Multi-level Crowding: Evidence from Global and Local Misreports

Mikel Jimenez (✉ mikeljimenez@gmail.com)
University of Haifa

Ruth Kimchi
University of Haifa

Amit Yashar
University of Haifa

Research Article

Keywords: Crowding, Visual periphery, Gestalt, Illusory Form, Grouping, Multi-level, Probabilistic mixture models, Object Recognition, Pooling, Misreport

Posted Date: January 31st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1279846/v1

License: ☺️ Ⓚ This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License
Abstract

Crowding refers to the inability to recognize objects in clutter, setting a fundamental limit on object recognition. Here, we investigated the processing level at which crowding occurs by exploring the type of crowding errors (global, local, or both). Twenty-seven observers estimated the orientation of a target when presented alone or surrounded by flankers (local shapes). Flankers were aligned to create an illusory rectangle (enhanced global configuration) or misaligned (reduced global configuration). We analyzed the error distributions by fitting probabilistic mixture models. Results showed that often participants misreported the orientation of a flanker instead of that of the target. Interestingly, in some trials the orientation of the global configuration was misreported. These results suggest that crowding occurs simultaneously across multiple levels of visual processing and crucially depends on the spatial configuration of the stimulus. Thus, crowding might be characterized as a bottleneck of visual object identification at different levels of representation.

Introduction

Recognition of objects is limited by their spacing. When objects are too close together, they become indistinguishable, a phenomenon known as crowding. Crowding sets a fundamental limit on conscious visual perception (Whitney & Levi, 2011) and impairs reading, eye and hand movements, visual search and other functions in typical peripheral and central amblyopic vision (Levi, 2008).

The study of crowding has gained increased attention in recent years, and our knowledge of crowding characteristics is substantial (Manassi & Whitney, 2018; Pelli & Tillman, 2008; Rosenholtz, Yu, & Keshvari, 2019; Whitney & Levy, 2011). For example, crowding impairs target identification but not detection (e.g., Livne & Sagi, 2007; Pelli, Palomares, & Majaj, 2004) and it depends on both target eccentricity and the distance between target and flankers (i.e., the critical spacing, 0.3-0.5 of target eccentricity; Bouma, 1970; Kooi, Toet, Tripathy, & Levi, 1994; Levi & Carney, 2009; Pelli et al., 2004; but see Manassi, Sayim, & Herzog, 2012; Vickery, Shim, Chakravarthi, Jiang, & Luedeman, 2009). In addition, crowding depends on the similarity between the target and the flankers, such that flankers more similar to the target (e.g., in color, shape, depth, spatial frequency or complexity) produce stronger crowding (Bernard & Chung, 2011; Chung, Levi, & Legge, 2001; Zhang, Zhang, Xue, Liu, & Yu, 2009).

Nonetheless, the underlying processes by which crowding occurs are still unclear. Most of the prevailing theories of crowding propose that crowding occurs because of the integration or “pooling” of low-level features at a single, relatively early stage of visual processing (e.g., He, Cavanagh, & Intriligator, 1996; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Pelli et al., 2004), in line with the classic hierarchical model of object recognition (Riesenhuber & Poggio, 1999; Serre, Oliva, & Poggio, 2007; DiCarlo, Zoccolan, & Rust, 2012). In these models, information proceeds from the processing of low-level visual features (such as spatial frequency, orientation or color) to the perception of coherent forms, shapes and complex objects. Importantly, the input of higher-level visual areas would be fully determined by the outputs of basic, lower-level feature detectors, and thus information lost at early stages of visual processing would
be irretrievably lost (Freeman & Simoncelli, 2011; Herzog, Sayim, Chicherov, & Manassi 2015). If crowding occurs due to the averaging of low-level visual features between target and flanks at an early visual stage, it would disrupt the formation of any higher-level target representation, and crowding would represent an early bottleneck of human vision.

However, in contrast to the prediction of a strict “pooling” or averaging mechanism, crowding errors often reflect reports of a flanker instead of the target – known as misreport errors (Ester, Klee, & Awh, 2014; Freeman, Chakravarthi, & Pelli, 2012; Harrison, & Bex, 2015; Shechter & Yashar, 2021; Yashar, Wu, Chen, & Carrasco, 2019). Interestingly, misreport errors lead to feature-binding errors - i.e., reporting the color of the target along with the orientation of a flanker (Yashar et al., 2019), indicating that crowding reflects a more complex integration failure than a simple averaging.

Classic models of crowding are also challenged by recent findings showing that crowding can occur at different levels of visual processing, not only between lower-level features but also between high-level representations of objects. Indeed, crowding may occur also for complex stimuli, like shapes (Kimchi & Pirkner, 2015; Pirkner & Kimchi, 2017), everyday objects (Wallace & Tjan, 2011) or faces (Martelli, Majaj, & Pelli, 2005; Louie, Bressler, & Whitney, 2007; Farzin, Rivera, & Whitney, 2009; see Manassi, & Whitney, 2018, for a review). For example, Kimchi and Pirkner (2015) examined whether crowding could occur at the object configural level in addition to feature- or part-level crowding. Results showed that crowding was weaker when the flanks were similar to the target’s local parts than when the flanks were similar to the target whole configuration. Herzog et al. (2015) showed that the same vertical flanks lose their crowding strength when becoming part of rectangles or good Gestalts. Interestingly, a recent study by Doerig, Bornet, Rosenholtz, Francis, Clarke and Herzog (2019) showed that crowding models that incorporate a grouping component strongly improve model performance. These findings suggest that crowding can occur at the global or configural levels and that grouping processes precede crowding. However, it is still unknown whether and how crowding processes operate at the local and global levels of stimulus representation.

In this study, we addressed this question by exploring whether and how misreport errors depend on the level of visual processing. Observers performed an orientation estimation task of a target (a black rectangle with two triangular cut-outs) when presented alone or surrounded by flanks in two conditions (Figure 1). In one condition, the flanks were aligned to create a coherent illusory rectangular configuration. Illusory – or Kanizsa – shapes (Kanizsa, 1979) refer to the perception of a shape defined by sharp illusory contours (see Spillmann & Dresp, 1995, for a review). In the second condition the flanks were misaligned, so no illusory shape was formed, yet inducers might be grouped according to Gestalt principles (e.g., Kimchi, 1998) forming a global shape, presumably weaker than the illusory one.

For each trial, we calculated the estimation error for orientation by subtracting the true value of the target from the estimation value and analyzed the error distributions by fitting probabilistic mixture models. If crowding occurs only at a lower level of visual processing, we expect crowding to be induced by local shapes, but not by global shapes. Namely, observers will misreport the orientation of a flanker, but not the
orientation of the global shape formed by the flankers, as the target orientation. If crowding occurs also at a higher level of visual processing, we expect crowding to be induced by local shapes, but also by global shapes (i.e., observers will misreport local as well as global orientations as the target orientation). Manipulating the strength of the global shape (illusory vs. grouped) would allow us to examine how different perceptual organization processes modulate crowding.

**Methods**

**Observers**

Twenty-seven graduate and undergraduate students (17 females: age range = 18 - 39 years, $M = 28.23$, $SD = 6.07$) from The University of Haifa participated in the experiment for either course credit or monetary payment (40 ILS per hour, around $12). The sample size was calculated on the basis of an a priori power analysis to detect a crowding effect with 80% power and a moderate effect size (0.5), given a .05 significance criterion. All observers had normal or corrected-to-normal visual acuity and normal color vision. None of them reported either attention deficits or epilepsy. A written informed consent was signed by the participants before the experiment. The method was carried out in accordance with the Declaration of Helsinki and was approved by the Human Ethics Committee of the University of Haifa.

**Apparatus**

The stimuli were displayed on a gamma-corrected 21-in CRT monitor (SGI, with 1280 × 960 resolution and 85-Hz refresh rate) connected to an iMac and were programmed in Matlab (The MathWorks, Inc., Natick, MA) using the Psychophysics Toolbox extensions (Kleiner, Brainard, & Pelli, 2007). An Eyelink 1000 Plus (SR Research, Ottawa, ON, Canada) system was used to monitor eye movements, and viewing distance was set to 57 cm using a chin rest. Observers used the mouse to report their responses.

**Stimuli**

The stimuli were presented on a grey background with a luminance level of 56 cd/m$^2$. The fixation display consisted of black cross subtending 0.3° at the center of the screen. The target display consisted of the fixation cross along with the target; a rectangular black (0 cd/m$^2$) shape subtending 1.1° in height and 0.9° in width, presented on the horizontal meridian, either on the left or on the right hemifield, and with 9° eccentricity. A triangular shape (0.1°) was cut out from the two sides of the rectangle (Fig. 1A). The target could appear alone (uncrowded condition) or surrounded by four ankers in two different anker configurations: ankers aligned and ankers misaligned conditions. The anker stimuli were identical to the target stimulus. The center-to-center target-anker spacing was 1.8°.

Target and flanker orientation were selected randomly from a circular parameter space, which consisted of 180 values evenly distributed between 1° and 180°. In both the flankers aligned and flankers
misaligned conditions, diagonal flanks had always the same orientation, therefore flanks were always rotated in pairs. In the flanks aligned condition, flanks’ triangular corners were aligned so they created an illusory rectangle (2° in height and 3° in width, see Fig. 1A).

**Procedure and design**

**Figure 1B** illustrates the trial sequence. We instructed observers to fixate on the fixation cross during the trial presentations. Each trial began with the presentation of the fixation display for 500 ms, which continued until the observer fixated consecutively in the fixation cross for 300 ms. Following observer’s fixation, the target display appeared for 200 ms. After the stimulus display, a blank screen was presented for 500ms, which was followed by a response screen. During the response display, observers estimated the target orientation by selecting a position for the target stimulus on an orientation wheel (see Fig. 1B). Following observer’s response, a blank inter-trial interval (ITI) appeared for 500ms.

Observers were presented with 200 trials for each of the three display conditions (uncrowded, flanks aligned and flanks misaligned), for a total of 600 trials per session. The experiment was divided in 10 blocks of 60 trials and lasted approximately 60 minutes. Display conditions were randomly mixed within each block. Observers were advised to take short rests between blocks.

In each trial, eye fixation was monitored using an eye tracker (see Apparatus). Trials in which fixation was broken (fixation window was < 2° for twenty subjects, and < 3° for seven subjects) were eliminated from the data and rerun at the end of the block.

**Models and analyses**

The estimation error for orientation reports was calculated by subtracting the true value of the target from the estimation value at each trial, such that zero indicated the target value. We calculated flanks’ values by subtracting the true value of the target from the values of the flanks in order to assess the contribution of the flanks to the error distribution. The error distributions were analyzed by fitting probabilistic mixture models, developed from both the standard model and the standard with misreport model (Bays, Catalao, & Husain, 2009). We compared three different models:

The Standard Mixture Model (Equation 1) with two components: a von Mises (circular) distribution that describes the probability density of reports around the target’s orientation, and a uniform distribution that describes the probability of reports that are unrelated to the target (guessing rate). Thus, the model includes two free parameters ($\gamma$, $\sigma$):

$$p(\theta) = (1 - \gamma)f(\theta) + \gamma \left( \frac{1}{n} \right)$$
being θ the value of the estimation error and γ the proportion of trials in which observers are randomly guessing (guessing rate). f(θ)σ is the von Mises distribution with a standard deviation of σ (variability; the mean was set to zero), and n the total number of possible values for the target’s orientation (in our case, 180).

The Local model (Equation 2) includes three free parameters. The model adds a misreporting component to the standard mixture model, which describes the probability of reporting the orientation of any of the four local-flankers to be the target:

\[
p(\theta) = (1 - \gamma - \beta) f(\theta)\sigma + \gamma \left( \frac{1}{180} \right) + \beta \frac{1}{m} \sum_i m_i f(\theta_i^\ast)\sigma
\]

where β the probability of reporting a flanker orientation as the target orientation, m represents the total number of nontarget items (four in the present study), and \( \theta_i^\ast \) is the error to the feature of the \( i \)th flanker. Notice that in this model, the von Mises distribution of the estimation errors \( [f(\theta)] \), describes the distribution when the observer correctly estimated the target’s feature; thus, its mean is zero. For \( f(\theta_i^\ast) \), which represents the distribution of estimating one flanker, the mean would be the orientation distance of the corresponding flanker to that of the target. Here, the variability of the distributions for each stimulus is assumed to be the same.

Finally, the Global-Local model (Equation 3) includes four free parameters. The model adds two misreporting components to the standard misreport model, which describe the probability of reporting the orientation of the global shape and the probability of reporting the orientation of any of the four local flankers to be the target:

\[
p(\theta) = (1 - \gamma - \beta_{\text{local}} - \beta_{\text{global}}) f(\theta)\sigma + \gamma \left( \frac{1}{180} \right) + \beta_{\text{local}} \frac{1}{m} \sum_i m_i f(\theta_i^\ast)\sigma + \beta_{\text{global}} f(\theta_g^\ast)\sigma
\]

where \( \beta_{\text{global}} \) is the probability of reporting the orientation of the global shape, \( \beta_{\text{local}} \) is the probability of reporting a flanker as the target, m is the total number of nontarget items (four in the present study), and \( \theta_i^\ast \) is the error to the feature of the \( i \)th flanker.

We used the MemToolbox (Suchow, Brady, Fougnie, & Alvarez, 2013) to fit the models to the individual data and then compared the Akaike information criterion with correction (AICc) to assess their fits.

Results

For each observer in each condition, we examined the bias of the errors by calculating the mean error. Mean error was close to zero in the uncrowded condition (\( M = -0.78, SD = 1.57 \)), the flankers aligned (\( M = 0.77, SD = 3.14 \)) and flankers misaligned (\( M = 1.12, SD = 3.68 \)) conditions. We then calculated precision as the inverse of the variance of the errors for each observer in each condition (Fig. 2A). We conducted a
one-way Analysis of Variance (ANOVA) on precision with display conditions (uncrowded, flankers aligned, flankers misaligned) as a within subject factor. A main effect on precision was observed, $F(2,52) = 130.59, p < 0.001, \eta^2_p = 0.83$, indicating significant differences in precision between the three display conditions. Pairwise comparisons (Bonferroni corrected) showed that precision was significantly higher when the target was presented uncrowded compared to when it was presented surrounded by flankers (aligned: $p < 0.001$; misaligned: $p < 0.001$). On the other hand, no differences in precision were found between the two flanker conditions ($p = .091$).

**Probabilistic models**

In the uncrowded condition, the standard mixture model described accurately the distribution of the errors (Fig. 2C). For each flanker condition we compared the two relevant models (e.g., the Local model and the Global-Local model) by calculating the Akaike information criterion with correction (AICc) for each observer. Figure 2b shows the mean AICc differences between the relevant models in each flanker condition. The Global-Local model outperformed (i.e., lower AICc value) the Local model in both the flankers aligned and flankers misaligned conditions [$t(26) = 2.28, p = .031$, Cohen's $d = 0.44$; $t(26) = 3.02, p = .006$, Cohen's $d = 0.58$, respectively]. Thus, we analyzed the fitted parameters of the best performing models (i.e., the standard mixture for the uncrowded display, and the Global-Local model for the two types of crowded displays).

We calculated target reporting rate ($P_T$) by subtracting the accumulative guessing rate and misreport rate from 1 (i.e., $P_T = 1 - \gamma$ for the standard mixture, $P_T = 1 - \gamma - \beta_{\text{global}} - \beta_{\text{local}}$ for the Global-Local model) for each fitted model. Figure 3 depicts the mean guessing rate ($\gamma$), variability ($\sigma$) and target reporting rate ($P_T$) of the fitted models in each condition. To assess the effect of crowding on performance, we conducted one-way, repeated measure ANOVAs on guessing rate, variability and target reporting rate as dependent variables, with display condition as a within subject factor. For guessing rate, a main effect of display condition [$F(2,38) = 4.99, p = 0.019, \eta^2_p = 0.16$] suggested differences between conditions (note that Greenhouse-Geisser corrections are applied on $p$-values and degrees of freedom when the sphericity assumption is violated). Pairwise comparisons showed that the guessing rate was significantly lower when the target was presented alone (uncrowded), than when it was flanked (aligned: $p = .015$; misaligned: $p = .024$), yet no differences were found between the two flanker conditions ($p = 1.000$). For variability, a main effect of crowding [$F(2,42) = 34.11, p < 0.001, \eta^2_p = 0.57$] also suggested differences between conditions. Similar to the pattern for guessing rate, variability in the orientation reports was significantly lower when the target was presented alone than when it was flanked (aligned: $p < 0.001$; misaligned: $p < 0.001$), yet no differences were found between flanker conditions ($p = .350$). For target reporting rate, a significant main effect was found [$F(2,52) = 34.11, p < 0.001, \eta^2_p = 0.57$], suggesting that target reporting rate significantly differed between conditions. Here, pairwise comparisons revealed significant differences between the three conditions: the probability of reporting the target orientation when the target was presented alone was significantly higher than when presented with flankers (aligned:
Results and Discussion

In the present study, we investigated the processing level at which crowding occurs by using crowded displays in which the ankers (local shapes) could form a global shape (with or without illusory contours) and exploring the extent to which crowding errors depend on these global-local levels. Our results show that observers misreported the orientation of both global and local shapes as the orientation of the local target, suggesting that crowding operates at various levels of visual processing.

The analyses of model parameters showed a common pattern of significant differences for uncrowded vs. crowded displays (Ester et al., 2014; Freeman et al., 2012; Harrison, & Bex, 2015; Shechter & Yashar, 2021; Yashar et al., 2019): Observers were more precise and reported the target orientation more accurately when the target was uncrowded than when it was surrounded by flankers; also, the error variability and guessing rate were significantly lower in the uncrowded condition. Interestingly, the probability of reporting the target when the flankers were aligned was significantly higher compared to the condition where the flankers were misaligned ($p = .012$).

Finally, we analyzed the misreport rates to the global and local orientations. Figure 4 depicts the probability of misreporting the global (i.e., either the illusory rectangle or the perceived global shape of the four ankers) or one of the local (i.e., one of the anker's) orientation in the two anker conditions.

The probability of misreporting either a global or a local orientation was significantly different from 0 in both anker conditions, as assessed by independent $t$-comparisons (all $ps < 0.001$), showing that participants misreported both the global and the local orientations as target orientation in the two anker conditions. In addition, results revealed no differences in the probability of misreporting the global shape between anker conditions [$t(26) = 0.37, p = .715$], but the probability of misreporting a local anker orientation was significantly higher when the ankers were misaligned [$t(26) = -2.94, p = .007$, Cohen's $d = -0.56$]. As a control, we explored the probability of misreporting the perpendicular (+ 90°) global orientation, which should have a value of 0 since this is the most distant orientation from the actual orientation of the global shape. Indeed, independent $t$-comparisons showed that misreporting the perpendicular global orientation was close to 0 in each anker condition, mean misreport rate of the global perpendicular orientation being 0.002 and 0.009 in the ankers aligned and flanks misaligned, respectively.

Discussion

In the present study, we investigated the processing level at which crowding occurs by using crowded displays in which the flankers (local shapes) could form a global shape (with or without illusory contours) and exploring the extent to which crowding errors depend on these global-local levels. Our results show that observers misreported the orientation of both global and local shapes as the orientation of the local target, suggesting that crowding operates at various levels of visual processing.

Importantly, we analyzed the misreport rates of the global and local orientations. If crowding occurs at a lower level of visual processing, we expected observers to misreport the orientation of a flanker (local orientation), but not the orientation of the global shape (illusory or grouped) formed by the flankers (global orientation), as the target orientation. Yet, if crowding occurs at a higher level of visual processing, we expected observers to misreport local as well as global orientations as the target orientation. The latter case is what we found. Model comparison showed that a model with both global and local
misreports produced a better fit than a model with only local misreports. Furthermore, model fitting revealed that observers significantly misreported both the global and local orientations as target orientation, showing that the global orientation of the flanker configuration is preserved in crowded displays, and suggesting that crowding may occur also at a higher level of visual processing between global and local stimulus representations.

The manipulation of the alignment of the flankers produced some interesting results. Even though the global orientation was equally misreported when the flankers where either aligned (enhanced global configuration) or misaligned (reduced global configuration), the probability of misreporting a local flanker orientation was significantly higher when the flankers were misaligned. This, together with the finding that the alignment of the flankers produced higher target reporting rates ($P_T$), suggests that the perception of the illusory shape modulated crowding effects, in line with previous findings showing that flankers might lose their crowding strength when becoming part of rectangles or good Gestalts (Herzog et al., 2015). Note that this modulation, however, is not manifested by a higher probability of misreporting the orientation of the illusory rectangular orientation, but rather, by a lower probability of misreporting flanker orientations as target orientation, indicating a reduction in the strength of crowding.

The finding that both the global orientation of the grouped flankers (flankers misaligned condition) and the orientation of the illusory rectangle (flankers aligned condition) produced similar misreport rates might be explained by recent evidence related to the processing of illusory shapes under restrictive visual conditions or when presented in peripheral vision. First, disentangling the perception of the illusory shape from the grouping of the local (flanker) elements is not easy when these configurations are presented under challenging visual conditions. This issue was explicitly studied by Jimenez and Montoro (2018), who displayed masked illusory (black “pacmen” and semicircles arranged in such a way that they produced horizontal or vertical illusory bars) and grouping (the same pacmen and semicircles rotated in a way that they did not produce illusory figures) primes that could be congruent or incongruent in their orientation with subsequent probe stimuli (vertical vs. horizontal bars). The authors found significant priming effects for illusory and grouping primes at different prime durations but, crucially, the magnitude of the priming effect was equal for the illusory and grouped primes. Thus, the dissociation of the grouped percept from that generated by the illusory shape is not accomplished easily under restrictive visual conditions. Second, it seems that retinal eccentricity plays a crucial role in illusory shape processing. Bakar, Liu, Conci, Elliott and Ioannides (2008) investigated this by presenting illusory shapes at central and peripheral visual field locations. Their behavioral results revealed that central stimulus presentations elicited faster responses than those presented at one of the four quadrants. In addition, magnetoencephalographic responses to illusory figures showed that for central presentations, specific responses to illusory figures peaked first in V1/V2 (96–101 ms), and then in the lateral occipital complex (LOC; 132–141 ms). For peripheral presentations, the relative modulation towards illusory figures was markedly reduced in V1/V2 and LOC while prominent activation peaks shifted to the fusiform gyrus (from 200 ms onwards). Thus, the processing of illusory shapes in the periphery required longer latencies and
the involvement of higher-level visual areas. Overall, this combined evidence might explain the absence of differences in misreporting the global orientation between flanker conditions observed in our results.

In sum, the present results suggest that crowding occurs simultaneously across multiple levels of visual processing, both at a lower level between local shapes but also at a higher level between global and local shapes. These results are in line with recent findings (Doerig et al., 2019) showing that the global stimulus configuration plays a crucial role in crowding, stressing the necessity of including grouping-like processes as a fundamental explanation of crowding. Since both local flanker information and global shape representations are preserved (not averaged) in crowded displays, our results provide a further challenge to low-level “pooling” or averaging models of crowding. The development of new pooling models (Bornet, Choung, Doerig, Whitney, Herzig, Manassi, 2021; Rosenholtz, Yu & Keshvari, 2019), which incorporate the possibility of multilevel crowding and account for complex target-flanker interactions, might lead to a better explanation of the phenomena of crowding in general, and the current results in particular.

**Declarations**

**Data Availability**

The methods used and the data analyzed in the present study are available in the Open Science Framework (OSF) repository, in the following link: [https://osf.io/f3cz8/?view_only=b47af6157a1941d48edbd3b0ecea10](https://osf.io/f3cz8/?view_only=b47af6157a1941d48edbd3b0ecea10).

**Author Contribution**

AY and RK developed the study concept. All authors contributed to the study design. Testing and data collection was performed by MJ. M.J and AY performed the data analysis and all authors contributed to result interpretation. MJ drafted the manuscript, and AY and RK provided critical revisions. All authors approved the final version of the manuscript for submission.

**Competing interests**

The author(s) declare no competing interests.

**References**

1. Bakar, A. A., Liu, L., Conci, M., Elliott, M. A., & Ioannides, A. A. (2008). Visual field and task influence illusory figure responses. *Human Brain Mapping, 29*(11), 1313-1326. [https://doi.org/10.1002/hbm.20464](https://doi.org/10.1002/hbm.20464)
2. Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision, 9*(10), 7-7. https://doi.org/10.1167/9.10.7

3. Bernard, J. B., & Chung, S. T. (2011). The dependence of crowding on flanker complexity and target–flanker similarity. *Journal of Vision, 11*(8), 1-1. https://doi.org/10.1167/11.8.1

4. Bornet, A., Choung, O. H., Doerig, A., Whitney, D., Herzog, M. H., & Manassi, M. (2021). Global and high-level effects in crowding cannot be predicted by either high-dimensional pooling or target cueing. *Journal of Vision, 21*(12), 1-10. https://doi.org/10.1167/jov.21.12.10

5. Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226*(5241), 177-178. https://doi.org/10.1038/226177a0

6. Chung, S. T., Levi, D. M., & Legge, G. E. (2001). Spatial-frequency and contrast properties of crowding. *Vision Research, 41*(14), 1833-1850. https://doi.org/10.1016/S0042-6989(01)00071-2

7. DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron, 73*(3), 415-434. https://doi.org/10.1016/j.neuron.2012.01.010

8. Doerig, A., Bornet, A., Rosenholtz, R., Francis, G., Clarke, A. M., & Herzog, M. H. (2019). Beyond Bouma's window: How to explain global aspects of crowding? *PLoS Computational Biology, 15*(5), e1006580. https://doi.org/10.1371/journal.pcbi.1006580

9. Ester, E. F., Klee, D., & Awh, E. (2014). Visual crowding cannot be wholly explained by feature pooling. *Journal of Experimental Psychology: Human Perception and Performance, 40*(3), 1022. https://doi.org/10.1037/a0035377

10. Farzin, F., Rivera, S. M., & Whitney, D. (2009). Holistic crowding of Mooney faces. *Journal of Vision, 9*(6), 18-18. https://doi.org/10.1167/9.6.18

11. Freeman, J., Chakravarthi, R., & Pelli, D. G. (2012). Substitution and pooling in crowding. *Attention, Perception, & Psychophysics, 74*(2), 379-396. https://doi.org/10.3758/s13414-011-0229-0

12. Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience, 14*(9), 1195-1201. https://doi.org/10.1038/nn.2889

13. Harrison, W. J., & Bex, P. J. (2015). A unifying model of orientation crowding in peripheral vision. *Current Biology, 25*(24), 3213-3219. https://doi.org/10.1016/j.cub.2015.10.052

14. He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383*(6598), 334-337. https://doi.org/10.1038/383334a0

15. Herzog, M. H., Sayim, B., Chicherov, V., & Manassi, M. (2015). Crowding, grouping, and object recognition: A matter of appearance. *Journal of Vision, 15*(6), 5-5. https://doi.org/10.1167/15.6.5

16. Intriligator, J., & Cavanagh, P. (2001). The spatial resolution of visual attention. *Cognitive Psychology, 43*(3), 171-216. https://doi.org/10.1006/cogp.2001.0755

17. Jimenez, M., & Montoro, P. R. (2018). Illusory Form Perception and Perceptual Grouping Operations under Conditions of Restricted Visual Awareness. *The Spanish Journal of Psychology, 21*. https://doi.org/10.1017/sjp.2018.47

18. Kanizsa, G. (1979). *Organization in vision: Essays on gestalt perception*. New York [etc.]: Praeger.
19. Kleiner, M., Brainard, D., & Pelli, D. (2007). What's new in Psychtoolbox-3?
20. Kimchi, R. (1998). Uniform connectedness and grouping in the perceptual organization of hierarchical patterns. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 1105-1118. https://doi.org/10.1037/0096-
21. Kimchi, R., & Pirkner, Y. (2015). Multiple level crowding: Crowding at the object parts level and at the object configural level. *Perception, 44*(11), 1275-1292. https://doi.org/10.1177/0301006615594970
22. Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision, 8*(2), 255-280. https://doi.org/10.1163/156856894x00350
23. Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research, 48*(5), 635-654. https://doi.org/10.1016/j.visres.2007.12.009
24. Levi, D. M., & Carney, T. (2009). Crowding in peripheral vision: Why bigger is better. *Current Biology, 19*(23), 1988-1993. https://doi.org/10.1016/j.cub.2009.09.056
25. Livne, T., & Sagi, D. (2007). Conguration influence on crowding. *Journal of Vision, 7*(2), 4-4. https://doi.org/10.1167/7.2.4
26. Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding: Selective interference between configural representations of faces in crowded scenes. *Journal of Vision, 7*(2), 24-24. https://doi.org/10.1167/7.2.24
27. Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision, 12*(10), 13-13. https://doi.org/10.1167/12.10.13
28. Manassi, M., & Whitney, D. (2018). Multi-level crowding and the paradox of object recognition in clutter. *Current Biology, 28*(3), R127-R133. https://doi.org/10.1016/j.cub.2017.12.051
29. Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision, 5*(1), 6-6. https://doi.org/10.1167/5.1.6
30. Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4*(7), 739-744. https://doi.org/10.1038/89532
31. Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*, 1136–1169. https://doi.org/10.1167/4.12.12
32. Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience, 11*(10), 1129-1135. https://doi.org/10.1038/nn.2187
33. Pirkner, Y., & Kimchi, R. (2017). Crowding and perceptual organization: Target's objecthood influences the relative strength of part-level and configural-level crowding. *Journal of Vision, 17*(11):7, 1–18. https://doi.org/10.1167/17.11.7
34. Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2*(11), 1019-1025. https://doi.org/10.1038/14819
35. Rosenholtz, R., Yu, D., & Keshvari, S. (2019). Challenges to pooling models of crowding: Implications for visual mechanisms. *Journal of vision, 19*(7), 15-15. https://doi.org/10.1167/jov.19.7.15

36. Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, USA, 104*(15), 6424–6429. https://doi.org/10.1073/pnas.0700622104

37. Shechter, A., & Yashar, A. (2021). Mixture model investigation of the inner–outer asymmetry in visual crowding reveals a heavier weight towards the visual periphery. *Scientific Reports, 11*(1), 1-12. https://doi.org/10.1038/s41598-021-81533-9

38. Spillmann, L., & Dresp, B. (1995). Phenomena of illusory form: Can we bridge the gap between levels of explanation? *Perception, 24*(11), 1333-1364. https://doi.org/10.1068/p241333

39. Suchow, J. W., Brady, T. F., Fougnie, D., & Alvarez, G. A. (2013). Modeling visual working memory with the MemToolbox. *Journal of Vision, 13*(10), 9-9. https://doi.org/10.1167/13.10.9

40. Vickery, T. J., Shim, W. M., Chakravarthi, R., Jiang, Y. V., & Luedeman, R. (2009). Supercrowding: Weakly masking a target expands the range of crowding. *Journal of Vision, 9*(2), 12-12. https://doi.org/10.1167/9.2.12

41. Wallace, J. M., & Tjan, B. S. (2011). Object crowding. *Journal of Vision, 11*(6), 19-19. https://doi.org/10.1167/11.6.19

42. Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences, 15*(4), 160-168. https://doi.org/10.1016/j.tics.2011.02.005

43. Yashar, A., Wu, X., Chen, J., & Carrasco, M. (2019). Crowding and binding: Not all feature dimensions behave in the same way. *Psychological Science, 30*(10), 1533-1546. https://doi.org/10.1177/0956797619870779

44. Zhang, J. Y., Zhang, T., Xue, F., Liu, L., & Yu, C. (2009). Legibility of Chinese characters in peripheral vision and the top-down influences on crowding. *Vision Research, 49*(1), 44-53. https://doi.org/10.1016/j.visres.2008.09.021

**Figures**
Figure 1

Stimulus type and display conditions. (A) The three display conditions. From top to bottom, U: Uncrowded, FA: Flankers Aligned, FM: Flankers Misaligned (B) Illustration of the sequence of events within a trial.
Figure 2

Error distributions, model comparison and precision. (A) Mean precision (inversed SD in radiance) of the errors in each display condition, U: uncrowded, FA: Flankers Aligned, FM: Flankers Misaligned. (B) Mean $\Delta$AICcs for each flanker condition. $\Delta$AICcs was calculated by subtracting the AICc of the Local model from that of the Global-Local model ($\text{AICc}_{\text{Global-local}} - \text{AICc}_{\text{Local}}$). Lower AICc indicates better model fit. In both conditions the global-local model outperformed the local model. (C) Frequency of errors (reported value minus true value of the target) for each of the three display conditions. Error bars represent standard error.
Figure 3

Model parameters from the fitting of the Global-Local model. (A) Guessing rate ($\gamma$), (B) Mean variability ($\sigma$), and (C) target reporting rate ($P_T$) for each display condition. U: uncrowded, FA: Flankers Aligned, FM: Flankers Misaligned. Error bars represent standard error.
Figure 4

Probability of misreporting the global shape orientation ($\beta_{global}$) or a local flanker orientation ($\beta_{local}$) as the target for flankers aligned and flankers misaligned conditions.