What crowding can tell us about object representations

Mauro Manassi*
Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
Department of Psychology, University of California, Berkeley, CA, USA

Sophie Lonchampt*
Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

Aaron Clarke
Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland
Department of Psychology, Bilkent University, Ankara, Turkey

Michael H. Herzog
Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

In crowding, perception of a target usually deteriorates when flanking elements are presented next to the target. Surprisingly, adding further flankers can lead to a release from crowding. In previous work we showed that, for example, vernier offset discrimination at 9° of eccentricity deteriorated when a vernier was embedded in a square. Adding further squares improved performance. The more squares presented, the better the performance, extending across 20° of the visual field. Here, we show that very similar results hold true for shapes other than squares, including unfamiliar, irregular shapes. Hence, uncrowding is not restricted to simple and familiar shapes. Our results provoke the question of whether any type of shape is represented at *any* location in the visual field. Moreover, small changes in the orientation of the flanking shapes led to strong increases in crowding strength. Hence, highly specific shape-specific interactions across large parts of the visual field determine vernier acuity.

Introduction

Object recognition is often thought to be feedforward and hierarchical (DiCarlo, Zoccolan, & Rust, 2012; Hubel & Wiesel, 1962; Hung, Kreiman, Poggio, & DiCarlo, 2005; Riesenhuber & Poggio, 1999; Serre, Kouh, Cadieu, & Knoblich, 2005; Serre, Kreiman, et al., 2007; Serre, Oliva, & Poggio, 2007; Thorpe, Delorme, & Van Rullen, 2001). The analysis of a visual scene starts with the extraction of basic features (e.g., lines and contours) in the early visual cortex and proceeds to more and more complex features (e.g., shapes, faces, and objects) in higher visual areas. Complex feature detectors are created by pooling outputs from more basic feature detectors. For example, a hypothetical square-detecting neuron receives input from neurons sensitive to its constituting vertical and horizontal lines. Accordingly, neural receptive field sizes along the processing hierarchy increase from step to step simply because a square covers more space than its constituting lines. Therefore, receptive fields need to be larger. One consequence of pooling is that neurons are sensitive to context. Hence, a prediction of such models is that elements neighboring a target element impair target processing because features of the target and flankers are pooled, and thus target information is lost. Indeed, this is the case for crowding (Flom, Heath, & Takahashi, 1963; Levi, 2008; Strasburger & Wade, 2015; Whitney & Levi, 2011). For this reason, pooling models have become the

Citation: Manassi, M., Lonchampt, S., Clarke, A., & Herzog, M. H. (2016). What crowding can tell us about object representations. *Journal of Vision, 16*(3):35, 1–13, doi:10.1167/16.3.35.

doi: 10.1167/16.3.35 Received March 16, 2015; published February 24, 2016 ISSN 1534-7362

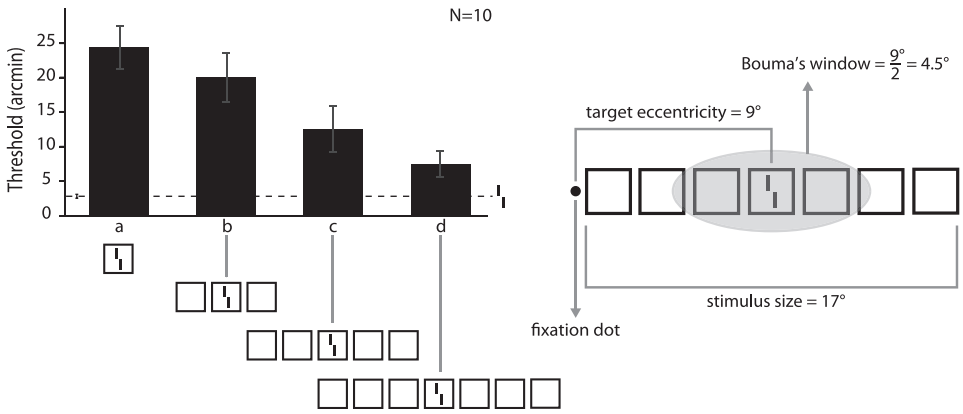

Figure 1. Left panel: Observers were asked to discriminate the offset direction of a vernier (dashed line). Thresholds increased when the vernier was embedded in a square (a). Thresholds gradually decreased when the number of flanking squares increased (b through d). Replotted from Manassi et al. (2013). Right panel: The stimulus configuration from panel d. Bouma's law states that elements interfere with vernier offset discrimination only within a region of 4.5°, which is half of the target eccentricity (9°). However, the outmost squares are presented beyond this region and still influence vernier thresholds.

standard in crowding research (Balas, Nakano, & Rosenholtz, 2009; Dakin, Cass, Greenwood, & Bex, 2010; Freeman, Chakravarthi, & Pelli, 2012; Freeman & Simoncelli, 2011; Greenwood, Bex, & Dakin, 2009, 2010; Parkes, Lund, Angelucci, Solomon, & Morgan, 2001; van den Berg, Roerdink, & Cornelissen, 2010; Wilkinson, Wilson, & Ellemberg, 1997).

Most models of crowding have three main characteristics in common. First, crowding occurs only in a restricted region according to Bouma's law, which states that only flanking elements within a window of about half the eccentricity of target presentation compromise target processing (Bouma, 1970; Pelli, 2008; Pelli, Palomares, & Majaj, 2004; Pelli & Tillman, 2008; Rosen, Chakravarthi, & Pelli, 2014). Second, flankers are treated as mere noise; therefore, increasing their number can only lead to increases in crowding strength (Parkes et al., 2001; Wilkinson et al., 1997). Third, crowding is feature specific; that is, crowding occurs only when target and flankers have the same color (Kooi, Toet, Tripathy, & Levi, 1994; Põder, 2007), orientation (Andriessen, 1976), or shape (Kooi et al., 1994; Nazir, 1992).

However, we have previously shown that none of these characteristics universally hold true (see Herzog & Manassi, 2015 and Herzog, Sayim, Chicherov, & Manassi, 2015 for a review; Malania, Herzog, & Westheimer, 2007; Manassi, Sayim, & Herzog, 2012, 2013; Saarela, Westheimer, & Herzog, 2010; Sayim, Westheimer, & Herzog, 2010). For example, we presented a vernier at 9° of eccentricity in the periphery. Performance strongly deteriorated when the vernier was surrounded by the outline of a square (Figure 1a). This is a classic crowding effect. However, when the vernier and the central square were flanked by further squares to the right and left, crowding was strongly reduced, almost to the unflanked level (Figure 1b through d; Manassi et al., 2013).

How can these results be explained? One scenario relies on explicit object representations. First, the squares are computed from their constituting lines. Next, the shape representations interact with each other (e.g., mutual inhibition), and then vernier acuity is determined. In a more dynamic model, all interactions occur more or less concurrently. Such a scenario requires explicit representations of squares at all positions in the visual field, pointing to fundamental questions about the nature of object representation. For example, if similar effects of uncrowding are also found with other shapes, including unfamiliar shapes, then the human brain needs to maintain zillions of object representations at one location of the visual field.

Explicit object representations may not necessarily be required to explain why adding shapes can decrease crowding. For example, if the visual system performs a Fourier decomposition of the incoming stimulus, then adding lines to the image can simplify the pattern in the Fourier domain. This is illustrated in Figure 2, where the top row shows two images, one with a single line and the other with many lines. The bottom row shows the respective Fourier transforms. The single line leads to Fourier energy distributed over a wide range of spatial frequencies (horizontal line through the center of the Fourier domain representation). The image with many lines, however, has energy at only two places: the center and the far left edge.

Alternatively, according to Balas et al. (2009) and Freeman and Simoncelli (2011), the visual system may extract complex features without full object representations. In models of texture processing, higher order structures of the stimuli—but not full object representations—are computed, which may be crucial for

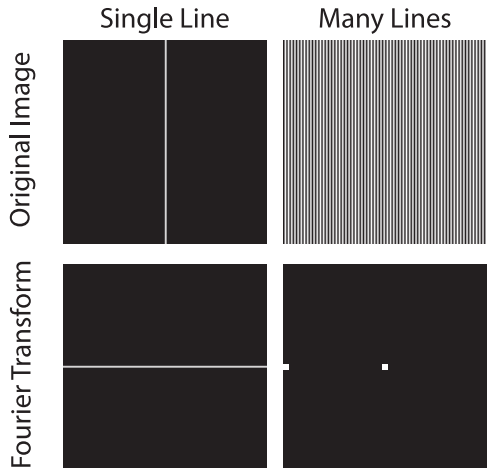

Figure 2. Top row: Original images. Bottom row: Fourier transform. In the Fourier domain, high spatial frequencies are represented at the center of the image and lower spatial frequencies are represented at increasing eccentricities from the center. Orientation is represented by position around the center. For a single line (top left image), the Fourier domain representation contains energy at many spatial frequencies (bottom left panel). For many lines (top right image), the Fourier domain representation is much simpler and contains energy at only two locations, shown by the white dots at the center and the far left (bottom right panel).

crowding. In other models of texture perception, only low-level cues may be important. For example, it may be that only the vertical lines making up the squares determine crowding strength. Of course, many more models are conceivable, such as combinations of the above models. Finally, regularities are often explained by the well-known Gestalt laws.

Here, we first show that uncrowding occurs with many shapes, including nonfamiliar and complex ones. Second, our results support our previous conclusions that simple models of crowding cannot explain uncrowding. Third, our results pose challenges that future models need to meet. For example, models need to explain how the human brain can code shapes at most locations in the visual field without suffering from the curse of dimensionality. Thus, our results point to very general questions about the representation of objects.

Materials and method

Observers

Participants were paid students of the École Polytechnique Fédérale de Lausanne. All observers had normal or corrected-to-normal vision with a visual acuity of 1.0 (corresponding to 20/20) or better in at least one eye, measured with the Freiburg Visual Acuity Test (Bach, 1996). Observers were told that they could quit the experiment at any time they wished. Participants signed an informed consent form and were informed about the general purpose of the experiment, which was approved by the local ethical committee. They were paid 20 CHF/hr for their participation.

Apparatus and stimuli

Stimuli were presented on a Philips (Amsterdam, The Netherlands) 201B4 cathode ray tube monitor, which was driven by a standard accelerated graphics card. Screen resolution was set to 1024×768 pixels at a 100-Hz refresh rate. The white point of the monitor was adjusted to D65. The color space was linearized by applying individual gamma correction to each color channel. Target and flankers consisted of white lines presented on a black background. The luminance of stimuli was 80 cd/m^2. A Minolta (Tokyo, Japan) CA-210 display color analyzer was used. All the experiments were programmed and run using Matlab 2012b (The MathWorks Inc., Natick, MA) with the Psychophysics Toolbox (Brainard, 1997) and Palamedes (Prins, 2009) routines.

Viewing distance was 75 cm. Observers were instructed to fixate a white dot (2-arcmin diameter). A vertical vernier embedded in various shape configurations was presented on the right visual field at 9° of eccentricity. Observers were asked to indicate the offset direction. The vernier consisted of two vertical lines (40 arcmin long) separated by a vertical gap of 4 arcmin. The stimulus duration was 150 ms. To reduce target position uncertainty, in Experiments 1 and 2 we added two vertical lines (40 arcmin long) 150 arcmin above and below the center of the target.

Procedure

An adaptive staircase procedure (QUEST; Watson & Pelli, 1983), as implemented by the Palamedes Psychometric Toolbox for Matlab (Prins, 2009), was used to determine the vernier offset for which observers reached 75% correct responses. We estimated both the threshold and the slope of the psychometric function (cumulative Gaussian) by means of maximum likelihood estimation, taking all data points into account (Wichmann & Hill, 2001). In order to avoid extremely large vernier offsets, we restricted the QUEST procedure to not exceed offsets of 33.32 arcmin (i.e., twice the starting value of 16.66 arcmin). If vernier offset thresholds were not stable across the experiment

(because of learning or fatigue), observers were screened out of the experiment.

Each condition was presented in separate blocks of 80 trials. All conditions were measured twice (i.e., 160 trials) and randomized individually for each observer. To compensate for possible learning effects, the order of conditions was reversed after each condition had been measured once. Observers were instructed to fixate the dot during the trial. After each response, the screen remained blank for a maximum period of 3 s, during which observers were required to make a response by pushing one of two buttons. Auditory feedback was provided after incorrect or omitted responses. The screen was blank for 500 ms between each response and the next trial.

Individual adjustment of stimulus configuration

In order to avoid floor and ceiling effects, we increased or reduced shape size (and consequently intershape spacing) individually for each observer. If the threshold in the single-shape condition was not at least three times higher than the unflanked vernier threshold, we reduced shape size by 85%. If the criterion was still not met, we reduced the ratio to 75%. Conversely, if the threshold was 33.32 arcmin in both single- and multishape conditions, we increased the size and spacing in the single-shape and multishape conditions to 115%, 132%, or 152%.

In Experiments 2 and 3, thresholds in the seven shape condition had to be at least 70% lower compared with the single-shape condition. If this criterion was not met, we increased or decreased the size of the shapes (and intershape spacing) by 85%, 115%, or 132%. In Experiment 3, we increased stimulus eccentricity from 9° to 11° (to 10° for one subject).

Fourier model

To investigate whether uncrowding can be explained by a simple Fourier model, we implemented a model following the approach of Hermens, Luksys, Gerstner, Herzog, and Ernst (2008) as follows.

First, for a left-offset image with k flankers—$I(x, y)_{L,k}$—and corresponding right-offset image—$I(x, y)_{R,k}$—we computed the Fourier transforms:

$$F(u,v)_{L,k} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x,y)_{L,k}\cdot e^{-2\pi i(ux+vy)}dx\,dy \quad (1)$$

$$F(u,v)_{R,k} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} I(x,y)_{R,k}\cdot e^{-2\pi i(ux+vy)}dx\,dy \quad (2)$$

Second, we took the Euclidian norm of the real and complex parts of $F(u, v)_{L,k}$ and $F(u, v)_{R,k}$ and normalized by the sum of the luminance values in the original image:

$$\tilde{F}(u,v)_{L,k} = \frac{||F(u,v)_{L,k}||_2}{\int_x\int_y I(x,y)_{L,k}dx\,dy} \quad (3)$$

$$\tilde{F}(u,v)_{R,k} = \frac{||F(u,v)_{R,k}||_2}{\int_x\int_y I(x,y)_{R,k}dx\,dy} \quad (4)$$

Third, we took the absolute value of the difference between the left and right $\tilde{F}$ values and integrated over all spatial frequencies:

$$\Delta\tilde{F}_k = \int_u\int_v |\tilde{F}(u,v)_{R,k} - \tilde{F}(u,v)_{L,k}|\,du\,dv \quad (5)$$

These differences indicate how different the left-offset image's Fourier transform is from the right-offset image's Fourier transform.

Fourth, to convert these differences to thresholds, we first flipped the values so that the largest difference corresponded to the lowest threshold (i.e., the best performance). For a threshold data set $X \in [x_0, x_1, \ldots, x_k, \ldots, x_n]$ comprising human thresholds for each of the k numbers of flankers for a given flanker type (e.g., the circles), we linearly rescaled $\Delta\tilde{F}$ to lie on roughly the same range r ($=\max_k X - \min_k X$) as the human threshold data:

$$\phi_k = \max_i(\Delta\tilde{F}_i) - \Delta\tilde{F}_k \quad (6)$$

$$\widetilde{\phi}_k = \frac{\phi_k - \min_i\phi_i}{\max_i\phi_i - \min_i\phi_i}\cdot r + 1 \quad (7)$$

We further constrained the model such that the response to the vernier alone (i.e., with zero flankers) exactly equals the mean subject threshold ($\bar{x}_0$) for this condition:

$$\widetilde{\phi}_k \leftarrow \frac{\widetilde{\phi}_k}{\widetilde{\phi}_0} - 1 + \bar{x}_0 \quad (8)$$

These obtained values now lie on the same range as the human threshold data and are constrained to have the same zero-flanker vernier offset discrimination threshold. An illustration of the four steps in the Fourier model is shown in Figure 3. This term $\widetilde{\phi}_k$ is plotted alongside the human data in Figures 4 and 6 in white bars.

Results

Experiment 1: Uncrowding with seven shapes

Crowding of a vernier can be strongly reduced by increasing the number of flanking squares (Figure 1; see

also Manassi et al., 2013). Here, we show that uncrowding occurs with other shapes as well. In addition, we tested whether Fourier analysis can account for the results.

Method

For each specific shape, different observers were tested in three conditions: vernier alone, vernier embedded in one shape, or vernier embedded in the central shape of an array of seven identical shapes. Depending on the shape, large vernier offsets may have overlapped with the shape outlines. For this reason, we changed the size of each shape. Accordingly, we also increased the spacing between shapes to avoid overlap. The shapes used in each experiment were as follows (*spacing* refers to the center-to-center distance between shapes):

- Circles: radius = 1.3°, spacing = 2.8°; four observers (one female, three males), two observers with 85% size (see Individual adjustment of stimulus configuration)
- Hexagons: radius = 1°, spacing = 2.2°; seven observers (two females, five males), one observer with 85% size and one observer with 115% size
- Octagons: radius = 0.91°, spacing = 2.2°; six observers (two females, four males), four observers with 115% size
- Four-pointed stars: inner radius = 0.92°, outer radius = 1.60°, spacing = 2.9°; five observers (two females, three males), one observer with 115% size
- Seven-pointed stars: inner radius = 0.98°, outer radius = 1.61°, spacing = 2.9°; seven observers (three females, four males), three observers with 115% size
- First irregular shape: horizontal and vertical axes = 1.72° and 2.22°, spacing = 2.2°; six observers (two females, four males), two observers with 115% size and three observers with 132% size
- Second irregular shape: horizontal and vertical axes = 2.72°, spacing = 2.7°; four observers (three females, one male), three observers with 115% size and one observer with 132% size

Results and discussion

When the vernier was embedded in a single shape, thresholds increased compared with the vernier-alone condition ($p < 0.05$). This is a classic crowding effect. When the vernier was flanked by three additional shapes on either side, thresholds decreased compared with the single-shape condition ($p < 0.05$, uncrowding).

For each shape, we found that flanker configurations increased discrimination thresholds significantly: circles, $F(2, 6) = 13.93$, $p < 0.01$, $\eta_p^2 = 0.82$; hexagons, $F(2, 12) = 43.64$, $p < 0.01$, $\eta_p^2 = 0.87$; octagons, $F(2, 10) = 30.65$, $p < 0.01$, $\eta_p^2 = 0.85$; four-pointed stars, $F(2, 8) = 56.85$, $p < 0.01$, $\eta_p^2 = 0.93$; seven-pointed stars, $F(2, 12) = 28.86$, $p < 0.01$, $\eta_p^2 = 0.82$; first irregular shape, $F(2, 10) = 52.43$, $p < 0.01$, $\eta_p^2 = 0.91$; second irregular shape, $F(2, 6) = 11.52$, $p < 0.01$, $\eta_p^2 = 0.79$. Tukey's post hoc tests were used for pairwise comparisons.

Uncrowding with circles (Figure 4a) rules out any explanation based on straight-line interactions. Uncrowding with more complex shapes such as hexagons, octagons, and stars (Figure 4b through e) shows that the visual system is sensitive to many types of shapes, even very complicated ones. Even highly unfamiliar, complex stimulus configurations such as irregular shapes (Figure 4f and g), which may not have been experienced by observers before, led to a decrease in crowding.

The Fourier model predictions (Figure 4, white bars) go in the opposite direction of the human data (Figure 4, black bars). Human thresholds decrease with increasing numbers of flankers, but the model thresholds increase. This is because as more flankers are added, the differences between the left- and right-offset vernier representations are reduced, making the discrimination task more difficult (see Figure 3 for an example of the Fourier spectrum with circles). Taken together, our results show that uncrowding occurs with any kind of shape we tested and that the spatial frequency content of the stimuli cannot account for the results.

Experiment 2: Uncrowding and shape orientation

In the first set of experiments, we showed that uncrowding occurs with many kinds of shapes. Here, we show that small changes in orientation can strongly affect uncrowding (Figure 5).

Method

First, we determined vernier offset discrimination thresholds with hexagons (Figure 5a). Five observers (two females, three males) participated in the experiment (three of them performed with the 115% size, and one performed with the 132% size). As before, we determined offset discrimination thresholds in the three conditions: vernier alone (dashed line), vernier embedded in a hexagon (Figure 5a), and vernier embedded in a hexagon flanked by six identical hexagons (Figure 5b; 0°). In four further conditions, the flanking hexagons were rotated by 2.5°, 5°, 10°, and 15° converging toward the central hexagon (Figure 5c through g).

Second, we tested the influence of mirroring the shapes on uncrowding (Figure 5b). For this purpose,

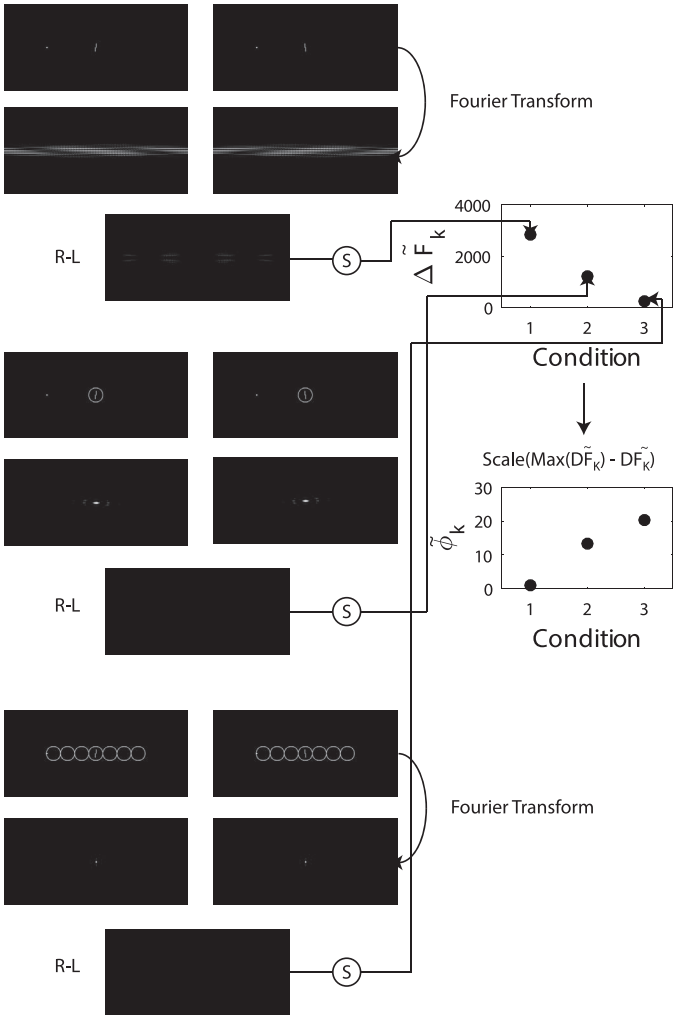

Figure 3. Illustration of the four steps in the Fourier model. For each raw image, we start with the left- and right-offset vernier stimuli (top of each group of three images). We next Fourier transform the image and normalize it by dividing by the sum of all values in the Fourier-transformed image (middle of each group of three images). Then we take the difference between the results for the right and left images (right left) and sum the differences over all spatial frequencies to get the points in the top graph on the right. These values are then flipped by subtracting them from the maximum value, and they are scaled to lie on the same range as the human data (bottom graph on the right).

we used the second type of irregular shape from Figure 4g. Six observers (three females, three males) participated in the experiment. Two of them performed with the 115% size, and three performed with the 132% size. As before, we tested the three main conditions (Figure 5h and i), rotated the flanking irregular shapes by 180° compared with the central shape (Figure 5j), or mirrored the flanking shapes (Figure 5k and l).

Results and discussion

As in the first set of experiments, when the vernier was embedded in a hexagon, thresholds increased compared with the vernier-alone condition (Figure 5a); paired t test: $t(4) = -10.92$, $p < 0.01$. When we added six hexagons with the same orientation, thresholds decreased compared with the single-shape condition (Figure 5a and b); paired t test: $t(4) = 5.45$, $p < 0.01$.

We performed a regression analysis on the individual data, regressing thresholds against the change in rotation of the flanking shapes (0°–30°). For each subject, this analysis yielded a slope and intercept of the regression line. We then performed t tests to determine whether the slope of the regression lines differed significantly from 0. When increasing the rotation of the flanking hexagons from 0° to 30°, thresholds gradually increased and uncrowding gradually disappeared (Figure 5b through g): slope = 121.91, $t(4) = 3.39$, $p = 0.02$. The more the flanking shapes were rotated compared with the central shape, the less the vernier was uncrowded.

In the second experiment (Figure 5b), we found a main effect of flanker configuration: $F(5, 25) = 16.96$, $p < 0.01$. Tukey's post hoc tests were used for pairwise comparisons. As in Figure 4g, when the vernier was embedded in the single irregular shape, thresholds increased compared with the vernier-alone condition (Figure 5h; $p < 0.05$). When we added six identical irregular shapes, thresholds decreased compared with the single-shape condition (Figure 5h and i; $p < 0.05$). When the flanking shapes were rotated by 180°, thresholds increased compared with the previous condition (Figure 5i and j; $p < 0.05$). When the flanking shapes were alternatingly mirrored, thresholds increased compared with the condition with seven identical shapes, although the difference was not statistically significant (Figure 5i through k). When the flanking shapes were all mirrored, thresholds increased compared with the condition with seven identical shapes (Figure 5i through l; $p < 0.05$).

Taken together, the results show that uncrowding is highly sensitive to small changes in shape orientation. Hence, the mechanism underlying uncrowding is not shape invariant.

Experiment 3: Uncrowding and patterns of shapes

Here, we show that uncrowding cannot easily be predicted by simple combinations of the Gestalt rules. It seems that complex shape interactions across large parts of the visual field can determine crowding.

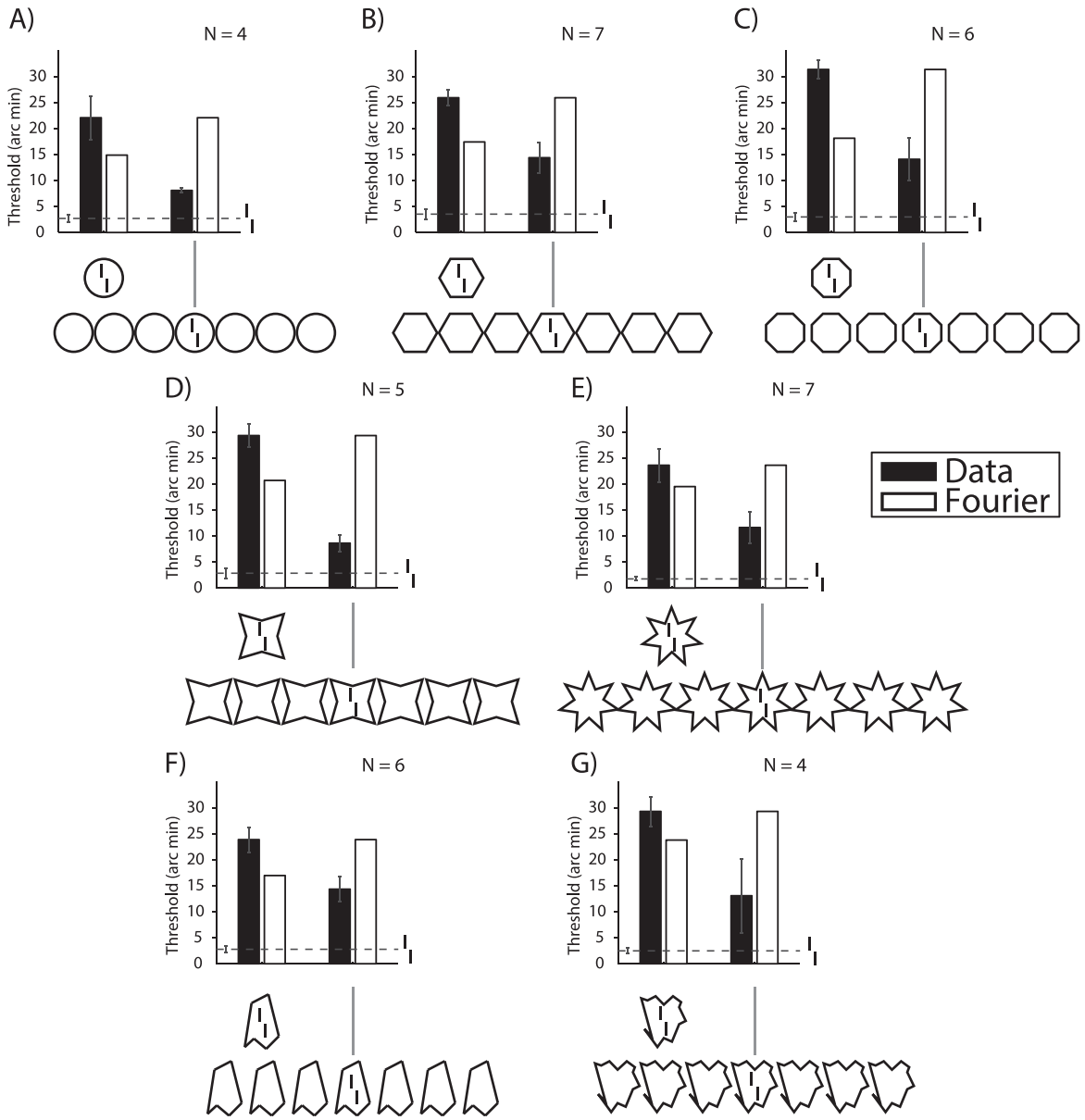

Figure 4. Uncrowding with several different kinds of shapes. Dashed lines show the thresholds for the vernier-alone condition. Black bars show vernier offset discrimination thresholds for the human data. Higher thresholds indicate stronger crowding. When the vernier was embedded in single shapes, thresholds increased compared with the single-vernier condition. When adding three identical flanking shapes on either side, thresholds decreased compared with the single-shape conditions. White bars indicate thresholds computed under a Fourier model. The Fourier model shows the opposite result (i.e., thresholds increase when increasing the number of shapes).

Method

We determined vernier offset discrimination thresholds with flanking patterns of squares and seven-pointed stars (Figure 6) and irregular shapes (Figure 7). Six observers (two females, four males) participated in Experiment 3A (Figure 6). Two of them performed with the 85% size, and two performed with the 75% size. Observers were presented with (a) a square, (b) an array of seven squares, (c) seven alternating squares and stars, (d) three rows of alternating squares and stars on a 3×7 grid, (e) a checkerboard of squares and stars on a 3×7 grid, (f) squares and stars arranged in an irregular fashion on a 3×7 grid, (g) condition d without the upper and lower central squares, and (h) seven alternating squares and stars with upper and lower central squares. Spacing between each shape was 2.2°. In addition, we applied our Fourier model to the stimuli.

Seven observers (three females, four males) participated in Experiment 3B (Figure 7). Four of them were presented with the 132% size. Instead of squares and

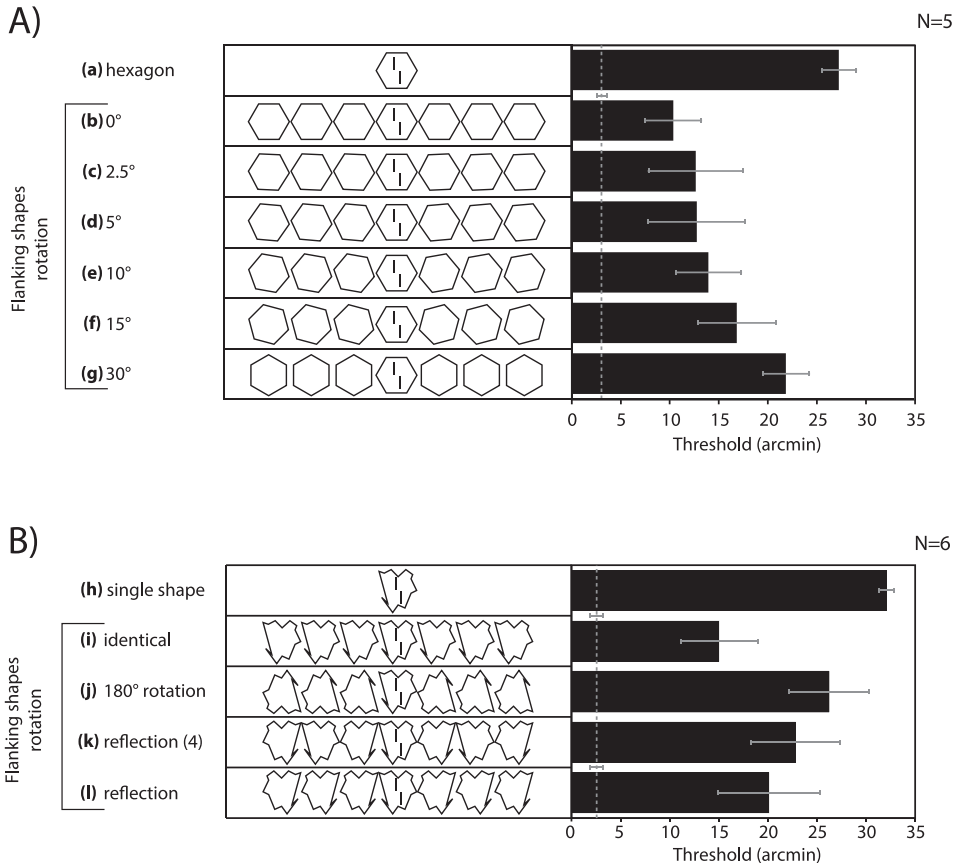

Figure 5. Uncrowding and shape similarity. Vertical white dashed lines show the thresholds for the vernier-alone condition. Experiment A: Thresholds increased compared with the vernier-alone condition (a). When the central hexagon was flanked by six identical hexagons, thresholds decreased compared with the single-hexagon condition (a, b). When the flanking hexagons were rotated by 2.5°, 5°, 10°, and 15°, thresholds gradually increased (c through g). Experiment B: When the vernier was embedded in an irregular shape, thresholds increased compared with the vernier-alone condition (h). When the central irregular shape was flanked by six identical irregular shapes, thresholds decreased compared with the single-shape condition (h, i). When the flanking shapes were rotated by 180°, thresholds increased compared with the previous condition (i, j). When the flanking shapes were all mirrored, thresholds increased slightly compared with the seven identical shapes condition (k, l).

stars, we presented the second irregular shape in Figure 4g (as a square) and the same shape rotated by 180° (as a star). Observers were presented with the same shape configurations as in the first experiment. Spacing between each shape was 2.9°.

Results

We found a significant main effect of flanker configuration on discrimination thresholds in both experiments; Figure 6: $F(8, 40) = 22.93$, $p < 0.001$; Figure 7: $F(8, 48) = 8.95$, $p < 0.001$. Tukey's post hoc tests were used for pairwise comparisons. As in Figure 1a, when the vernier was embedded in the square, thresholds increased compared with the vernier-alone condition (Figure 6a; $p < 0.05$). When the vernier was flanked by three additional squares on each side, thresholds decreased compared with the previous condition (Figure 6b; $p < 0.05$). When the vernier was embedded in an array of seven alternating squares and stars, thresholds were as high as in the single-shape condition (Figure 6a and c). When the vernier was embedded in three rows of alternating squares and stars, thresholds strongly decreased compared with the previous condition (Figure 6c and d; $p < 0.05$). When the vernier was embedded in other configurations of squares and stars, thresholds remained as high as in the single-shape condition (Figure 6a and e through h).

The Fourier model predictions (Figure 6, white bars) strongly differ compared with the human data (Figure 6, black bars). Human thresholds show uncrowding in Figure 6b and d, whereas the model thresholds increase in all conditions.

In Figure 6, when the vernier was embedded in an irregular shape, thresholds increased compared with the vernier-alone condition (Figure 7a). In all other conditions, thresholds decreased compared with the single-shape condition (Figure 7a vs. b through h; $p <$

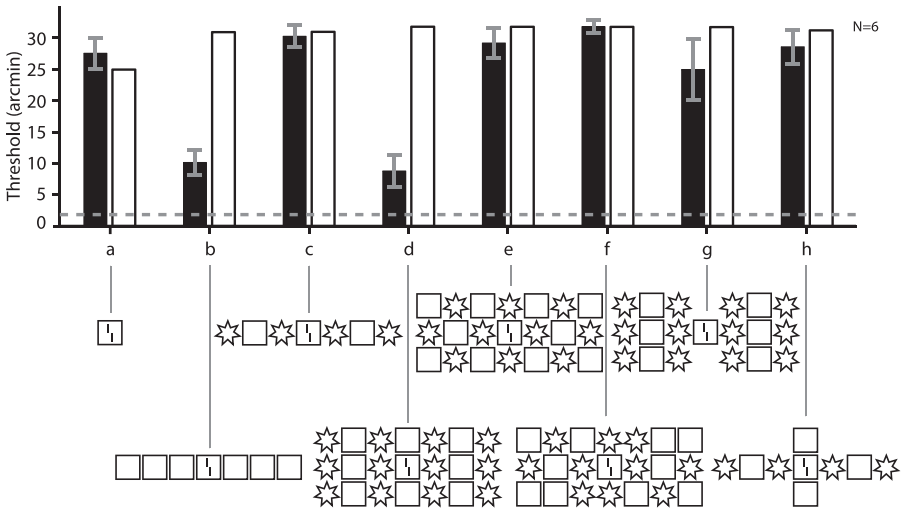

Figure 6. Patterns of squares and stars. The dashed line shows the vernier-alone condition. When the vernier was embedded in a square, thresholds increased compared with the vernier-alone condition (a). When the square was flanked by three squares on each side, thresholds decreased compared with the single-shape condition (a, b). When the central square was embedded in an array of alternating squares and seven-pointed stars, thresholds were as high as in the single-shape condition (a, c). When an identical array of alternated shapes was added on the top and bottom, thresholds decreased compared with the previous condition (c, d). In all the other conditions, thresholds were as high as in the single-shape condition (a and e through h). White bars indicate thresholds computed by the Fourier model. Thresholds increased when the number of flanking shapes increased.

0.05; comparisons a vs. d, a vs. g, and a vs. h were not significantly different, probably because of the small sample size).

Discussion

As we showed in the second set of experiments, small changes in shape orientation can strongly determine crowding strength (Figure 5). In line with this notion, uncrowding occurred with seven identical squares (Figure 6a and b) and vanished with alternating squares and stars (Figure 6c). How can we explain uncrowding in Figure 6d? The central row of alternating squares and stars was identical to the condition in Figure 6c; however, crowding strength strongly differed.

We propose that a *regular* pattern of dissimilar shapes led to the uncrowding of the vernier. As a control, we checked whether uncrowding is due to the

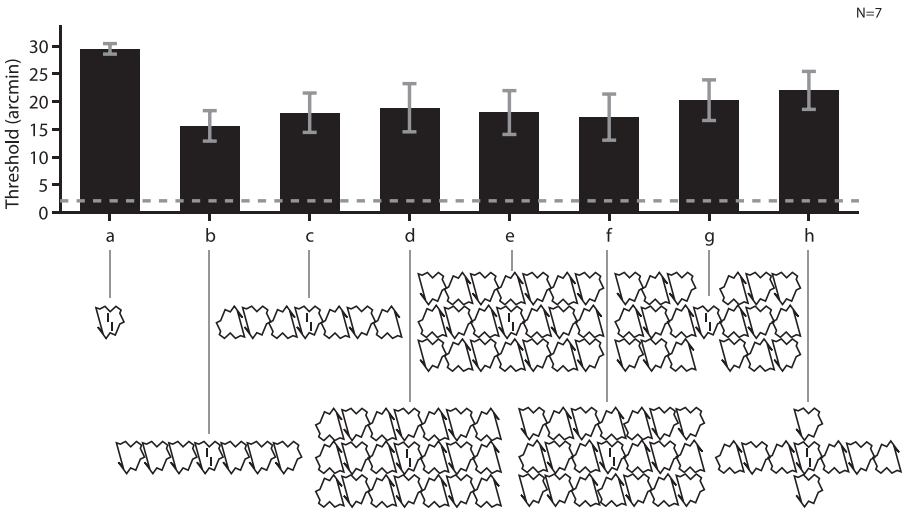

Figure 7. Patterns of irregular shapes. The dashed line shows the vernier-alone condition. When the vernier was embedded in an irregular shape, thresholds increased compared with the vernier-alone condition (a). When the square was flanked by three irregular shapes on each side, thresholds decreased compared with the single-shape condition (a, b). Thresholds remained on the same level with all the other depicted stimulus configurations (c through j).

central or flanking shape columns (Figure 6g and h), but crowding remained strong in both conditions. Hence, the *global* shape configuration led to uncrowding of the vernier.

It should be mentioned that only few, special patterns lead to uncrowding. For example, when presenting a checkerboard of squares and stars (Figure 6e) or the same shapes arranged in an irregular fashion (Figure 6f), crowding remained strong.

Figure 7 further supports this hypothesis. Uncrowding always occurred with increasing numbers of flanking shapes despite their overall configuration (Figure 7a vs. b through h). Our results show that elements presented well outside Bouma's window (Bouma, 1970; Pelli, 2008; Pelli & Tillman, 2008; Rosen et al., 2014) can modulate crowding strength on a vernier. We propose that crowding strength on a single element can be determined only by taking all the other elements and their overall configuration into account.

General discussion

Crowding characteristics and models

Most theories of crowding propose that (a) crowding occurs only within a restricted region (namely Bouma's window), (b) adding flankers does not improve performance, and (c) crowding is feature specific (i.e., crowding occurs only between similar features such as color, orientation, and shape). In line with previous studies (Malania et al., 2007; Manassi et al., 2012, 2013; Manassi, Hermens, Francis, & Herzog, 2015; Põder, 2007; Saarela et al., 2010), we have shown here that none of these characteristics are crucial for crowding (for reviews see Herzog et al., 2015; Herzog & Manassi, 2015). Adding flankers outside Bouma's window can improve performance contrary to theories a and b. It is important to note that adding elements outside Bouma's window can also increase crowding (Manassi et al., 2012; Rosen & Pelli, 2015; Saarela et al., 2010; Vickery, Shim, Chakravarthi, Jiang, & Luedeman, 2009). Hence, neither Bouma's window nor the number or extent of flankers determines crowding. Crowding is not feature specific because, for example, strong uncrowding occurred with circles, which share very few low-level features with the vernier and, in particular, do not contain straight lines (Figure 4a). Hence, crowding and uncrowding are not restricted to simple interactions, such as line–line feature detector inhibition. One might argue that uncrowding occurs only with simple, familiar shapes. However, uncrowding also occurs with irregular and unfamiliar shapes (Figure 4f and g).

Our results support our previous conclusions that simple pooling and substitution models cannot explain crowding because adding elements should not improve performance (for in-depth reviews see Herzog et al., 2015; Herzog & Manassi, 2015). In general, we think that simple, low-level interactions cannot explain crowding and uncrowding. It seems that low-level vernier acuity is determined by the overall high-level spatial configurations of elements across large parts of the visual field. High-level processing determines low-level processing as much as the other way around. Hence, one needs to take into account the entire visual scene to predict fine-grained vernier acuity. This conclusion is particularly supported by Experiment 3A (Figure 6), where the configuration of all stimulus elements, distributed across large parts of the visual field, is crucial.

On the level of perceptual organization, we proposed that crowding can be best explained in terms of perceptual organization and grouping (Manassi et al., 2012, 2013, 2015). Crowding is strong only when the target groups with the flankers. When the target ungroups from the flankers, crowding is weak.

Even though grouping and perceptual organization seem to be crucial to explain crowding, we suggest that simple combinations of Gestalt rules are unable to explain our results (e.g., Kubovy & van den Berg, 2008). For example, uncrowding occurred when the row of squares and stars in Figure 6c was added on top of and below the central row (Figure 6d). It is unclear how simple Gestalt laws can account for these results. Importantly, the creation of three vertical squares is not sufficient for uncrowding because these three squares do not lead to uncrowding when embedded in different configurations (Figure 6h; see also Figure 1b). As a further example, the configuration in Figure 6e is much more symmetric than the one in Figure 6f, but performance is roughly the same. In general, why should uncrowding occur at all when more elements are added from a perspective of basic Gestalt rules?

Crowding and object representations

Our results provoke the question of at which level neural processing occurs and how objects are represented in the human brain. As mentioned, crowding and uncrowding seem to occur with all types of shapes in a more or less similar way. Experiments 2A and 2B (Figure 5) show that small changes in orientation lead to strong changes in performance. For example, turning the flanking hexagons by only 10° strongly increased crowding (Figure 5b vs. e). Hence, uncrowding depends not only on the repetition of the same shape but also on the shapes' exact orientation. For uncrowding it seems to be crucial that the human brain represents stimuli with great detail and on a level where position and orientation invariance are not yet reached. For models based on convergent coding (grandmother cell coding), our results

may imply that at each location of the visual field there are neurons coding for a hexagon of any given orientation. The same is true for other shapes, even irregular ones. Hence, there must be a large number of shape detectors represented at most locations in the visual field, as proposed by models of ultrafast object recognition (Crouzet, Kirchner, & Thorpe, 2010; Guyonneau, Kirchner, & Thorpe, 2006; Kirchner & Thorpe, 2006). In addition, not only are a large number of detectors necessary, but also the appropriate neural wiring allowing for the interactions leading to crowding and uncrowding. Whether such a scenario is possible from a combinatorial point of view remains to be seen.

However, it remains an open question whether all sorts of shapes are treated the same way by the visual system. For example, the exact spatial configuration seems to matter in Figure 6, where only one of the configurations leads to strong uncrowding (Figure 6d), while shuffling the elements of this configuration always leads to strong crowding (Figure 6f). However, this is not true for irregular shapes. All shuffled versions lead to uncrowding with little changes in performance (Figure 7). Hence, it may be that different shapes and combinations of shapes are treated differently by the human brain. For some of them, the exact orientation of the shapes may not matter. Hence, it may be important to study large sets of data to determine what level of detail is crucial for spatial processing for particular classes of shapes.

In general, it is surprising that small changes in orientation can lead to strong differences in performance, which implies that the brain represents these details with high precision. On the other hand, why does crowding occur at all when fine details are well represented?

Instead, on a level of explicit object and shape representation, our results may be explained at a midlevel, texture-related stage that picks up higher order structures (Balas et al., 2009; Freeman & Simoncelli, 2011). Texture models can operate on very different levels of representation, such as based on the statistics of orientations and other basic features (Julesz, 1981; Portilla & Simoncelli, 2000; Renninger & Malik, 2004). These models may be challenged by the fact that small changes matter, as in Experiments 2A and 2B (Figure 5). Hence, it may well be that crowding occurs at higher levels, such as the level of protoshapes.

Clearly, regularity seems to matter for crowding and uncrowding. However, there are many types of regularities, and it will not be an easy task to determine exactly what types of regularity matter. For example, the configuration in Figure 6d (regular sequence of triplets) leads to uncrowding, whereas the configuration in Figure 6e (regular checkerboard sequence) does not.

Fourier models of early vision are highly sensitive to regularities in the stimulus configuration. We have applied a very basic model to show that there are no obvious differences in the spectra between stimuli that lead to either crowding or uncrowding (Figures 4 and 6). In a previous publication, we performed an exhaustive search over the possible space of bandpass Fourier models for the stimuli shown in Figure 1 and other stimuli. We did not find a robust match between performance and model behavior using these stimuli (Clarke, Herzog, & Francis, 2014).

As a final option, it may be that there are special, emergent configurations (Pomerantz & Portillo, 2011) that lead to uncrowding and cannot be described by simple rules. That is why they are emergent (Pomerantz, Sager, & Stoever, 1977).

As shown here, crowding can be an effective tool for probing the nature of object representations, particularly for showing on which level(s) models need to operate. For such an enterprise, large sets of data are needed, which can be obtained only by large-scale studies. We would like to mention that the research area of regularity, texture, emergent configurations, and so on is rather under-investigated at the moment. In addition, most prior research has used subjective measures, such as pointing to textures and regularities in the image. Crowding offers the possibility of obtaining both objective performance measures (vernier acuity) and subjective measures about how elements group, whether there are subtextures in an ensemble of elements, or whether symmetries or regularities are subjectively visible. For example, in Experiment 3A we showed that the configuration in Figure 6d led to good performance but that the configuration in Figure 6e did not, indicating that the symmetries in the latter configuration did not play an important role in crowding. It would have been interesting to test how visible these regularities are subjectively and to correlate these ratings with crowding performance.

Keywords: crowding, grouping, object recognition, shape perception

Acknowledgments

We thank Marc Repnow for technical support. This work was supported by the Swiss National Science Foundation Project "Basics of visual processing: What crowds in crowding?" Mauro Manassi was supported in part by the Swiss National Science Foundation fellowship P2ELP3_158876.

*MM and SL contributed equally to this article.
Commercial relationships: none.
Corresponding author: Mauro Manassi.
Email: mauro.manassi@berkeley.edu.
Address: Laboratory of Psychophysics, Brain Mind Institute, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

References

Andriessen, J. (1976). Eccentric vision: Adverse interactions between line segments. *Vision Research, 16,* 71–78.

Bach, M. (1996). The Freiburg visual acuity test—Automatic measurement of visual acuity. *Optometry and Vision Science, 73,* 49–53.

Balas, B., Nakano, L., & Rosenholtz, R. (2009). A summary-statistic representation in peripheral vision explains visual crowding. *Journal of Vision, 9*(12):13, 1–18, doi:10.1167/9.12.13. [PubMed] [Article]

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature, 226,* 177–178.

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

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,* 1–12.

Crouzet, S. M., Kirchner, H., & Thorpe, S. J. (2010). Fast saccades toward faces: Face detection in just 100 ms. *Journal of Vision, 10*(4):16, 1–17, doi:10.1167/10.4.16. [PubMed] [Article]

Dakin, S., Cass, J., Greenwood, J., & Bex, P. (2010). Probabilistic, positional averaging predicts object-level crowding effects with letter-like stimuli. *Journal of Vision, 10*(10):14, 1–16, doi:10.1167/10.10.14. [PubMed] [Article]

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron, 73,* 415–434.

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

Freeman, J., Chakravarthi, R., & Pelli, D. G. (2012). Substitution and pooling in crowding. *Attention, Perception, & Psychophysics, 74,* 379–396.

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

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

Greenwood, J., Bex, P., & Dakin, S. (2010). Crowding changes appearance. *Current Biology, 20,* 496–501.

Guyonneau, R., Kirchner, H., & Thorpe, S. J. (2006). Animals roll around the clock: The rotation invariance of ultrarapid visual processing. *Journal of Vision, 6*(10):1, 1008–1017, doi:10.1167/6.10.1. [PubMed] [Article]

Hermens, F., Luksys, G., Gerstner, W., Herzog, M. H., & Ernst, U. (2008). Modeling spatial and temporal aspects of visual backward masking. *Psychological Review, 115,* 83–100.

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, 1–18, doi:10.1167/15.6.5. [PubMed] [Article]

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

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310,* 863–866.

Julesz, B. (1981). Textons, the elements of texture perception, and their interactions. *Nature, 290,* 91–97.

Kirchner, H., & Thorpe, S. J. (2006). Ultra-rapid object detection with saccadic eye movements: Visual processing speed revisited. *Vision Research, 46,* 1762–1776.

Kooi, F., Toet, A., Tripathy, S., & Levi, D. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision, 8,* 255–279.

Kubovy, M., & van den Berg, M. (2008). The whole is equal to the sum of its parts: A probabilistic model of grouping by proximity and similarity in regular patterns. *Psychological Review, 115,* 131–154.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A minireview. *Vision Research, 48,* 635–654.

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

Manassi, M., Hermens, F., Francis, G., & Herzog, M. H. (2015). Release of crowding by pattern completion. *Journal of Vision, 15*(8):16, 1–15, doi:10.1167/15.8.16. [PubMed] [Article]

Manassi, M., Sayim, B., & Herzog, M. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision, 12*(10):13, 1–14, doi:10.1167/12.10.13. [PubMed] [Article]

Manassi, M., Sayim, B., & Herzog, M. (2013). When crowding of crowding leads to uncrowding. *Journal*

of Vision, 13(13):10, 1–10, doi:10.1167/13.13.10. [PubMed] [Article]

Nazir, T. (1992). Effects of lateral masking and spatial precueing on gap-resolution in central and peripheral vision. *Vision Research, 32,* 771–777.

Parkes, L., Lund, J., Angelucci, A., Solomon, J., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience, 4,* 739–744.

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

Pelli, D., Palomares, M., & Majaj, N. (2004). Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision, 4*(12): 12, 1136–1169, doi:10.1167/4.12.12. [PubMed] [Article]

Pelli, D., & Tillman, K. (2008). The uncrowded window of object recognition. *Nature Neuroscience, 11,* 1129–1135.

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

Pomerantz, J. R., & Portillo, M. C. (2011). Grouping and emergent features in vision: Toward a theory of basic gestalts. *Journal of Experimental Psychology: Human Perception and Performance, 37,* 1331–1349.

Pomerantz, J. R., Sager, L. C., & Stoever, R. J. (1977). Perception of wholes and of their component parts: Some configural superiority effects. *Journal of Experimental Psychology: Human Perception and Performance, 3,* 422–435.

Portilla, J., & Simoncelli, E. P. (2000). A parametric texture model based on joint statistics of complex wavelet coefficients. *International Journal of Computer Vision, 40,* 49–70.

Prins, N., & Kingdom, F. A. A. (2009). Palamedes: Matlab routines for analyzing psychophysical data. Available at http://www.palamedestoolbox.org

Renninger, L. W., & Malik, J. (2004). When is scene identification just texture recognition? *Vision Research, 44,* 2301–2311.

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience, 2,* 1019–1025.

Rosen, S., Chakravarthi, R., & Pelli, D. G. (2014). The Bouma law of crowding, revised: Critical spacing is equal across parts, not objects. *Journal of Vision, 14*(6): 10, 1–15, doi:10.1167/14.6.10. [PubMed] [Article]

Rosen, S., & Pelli, D. G. (2015). Crowding by a repeating pattern. *Journal of Vision, 15*(6):10, 1–9, doi:10.1167/15.6.10. [PubMed] [Article]

Saarela, T., Westheimer, G., & Herzog, M. (2010). The effect of spacing regularity on visual crowding. *Journal of Vision, 10*(10):17, 1–7, doi:10.1167/10.10.17. [PubMed] [Article]

Sayim, B., Westheimer, G., & Herzog, M. H. (2010). Gestalt factors modulate basic spatial vision. *Psychological Science, 21,* 641–644.

Serre, T., Kouh, M., Cadieu, C., Knoblich, U., Kreiman, G., & Poggio, T. (2005). *A theory of object recognition: computations and circuits in the feedforward path of the ventral stream in primate visual cortex* (No. AI MEMO-2005-036). Cambrige, MA: MIT Center for Biological and Computational Learning.

Serre, T., Kreiman, G., Kouh, M., Cadieu, C., Knoblich, U., & Poggio, T. (2007). A quantitative theory of immediate visual recognition. *Progress in Brain Research, 165,* 33–56.

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, USA, 104,* 6424–6429.

Strasburger, H., & Wade, N. J. (2015). James Jurin (1684-1750): A pioneer of crowding research? *Journal of Vision, 15*(1):9, 1–7, doi:10.1167/15.1.9. [PubMed] [Article]

Thorpe, S., Delorme, A., & Van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks, 14,* 715–725.

van den Berg, R., Roerdink, J. B. T. M., & Cornelissen, F. W. (2010). A neurophysiologically plausible population code model for feature integration explains visual crowding. *PLoS Computational Biology, 6*(1), e1000646.

Vickery, T., Shim, W., Chakravarthi, R., Jiang, Y., & Luedeman, R. (2009). Supercrowding: Weakly masking a target expands the range of crowding. *Journal of Vision, 9*(2):12, 1–15, doi:10.1167/9.2.12. [PubMed] [Article]

Watson, A., & Pelli, D. (1983). QUEST: A Bayesian adaptive psychometric method. *Attention, Perception, & Psychophysics, 33,* 113–120.

Whitney, D., & Levi, D. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences, 15,* 160–168.

Wichmann, F. A., & Hill, N. J. (2001). The psychometric function: I. Fitting, sampling, and goodness of fit. *Perception & Psychophysics, 63,* 1293–1313.

Wilkinson, F., Wilson, H., & Ellemberg, D. (1997). Lateral interactions in peripherally viewed texture arrays. *Journal of the Optical Society of America A Optics, Image Science, and Vision, 14,* 2057–2068.