Dissecting (un)crowding

Oh-Hyeon Choung

Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland ✉

Alban Bornet

Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland ✉

Adrien Doerig

Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland
Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands ✉

Michael H. Herzog

Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland ✉

In crowding, perception of a target deteriorates in the presence of nearby flankers. Surprisingly, perception can be rescued from crowding if additional flankers are added (uncrowding). Uncrowding is a major challenge for all classic models of crowding and vision in general, because the global configuration of the entire stimulus is crucial. However, it is unclear which characteristics of the configuration impact (un)crowding. Here, we systematically dissected flanker configurations and showed that (un)crowding cannot be easily explained by the effects of the sub-parts or low-level features of the stimulus configuration. Our modeling results suggest that (un)crowding requires global processing. These results are well in line with previous studies showing the importance of global aspects in crowding.

Introduction

In crowding, perception of a target strongly deteriorates when embedded in context (review: Herzog, Thunell, & Ögmen, 2016; Levi, 2008; Pelli & Tillman, 2008; Strasburger, 2020). Crowding is the standard situation in everyday vision because elements are rarely encountered in isolation. Crowding is stronger when the target and the flankers share similar features, such as same contrast polarity (Kooi, Toet, Tripathy, & Levi, 1994), color (Kennedy & Whitaker, 2010; Põder, 2007; van den Berg, Roerdink, & Cornelissen, 2007), orientation (Andriessen & Bouma, 1976; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; Wilkinson, Wilson, & Ellemberg, 1997), motion (Bex & Dakin, 2005; Gheri, Morgan, & Solomon, 2007), spatial frequency (Chung, Levi, & Legge, 2001; Põder & Wagemans, 2007), etc. It is often argued that only flankers within a certain spatial window (Bouma's window) around the target deteriorate performance (Bouma, 1970; Bouma, 1973; Levi, 2008; Strasburger, Harvey, & Rentschler, 1991; Weymouth, 1958). Crowding has specific characteristics. For example, flankers in the radial orientation interfere stronger than flankers in the tangential orientation (radial-tangential anisotropy; Chung, 2013; Greenwood, Szinte, Sayim, & Cavanagh, 2017; Kwon, Bao, Millin, & Tjan, 2014; Malania, Pawellek, Plank, & Greenlee, 2020; Toet & Levi, 1992), which was explained by elliptic receptive fields in early visual areas (Hubel, Wiesel, & Stryker, 1978; Silson, Reynolds, Kravitz, & Baker, 2018; Toet & Levi, 1992) or by an uneven sampling density in the early visual cortex (Kwon & Liu, 2019; Motter & Simoni, 2007).

Accordingly, crowding is traditionally explained by local, feature-specific interactions between the neural representations of the target and its direct neighbors. For example, neurons sharing the same orientation may interact with each other through lateral inhibition, feedforward pooling, etc. (e.g., Dakin, Cass, Greenwood, & Bex, 2010; Greenwood, Bex, &
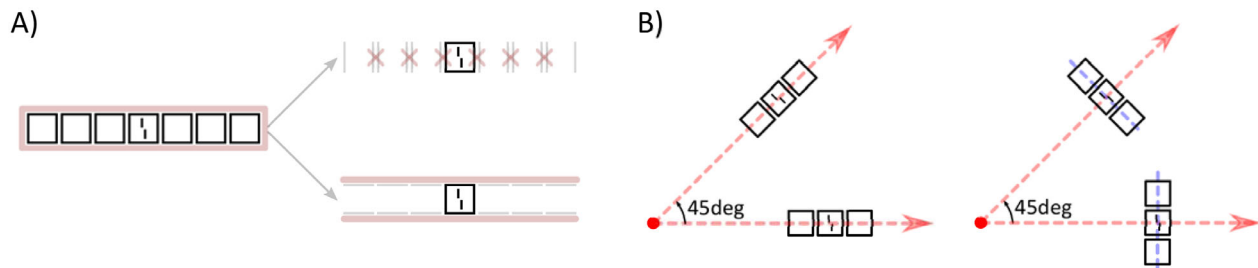
Figure 1. Experimental conditions to test low-level impacts on (un)crowding. (A) Experiment 1: Dissecting global configurations to iso-target (upper) and ortho-target (lower) flankers to test if low-level interactions can explain uncrowding. For example, line-line detector inhibitions (iso-target; upper) such as divisive normalization may suppress the center square (Carandini & Heeger, 2012; Coen-Cagli et al., 2015) so that the target uncrowds from the flanker. Alternatively, contour-contour interactions (ortho-target; lower) may create an illusory contour, which can group the flankers together and segment them out from the target (Clarke, Herzog, et al., 2014; Doerig et al., 2019; Francis et al., 2017). (B) Experiment 1 & 2: Radial (left)-tangential (right) anisotropic effects on uncrowding either in cardinal (0°) or oblique (45°) orientations. Here, red dots represent the fixation point, red dotted line represents the radial axis, and blue dotted line represents the tangential axis.

Dakin, 2009; Greenwood et al., 2017; Parkes et al., 2001; Pelli, 2008; Rosenholtz, Huang, Raj, Balas, & Ilie, 2012; Solomon, Felisberti, & Morgan, 2004). In all these models, target information is irretrievably lost at the early stages of visual processing. Thus this kind of crowding research is very much in the spirit of an atomistic view of visual processing, where basic, local processing precedes more complex processing.

However, all these explanations break down when the target is presented with complex, instead of simple flanker configurations (e.g., Livne & Sagi, 2007; Manassi, Lonchampt, Clarke, & Herzog, 2016; Põder, 2007; Saarela, Sayim, Westheimer, & Herzog, 2009; Sayim, Westheimer, & Herzog, 2010; Yeotikar, Khuu, Asper, & Suttle, 2011). For example, a vertical Vernier (target) is presented, and participants indicate whether the lower segment is offset either to the left or right compared to the upper one (Figure 1). Performance is good when the Vernier is presented alone but strongly deteriorates when surrounded by a square, a classic crowding effect. Traditionally, the deterioration may be explained by interactions between the vertical lines of the square and the Vernier. However, adding more squares does not further deteriorate performance. Instead, performance improves as more squares are added, approaching the performance level of the unflanked Vernier condition (uncrowding). Manassi, Sayim, and Herzog (2013) proposed that the Vernier target is released from crowding, because the additional flanking squares suppress the central square surrounding the Vernier. Uncrowding effects go well beyond Bouma's window and depend on the configuration of the entire stimulus, across more or less the entire visual field (Chicherov, Plomp, & Herzog, 2014; Chicherov & Herzog, 2015; Doerig, Bornet, Rosenholtz, Francis, Clarke, & Herzog, 2019; Herzog, Sayim, Chicherov, & Manassi, 2015; Herzog, Sayim, Chicherov, & Manassi, 2016; Herzog & Manassi, 2015; Malania, Herzog, & Westheimer, 2007; Manassi, Sayim, & Herzog, 2012; Manassi et al., 2013; Manassi, Hermens, Francis, & Herzog, 2015; Manassi et al., 2016; Saarela et al., 2009; Sayim, Westheimer, & Herzog, 2008; Sayim et al., 2010; Sayim et al., 2011).

Obviously, local approaches are of no avail. Contextual information across large parts of the visual field needs to be taken into account. Accordingly, models that go beyond spatially confined processing are needed. On the one hand, two-stage models propose that visual elements are first parsed in different groups, and then crowding occurs only within these groups. For example, grouping may arise from the integration of low-level features (*Laminart model*: Francis, Manassi, & Herzog, 2017) or from the competitions between different object-level representations of visual content (*Capsule networks*: Doerig, Bornet, Choung, & Herzog, 2020; Sabour, Frosst, & Hinton, 2017). On the other hand, Rosenholtz, Yu, and Keshvari (2019) suggested a one-stage pooling model, the texture tiling model (*TTM*), which may account for the complex effect of global configurations, despite its local nature. The main claim is that pooling a sufficiently large number of low-level image statistics (High Dimensional (HD) pooling) can preserve sufficient information about complex configurations, which can be used at the decision-making stage.

Doerig and colleagues (Doerig, Bornet, et al., 2020; Doerig et al., 2019; Doerig, Schmittwilken, et al., 2020) showed with extensive comparisons that grouping and segmentation processes are crucial for (un)crowding. In contrast, Rosenholtz and colleagues (2019) proposed that HD pooling can explain uncrowding (but see Bornet et al., same volume). However, in both approaches, it is unclear which aspects of the configuration may impact (un)crowding and to what extent low-level interactions within sub-parts can explain complex global processing.

Moreover, it is currently unclear to what extent low-level features and properties, such as the target orientation or the radial-tangential anisotropy, contribute to (un)crowding.

To study whether (un)crowding is truly a global phenomenon or instead can be explained by how sub-parts of the stimulus interfere, we systematically dissected the holistic configuration, as depicted in Figure 1A. We tested whether parts of the squares, such as their vertical lines, can explain uncrowding or if global processing of the good Gestalt of squareness is needed (Figure 1A, experiment 1). For example, line-line detector inhibition (Figure 1A upper) by divisive normalization may suppress the center square (Carandini & Heeger, 2012; Coen-Cagli, Kohn, & Schwartz, 2015). Alternatively, contour-contour interactions (Figure 1A lower) may create an illusory contour, which can group the flankers together and segment them out from the target (Clarke, Herzog, & Francis, 2014; Doerig et al., 2019; Francis et al., 2017).

As mentioned, crowding is stronger with flankers in the radial orientation than in the tangential orientation. Here, we tested whether there is such an anisotropy also in uncrowding. We presented arrays of squares in cardinal (experiment 1; Figure 1B, horizontal arrows) or oblique (experiment 2; Figure 1B, 45° arrows) orientation, and aligned the squares either along the radial (Figure 1B left) or tangential (Figure 1B right) direction. Also, we varied the target orientations (vertical, horizontal, or ±45°) to further assess low-level flanker-target interactions.

Then, we tested to what extent crowding by the flanking squares on the central square determines crowding by the central square on the Vernier (experiment 3). Does "crowding of crowding" lead to uncrowding? Finally, we tested which modeling approaches best suit these results by comparing models based on grouping and segmentation versus HD pooling.

Materials and methods

Participants

Thirty-eight participants took part in four experiments (Experiment 1: 11 [one participant was excluded from 12 initially recruited participants], Experiment 2: 10 [five excluded from 15], Experiment 3a: 7 [three excluded from 10], Experiment 3b: 10). Overall, nine participants were excluded right after the calibration session, because they did not show strong crowding in the one square condition, which is a prerequisite to test for a release of crowding to avoid the ceiling effect (see *Calibration session*). Hence, we retained the data of 38 participants from 47 initially recruited participants (mean age: 23 ± 3.7, 17 females, two left-handed, eight with left eye dominance by

the Miles test (1930)). All participants had normal or corrected to normal visual acuity in the Freiburg Visual Acuity Test, as indicated by a binocular score greater than 1.0 (Bach, 1996). Participants gave written consent before the experiment. All experiments were conducted following the Declaration of Helsinki except for the preregistration (World Medical Organization, 2013) and were approved by the local ethics committee (Commission d'éthique du canton de Vaud).

Apparatus

Stimuli were displayed on a gamma-calibrated 24-inch ASUS VG248QE LCD monitor (1920 × 1080 px, 120 Hz). The room was dimly illuminated (0.5 lux). The viewing distance was 75 cm, and the participant's chin and forehead were positioned on a chin-rest. Responses were collected using hand-held push buttons. In experiment 2, participants' eye movements were tracked with a The Eye Tribe eye tracker (60 Hz sampling frequency; The Eye Tribe, Copenhagen, Denmark), and stimuli were displayed only when participants adequately fixated.

Stimuli

Stimuli were white (100 cd/m^2), presented on a black background with luminance below 0.3 cd/m^2. Participants were asked to fixate on a red fixation dot (diameter of 8 arcmin, 20 cd/m^2). Stimuli were presented for 150 ms. When no response was registered within 3 seconds, the trial was repeated randomly within the same block. A feedback tone was given for incorrect responses (600 Hz) and omissions (300 Hz). Vernier stimuli were composed of two vertical/horizontal/45° clockwise or counter-clockwise tilted bars (depending on conditions; see below). Each bar was 40 arcmin long, 1.8 arcmin wide (anti-aliased), and separated by a 4 arcmin gap. Left/right offsets of vertical Verniers, up/down offsets of horizontal Verniers, or closer/further from the fixation dot offsets of 45° tilted Verniers, were balanced within a block. Flankers were combinations of squares and lines. In the Vernier discrimination tasks in experiments 1, 2, and 3a, the square and the distance between the squares were individually calibrated as described in *Procedures*. Before the calibration, squares and lines were composed of 120 arcmin long lines and were separated by 30 arcmin; thus the center-to-center distance between two flankers was 150 arcmin. For the aspect ratio discrimination tasks in experiments 3a and 3b, stimuli dimensions were identical for all participants; squares and lines were composed of 96 arcmin long lines and were separated by 24 arcmin.

Except for the 45° tilted conditions in experiment 2, each configuration was presented at the center of the screen, and the fixation dot was presented at an

eccentricity of 9° to the left. In the 45° conditions of experiment 2, each configuration was presented 2° to the right and up from the center. The fixation dot was $\frac{7}{\sqrt{2}}$° to the left and down from the target presentation position. Hence, the target eccentricity was 7°. Psychophysics Toolbox was used to present the stimuli (Brainard, 1997; Kleiner, Brainard, & Pelli, 2007; Pelli & Vision, 1997).

Procedures

General procedure

Different flanking configurations were tested in blocks of 100 trials. To reduce target-location uncertainty, only the target was presented alone for 150 ms at the beginning of each block. We used the PEST (Parameter Estimation by Sequential Testing) stair-case procedure (Taylor & Creelman, 1967). In PEST, test levels are changed step-wise based on the recent response history. The current test level is only changed when the hit rate for this test level lies, with some certainty, above or below the threshold criterion of 75%. The test levels are changed to make the hit rate converge to 75%, thereby boxing the threshold. After a fixed number of trials (100), we ended the procedure and took the threshold from the psychometric function that was fitted to the data post-hoc (details in *Data analysis*). We randomized the order of experimental conditions across participants. In experiments 1, 2, and 3a, participants went through a calibration session to adjust flanker size individually (see *calibration*).

Calibration session

Before the experimental conditions of experiments 1, 2, and 3a (not including Experiment 3b), 37 (Experiment 1: 12, Experiment 2: 15, Experiment 3a: 10) initially recruited participants went through a calibration session to avoid floor and ceiling effects. First, we familiarized participants with the peripheral Vernier task, where only a Vernier target was presented (160 trials, Vernier alone condition). If the Vernier offset threshold was smaller than 200 arcsec, the participant proceeded to the next condition. Otherwise (threshold larger than 200 arcsec), the same block was repeated to familiarize with the stimuli. Thirteen among 37 participants repeated the familiarization block. Second, up to 7 blocks with a Vernier surrounded by one square (80 trials/block) were tested to find the spatial parameters so that thresholds were at least six times larger than in the Vernier alone condition (mean threshold for the Vernier alone condition: 175.9 ± 11.2 arcsec, for the one square condition: 1099.0 ± 46.3 arcsec). For this, we reduced the square size gradually. We excluded participants whose thresholds were still below the criterion even after reducing the

square size to 70% of the original size (120 arcmin). In total, nine from the initial 37 participants were excluded right after the calibration session; thus the excluded participants did not continue the main experiment. The side length of the squares varied between 84 to 114 arcmin, depending on participants. Accordingly, the square-to-square distance for the experimental conditions with multiple squares varied between 21 to 28.5 arcmin.

Common (pooled) conditions

In experiments 1, 2, and 3a (not including Experiment 3b), seven flanker configurations were commonly used for the Vernier discrimination tasks. These flanker configurations were used to test how low-level features interact and their influence on (un)crowding (see Figure 2). Flanker configurations were arranged vertically or horizontally to test the impact of the radial-tangential anisotropy. In addition, the Vernier targets were in vertical or horizontal orientations to observe the interactions between flanker-target orientations. The configurations were as follows: Vernier alone, Vernier with one square, three vertically or horizontally aligned squares, seven vertically or horizontally aligned squares, and 35 (5 × 7) square grid configurations. For each configuration, the Vernier target was either vertically or horizontally oriented. Therefore, overall 14 conditions were tested (Figure 2). The data from the three experiments were pooled (no participant participated in more than one experiment).

Experiment 1

Eleven participants completed the experiment. We tested the seven aforementioned common configurations and the partial square configurations to investigate possible low-level interactions in uncrowding. As shown in Figure 3, the partial square configurations had the same number of flanker elements as the common configurations, but only the vertical bars of the squares or only the horizontal bars of the squares. Vernier targets were either vertical or horizontal. Participants were asked to report the Vernier offset direction. For the vertical Vernier, the task was to report whether the lower bar was offset to the left (left button) or right (right button) compared to the upper bar. For the horizontal Vernier, the task was to report whether the right bar was on the top (left button) or bottom (right button) compared to the left bar.

Experiment 2

Ten participants completed the experiment. To test whether uncrowding is universal despite the oblique orientations, in addition to the 14 common conditions (7 flanker configurations × 2 Vernier target
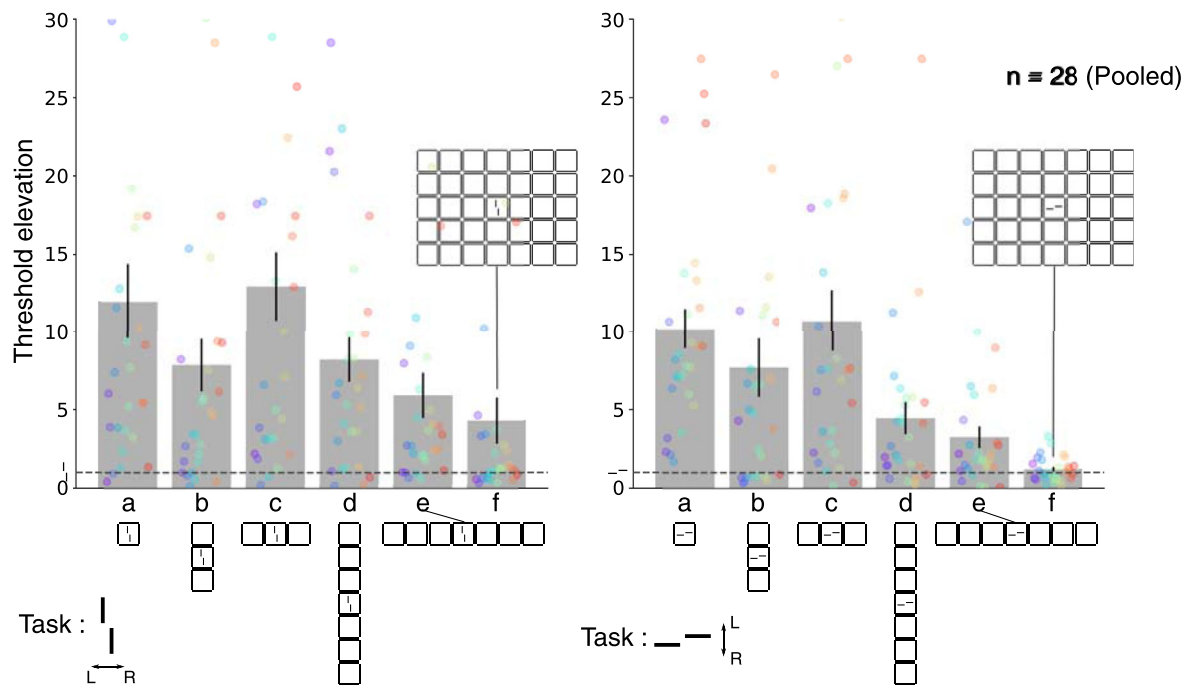
Figure 2. Pooled conditions. The y-axis shows mean threshold elevation ($\pm$ SEM) relative to the unflanked (Vernier alone) condition (gray dotted lines equal to 1). Larger thresholds represent poor performance (strong crowding), and smaller thresholds represent good performance (weak crowding). Also, performance improves the more squares are presented, independently of the flanker and the Vernier orientation; vertical Vernier left and horizontal Vernier right panel. Colored dots show individual data points.

orientations), we tested configurations with the flanker configuration tilted by 45° and the Vernier tilted by either ±45° (stimuli details in Figure 4). For the 45° counterclockwise rotated conditions, the task was to report whether the Vernier bar further away from the fixation dot (outer bar) was offset to the left or right compared to the bar closer to the fixation dot (inner bar). For the 45° clockwise rotated conditions, the task was to report whether the inner bar was offset to the top (left) or bottom (right) compared to the outer bar. Each trial started only if the participants kept their eyes fixated on the fixation dot for 150 ms.

Experiment 3

In experiment 3a, seven participants completed the experiment. To test whether uncrowding can be explained by crowding of the flanker squares on the center square, a square aspect ratio discrimination task was tested with the seven common configurations, in addition to the Vernier offset discrimination task. For this task, participants were asked to discriminate whether the width or the height of the central square was longer (hence, strictly speaking, the central square was a rectangle). For vertically aligned squares conditions (Figures 5a, 5c, 5e), the height was adjusted and the width for horizontally aligned squares conditions (Figures 5b, 5d, 5f, 5g).

In experiment 3b, 10 participants were tested in the aspect ratio discrimination task as in experiment 3a, but with 2 additional configurations (five vertically or horizontally aligned squares) and with/without the Vernier presentation in the center square. To avoid overlap between the Vernier and the target square, we reduced the Vernier length. Vernier bars were 20 arcmin long, separated by a gap of 2 arcmin. The various conditions are shown in Figure 5.

Data analysis

We fitted a cumulative Gaussian function (psychometric function) to the data (tested levels and hit rates) and determined the Vernier offset or the square size (Experiment 3) for which 75% correct responses were reached (threshold). Psignifit 3 python toolbox (Fründ, Haenel, & Wichmann, 2011) was used for the fitting. High thresholds indicate inferior performance, and low thresholds indicate good performance. Next, we divided the threshold in each condition by the threshold in the Vernier alone condition (threshold elevation). Data were log-transformed to bring the data closer to normality. No obvious violation was detected by visual inspection.

Using R (R Core Team, 2019) and *lme4* package (Bates, Machler, Bolker, & Walker, 2015), we computed
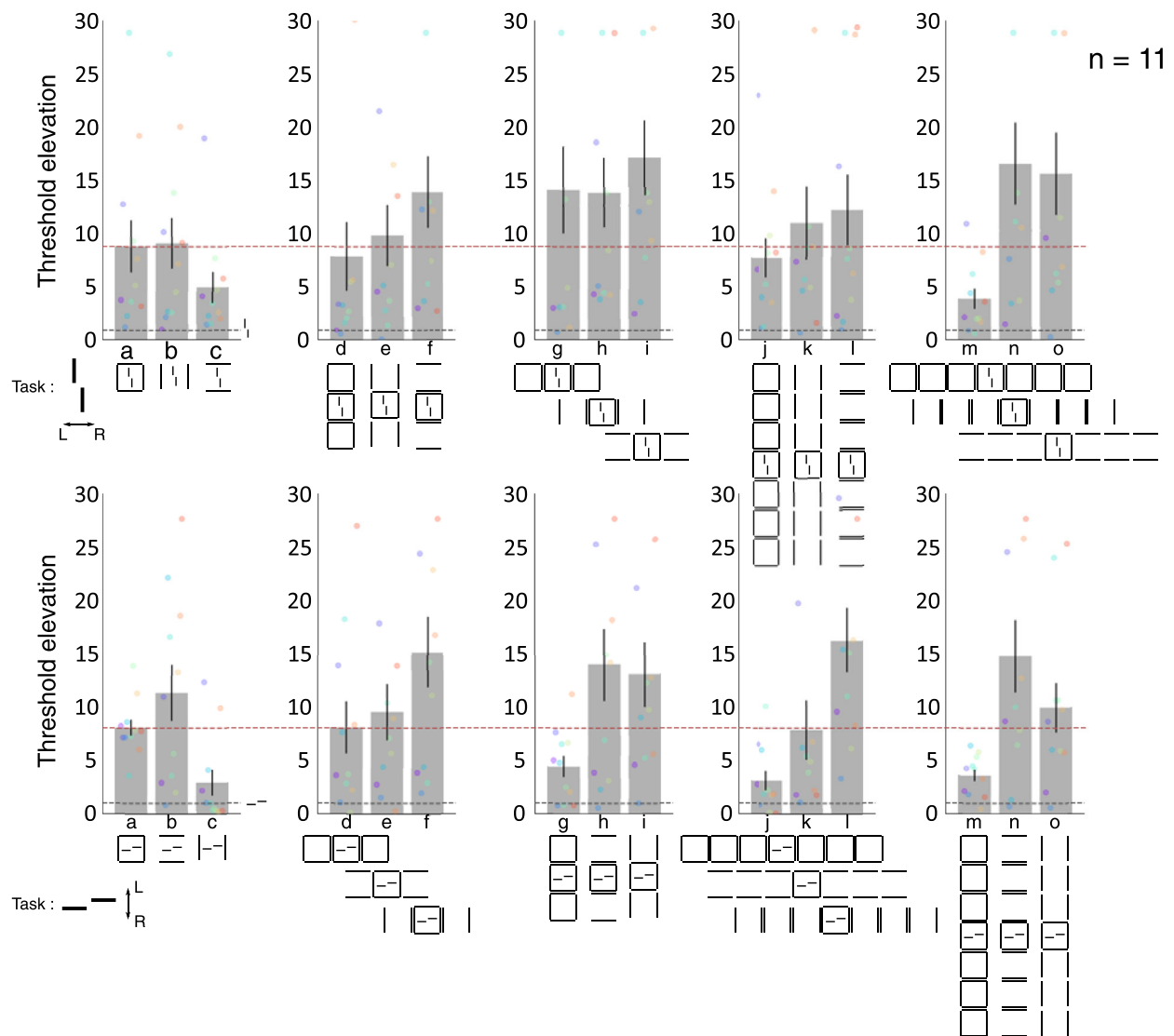
Figure 3. Experiment 1. Systematic dissection of flanker configurations with a vertical (top) or horizontal (bottom) Vernier target. The y-axis shows threshold elevation relative to the unflanked (Vernier alone) condition. In the 1-flanker conditions (a, b, & c: crowding conditions), iso-target flankers lead to the same performance deterioration as the complete square (b vs. a). In the three- and seven-flankers conditions, complete squares (d, g, j, & m) lead to better performance than the iso-target flankers (e, h, k, & n) or ortho-target flankers (f, i, l, & o). Bars and error bars represent Mean ± SEM, colored dots represent individual data points. Red dotted lines show the performance of the 1 square condition.

linear mixed-effects models (LMM) to account for dependent variables and random variations because of individual differences. The fixed and random effects are specified for each experiment (see *Results* for specifications of each experiment). The model significance (*p* value) was obtained through likelihood ratio tests (chi square) χ^2 by comparing nested models. For each fitted model, using *MuMIn* package (Barton, 2020), we computed the effect size (r^2), that is, the explained variance, when including (conditional r_c^2) and excluding (marginal r_m^2) the random effects (Johnson, 2014; Nakagawa et al., 2017; Nakagawa & Schielzeth, 2013). Post-hoc multiple comparisons

(Tukey's HSD test) of means were computed with the *multcomp* package (Hothorn, Bretz, & Westfall, 2008).

Model comparisons

We simulated the conditions of experiment 1 (Figure 3) with Capsule Networks (Doerig et al., 2020, https://github.com/adriendoerig/Capsule-networks-as-recurrent-models-of-grouping-and-segmentation), the Laminart model (Doerig, Bornet, et al., 2019, https://bitbucket.org/albornet/laminart/) and the texture tiling model (TTM;
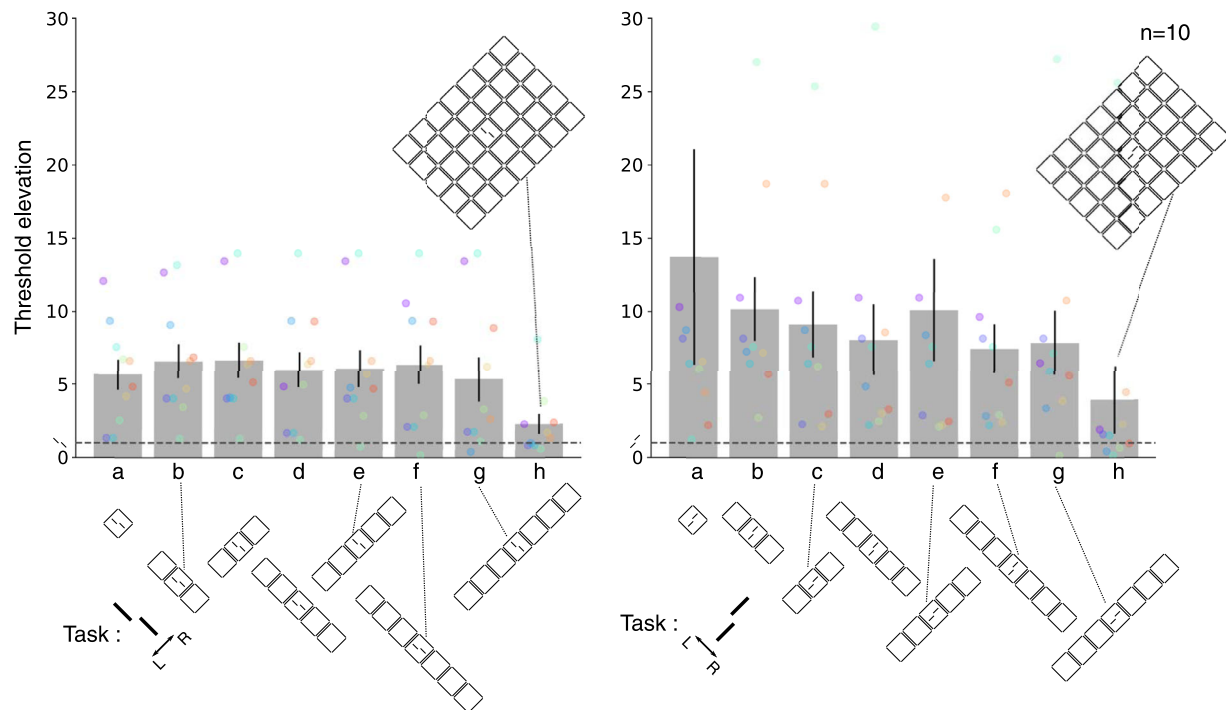
Figure 4. Experiment 2. The left panel shows the −45° rotated Vernier conditions (tangential direction), and right the +45° rotated Vernier conditions (radial direction). The y-axis shows threshold elevation relative to the unflanked (Vernier alone) condition. Performance was poor in most conditions (a–g), regardless of the radial (c, e, g) or tangential (b, d, f) alignments, except with the 35 squares grid (h). Bars and error bars represent Mean ± SEM, colored dots represent individual data points.

Rosenholtz et al., 2019, https://dspace.mit.edu/handle/1721.1/121152). Capsule networks were trained to recognize Verniers, groups of squares, groups of horizontal bars, and groups of vertical bars presented in isolation (i.e., there were only flankers or the Vernier). After training, the Capsule Network was tested on the different crowding conditions. The model performance was obtained by the percentage of error as in Doerig et al. (2020). Performance of the Laminart model was obtained as in Francis et al. (2017). Performance of TTM was obtained by using an algorithm that matches left and right Vernier templates to mongrels generated by the model (see Bornet et al., Same volume). Model-specific parameters, conditions, and algorithms are available online (https://github.com/Ohyeon5/dissecting_uncrowding).

Results

Pooled conditions. Crowding decreases with the number of squares in vertical and horizontal orientations

In experiments 1, 2, and 3a, we tested the same seven conditions (Vernier alone, 1-square, 3-squares,

7-squares either vertically or horizontally aligned, and 35-squares grid). We pooled the data of 28 participants for these conditions.

We mainly replicated previous findings (Manassi et al., 2013; Manassi et al., 2016). When the vertical Vernier was surrounded by a single square, thresholds strongly increased as aimed for. Contrary to previous findings (Manassi et al., 2013; Manassi et al., 2016), however, adding only a square both on the left and right of the central square in the horizontally aligned condition did not substantially improve performance (Figure 2 left.c). Although adding three squares on the left and right (seven squares, horizontally aligned condition) led to a strong decrease of crowding (Figure 2 left.e), which is in line with previous findings. When squares were vertically aligned (three and seven squares conditions, Figure 2 left.b & 1 left.d), we also found a decrease of crowding. The same pattern of results holds true when the target Vernier was horizontal (Figure 2 right). Performance was best with the 35-squares grid (Figure 2f left & right).

To analyze the relation between threshold elevation and configuration, we computed an LMM with the number of squares in the vertical or horizontal dimension as fixed effects. For example, the horizontally aligned three-squares condition was coded as having three squares in the horizontal dimension and one in the vertical dimension, respectively. Individual participants
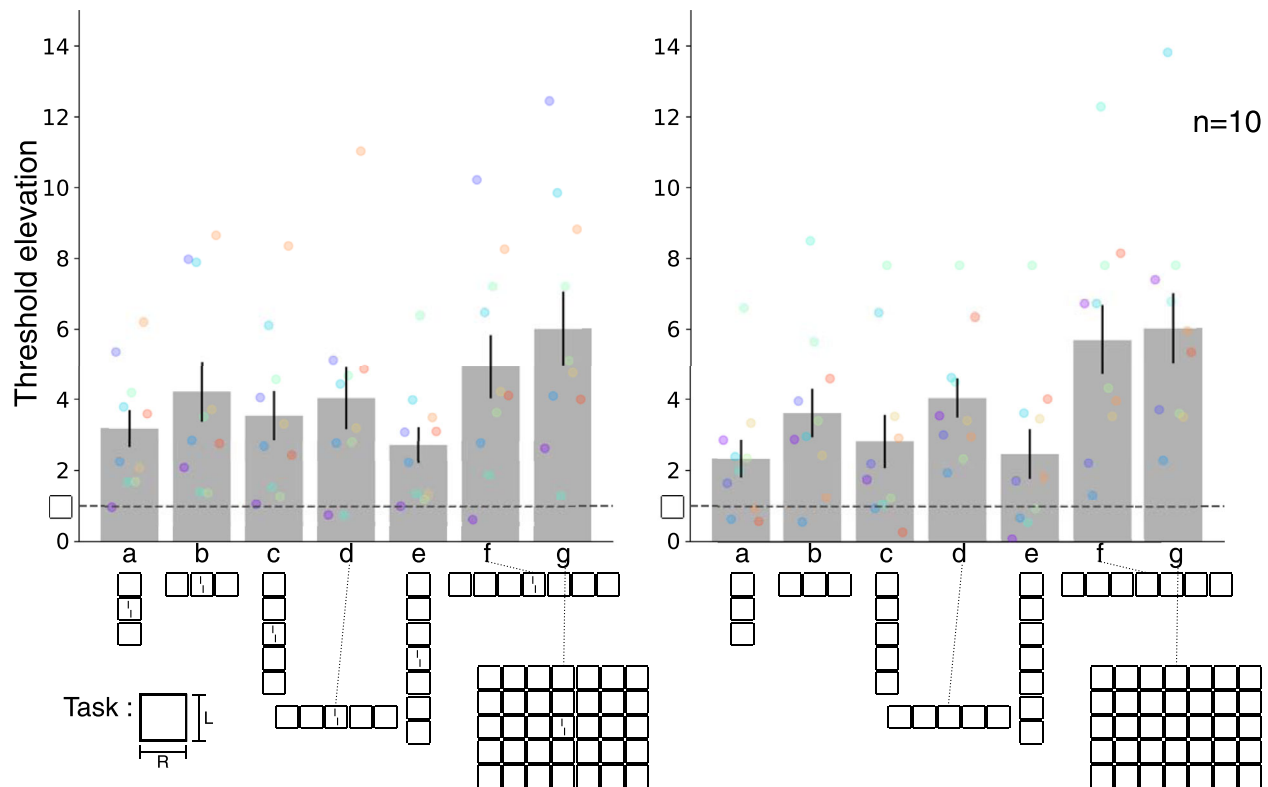
Figure 5. Experiment 3b. The center square aspect ratio discrimination task with (left) and without (right) Vernier presentation. Performance deteriorated (the target was more crowded) as the number of squares increased in the horizontal dimension, independent of whether or not the Vernier was presented. The y-axis shows threshold elevation relative to the one square condition. Mean ± SEM, colored dots represent individual data points. Note the change of y-axis scaling.

and target orientations were considered as random intercepts. We found no significant interaction between the two fixed effects (likelihood ratio test between an additive and an interaction model: $\chi^2(1) = 0.774$, $p = 0.379$). Both fixed effects showed significant differences (horizontal: $\chi^2(1) = 60.980$, $p < 0.001$; vertical: $\chi^2(1) = 36.985$, $p < 0.001$). The negative parameter estimates in both dimensions (in Supplementary Table S1) show that thresholds significantly decreased when the number of squares increased, which means performance improved. The model explains 38.9% of the variance and only 17.7% when not accounting for the random effects ($r_m^2 = 0.177$, $r_c^2 = 0.389$). In addition, we only found a marginal significance of target orientation as a random intercept ($\chi^2(1) = 3.853$, $p = 0.049$). The difference of explained variance by the models with and without the target orientation as a random intercept is only 1.7% ($r_c^2 = 0.389$, $r_c^2 = 0.372$). It seems that qualitative results are similar for both target orientations.

Next, we ran an LMM with only the three and seven flankers configurations to see the possible differences between horizontally and vertically aligned flankers.

We included one more fixed effect, namely, flanker orientation (vertical or horizontal); thus the LMM had three fixed effects and two random intercepts. Interestingly, the fixed effect of the flanker orientations and the number of squares in the horizontal dimension was significant, but not the number of squares in the vertical dimension (flanker orientations: $\chi^2(1) = 9.775$, $p < 0.01$, horizontal: $\chi^2(1) = 21.039$, $p < 0.001$, vertical: $\chi^2(1) = 0.152$, $p = 0.697$). This result indicates that increasing the number of squares in the horizontal dimension improves performance gradually. In contrast, the performance improvement by increasing the number of squares in the vertical dimension was not gradual. Also, post-hoc Tukey's HSD comparison showed that performance improvement with horizontally oriented flankers was significantly better than the vertically oriented ones ($z = 3.166$, $p < 0.01$). Interactions among fixed effects were not tested because of the dependency. The detailed estimates are reported in Supplementary Table S2. In summary, crowding decreases with the number of squares roughly independent of target and flanker array orientations, despite minor differences.

Experiment 1. Uncrowding cannot be explained by the addition of parts

As shown, regardless of the stimuli orientation, the more flanking squares there are, the better performance is (uncrowding). A major question is to what extent uncrowding depends on low-level interactions (Figure 1A), such as contour-contour interactions, line-line detector inhibitions, or rather on holistic aspects, such as the good Gestalt of squareness. In other words, can low-level interactions release the strong crowding by the single square around the Vernier target? Here, we systematically dissected the configurations of Figure 2 in three different ways (Figure 3): complete square, only the vertical bars of the squares, only the horizontal bars of the squares. Flankers were either horizontally or vertically oriented, and the number of flankers was one, three, to seven flankers. Overall, there were 15 flanker configurations. Note that the central square was always a complete square in the conditions with multiple flankers. The Vernier target was either vertical or horizontal. We call the partial squares whose lines have the same orientation as the target iso-target-flankers and those whose lines are orthogonal ortho-target-flankers.

In the one flanker conditions, performance was worst when the Vernier was flanked by lines of the same orientation, e.g., vertical lines for the vertical Vernier. The LMM was computed with flanker configurations (complete square, iso-, or ortho- target flanker) as a fixed effect and individual participants and target orientations as random intercepts. The fixed effect was significant when compared with an intercept only model ($\chi^2(2) = 19.470$, $p < 0.001$). Post-hoc Tukey's HSD comparisons indicated that performance with ortho-target-flankers was significantly better than with iso-target flankers (Figure 3 top & bottom. b vs. c; $z = 4.135$, $p < 0.001$) and the complete square (Figure 3 top & bottom.a vs. c; $z = 4.235$, $p < 0.001$). We found no significant difference between iso-target flankers and the complete squares conditions (Figure 3 top & bottom a vs. b; $z = 0.099$, $p < 0.995$). The model explains 34.7% of the variance and 23.5% when not accounting for the random effects ($r_m^2 = 0.235$, $r_c^2 = 0.347$). The detailed parameter estimates are shown in Supplementary Table S3.

In the three and seven flankers conditions, the complete square conditions always showed better performance independently of the number of flankers or flanker orientations. The LMM with a fixed effect of flanker configuration and random intercepts of individual participants and target orientation showed a significant fixed effect (Likelihood ratio test compared with the intercepts only model; $\chi^2(2) = 49.696$, $p < 0.001$). Post-hoc Tukey's HSD tests showed a significantly better performance with the complete squares than the other two partial square configurations (complete squares vs. iso -target flankers, $z = 5.060$, $p < 0.001$; complete squares vs. ortho-target flankers, $z = 7.220$, $p < 0.001$). Although there appears to be a trend of the flankers to crowd more in the ortho-target flanker conditions than in the iso-target flanker conditions, evidence is not strong enough to make firm claims (Figure 3; iso- vs. ortho-target flankers, $z = 2.160$, $p = 0.078$). In addition, even if the effect were significant, the effect size is much smaller than the effect size of crowding vs. uncrowding. This shows that even though iso-target flankers may have a minor influence, it is not the main driving force. The model explains 45.1% of the variance, but only 11.5% when not accounting for the random effects ($r_m^2 = 0.115$, $r_c^2 = 0.451$). The detailed parameter estimates are shown in Supplementary Table S4.

Experiment 2. Oblique orientations

In the pooled conditions, we found no clear differences between vertical and horizontally arranged arrays of squares. Uncrowding seems not to reveal a radial-tangential anisotropy in cardinal orientation, further indicating that low-level aspects, such as the shape of receptive fields in early visual areas, are less important than the overall shape of the configuration. Then what about when a stimulus is presented in oblique orientation (Figure 1B, 45° arrows)? It is well known that stimuli in cardinal orientations lead to significantly better performance than oblique ones in many visual paradigms (Li, Peterson, & Freeman, 2003; Mach, 1860; Westheimer, 2005) because more neurons are tuned to the cardinal axes (Bauer, Owens, Thomas, & Held, 1979; Furmanski & Engel, 2000; Xu, Collins, Khaytin, Kaas, & Casagrande, 2006) or there is an uneven sampling density in the early visual cortex (cortical magnification; Kwon & Liu, 2019; Motter & Simoni, 2007). Oblique orientations may lead to different (un)crowding. There can be three scenarios: (1) a crowding anisotropy between radially (+45°) and tangentially (−45°) arranged array of squares, unlike for cardinal orientation, (2) a similar behavioral pattern as in cardinally oriented stimuli, but with mere performance deterioration, that is, performance improves (uncrowding) as the number of squares increases in either + or −45° direction but not as much as in the cardinal orientation, (3) a completely different behavioral pattern. Here, we tested performance for +45° rotated Verniers in either tangential or radial direction.

Vernier discrimination of the unflanked target was substantially harder in oblique orientations than in the vertical and horizontal orientations (cardinal). Hence, we computed an LMM with stimulus orientations

(cardinal or oblique) as the fixed effect and individual participants and the target orientations (vertical, horizontal, −45°, or +45°) as random intercepts. The fixed effect was significant (likelihood ratio test with the intercept only model; $\chi^2(1) = 9.251, p < 0.01$). Tukey's HSD post-hoc test shows that the performance of the Vernier alone condition with oblique orientations was significantly worse than with cardinal orientations (cardinal vs. oblique; $z = 4.753, p < 0.001$). Detailed estimates are presented in Supplementary Table S5.

Contrary to the vertical or horizontal orientations, we did not find a gradual performance improvement as the number of squares increased in one of two orientations. However, there was strong uncrowding in the 35 squares grid configuration, independent of stimulus orientation. An LMM with the number of squares and individual participants as random intercept showed a significant difference for the number of squares (likelihood ratio test with the intercept only model; $\chi^2(1) = 32.148, p < 0.001$). The model explains 26.5% of the variance and 16.6% when not accounting for the random effects ($r_m^2 = 0.166, r_c^2 = 0.265$). The detailed parameter estimates are presented in Supplementary Table S6.

Indeed, it seems that Vernier discrimination is substantially harder in oblique than in cardinal orientations (oblique effect). Also, there is no obvious uncrowding for arrays oriented along the 45° axis. However, for the 35-square grid, neither the oblique orientation of the grid as such nor of the single squares seems to matter. There is clear-cut and strong uncrowding.

Experiment 3. Is uncrowding "crowding of crowding"?

The above experiments showed that uncrowding depends on holistic aspects rather than low-level interactions, regardless of the orientations. It has been suggested crowding is reduced when flankers are suppressed by themselves (Manassi et al., 2013), by flanker awareness (Wallis & Bex, 2011), or by masking (Chakravarthi & Cavanagh, 2009). Especially, Manassi and colleagues (2013) showed that uncrowding of the Vernier is a consequence of mutual crowding of the squares: Vernier crowding is weak when the central square is crowded by other squares and strong when the square is weakly or not at all crowded. So then, can crowding of crowding fully explain uncrowding?

Consistent with Manassi and colleagues (2013), we found that the aspect ratio of the center square is harder to discriminate as the number of flanking squares increases. Thus the center square was highly crowded by the additional flankers. In addition, we found a crowding anisotropy between horizontally

versus vertically aligned squares. The central square was strongly crowded by adding more squares in the horizontal dimension (radial) but not in the vertical dimension (tangential). The LMM was computed with the number of squares in horizontal or vertical dimensions as fixed effects. We coded each flanker as we did in the pooled conditions, that is, three-squares vertically aligned condition as three squares in the vertical dimension and one in the horizontal dimension, respectively. Individual participants and Vernier presentation (experiment 3b only) were considered as random intercepts. LMMs were applied to experiments 3a and 3b separately.

In experiment 3a (Supplementary Figure S1), the two fixed effects had no significant interaction ($\chi^2(1) = 1.159, p = 0.282$). The number of squares in the horizontal dimension had a significant effect ($\chi^2(1) = 13.319, p < 0.001$), whereas the effect in the vertical dimension was not significant ($\chi^2(1) = 3.452, p = 0.063$). In addition, the explained variance difference between the full model with both fixed effects and the nested model without the effect of the vertical dimension was small, only 3.8% (full model: $r_m^2 = 0.181, r_c^2 = 0.649$, reduced model: $r_m^2 = 0.143, r_c^2 = 0.602$). Therefore the number of squares in the vertical dimension may not be a good predictor of the crowding level, whereas the number of squares in the horizontal dimension is a good one. In other words, the number of squares in the horizontal (radial) dimension impacts crowding more than those in the vertical (tangential) dimension, which can be related to the radial-tangential anisotropy of the crowding. The detailed parameter estimates are presented in Supplementary Table S7.

The same results hold for experiment 3b (Figure 5). Two fixed effects showed no significant interaction ($\chi^2(1) = 12.682, p = 0.102$). The number of squares in the horizontal dimension was significant ($\chi^2(1) = 27.387, p < 0.001$), but not for the vertical orientation ($\chi^2(1) = 0.116, p = 0.733$). Again, the explained variance difference was tiny. The difference between the full model, including both effects and the reduced model excluding the effect in the vertical dimension, was only 0.05% (full model: $r_m^2 = 0.163, r_c^2 = 0.423$, reduced model: $r_m^2 = 0.163, r_c^2 = 0.423$). The detailed parameter estimates are shown in Supplementary Table S8.

To explicitly test that crowding was stronger for horizontally aligned squares than for vertically aligned squares, we computed another LMM. Here, we used the 3, 5, or 7 square conditions only and considered the number of squares and the flanker alignment orientation as fixed effects and the same random intercepts as the tested model. The two fixed effects showed no significant interactions ($\chi^2(1) = 3.264\ p = 0.071$). The flanker alignment orientation showed a significant effect ($\chi^2(1) = 17.848, p < 0.001$). The number of squares had no significant effect ($\chi^2(1) = $

0.122, $p = 0.727$). Although the interaction model showed no significant effect, there was a trend for an interaction, that is, crowding increased with the number of squares in the horizontal but not clear in the vertical orientation. However, the interaction is minor compared to the effect of the flanker alignment orientation. This minor interaction may be a reason why the fixed effect of the number of squares did not show significance. Post-hoc Tukey's HSD tests showed a significantly stronger crowding for the horizontally aligned squares than for the vertically aligned squares (horizontal vs. vertical; $z = 4.404$, $p < 0.001$). The detailed parameter estimates are shown in Supplementary Table S9. The results are consistent with the well-known crowding radial-tangential anisotropy (Toet & Levi, 1992).

In addition, the presentation of a Vernier in the central square does not affect performance (experiment 3b, Figure 5 left vs. right), i.e., crowding was not due to target location uncertainty. We used the LMM with two fixed effects, namely, the number of squares in each dimension and random intercepts for individual participants and Vernier presentation (the same LMM as applied to experiment 3b, Supplementary Table S8). The likelihood ratio test showed no significant difference between the full model and the model excluding Vernier presentation ($\chi^2(1) = 0.167$, $p = 0.683$). Also, the explained variance difference between both models was little, only 0.5% (full model: $r_m^2 = 0.163$, $r_c^2 = 0.423$, reduced model: $r_m^2 = 0.164$, $r_c^2 = 0.418$).

In summary, flankers aligned in the horizontal (radial) dimension crowd stronger than in the vertical (tangential) dimension. However, such an anisotropy was not reflected in the Vernier discrimination task; that is, the Vernier performance was not better in horizontally aligned squares than in vertically aligned squares (Figure 2, further discussion in *Discussion*).

Models. Model comparison suggests that object-based grouping is needed to explain uncrowding

In the above experiments, we showed that uncrowding happens regardless of orientation and depends on holistic, rather than local, aspects of the stimulus. Here, we tested three models, which take global aspects into account but are based on different premises. Capsule networks and the Laminart model are two-stage models, in which elements are first parsed into different groups, and then interference occurs only within the groups. Capsule networks group elements on the basis of object-level routing by agreement (for details, see Doerig et al., 2020; Sabour et al., 2017), whereas the Laminart model groups elements on the basis of

low-level features (for details, see Francis et al., 2017; Bornet et al., 2019). The TTM model is a one-stage model that pools many low-level features computed over pooling regions whose size grows with eccentricity (for details, see Rosenholtz et al., 2019). We tested the vertical Vernier target conditions of experiment 1. Here, we only show results obtained with the horizontally aligned flanker conditions (Figure 6). The model results for the vertically aligned flanker conditions are comparable (Supplementary Figure S4).

Capsule Network reproduced the general human behavior pattern well, that is, performance improved when adding more squares (Figures 6Ad, 6Ag; red bars) and deteriorated when adding either the iso- or ortho-target flankers (Figures 6Ae, 6Af, 6Ah, 6Ai; gray bars). Note that there were still minor performance differences, for example, human performance for only vertical lines (Figure 6Eb) was equally bad as in the one-square condition, but the model performance was much better (Figure 6Db vs. Figure 6Ab). The Laminart model partially reproduced the human behavior, that is, performance improved when adding more squares (Figures 6Bd, 6Bg; red bars) and deteriorated when adding the iso-target flankers (Figures 6Be, 6Bh). However, unlike humans, model performance improved when adding ortho-target flankers (Figures 6Bf, 6Bi). In both models, the performance of the complete square conditions could not be explained by simply adding the performances of the iso- and ortho-target flankers conditions, that is, the performance of Figure 6Ad was smaller than Figure 6Ae or Figure 6Af.

The TTM (1-stage model) could not reproduce the human behavior, that is, adding more squares deteriorated performance, and performances with iso- or ortho-target flankers were better than in the complete squares conditions (Figure 6C). Moreover, performance in the complete squares conditions could more or less be explained by adding the performance levels of the corresponding iso- and ortho-target flankers conditions (i.e., the performance of Figure 6Cd was roughly equal to Figure 6Ce plus Figure 6Cf).

In addition, we trained three control networks using the exact same training procedure as we used for the Capsule networks: a feedforward CNN and two recurrent CNNs. These networks had the same number of layers and neurons, and the only differences were in the connectivity between the neurons in the last two layers. This allowed us to control (1) whether the training regime is sufficient to explain the experimental results, even without recurrent grouping and segmentation, and (2) whether any kind of recurrence is sufficient vs. whether specific grouping and segmentation processing of Capsule networks is needed. The results clearly show that these control networks do not reproduce our results, supporting our claim that grouping and segmentation processes are needed (Supplementary Figure S2) for uncrowding
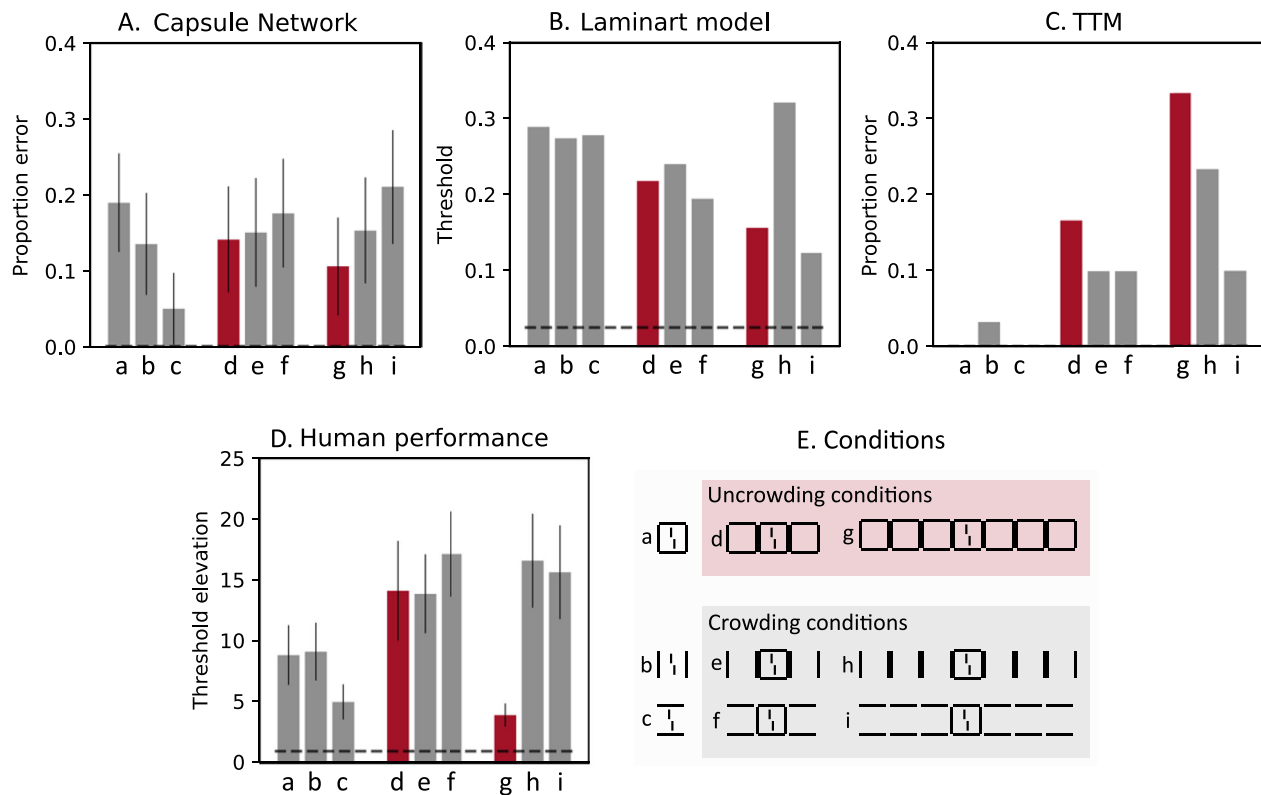
Figure 6. Model performance: percent error for Capsule networks and TTM, Vernier offset thresholds for the Laminart model. For both measures, larger values indicate worse performances. Red bars represent conditions leading to uncrowding in humans (good performance), and gray bars represent crowding (poor performance). Gray dashed lines show the model performance for the Vernier only condition. (A) Performance of Capsule Networks. We averaged the proportion of errors from 10 separately trained networks (mean ± SEM). (B) Performance of the Laminart model. We used an inference mechanism as described in Francis et al. (2017), and averaged the results over 20 runs per condition. (C) Performance of the TTM. We created 15 mongrels per condition and per offset direction (in total, 30 mongrels per condition) and determined the proportion of errors using a template matching algorithm. (D) Human performance reordered from Figure 3. (E) Conditions tested. Vertically aligned flanker conditions were also tested and presented in Supplementary Figure S4.

(Figure 3). Moreover, the Laminart model and TTM were tested with different parameters. Changes in the model parameters did not lead to obvious differences (Supplementary Figure S3). Hence, in summary, our results favor the two-stage models over the one-stage model.

Discussion

Crowding is at the heart of vision research as elements are rarely encountered in isolation. However, even after a century of research (e.g., Korte, 1923; Ehlers, 1936; Flom, Heath, & Takahashi, 1963; Bouma, 1970), the mechanisms underlying crowding are still largely unknown and controversially discussed. Classically, crowding was explained by local models, where only *neighboring* elements with *similar* features interact with each other, for example, via lateral

inhibition (Carandini & Heeger, 2012). Alternatively, the outputs of the neurons may be pooled (Dakin et al., 2010; Greenwood et al., 2009; Greenwood et al., 2017; Parkes et al., 2001; Pelli, 2008; Rosenholtz et al., 2012; Rosenholtz et al., 2019), features may be substituted (Huckauf & Heller, 2002; Strasburger, 2005; Strasburger et al., 1991), or crowding may be mediated by top-down processes (He, Cavanagh, & Intriligator, 1996; Montaser-Kouhsari & Rajimehr, 2005; Tripathy & Cavanagh, 2002; Yeshurun & Rashal, 2010). Such models were motivated by experiments showing that, for example, crowding strongly decreases when target and flanker have different contrast polarity (Kooi et al., 1994), color (Kennedy & Whitaker, 2010; van den Berg et al., 2007), motion (Bex & Dakin, 2005), and more. Likewise, crowding decreases when flankers are moved away from the target, which is often described by Bouma's law stating that flankers interfere only within a window of half the eccentricity around the target (Bouma, 1973).

However, all these explanations fall apart when more flankers are presented. Flankers outside Bouma's window can suppress crowding up to the performance level of the unflanked target (Figure 2f, grid condition). Similar effects have been shown previously with various stimuli such as Verniers (Manassi et al., 2012; Manassi et al., 2013; Manassi et al., 2015; Manassi et al., 2016; Sayim et al., 2010), Gabors (Levi & Carney, 2009; Livne & Sagi, 2007; Maus, Fischer, & Whitney, 2011; Saarela et al., 2009; Saarela & Herzog, 2008), shapes (Kimchi & Pirkner, 2015), letters (Reuther & Chakravarthi, 2014; Saarela et al., 2010), textures (Herrera-Esposito, Coen-Cagli, & Gomez-Sena, 2020), as well as in haptics (Overvliet & Sayim, 2016) and audition (Oberfeld & Stahn, 2012). Feature similarity is important but not decisive because strong crowding can also occur with flankers of different contrast polarity and color (Manassi et al., 2012; Sayim et al., 2008). What matters is the configuration (Livne & Sagi, 2007) of potentially all elements across large parts of the visual field (Herzog et al., 2016; Herzog & Manassi, 2015). For this reason, simple local (pooling) models have been largely but not fully abandoned. For example, Greenwood and colleagues (2020) take configuration effects in crowding as "modulations" of a local pooling mechanism. However, we think that strong effects such as uncrowding, going from good performance for a single target to strong crowding with a single square (threshold elevation of 12, Figure 2) to uncrowding with the 35 squares (threshold elevation of 4, Figure 2), are beyond what can be called a modulation.

Here, we have tested to what extent the overall configuration plays a role in crowding by dissecting uncrowding configurations systematically. First, we reproduced previous findings of uncrowding with an increasing number of elements. Performances in the 35 square grid condition were about at the same level as in the Vernier alone condition (Figure 2f). Importantly, despite minor differences, uncrowding occurs for both the horizontal and vertical arranged flankers, also for horizontal and vertical Vernier targets (Figures 2 and 3). Note that, unlike previous work (Manassi et al., 2013; Manassi et al., 2016), the three horizontally aligned squares did not show a clear performance improvement compared to the single square condition. We do not have an explanation for this beyond noises. The oblique configuration showed a different behavioral pattern (Figure 4). For example, increasing the number of squares in one of the ±45° orientations did not improve performance (Figure 4, conditions a vs. c, e, and g, or conditions a vs. b, d, and f). Surprisingly, when the entire 35-square grid was presented, the performance was as good as for the cardinal orientations (Figure 4h left & right, 35-square grid condition). Livne and Sagi (2011) showed that obliquely oriented and positioned flankers crowd stronger than cardinally oriented flankers. In addition, the obliquely presented stimuli

made various visual tasks significantly harder, including orientation discrimination (Bouma & Andriessen, 1968), orientation discrimination under crowding (Livne & Sagi, 2011), Vernier discrimination (Saarinen & Levi, 1995; Westheimer, 2005), motion discrimination (Ball & Sekuler, 1982; Coletta, Segu, & Tiana, 1993), orientation detection (Attneave & Olson, 1967), and more, likely because of neuronal preferences of cardinal orientations in low-level visual areas (Bauer et al., 1979; Furmanski & Engel, 2000; Li et al., 2003; Xu et al., 2006). Hence, similarly, we expected performance would deteriorate with the oblique flankers while keeping crowding characteristics similar to the cardinal flankers. However, the results were not as expected. There was no performance improvement (uncrowding) when increasing the number of squares in either ±45 orientations but only with the 35-square grid (Figure 4). This may be because the grouping cue was too weak with three, five, or seven squares for one of the ±45° orientations. Nevertheless, when a stronger grouping cue was provided by the grid of 35 squares, performance was good (uncrowding) regardless of orientation, approaching the performance in Vernier alone conditions (and comparable to the cardinal stimuli's performance). Therefore our results once again argue for complex spatial interactions, which most existing models cannot capture easily.

Second, crowding and uncrowding in the multisquare conditions cannot be explained by local interactions of its subparts—the configuration matters. Alternatively, local interactions were important for the single flanker conditions (Figures 3a, 3b, & 3c, upper & bottom panels; Supplementary Table S3). The LMM and post-hoc comparisons showed that ortho-target flanker conditions had significantly better performance, whereas iso-target flanker conditions had comparable performances to complete square conditions. More specifically, performance for the vertical Vernier surrounded by the square is as poor as for the vertical lines of the square only (Figure 3 upper panel a & b). This result may be taken as support that only neurons of similar orientation interact. However, there is still some effect of the horizontal lines too, which may be considered an unspecific effect (Figure 3 upper panel c). This effect is even more pronounced for the horizontal Vernier since the horizontal lines crowd more than the square (Figure 3 lower panel a & b). On the other hand, (un)crowding in the multi-square conditions showed clearly a different pattern (Figures 3d–3o upper & bottom panels; Supplementary Table S4). In general, conditions with complete squares lead to better or equal performance than conditions with parts of a square only, except for an iso-target condition (Figure 3 h, upper panel), indicating good Gestalt matters. In the three and seven squares conditions, post-hoc Tukey's HSD test after an LMM analysis showed that complete squares flanker configurations led to significantly

better performance than iso-target and ortho- target flanker configurations. Therefore our results imply that, unlike complete squares, parts of the squares cannot release crowding by low-level interactions, such as contour-contour interaction or line-line detector inhibition.

Third, as a side note, the results in Figure 3 conditions b and c also show that participants did not perceive the task as a bisection task. In other words, participants did not discriminate the Vernier offset relative to the bisector (bisection cue) of two parallel bars, as in a bisection task (e.g., Clarke et al., 2014). Since performance with iso-target flankers was worse in this condition than with the ortho-target flankers, no bisection cue can be used (to be more precise: if there were a bisection cue, it must be much weaker than other mechanisms involved).

Fourth, there were complex interactions between Vernier orientation and square configuration orientation, which cannot easily be explained by a single, local mechanism, except for the single iso- and ortho-target flankers, which are in accordance with the predictions of most local models. However, for the more complex configurations, Vernier orientation did not matter. For example, performance for the vertical and horizontal Vernier showed a very similar pattern independent of Vernier orientation for the horizontally arranged squares: strong crowding for one and three squares and strong improvement of the seven squares conditions. The random intercept of target orientation only had a marginal significance; also, the explained variance with and without the random intercept was only 1.7% (in Pooled conditions). However, qualitatively, there was a trend of an effect of the vertically oriented square array with Vernier orientation. There is only a weak improvement, if at all, for the vertical Vernier as the number of squares increases (Figure 2 left b vs. d, weaker in Figure 3 upper d vs. j), but there is a clear improvement for the horizontal target (Figure 2 right b vs. d, weaker in Figure 3 lower g vs. m). Again, this latter effect cannot be explained by an increase in the number of horizontal lines because performance deteriorates the more horizontal lines there are. Thus mutual inhibition between the horizontal lines is not a viable explanation. Finally, rotating the entire configuration showed a different behavioral pattern, that is, increasing the number of squares only in one orientation (either ±45°) did not improve performance, but significant performance improvements were observed with 35-square grid conditions. However, again, Vernier orientation did not matter.

Fifth, we reproduced the previous finding that the mutual crowding of the squares increases with the number of flanking squares (Manassi et al., 2013). In addition, we found a radial-tangential anisotropy (Chung, 2013; Greenwood et al., 2017; Kwon et al., 2014; Malania et al., 2020; Toet & Levi, 1992). The target square in the horizontally aligned squares was more crowded than in the vertically aligned squares (Figure 5 & Supplementary Figure S1). However, such anisotropy is not reflected in the Vernier discrimination task, that is, the Vernier performance was not better in horizontally aligned squares than in vertically aligned squares (Figure 2). If uncrowding can be simply explained by "crowding of crowding" as Manassi and colleagues (2013) suggested, stronger crowding in horizontally aligned squares would have induced better segregation of squares from Vernier target, hence, better performance with horizontally aligned squares than with vertically aligned squares. However, this was not the case. In the 3-squares conditions, to the contrary, vertically aligned squares led to better performance than the horizontally aligned squares (Figures 2b vs. 2c), but not in the seven-squares conditions (Figures 2d vs. 2e). The results again suggest that uncrowding is not a single process but rather a complex problem with many factors involved.

Sixth, we found no significant differences with and without the Vernier stimulus in the center square indicating the target position (the explained variance difference between with and without the random intercept of Vernier presentation was small, only 0.5%; Figure 5 left vs. right). This result indicates that performance deterioration does not come from location uncertainty.

In the current work, we did not compute statistics for all possible comparisons between conditions in experiments to avoid multiple testing and because they were not part of our main research question. The majority of the analyzed comparisons show that the holistic structure matters (e.g., Experiment 1, complete square conditions vs. partial square conditions).

Whereas certain types of element-element interactions might explain single conditions, it seems that the entirety of findings resists such an explanation. Likely, there are many mechanisms in operation, and these mechanisms may be found more on an implicit statistical level than by explicit element-element interactions similar to the processing of CNNs where single neurons code for a large number of stimulus features. For this reason, we subjected our data to two 2-stage models, which take large-scale configural interactions into account (Laminart and Capsule networks), and a 1-stage model, which was proposed to account for complex configurational effects with high-dimensional pooling (HD pooling) and in the decision process. Other models, such as classic CNNs, epitomes, Fourier analysis, etc., failed with the basic crowding conditions and were not considered here (Doerig et al., 2019).

The TTM did not show uncrowding when adding more squares. Albeit its ability to pool a large number of features (HD pooling), the information of the

target Vernier and the precise flanker structure was irretrievably lost. Whereas TTM is an excellent model for textural processing and summary statistics, we suggest that TTM misses a flexible segmentation stage, which segments visual scenes in multiple groups depending on the configuration. The TTM, as a 1-stage model, does not have a flexible segmentation stage and thus treats fine details of all elements equally. For this reason, it erases small details, which makes a major difference for the human system and leads to qualitatively different results (Wallis, Funke, Ecker, Gatys, Wichmann, & Bethge, 2019). In addition, there is a similar problem with the pooling regions. As shown in the experiments, changes across large parts of the visual field matter. For example, the outmost squares strongly matter but are 8.5° away from the Vernier target. Using wider filters to take this information into account would strongly compress the target. Hence, further detailed information is crucial, thus, more flexible architectures are needed.

The Laminart model reproduced human performance when more square flankers were added (uncrowding) but unexpectedly showed uncrowding in the iso-target conditions (Figures 6Bf & 6Bi), whereas human participants showed strong crowding (Figures 6Df & 6Di). Capsule networks reproduced the results best as they take explicit object representations into account, suggesting that object-level segmentation is needed to fully account for the complex effects of configuration. However, Capsule networks were trained for the specific stimuli and task, whereas TTM and Laminart were not adapted. Nonetheless, the human-like performance of Capsule networks was not due to the training process, since the control networks, without grouping and segmentation process, using the same training procedure, could not reproduce the human performance (Supplementary Figure S2). Thus these results support that object-based grouping and segmentation processes are crucial to explain human behavior.

We believe that our results show that flexible segmentation and grouping are critical for human vision (as do Capsule networks and Laminart model). In natural conditions, nearby elements on the retina may not be nearby in the outer world because they may be located at very different depth planes (perceptual groups). For example, a mesh fence in front of a house leads to overlapping contours of the fence and the house in the early visual areas. A flexible grouping and segmentation stage first groups these contours with each other before any interaction occurs across the depth planes. Crowding occurs when the individual contours within the depth plane may be suppressed to see the wholes, such as the fence and the house. No crowding should occur between contours that do not belong to the same depth plane. Indeed, crowding does not occur when the target and the flankers belong to different depth planes, even though they lie at nearby locations in retinal coordinates (Astle, McGovern, & McGraw, 2014; Kooi et al., 1994; Sayim et al., 2008). In our experiments, a single square and Vernier are grouped as one object, that is, they belong to a single depth plane. In contrast, with the number of squares, square flankers are grouped, and the Vernier is assigned to a different group, either because it is perceived as belonging to a different depth plane or to a different object in the same depth plane.

In summary, we are still quite far from understanding and explaining the major characteristics of crowding. A model that can explain the major characteristics of crowding, in a nutshell, does not exist yet. We are, however, optimistic that such a model exists, since crowding shows universal characteristics across all types of stimuli (Herrera-Esposito et al., 2020; Herzog et al., 2015; Kimchi & Pirkner, 2015; Levi & Carney, 2009; Pelli, Palomares, & Majaj, 2004; Reuther & Chakravarthi, 2014; Saarela et al., 2009; van den Berg et al., 2007; Wallace & Tjan, 2011), tasks (Farzin, Rivera, & Whitney, 2009; Fischer & Whitney, 2011; Yeh, He, S., & Cavanagh, 2012), and modalities (Oberfeld & Stahn, 2012; Overvliet & Sayim, 2016). Understanding crowding may unearth the strategies that are used to make sense of the outer world.

Commercial relationships: none.
Corresponding author: Oh-Hyeon Choung.
Email: oh-hyeon.choung@epfl.ch.
Address: EPFL SV BMI LPSY, Station 19, CH-1015 Lausanne, Switzerland.

References

Andriessen, J.J., & Bouma, H. (1976). Eccentric vision: Adverse interactions between line

segments. *Vision Research, 16*(1), 71–78, https://doi.org/10.1016/0042-6989(76)90078-X.

Astle, A. T., McGovern, D. P., & McGraw, P. V. (2014). Characterizing the role of disparity information in alleviating visual crowding. *Journal of Vision, 14*(6), 8–8, https://doi.org/10.1167/14.6.8.

Attneave, F., & Olson, R. K. (1967). Discriminability of stimuli varying in physical and retinal orientation. *Journal of Experimental Psychology, 74*(2p1), 149.

Bach, M. (1996). The Freiburg Visual Acuity Test—Automatic Measurement of Visual Acuity. *Optometry and Vision Science, 73*(1), 49–53.

Ball, K., & Sekuler, R. (1982). A specific and enduring improvement in visual motion discrimination. *Science, 218*(4573), 697–698.

Barton, K. (2020). *MuMIn: Multi-Model Inference*, https://CRAN.R-project.org/package=MuMIn.

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1–48, https://doi.org/10.18637/jss.v067.i01.

Bauer, J. A., Owens, D. A., Thomas, J., & Held, R. (1979). Monkeys show an oblique effect. *Perception, 8*(3), 247–253.

Bex, P. J., & Dakin, S. C. (2005). Spatial interference among moving targets. *Vision Research, 45*(11), 1385–1398.

Bornet, A., Choung, O.-H., Doerig, A., Whitney, D., Herzog, M. H., & Manassi, M. Global and high-level effects in crowding cannot be predicted by either high-dimensional pooling or target cueing. Manuscript submitted for publication.

Bornet, A., Doerig, A., Herzog, M. H., Francis, G., & der Burg, E. V. (2021). Shrinking Bouma's window: How to model crowding in dense displays. *PLOS Computational Biology, 17*(7), e1009187, https://doi.org/10.1371/journal.pcbi.1009187.

Bornet, A., Kaiser, J., Kroner, A., Falotico, E., Ambrosano, A., Cantero, K., Herzog, M. H., . . . Francis, G. (2019). Running large-scale simulations on the Neurorobotics Platform to understand vision-the case of visual crowding. *Frontiers in Neurorobotics, 13*, 33.

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

Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Research, 13*(4), 767–782.

Bouma, H., & Andriessen, J. J. (1968). Perceived orientation of isolated line segments. *Vision Research, 8*(5), 493–507.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision, 10*(4), 433–436, https://doi.org/10.1163/156856897x00357.

Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience, 13*(1), 51–62, https://doi.org/10.1038/nrn3136.

Chakravarthi, R., & Cavanagh, P. (2009). Recovery of a crowded object by masking the flankers: Determining the locus of feature integration. *Journal of Vision, 9*(10), 4–4, https://doi.org/10.1167/9.10.4.

Chicherov, V., & Herzog, M. H. (2015). Targets but not flankers are suppressed in crowding as revealed by EEG frequency tagging. *NeuroImage, 119*, 325–331, https://doi.org/10.1016/j.neuroimage.2015.06.047.

Chicherov, V., Plomp, G., & Herzog, M. H. (2014). Neural correlates of visual crowding. *NeuroImage, 93*, 23–31, https://doi.org/10.1016/j.neuroimage.2014.02.021.

Chung, S. T. L. (2013). Cortical Reorganization after Long-Term Adaptation to Retinal Lesions in Humans. *Journal of Neuroscience, 33*(46), 18080–18086, https://doi.org/10.1523/JNEUROSCI.2764-13.2013.

Chung, S. T. L., 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.

Clarke, A. M., Grzeczkowski, L., Mast, F. W., Gauthier, I., & Herzog, M. H. (2014). Deleterious effects of roving on learned tasks. *Vision Research, 99*, 88–92, https://doi.org/10.1016/j.visres.2013.12.010.

Clarke, A. M., Herzog, M. H., & Francis, G. (2014). Visual crowding illustrates the inadequacy of local vs. Global and feedforward vs. Feedback distinctions in modeling visual perception. *Frontiers in Psychology, 5*, 1193, https://doi.org/10.3389/fpsyg.2014.01193.

Coen-Cagli, R., Kohn, A., & Schwartz, O. (2015). Flexible gating of contextual influences in natural vision. *Nature Neuroscience, 18*, 1648.

Coletta, N. J., Segu, P., & Tiana, C. L. (1993). An oblique effect in parafoveal motion perception. *Vision Research, 33*(18), 2747–2756.

Dakin, S. C., Cass, J., Greenwood, J. A., & Bex, P. J. (2010). Probabilistic, positional averaging predicts object-level crowding effects with letter-like stimuli. *Journal of Vision, 10*(10), 14–14.

Doerig, A., Bornet, A., Choung, O. H., & Herzog, M. H. (2020). Crowding reveals fundamental differences in local vs. Global processing in humans and machines. *Vision Research, 167*, 39–45.

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.

Doerig, A., Schmittwilken, L., Sayim, B., Manassi, M., & Herzog, M. H. (2020). Capsule networks as recurrent models of grouping and segmentation. *PLOS Computational Biology, 16*(7), e1008017.

Ehlers, H. (1936). V: The movements of the eyes during reading. *Acta Ophthalmologica, 14*(1–2), 56–63.

Farzin, F., Rivera, S. M., & Whitney, D. (2009). Holistic crowding of Mooney faces. *Journal of Vision, 9*(6), 18.1–18.15, https://doi.org/10.1167/9.6.18.

Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology, 106*(3), 1389–1398.

Flom, M. C., Heath, G. G., & Takahashi, E. (1963). Contour interaction and visual resolution: Contralateral effects. *Science, 142*(3594), 979–980.

Francis, G., Manassi, M., & Herzog, M. H. (2017). Neural dynamics of grouping and segmentation explain properties of visual crowding. *Psychological Review, 124*(4), 483–504, https://doi.org/10.1037/rev0000070.

Fründ, I., Haenel, N. V., & Wichmann, F. A. (2011). Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision, 11*(6), 16–16, https://doi.org/10.1167/11.6.16.

Furmanski, C. S., & Engel, S. A. (2000). An oblique effect in human primary visual cortex. *Nature Neuroscience, 3*(6), 535–536.

Gheri, C., Morgan, M. J., & Solomon, J. A. (2007). The relationship between search efficiency and crowding. *Perception, 36*(12), 1779–1787.

Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2009). Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences, 106*(31), 13130–13135.

Greenwood, J. A., & Parsons, M. J. (2020). Dissociable effects of visual crowding on the perception of color and motion. *Proceedings of the National Academy of Sciences, 117*(14), 8196–8202.

Greenwood, J. A., Szinte, M., Sayim, B., & Cavanagh, P. (2017). Variations in crowding, saccadic precision, and spatial localization reveal the shared topology of spatial vision. *Proceedings of the National Academy of Sciences, 114*(17), E3573–E3582, https://doi.org/10.1073/pnas.1615504114.

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.

Herrera-Esposito, D., Coen-Cagli, R., & Gomez-Sena, L. (2020). Flexible contextual modulation of naturalistic texture perception in peripheral vision. *BioRxiv*, 2020.01.24.918813, https://doi.org/10.1101/2020.01.24.918813.

Herzog, M. H., & Manassi, M. (2015). Uncorking the bottleneck of crowding: A fresh look at object recognition. *Current Opinion in Behavioral Sciences, 1*, 86–93.

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, https://doi.org/10.1167/15.6.5.

Herzog, M. H., Thunell, E., & Ögmen, H. (2016). Putting low-level vision into global context: Why vision cannot be reduced to basic circuits. *Vision Research, 126*, 9–18, https://doi.org/10.1016/j.visres.2015.09.009.

Hothorn, T., Bretz, F., & Westfall, P. (2008). Simultaneous Inference in General Parametric Models. *Biometrical Journal, 50*(3), 346–363, https://doi.org/10.1002/bimj.200810425.

Hubel, D. H., Wiesel, T. N., & Stryker, M. P. (1978). Anatomical demonstration of orientation columns in macaque monkey. *Journal of Comparative Neurology, 177*(3), 361–379.

Huckauf, A., & Heller, D. (2002). What various kinds of errors tell us about lateral masking effects. *Visual Cognition, 9*(7), 889–910.

Johnson, P. C. (2014). Extension of Nakagawa & Schielzeth's R2GLMM to random slopes models. *Methods in Ecology and Evolution, 5*(9), 944–946, https://doi.org/10.1111/2041-210X.12225.

Kennedy, G. J., & Whitaker, D. (2010). The chromatic selectivity of visual crowding. *Journal of Vision, 10*(6), 15–15.

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.

Kleiner, M., Brainard, D., & Pelli, D. (2007). *What's new in Psychtoolbox-3?*

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), 255–279.

Korte, W. (1923). Über die Gestaltauffassung im indirekten Sehen. *Zeitschrift Für Psychologie, 93*, 17–82.

Kwon, M., Bao, P., Millin, R., & Tjan, B. S. (2014). Radial-tangential anisotropy of

crowding in the early visual areas. *Journal of Neurophysiology, 112*(10), 2413–2422, https://doi.org/10.1152/jn.00476.2014.

Kwon, M., & Liu, R. (2019). Linkage between retinal ganglion cell density and the nonuniform spatial integration across the visual field. *Proceedings of the National Academy of Sciences, 116*(9), 3827–3836, https://doi.org/10.1073/pnas.1817076116.

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.

Levi, D. M., & Carney, T. (2009). Crowding in peripheral vision: Why bigger is better. *Current Biology, 19*(23), 1988–1993.

Li, B., Peterson, M. R., & Freeman, R. D. (2003). Oblique effect: A neural basis in the visual cortex. *Journal of Neurophysiology, 90*(1), 204–217.

Livne, T., & Sagi, D. (2007). Configuration influence on crowding. *Journal of Vision, 7*(2), 4, https://doi.org/10.1167/7.2.4.

Livne, T., & Sagi, D. (2011). Multiple levels of orientation anisotropy in crowding with Gabor flankers. *Journal of Vision, 11*(13), 18–18, https://doi.org/10.1167/11.13.18.

Mach, E. (1860). Ueber das Sehen von Lagen und Winkeln durch die Bewegung des Auges. *Sitzungsberichte Der Math Cl Der Kais Akad Der Wissenschaften, 42*, 215–224.

Malania, M., Herzog, M. H., & Westheimer, G. (2007). Grouping of contextual elements that affect vernier thresholds. *Journal of Vision, 7*(2), 1–1, https://doi.org/10.1167/7.2.1.

Malania, M., Pawellek, M., Plank, T., & Greenlee, M. W. (2020). Training-Induced Changes in Radial–Tangential Anisotropy of Visual Crowding. *Translational Vision Science & Technology, 9*(9), 25–25, https://doi.org/10.1167/tvst.9.9.25.

Manassi, M., Hermens, F., Francis, G., & Herzog, M. H. (2015). Release of crowding by pattern completion. *Journal of Vision, 15*(8), 16–16, https://doi.org/10.1167/15.8.16.

Manassi, M., Lonchampt, S., Clarke, A., & Herzog, M. H. (2016). What crowding can tell us about object representations. *Journal of Vision, 16*(3), 35, https://doi.org/10.1167/16.3.35.

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.

Manassi, M., Sayim, B., & Herzog, M. H. (2013). When crowding of crowding leads to uncrowding. *Journal of Vision, 13*(13), 10–10, https://doi.org/10.1167/13.13.10.

Maus, G. W., Fischer, J., & Whitney, D. (2011). Perceived Positions Determine Crowding. *PLOS ONE, 6*(5), e19796, https://doi.org/10.1371/journal.pone.0019796.

Miles, W. R. (1930). Ocular dominance in human adults. *The Journal of General Psychology, 3*(3), 412–430.

Montaser-Kouhsari, L., & Rajimehr, R. (2005). Subliminal attentional modulation in crowding condition. *Vision Research, 45*(7), 839–844, https://doi.org/10.1016/j.visres.2004.10.020.

Motter, B. C., & Simoni, D. A. (2007). The roles of cortical image separation and size in active visual search performance. *Journal of Vision, 7*(2), 6–6, https://doi.org/10.1167/7.2.6.

Nakagawa, S., Johnson, P. C. D., & Schielzeth, H. (2017). The coefficient of determination R2 and intra-class correlation coefficient from generalized linear mixed-effects models revisited and expanded. *Journal of the Royal Society Interface, 14*(134), 20170213, https://doi.org/10.1098/rsif.2017.0213.

Nakagawa, S., & Schielzeth, H. (2013). A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution, 4*(2), 133–142, https://doi.org/10.1111/j.2041-210x.2012.00261.x.

Oberfeld, D., & Stahn, P. (2012). Sequential grouping modulates the effect of non-simultaneous masking on auditory intensity resolution. *PloS One, 7*(10), e48054.

Overvliet, K. E., & Sayim, B. (2016). Perceptual grouping determines haptic contextual modulation. *Vision Research, 126*, 52–58, https://doi.org/10.1016/j.visres.2015.04.016.

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.

Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current Opinion in Neurobiology, 18*(4), 445–451.

Pelli, D. G., Palomares, M., & Majaj, N. J. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*(12), 12, https://doi.org/10.1167/4.12.12.

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.

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

Põder, E. (2007). Effect of colour pop-out on the recognition of letters in crowding conditions. *Psychological Research, 71*(6), 641–645.

Põder, E., & Wagemans, J. (2007). Crowding with conjunctions of simple features. *Journal of Vision, 7*(2), 23–23.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing, https://www.R-project.org/.

Reuther, J., & Chakravarthi, R. (2014). Categorical membership modulates crowding: Evidence from characters. *Journal of Vision, 14*(6), 5–5, https://doi.org/10.1167/14.6.5.

Rosenholtz, R., Huang, J., Raj, A., Balas, B. J., & Ilie, L. (2012). A summary statistic representation in peripheral vision explains visual search. *Journal of Vision, 12*(4), 14–14.

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/10.1167/19.7.15.

Saarela, T. P., & Herzog, M. H. (2008). Time-course and surround modulation of contrast masking in human vision. *Journal of Vision, 8*(3), 23–23, https://doi.org/10.1167/8.3.23.

Saarela, T. P., Sayim, B., Westheimer, G., & Herzog, M. H. (2009). Global stimulus configuration modulates crowding. *Journal of Vision, 9*(2), 5–5, https://doi.org/10.1167/9.2.5.

Saarela, T. P., Westheimer, G., & Herzog, M. H. (2010). The effect of spacing regularity on visual crowding. *Journal of Vision, 10*(10), 17–17, https://doi.org/10.1167/10.10.17.

Saarinen, J., & Levi, D. M. (1995). Orientation anisotropy in vernier acuity. *Vision Research, 35*(17), 2449–2461.

Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. arXiv preprint arXiv:1710.09829.

Sayim, B., Westheimer, G., & Herzog, M. H. (2008). Contrast polarity, chromaticity, and stereoscopic depth modulate contextual interactions in vernier acuity. *Journal of Vision, 8*(8), 12–12, https://doi.org/10.1167/8.8.12.

Sayim, B., Westheimer, G., & Herzog, M. H. (2010). Gestalt Factors Modulate Basic Spatial Vision. *Psychological Science, 21*(5), 641–644, https://doi.org/10.1177/0956797610368811.

Sayim, B., Westheimer, G., & Herzog, M. H. (2011). Quantifying target conspicuity in contextual modulation by visual search. *Journal of Vision, 11*(1), 6–6, https://doi.org/10.1167/11.1.6.

Silson, E. H., Reynolds, R. C., Kravitz, D. J., & Baker, C. I. (2018). Differential Sampling of Visual Space in Ventral and Dorsal Early Visual Cortex. *The Journal of Neuroscience, 38*(9), 2294–2303, https://doi.org/10.1523/JNEUROSCI.2717-17.2018.

Solomon, J. A., Felisberti, F. M., & Morgan, M. J. (2004). Crowding and the tilt illusion: Toward a unified account. *Journal of Vision, 4*(6), 9–9, https://doi.org/10.1167/4.6.9.

Strasburger, H. (2005). Unfocussed spatial attention underlies the crowding effect in indirect form vision. *Journal of Vision, 5*(11), 8–8, https://doi.org/10.1167/5.11.8.

Strasburger, H. (2020). Seven Myths on Crowding and Peripheral Vision. *I-Perception, 11*(3), 2041669520913052, https://doi.org/10.1177/2041669520913052.

Strasburger, H., Harvey, L. O., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Perception & Psychophysics, 49*(6), 495–508.

Taylor, M. M., & Creelman, C. D. (1967). PEST: Efficient estimates on probability functions. *Journal of the Acoustical Society of America, 41*(4A), 782–787, https://doi.org/10.1121/1.1910407.

Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research, 32*(7), 1349–1357, https://doi.org/10.1016/0042-6989(92)90227-A.

Tripathy, S. P., & Cavanagh, P. (2002). The extent of crowding in peripheral vision does not scale with target size. *Vision Research, 42*(20), 2357–2369, https://doi.org/10.1016/S0042-6989(02)00197-9.

van den Berg, R., Roerdink, J. B., & Cornelissen, F. W. (2007). On the generality of crowding: Visual crowding in size, saturation, and hue compared to orientation. *Journal of Vision, 7*(2), 14–14.

van der Burg, E., Olivers, C. N. L., & Cass, J. (2017). Evolving the keys to visual crowding. *Journal of Experimental Psychology: Human Perception and Performance, 43*(4), 690–699, https://doi.org/10.1037/xhp0000337.

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

Wallis, T. S. A., & Bex, P. J. (2011). Visual crowding is correlated with awareness. *Current Biology : CB, 21*(3), 254–258, https://doi.org/10.1016/j.cub.2011.01.011.

Wallis, T. S. A., Funke, C. M., Ecker, A. S., Gatys, L. A., Wichmann, F. A., & Bethge, M. (2019). Image content is more important than Bouma's Law for scene metamers. *ELife, 8*, e42512.

Westheimer, G. (2005). Anisotropies in peripheral vernier acuity. *Spatial Vision, 18*(2), 159–167.

Weymouth, F. W. (1958). Visual sensory units and the minimal angle of resolution. *American Journal of Ophthalmology, 46*(1), 102–113.

Wilkinson, F., Wilson, H. R., & Ellemberg, D. (1997). Lateral interactions in peripherally viewed texture arrays. *Journal of the Optical Society of America A, 14*(9), 2057, https://doi.org/10.1364/JOSAA.14.002057.

World Medical Association (2013). Declaration of Helsinki: Ethical principles for medical research involving human subjects. *Journal of the American Medical Association, 310*(20), 2191–2194, https://doi.org/10.1001/jama.2013.281053.

Xu, X., Collins, C. E., Khaytin, I., Kaas, J. H., & Casagrande, V. A. (2006). Unequal representation of cardinal vs. Oblique orientations in the middle temporal visual area. *Proceedings of the National Academy of Sciences, 103*(46), 17490–17495.

Yeh, S.-L., He, S., & Cavanagh, P. (2012). Semantic priming from crowded words. *Psychological Science, 23*(6), 608–616, https://doi.org/10.1177/0956797611434746.

Yeotikar, N. S., Khuu, S. K., Asper, L. J., & Suttle, C. M. (2011). Configuration specificity of crowding in peripheral vision. *Vision Research, 51*(11), 1239–1248, https://doi.org/10.1016/j.visres.2011.03.016.

Yeshurun, Y., & Rashal, E. (2010). Precueing attention to the target location diminishes crowding and reduces the critical distance. *Journal of Vision, 10*(10), 16–16, https://doi.org/10.1167/10.10.16.