with a resolution of 1920 x 1080. Stimuli comprised black contours presented on a white background. (B) The three stimulus conditions used in our experiment. The leftmost two conditions reproduce Harrison and Bex [3]. In the third condition, we added four additional flankers, equally spaced from the target, with an identically-oriented gap. Each observer completed 200 trials per condition, randomly intermixed, following 75 practice trials. (C) Experimental results. Perceptual error, as in Harrison and Bex [3], is defined as the standard deviation of a Von Mises function fit to the distribution of errors across trials. The mean of four human observers is plotted in light grey ± 1 SEM, along with their individual data. All four observers demonstrated lower error in the five flanker condition than the one flanker condition. In dark grey, we reproduce the model of Harrison and Bex [3] with the adjustable parameter ω set at 3.5 to best fit the human results (whereas Harrison and Bex [3] used a value of 2). (D) Because ω is arbitrarily chosen to maximize fit, we simulated the effect of choosing different values for this parameter. Critically, the model never produces lower error in the five flanker condition than the one flanker condition.

In crowding, the perception of an object deteriorates in the presence of nearby elements. Obviously, crowding is a ubiquitous phenomenon, as elements are rarely seen in isolation. One of the main characteristics of crowding is that the elements themselves are not rendered invisible, but their features are *averaged* [1] or *substituted* [2] with those of neighboring elements. Recently, Harrison and Bex [3] presented “A Unifying Model of Orientation Crowding in Peripheral Vision”, which elegantly explains these two characteristics of crowding with one unifying mechanism. They tested their model using a new crowding paradigm and demonstrated an excellent match between human and model results. A key prediction of their model is that a higher number of flankers leads to stronger crowding, simply because more non-target features contribute to the model’s output and thus deteriorate performance. However, several recent studies have shown that increasing the number of flankers can actually improve performance [4–9]. Using the same experimental design as Harrison and Bex [3], we report here that adding more flankers can also improve performance in their paradigm, whereas their model predicts the opposite result. We propose that a truly unified model of crowding must include a grouping stage.

As in Harrison and Bex [3], we presented a Landolt C in isolation with its gap orientation randomly chosen (no flanker condition) or a Landolt C with a second ring containing a randomly-oriented gap surrounding it (one flanker condition). We added a third condition, in which five rings surrounded the target (five flanker condition). The gap orientation of these flankers was randomly chosen but always aligned (Figure 1B). All conditions were randomly interleaved and observers reported the location of the target gap using an adjustment response (Figure 1A). On average, errors in the no-flanker condition (M = 12.72) were lower than in the one flanker condition (M = 26.87; $t(3) = 3.82$, $p = 0.03$, two-tailed t-test), replicating Harrison and Bex [3] well. Errors in the five-flanker condition (M = 18.66) were lower than in the one-flanker condition ($t(3) = 4.08$, $p = 0.027$, two-tailed t-test), and the direction of this result was consistent in all four observers.

Using their code, we simulated the responses of Harrison and Bex’s [3] model and found that, for the five flanker condition, the model predicts the opposite of the data: performance is worse for the five flanker than the one flanker condition (Figure 1C). The model contains an adjustable parameter, ω, which determines the region of integration, and when we systematically varied this parameter, model performance was always worse in the five flanker than the one flanker condition (Figure 1D).

Why does Harrison and Bex’s [3] model fail to explain the results in the five flanker condition? They presented flanker rings at different distances and found decreased performance with each ring (with crowding strength decreasing with distance). In our five flanker condition, we essentially presented all of these rings simultaneously. In their model, outputs are determined by sampling from a probability distribution corresponding to a weighted sum of target and flanker information (with weights depending on flanker distance). Hence, by design, all five flankers contribute detrimentally to the model’s responses, which stands in *qualitative* contrast to the experimental data. Clearly, human perception in this paradigm is not strictly linear and, thus, performance on the entire stimulus cannot be predicted by a simple summation of the performance levels of its parts.

We propose that the overall stimulus configuration and grouping play a crucial role in crowding [8,9]. Specifically, when the target *ungroups* from the flankers, performance is improved, and only when the target and flankers *group* is crowding strong. However, grouping does not explain why performance deteriorates in crowding. Instead, grouping specifies which elements are prone to crowd each other (see [8], p12, point 6). For this reason, we propose that a unified model of crowding needs both a grouping stage and a mechanism to account for the detrimental effects, such as the one proposed by Harrison and Bex [3]. Further research will explore how such a model may be constructed.

ACKNOWLEDGEMENTS

We thank Harrison and Bex for sharing their model code. This work was supported by the Swiss National Science Foundation (SNF) Sinergia project: “Learning from delayed and sparse feedback”.

REFERENCES

1. Pelli, D.G., Palomares, M., and Majaj, N.J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. J. Vis. *4*, 1136–1169.
2. Strasburger, H., Harvey, Jr., L.O., and Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. Percept. Psychophys. *49*, 495–508.
3. Harrison, W.J., and Bex, P.J. (2015). A unifying model of orientation crowding in peripheral vision. Curr. Biol. *25*, 3213–3219.
4. Banks, W.P., Larson, D.W., and Prinzmetal, W. (1979). Asymmetry of visual interference. Percept. Psychophys. *25*, 447–456.
5. Põder, E. (2006). Crowding, feature integration, and two kinds of “attention”. J. Vis. *6*, 163–169.
6. Saarela, T.P., Sayim, B., Westheimer, G., and Herzog, M.H. (2009). Global stimulus configuration modulates crowding. J. Vis. *9*, 1–11.
7. Levi, D.M., and Carney, T. (2009). Crowding in peripheral vision: why bigger is better. Curr. Biol. *19*, 1988–1993.
8. Herzog, M.H., Sayim, B., Chicherov, V., and Manassi, M. (2015). Crowding, grouping, and object recognition: A matter of appearance. J. Vis. *15*, 1–18.
9. Manassi, M., Sayim, B., and Herzog, M.H. (2012). Grouping, pooling, and when bigger is better in visual crowding. J. Vis. *12*, 1–14.

Laboratory of Psychophysics, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.
E-mail: matt.pachai@epfl.ch

Correspondence

Reply to Pachai *et al.*

William J. Harrison[1,2] and Peter J. Bex[3]

Peripheral vision is fundamentally limited by the spacing between objects. When asked to report a target’s identity, observers make erroneous reports that sometimes match the identity of a nearby distractor and sometimes match a combination of target and distractor features. The classification of these errors has previously been used to support competing ‘substitution’ [1] or ‘averaging’ [2] models of the phenomenon known as ‘visual crowding’. We recently proposed a single model in which both classes of error occur because observers make their reports by sampling from a biologically-plausible population of weighted responses within a region of space around the target [3]. It is critical to note that there is no probabilistic substitution or averaging process in our model; instead, we argue that neither substitution nor averaging occur, but that these are misclassifications of the distribution of reports that emerge when a population response distribution is sampled. This is a fundamentally different way of thinking about crowding, and on this basis we claim to have provided a mechanism unifying categorically distinct perceptual errors. Our goal was not to model all crowding phenomena, such as the release from crowding when target and flanks differ in color or depth [4]. Pachai *et al.* [5] have suggested that our model is not unifying because it inaccurately predicts perceptual performance for a particular stimulus. Although we agree that our model does not predict their data, this specific demonstration overlooks the critical aspect of the model: perceptual reports are drawn from a weighted population code. We show that Pachai *et al*.’s [5] own data actually provide evidence for the population code we have described [3], and we suggest a biologically-plausible analysis of their stimuli that provides a computational basis for their ‘grouping’ account of crowding.

In both Pachai *et al*.’s [5] work and our original study [3], following the presentation of a randomly oriented target Landolt C in peripheral vision, observers adjusted a foveal Landolt C so that it matched the target orientation. The target was presented alone (Figure 1A, top panel), or surrounded by a larger and independently oriented Landolt C (Figure 1B, top panel). In close agreement with our data [3], the mean error in their observers’ reports is greater with the flanker than without. They further show that our model provides a good fit to these behavioural data. Pachai *et al.* [5] included a second flanker condition, in which the target was flanked by five concentric distractors, all with the same orientation (Figure 1C, top panel). For this condition, our model generally predicts that observers’ performance should be worse than in the one-flanker condition. In contrast to this prediction, performance improved in this new condition. Pachai *et al.* [5] conclude that our model is fundamentally limited because it predicts crowding according to the distance between target and flankers. They advance a ‘grouping’ explanation, in which crowding is released because the flankers are somehow grouped independently of the target. Notice, however, they still found crowding even in their grouped-flanker condition. In Figure S1 in the Supplemental Information, we provide a demonstration that shows, consistent with our approach, crowding in this condition also relies on the distance between target and inner flanker. Nevertheless, we believe that the conclusions of Pachai *et al.* [5] overlook the critical aspect of our model.

In our model [3], populations of neurons coding a target’s orientation also code information about flanking distractors and this contamination leads to perceptual errors (details given in [3]). Shown in the top panel of Figure 1E are the trial-by-trial errors made by observers in the one-flanker condition of Pachai *et al.* [5]. These data are very well captured by our population code model [3]. We think that a simple extension of