Crowding, grouping, and object recognition: A matter of appearance

4 authors, including:

Mauro Manassi
University of Aberdeen
19 PUBLICATIONS **315** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Crowding View project

Crowding, grouping, and object recognition: A matter of appearance

Michael H. Herzog

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

Bilge Sayim

Laboratory of Experimental Psychology, University of Leuven, Leuven, Belgium ✉ ⌂

Vitaly Chicherov

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

Mauro Manassi

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

In crowding, the perception of a target strongly deteriorates when neighboring elements are presented. Crowding is usually assumed to have the following characteristics. (a) Crowding is determined only by nearby elements within a restricted region around the target (Bouma's law). (b) Increasing the number of flankers can only deteriorate performance. (c) Target-flanker interference is feature-specific. These characteristics are usually explained by pooling models, which are well in the spirit of classic models of object recognition. In this review, we summarize recent findings showing that crowding is not determined by the above characteristics, thus, challenging most models of crowding. We propose that the spatial configuration across the entire visual field determines crowding. Only when one understands how all elements of a visual scene group with each other, can one determine crowding strength. We put forward the hypothesis that appearance (i.e., how stimuli look) is a good predictor for crowding, because both crowding and appearance reflect the *output* of recurrent processing rather than interactions during the initial phase of visual processing.

Introduction

What is crowding?

In crowding, perception of a target is compromised by neighboring elements. For example, identification of a peripherally presented letter can strongly deteriorate in the presence of flanking letters (Flom, Heath, & Takahashi, 1963; Bouma, 1970; Strasburger, Harvey, & Rentschler, 1991; Levi, 2008). Hence, crowding is a key issue in reading (Legge, 2007). For this reason, crowding research started off as reading research. In the meantime, crowding has become a tool to investigate object recognition in general.

Detection of the target *itself* is unaffected in crowding. In fact, it is very easy to tell whether or not the target is present. Crowding impairs the *identification* of target features (Pelli, Palomares, & Majaj, 2004). Crowding was first thought to occur only for simple stimuli, like lines, verniers, letters, and Gabors (Andriessen & Bouma, 1976; Levi, Klein, & Aitsebaomo, 1985; Levi, Hariharan, & Klein, 2002; Harp, Bressler, & Whitney, 2007). However, recent research has shown that crowding occurs also with more complex stimuli, like faces (Martelli, Majaj, & Pelli, 2005; Louie, Bressler, & Whitney, 2007; Farzin, Rivera, & Whitney, 2009), objects (Wallace & Tjan, 2011), and
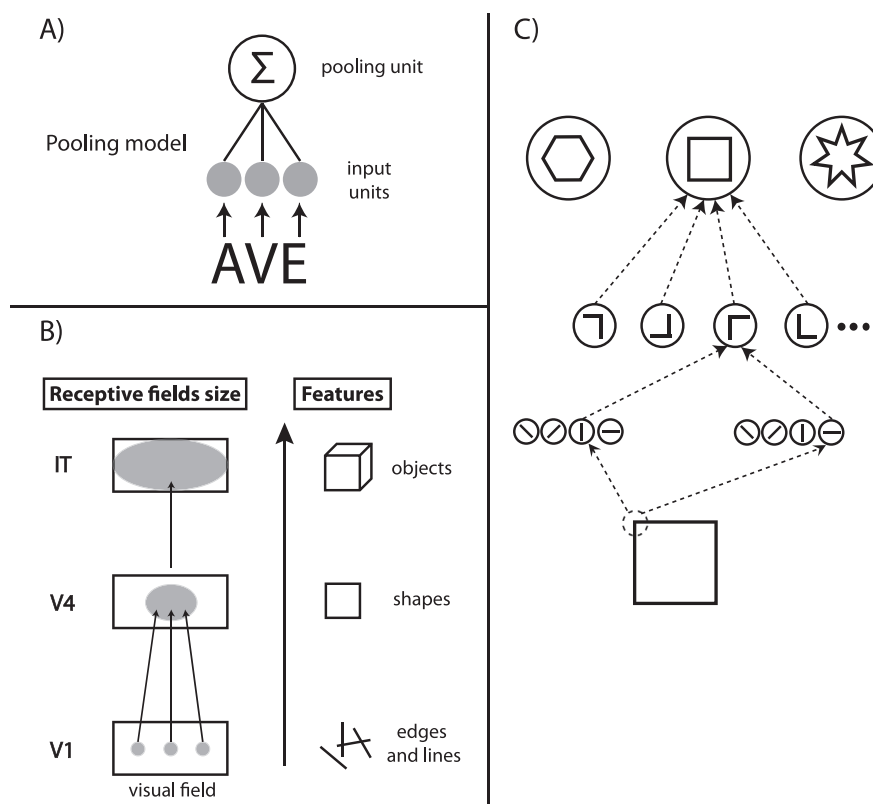
Figure 1. (A) Basic pooling model. Elements (e.g., letters A, V, and E) activate input units that subsequently feed into a pooling unit. Because of the larger receptive field of the pooling unit, the features of the letters are jumbled. (B) Neurophysiology. Neurons in V1 are sensitive to simple features such as edges and lines. In higher visual areas, neurons are sensitive to more and more complex features, such as simple shapes in V4 and objects in IT. Receptive field sizes increase from lower visual areas to higher visual areas. (C) Hierarchical models of object recognition formalize the neurophysiological findings (see, e.g., Riesenhuber & Poggio, 1999). Stimulus processing starts with the analysis of very simple features (edges and lines) and proceeds to more and more complex visual representations (shapes). A hypothetical "square neuron" receives input from neurons tuned to angles, which in turn receive inputs from basic line detectors. Receptive field sizes increase as they integrate more and more information across the visual field. At each step in the hierarchy, only signals from the previous areas are combined. Responses in higher areas are fully determined by the input from lower areas. Information lost on early stages is irretrievably lost.

biological motion (Ikeda, Watanabe, & Cavanagh, 2013).

Crowding is mainly studied in peripheral vision. A hallmark of crowding is Bouma's law, which states that flankers deteriorate performance only when presented within a restricted window around the target (Bouma, 1970; Pelli & Tillman, 2008; Rosen, Chakravarthi, & Pelli, 2014). The size of the window is about half the eccentricity of the target (correspondingly roughly to 6 mm on the primary visual cortex; Tripathy & Levi, 1994; Pelli, 2008; Pelli & Tillman, 2008). For this reason, in the last 40 years, crowding has been probed mainly by elements directly neighboring the target. Crowding shows characteristic anisotropies. When the flankers are arranged radially relative to the fixation point, crowding is much stronger compared to when the flankers are arranged tangentially (Toet & Levi, 1992). Crowding is asymmetric: Flankers presented away from the fovea crowd stronger than flankers closer to it (Bouma, 1973; Bex, Dakin, & Simmers, 2003).

Crowding is often thought to be feature specific, that is, crowding occurs only between similar features. For example, crowding is strong when target and flankers have the same color. Crowding strongly reduces when colors are different (Kooi, Toet, Tripathy, & Levi, 1994; Põder, 2007; Scolari, Kohnen, Barton, & Awh, 2007). Similar results were found for orientation (Andriessen & Bouma, 1976), spatial frequency (Chung, Levi, & Legge, 2001), shape (Nazir, 1992; Kooi, Toet, Tripathy, & Levi, 1994), and faces (Louie et al., 2007; Farzin et al., 2009).

Models of crowding

The most popular models of crowding are pooling models, where elements are first analyzed by neurons

with small receptive fields (Figure 1A; Wilkinson, Wilson, & Ellemberg, 1997; Pelli et al., 2004; Greenwood, Bex, & Dakin, 2009, 2010; Van Den Berg, Roerdink, & Cornelissen, 2010; Zahabi & Arguin, 2014). These neurons project then to neurons on the next level of the visual hierarchy with larger receptive fields, pooling information from the low-level neurons. Because of pooling, features of nearby elements are jumbled and target identification deteriorates (Figure 1A). Since receptive field sizes increase with eccentricity, Bouma's window increases as well. In other pooling models, target and flanker signals are averaged (Parkes et al., 2001) or merged in textural representations by summary statistics (Balas et al., 2009; Freeman & Simoncelli, 2011). All these models have in common that they are based on *local* interactions within Bouma's window. Pooling was proposed to occur as early as V1/V2 (Pelli, 2008; Freeman & Simoncelli, 2011) but higher areas, such as in V4, were proposed too (Levi, 2008; Liu, Jiang, Sun, & He, 2009).

Pooling models were made in the spirit of classic hierarchical models of object recognition (see Figure 1B, C; Riesenhuber & Poggio, 1999; Thorpe, Delorme, & Van Rullen, 2001; Hung, Kreiman, Poggio, & DiCarlo, 2005; Serre, Oliva, & Poggio, 2007; Serre, Wolf, Bileschi, Riesenhuber, & Poggio, 2007; DiCarlo, Zoccolan, & Rust, 2012). For example, activity of a hypothetical "square neuron" is fully determined by the input from neurons tuned to the constituent vertical and horizontal lines of the square. Because information processing proceeds from low-level features to complex figures, the input of higher level feature detectors is fully determined by the outputs of basic feature detectors. Information lost on early stages is irretrievably lost (Freeman & Simoncelli, 2011). For this reason, crowding is often viewed as a (low-level) "bottleneck" of vision (Levi, 2008) or a "fundamental limit for object recognition" (Whitney & Levi, 2011).

In substitution models, crowding occurs because features of the flankers are confused with features of the target (Krumhansl & Thomas, 1977; Strasburger et al., 1991; Huckauf & Heller, 2002; Zhang, Zhang, Liu, & Yu, 2012; Ester, Klee, & Awh, 2014; Ester, Zilber, & Serences, 2015). In models of limited attentional resolution, target and flankers are accurately processed by neurons in (early) visual areas. However, the flankers hinder attention's access to the target features (He, Cavanagh, & Intriligator, 1996; Intriligator & Cavanagh, 2001). Attentional read-out windows are affected by flankers similarly to neurons pooling information in feedforward processing (He et al., 1996; Intriligator & Cavanagh, 2001). Recently, eye movements were shown to modify crowding (Harrison, Mattingley, & Remington, 2013; Harrison, Retell, Remington, & Mattingley, 2013; van Koningsbruggen & Buonocore, 2013; Harrison & Bex, 2014; Wolfe &

Whitney, 2014; see also Kowler, Anderson, Dosher, & Blaser, 1995; Deubel & Schneider, 1996). Nandy and Tjan (2012) proposed a model in which image displacements during saccades yield an inappropriate wiring of lateral connections in V1, leading to crowding. This model predicts that crowding occurs only within Bouma's window and the radial–tangential and inner–outer asymmetries.

Here, we will summarize and review recent findings, which show that crowding is determined by configural factors that cannot easily be explained by most models of crowding.

Selective review of previous findings

Crowding, pooling, & Bouma's window

Crowding is traditionally characterized by target–flanker interactions, which are (a) deleterious (e.g., pooling and substitution), (b) locally confined (Bouma's window), and (c) feature-specific.

Adding flankers can decrease crowding

Already in 1979, Banks, Larson, and Prinzmetal showed that bigger *can* be better. They presented a target letter flanked by a single letter. When more flanking letters were added, target identification improved compared to the single flanking letter condition (see also Wolford & Chambers, 1983; Banks & White, 1984). These results were forgotten for more than 30 years. Recently, Malania, Herzog, and Westheimer (2007) and Manassi, Sayim, and Herzog (2012) showed when bigger is better. Adding flanking lines improved performance in a vernier discrimination task when the lines were shorter or longer than the vernier. However, there was no change in performance when lines were of the same size as the vernier (Malania et al., 2007; Manassi et al., 2012). We argued that increasing the number of short or long lines leads to increased flanker–flanker and diminished target–flanker grouping. Crowding decreases. For same length lines, grouping does not change with the number of flankers, hence, performance does not change either. Similar results were found for lines differing in color (Põder, 2006). Not only more, but also longer and bigger flankers can improve performance (fovea: Malania et al., 2007; periphery: Levi & Carney, 2009; Saarela, Sayim, Westheimer, & Herzog, 2009; Manassi et al., 2012). Hence, target–flanker interactions are not always deleterious.
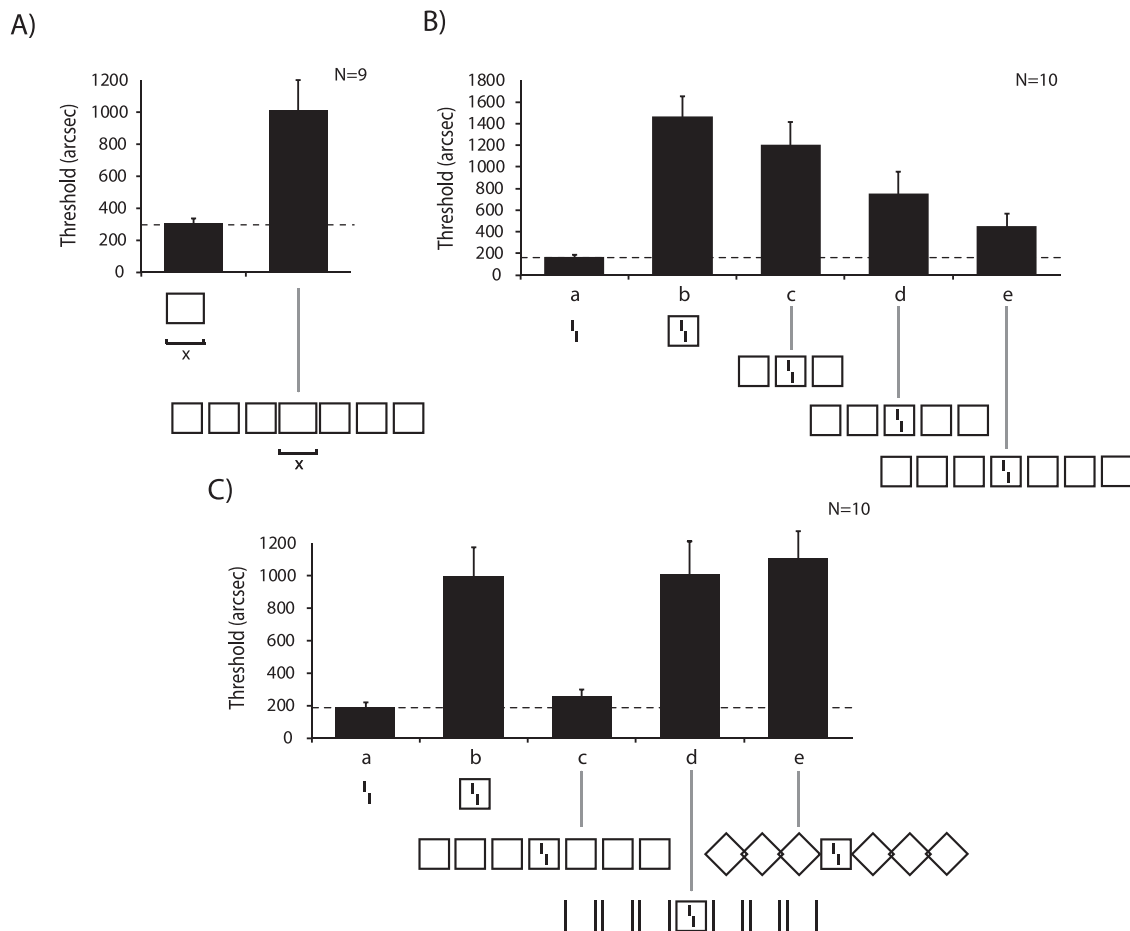
Figure 2. (A) Observers were asked to discriminate whether a rectangle was wider along the horizontal (*x*) or vertical (*y*) axis. We determined the threshold width for which 75% correct responses were obtained. When the rectangle was flanked by three squares on each side, performance strongly deteriorated compared to when presented alone. This is a classic crowding effect. (B) Next, we asked observers to discriminate whether a vernier was offset to the left or right (a). We determined the offset size for which 75% correct responses occurred (left bar and dashed line). Performance deteriorated (i.e., thresholds increased) when the vernier was surrounded by a square (b). This is another classic crowding effect. Surprisingly, vernier discrimination improved when we combined the two conditions. Performance improved gradually, with the more squares that were presented. Best performance occurred with 2 × 3 contextual squares. In this condition, the fixation dot is close to the leftmost square and the rightmost square is at 17.5° (i.e., well outside Bouma's window). (C) First, we repeated the basic conditions (a–c). Next, crowding was strong when we removed the horizontal lines of the flanking squares (d) or rotated the flanking squares by 90° (e). Data from (d) and (e) were collected in different experiments with different observers and are shown here together to ease presentation. In part (B), we adjusted square size individually to enhance effects. This explains the higher thresholds compared to (C). Modified from Manassi et al. (2013).

Elements outside Bouma's window can modulate crowding

At 9° of eccentricity, we presented a square and varied the width slightly (making it a rectangle). Observers indicated whether the width was greater or less than the height. When we added three squares on each side, performance strongly deteriorated (Figure 2A; Manassi, Sayim, & Herzog, 2013). This is a classic crowding effect. Next, we presented a vernier. When the vernier was surrounded by the outline of a square, performance strongly deteriorated (Figure 2B-b, C-b). This is another classic crowding effect. Next, we combined the two conditions. One might expect that

first, the central square strongly crowds the vernier. Then, the neighboring squares crowd the central square. Hence, crowding should become even stronger and performance should deteriorate further (super-crowding; Vickery, Shim, Chakravarthi, Jiang, & Luedeman, 2009). However, the opposite was the case. Crowding of crowding led to *un*crowding, that is, a release from crowding. Performance was almost at the same level as in the vernier alone condition (Figure 2B-e, C-c; Manassi et al., 2013). This experiment provides further evidence that more can be better. Most importantly, uncrowding increased smoothly with the number of squares, that is, performance

gradually improved as more squares were added (Figure 2B). The 2 × 3 squares to the right and left of the central square make up large parts of the right visual field. The seven squares range from 0.5° to 17.5° of eccentricity, whereas Bouma's window ranges only from 4.5° to 13.5°. Hence, our results show that elements outside Bouma's window can strongly decrease crowding (see also Manassi et al., 2012; Harrison & Bex, 2014; Sayim, Greenwood, & Cavanagh, 2014). Elements outside Bouma's window can also increase crowding (Vickery et al., 2009; Manassi et al., 2012; Chanceaux & Grainger, 2013; Rosen & Pelli, in press), and crowding can even occur when target and flankers are presented in opposite hemifields (see Harrison, Retell et al., 2013).

High-level processing determines low-level processing

The release from crowding in the multisquare conditions cannot be explained by the vertical lines making up the contextual squares because when we omitted their horizontal lines, crowding was as strong as in the single square condition (Figure 2C-d). We propose that the human brain first computes the shapes of the squares from their constituent lines. Then, square–square interactions determine processing of the vernier. Hence, high-level processing determines low-level processing in the sense that wholes can determine perception of and performance on their parts. The following experiment supports this notion further. When we rotated the contextual squares by 90°, creating diamonds, crowding was strong because the contextual diamonds single out the central square, which thus ungroups from the diamonds (Figure 2C-e). Hence, crowding is not feature specific. Particularly, high-level features can interfere with low-level ones. Importantly, our claim is independent of the neural basis in the sense that we do not propose that the square shapes are computed in higher cortical areas and the vernier, for example, in V1. Our proposition is on a truly phenomenological level about wholes and their constituent elements.

Grouping cues

Clearly simple models of crowding cannot explain our results. On a phenomenological level, we proposed that crowding occurs only when target and flankers group. When the target ungroups from the flanker configurations, crowding is weak. Hence, to understand crowding, we need to understand grouping (Malania et al., 2007; Manassi et al., 2013). Here, we show that crowding and the release from crowding can depend on many grouping cues, including target–flanker similarity on various levels (low level: Malania et al., 2007;

Manassi et al., 2012; high level: Manassi et al., 2013), good Gestalt (Sayim, Westheimer, & Herzog, 2010; Manassi et al., 2012), regularity (Saarela, Westheimer, & Herzog, 2010; Manassi et al., 2012), and contour integration (Livne & Sagi, 2007, 2010; Chakravarthi & Pelli, 2011).
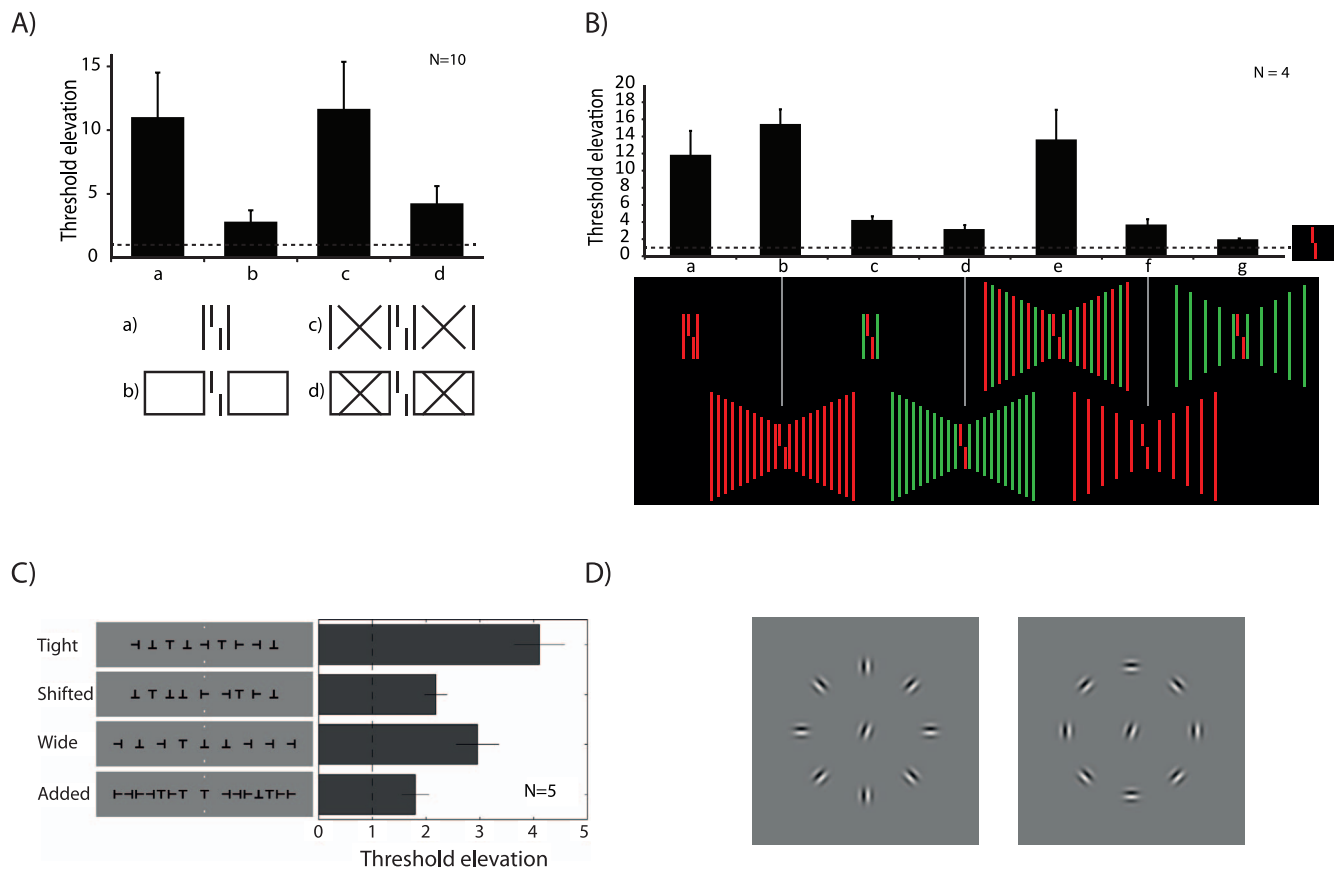
Good Gestalt

We presented a vernier flanked by one equal-length line on each side (Figure 3A-a). Performance strongly deteriorated compared to when the vernier was presented without flankers (Figure 3A-a; Sayim et al., 2010; Manassi et al., 2012). When the lines became part of a rectangle, performance improved (Figure 3A-b). When we crossed the horizontal lines, performance strongly deteriorated (Figure 3A-c). Performance improved again when we added horizontal lines, creating rectangles with crosses within (Figure 3A-d). We propose that when the lines become part of a rectangle, the entire rectangle ungroups from the vernier and, hence, performance improves (Manassi et al., 2012).

Regularity

We presented a red vernier flanked by one or 10 green line(s) on each side (Figure 3B-c, d; Manassi et al., 2012). Crowding was weak regardless of the number of flankers because of color dissimilarity (see also Kooi et al., 1994). When, however, every second line was red, creating a pattern of alternating red–green lines, crowding was as strong as when only red lines were presented (Figure 3B-e). The increase in crowding cannot be explained by the red lines in the alternating pattern themselves because crowding was weak when the green lines were absent (Figure 3B-f). Similar results were also found with letters (Rosen & Pelli, in press).

As another example of regularity, we presented a target letter T in one of the four cardinal orientations, flanked by four Ts on each side (Figure 3C; Saarela et al., 2010). Performance deteriorated compared to the condition in which the target T was presented without flankers, that is, a classic crowding effect (Figure 3C, *tight*). When we increased the spacing between the target and the two directly flanking Ts, performance improved (Figure 3C, *shifted*). Next, we increased the interflanker spacing between the remaining Ts. All letters were now equally spaced. Performance decreased, even though all letters are further away from the target T than in the second condition (Figure 3C, *wide* vs. *shifted*). Adding further Ts between the flankers improved performance (Figure 3C, *added*). Very similar results were also found for Gabors (Saarela et al., 2010). We suggest that when the directly neighboring flankers are moved away from the target,

Figure 3. Crowding and uncrowding depend on many grouping cues (for demonstrations see Figure 5). (A) Good Gestalt. A vernier flanked by two lines of the same length yields high thresholds, that is, strong crowding (a). When the two lines are integrated in a rectangle, thresholds strongly decrease (b). Crossing the horizontal lines of the rectangle increases thresholds, similar to the single lines condition (c). Closing the rectangle by additional horizontal lines reduces crowding again (foveal vision: Sayim et al., 2010; peripheral vision: Manassi et al., 2012). The dashed line indicates performance for the unflanked vernier. (B) Pattern regularity. Thresholds for a red vernier flanked by single red lines (a) and 10 red lines (b) are high compared to the unflanked vernier condition (shown by the dashed line). When the flankers are green (c–d), thresholds are much lower. A grating with alternating red and green lines leads to high thresholds (e). The red (f) and green (g) parts of the alternating grating themselves crowd very little. Only when parts of the alternating grating are *combined*, do they form a pattern that leads to strong crowding (adapted from Manassi et al., 2012). (C) Spacing regularity. Observers discriminated the orientation of a central letter T. Threshold elevation is high when the spacing between flanking letters is small and regular (*tight* condition). Increasing the spacing between the target and the innermost flankers decreased crowding (*shifted* condition). Crowding increased when we increased the spacing between the remaining flankers creating a regular pattern (*wide* condition). Adding more flankers in the gaps between the flankers (*added* condition) decreased crowding again (modified from Saarela et al., 2010). (D) Contour integration. Gabor orientation discrimination is weak when the central Gabor is surrounded by radially arranged flankers. When the flankers make up a smooth contour, crowding is reduced (adapted from Livne & Sagi, 2007).

crowding reduces simply because of distance, in accordance with most findings in crowding (*shifted* condition). When all elements are moved further away, a regular grating is created and strong target–flanker grouping occurs again. Crowding increases (*wide* condition). Again, the sheer number of flankers is not predictive for crowding. Also the distance of the flankers to the target is not predictive, which can be seen as a violation of Bouma's law. When additional flankers are added (*added* condition), the flankers now group with each other more strongly because of proximity.

Contour integration

Orientation discrimination of a target Gabor was better when Gabor flankers were arranged in a smooth contour surrounding the target, compared to when flankers did not make up a contour (Figure 3D; Livne & Sagi, 2007; see also Chakravarthi & Pelli, 2011; Yeotikar, Khuu, Asper, & Suttle, 2011).
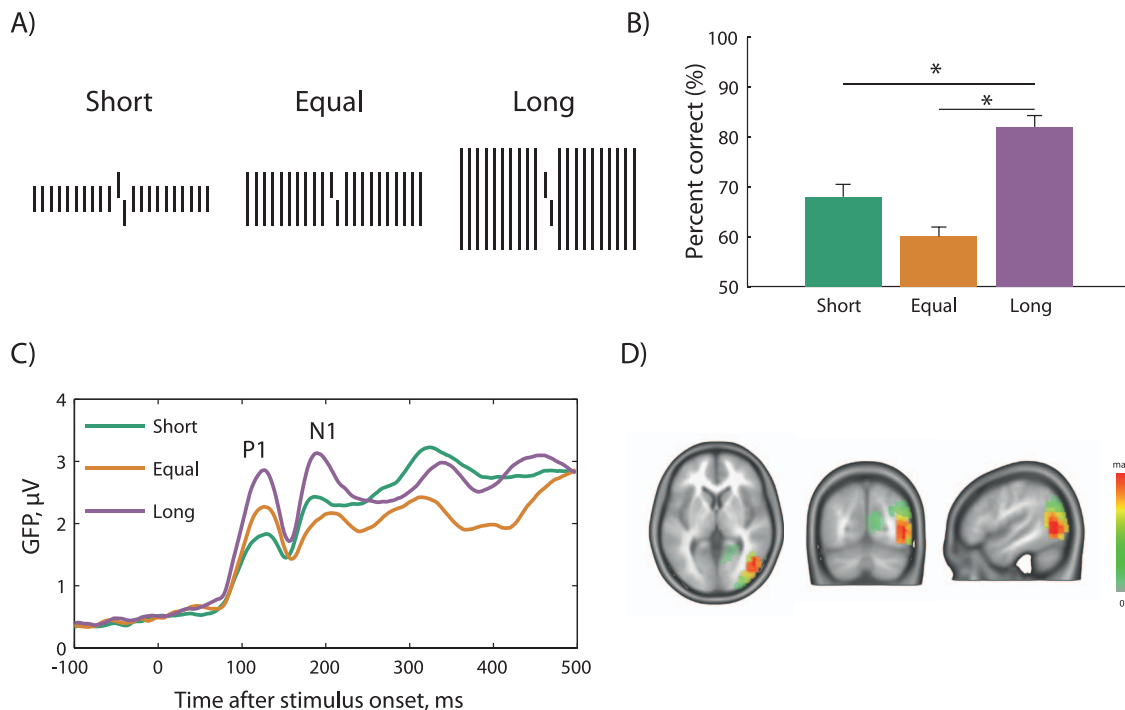
Figure 4. Electrophysiological correlates of crowding. (A) A vernier target was presented in the fovea and flanked by arrays of short, equal length, or long lines. (B) Accuracy was highest for long, intermediate for short, and worst for equal length flankers in line with our grouping hypothesis. (C) Event-related potentials were recorded and global field power (GFP) computed, which reflects overall brain activity. The time axis is referenced to stimulus onset. The early visual response (the P1 component) reflects flanker length. P1 amplitudes are highest for long flankers, intermediate for equal length, and lowest for short flankers. Crowding strength is (inversely) reflected in the N1 component around 180–200 ms, which is highest for long, intermediate for short, and lowest for equal length flankers. Hence, it seems that it takes about 50–80 ms to transform the initial encoding into an object-based perceptual code. (D) Source localization in the N1 time window. The color scale reflects activation differences in the brain associated with crowding (difference between the long flanker and equal length flanker conditions). Particularly, sources in the lateral occipital and posterior temporal and parietal areas reflect crowding strength. Sources in the V1 do not contribute significantly. Modified from Chicherov et al. (2014).

The dynamics of crowding: Electrophysiological correlates

What are the neural correlates of crowding? Based on the traditional characteristics of crowding, most EEG and fMRI studies have targeted low-level interactions and found that flankers suppress target-related brain activity and that suppression decreases with target–flanker separation (Chen et al., 2014; Millin et al., 2014). The strongest effects of crowding occurred in V1 and V2 in accordance with the idea that crowding occurs at the earliest stages of vision (Chen et al., 2014; Millin et al., 2014). Likewise, suppression is larger for radial than tangential flankers (Chen et al., 2014; Kwon, Bao, Millin, & Tjan, 2014).

We used stimuli suitable to understand the role of grouping in crowding and recorded high-density EEG determining global field power (GFP; Figure 4; Chicherov, Plomp, & Herzog, 2014). GFP is the standard deviation of the potentials across all electrodes and thus is a measure of global brain activity. As

in Malania et al. (2007), a vernier was flanked by arrays of lines, which were shorter, of equal length, or longer than the vernier. Performance was worst for the equal length lines and best for the longer lines because, as we argued, grouping is strongest for equal length flankers and weaker for shorter and longer flankers. The P1 component of the GFP correlated mainly with stimulus size or, likewise, the overall amount of light, that is, the P1 was highest for the long, medium for equal length, and lowest for the shorter flankers. In the N1 component, the shorter lines led to higher GFP amplitudes than the equal length flankers because, as we propose, the N1 reflects the perceptual organization of the entire stimulus configuration. For this reason, highest GFP amplitudes occurred for the most clearly segregated configuration with the long flankers, intermediate amplitudes for short flankers, and lowest for the equal length condition where all elements group. N1 amplitudes and performance correlated highly. In control experiments, we showed that GFP does not simply reflect the performance level as such, but truly

the spatial configuration (Chicherov et al., 2014). Next, we localized the cortical sources that correlated significantly with crowding and found that only high-level visual areas (lateral occipital cortex, and neighboring temporo-parietal cortices) reflected crowding, whereas activities in V1 were similar in crowding and uncrowding conditions (Figure 4D). Hence, it seems that the brain first encodes the stimulus based on its low-level features and then converts this code to a code based on perceptual grouping, which correlates well with performance and, as we would like to argue, appearance.

In this line Anderson, Dakin, Schwarzkopf, Rees, and Greenwood (2012) found that BOLD responses reflect the appearance of crowded stimuli. As in our study, high-level visual areas (in this case, V4) reflected crowding much better than the early visual areas V1 and V2. Likewise, Freeman, Donner, and Heeger (2011) showed that BOLD suppression in crowding is strongest in V4, intermediate in V3 and V2, and absent or weak in V1. In Joo, Boynton, and Murray (2012), the BOLD suppression was more robust (less variable) in V3 than in V2, and more robust in V2 than in V1, although magnitudes of the suppression were similar in the three areas. Overall, there is accumulating evidence that high-level visual areas reflect crowding better than low-level areas in studies where complex grouping and appearance determine crowding strength.

It seems that active target processing or attention to the target are important in crowding, that is, processing does not occur fully automatically. As mentioned, when observers discriminated the vernier in the equal length condition, the N1 component was strongly suppressed (Chicherov et al., 2014). However, when the task was to discriminate the length of flankers, suppression was much weaker (Chicherov et al., 2014). Likewise, Chen et al. (2014) found that there was little or no suppression when there was no attention to the stimuli (however, see Millin et al., 2014).

Discussion

Characteristics of crowding

Crowding is traditionally characterized by target-flanker interactions, which are (a) deleterious, (b) locally confined (Bouma's window), and (c) feature-specific. Most research has accordingly presented single flankers similar to the target, in its direct neighborhood (Pelli et al., 2004; Pelli & Tillman, 2008).

However, as reviewed here, (a) adding flankers does not always deteriorate performance: "bigger *can* be better" (Banks, Larson, & Prinzmetal, 1979; Wolford & Chambers, 1983; Põder, 2006; Levi & Carney, 2009;

Manassi et al., 2012). (b) Flankers well outside Bouma's region can increase (Vickery et al., 2009; Manassi et al., 2012; Chanceaux & Grainger, 2013; Harrison, Retell, et al., 2013; Rosen & Pelli, in press), but also decrease crowding strength (Malania et al., 2007; Manassi et al., 2012, 2013; Harrison & Bex, 2014; Sayim, Greenwood, & Cavanagh, 2014). Hence, crowding is not restricted to local interactions. It seems that crowding strength depends on the spatial config-uration of the entire stimulus, that is, on all elements in the visual field (or at least large parts of it). Since crowding can be modulated by elements far outside Bouma's region, the link between crowding, receptive fields size and cortical magnification factor needs to be rethought (Pelli, 2008; Pelli & Tillman, 2008). In addition, attention and task setting can change crowding strength, arguing against a *fixed* window of interaction (Huckauf, 2007; Yeshurun & Rashal, 2010; Whitney & Levi, 2011). (c) Crowding and uncrowding are not restricted to target-flankers interactions on the same level (same color, orientation, faces, etc.). For example, the very same vertical flankers in Figure 3A lose their crowding power when becoming part of rectangles, that is, good Gestalts. High-level feature processing interferes with low-level feature processing, in the sense that the whole determines performance on its parts as much as the other way around.

Bottlenecks

It was proposed that crowding is a bottleneck of low-level vision in the sense that, for example, the spatial relationships between nearby lines are "lost" when V1 signals are pooled by higher level neurons (Levi, 2008; Pelli, 2008; Pelli & Tillman, 2008). This is not true for three reasons. First, adding flankers should "cork" the bottleneck but, as mentioned, more flankers can even reduce crowding. Second, many features, including low and high level ones, "survive" crowding (He et al., 1996; Faivre & Kouider, 2011; Fischer & Whitney, 2011; Kouider, Berthet, & Faivre, 2011; Yeh, He, & Cavanagh, 2012; see also Whitney & Levi, 2011). Third and most importantly, the visual system can have great spatial resolution in certain but not other crowding situations, even though the low level features are identical. For example, the very same flanking lines in Figure 3A-a that exert strong crowding when presented alone "lose" their crowding power when they are parts of rectangles (Figure 3A-b, d). The same is true for the central square in Figure 2.

Here, the question arises: Why does performance deteriorate at all? Why is the visual system giving up excellent spatial resolution in some but not other conditions? Our working hypothesis is that crowding is not a bottleneck in the sense that information is lost

when passing information from one processing stage to another (Herzog & Manassi, 2015). We propose rather that elements are invisible because the brain renders wholes visible at the expense of the visibility of single elements (see Recurrent processing and appearance section; Herzog, Hermens, & Öğmen, 2014).

Grouping

We propose that only when the target groups with the flankers is crowding strong. Subjective ratings about "target–flanker standing out," a measure of grouping, showed good correlations with crowding strength (Wolford & Chambers, 1983; Malania et al., 2007; Saarela et al., 2009; Manassi et al., 2012). Similarly, pop out always leads to uncrowding; however, the opposite is not true (Sayim, Westheimer, & Herzog, 2011; but see Felisberti, Solomon, & Morgan, 2005). An element may not pop out from distractors, still, crowding may be weak because distractors are remote. Along the same lines, Dakin, Greenwood, Carlson, and Bex (2011) showed that the apparent position of flankers, not the physical one, determines crowding strength (see also Maus, Fischer, & Whitney, 2011; Wallis & Bex, 2011). Hence, only when one knows how the elements of a visual scene group, can one determine crowding strength. In this sense, traditional crowding research seems to have studied grouping under impoverished conditions by using single flanker conditions and manipulating only basic cues such as target-flanker proximity and similarity.

Importantly, grouping is necessary but not sufficient for crowding. For example, three remote red lines may group with each other but not necessarily crowd each other because grouping can operate over larger spatial scales than crowding (see also Sayim & Cavanagh, 2013). Tannazzo, Kurylo, and Bukhari (2014) showed that grouping by basic Gestalt laws occurs up to 40° of eccentricity. As a final point, flankers with high luminance or contrast can strongly deteriorate performance even though they do not group with a target of a lower luminance or contrast (Chung et al., 2001; Felisberti et al., 2005; Rashal & Yeshurun, 2014). We would like to argue that in this case masking rather than crowding mechanisms are in operation since the visibility of the target itself is compromised, rather than discrimination of its features.

Many grouping cues are involved in crowding (for demonstrations see Figure 5). For example, crowding and uncrowding depend on low-level color or length (dis)similarities (Kooi et al., 1994; Malania et al., 2007), figural (dis)similarity (Manassi et al., 2013), on three-dimensional cues (Sayim, Westheimer, & Herzog, 2008), spacing regularity (Saarela et al., 2010), contour

grouping (Livne & Sagi, 2007; Chakravarthi & Pelli, 2011), good Gestalt (Figure 3A; Sayim et al., 2010; Manassi et al., 2012), and higher order regularities (Figure 3B; Sayim et al., 2008; Manassi et al., 2012; Rosen & Pelli, in press). Bouma's law can be reinterpreted as grouping by proximity.

How do all these factors depend on each other? A century ago, the Gestaltists proposed a program in which the combination of basic Gestalt rules, such as proximity and similarity, leads to the ultimate factor of good Gestalt (Wertheimer, 1922, 1923). The program has largely failed and Gestalt research has fallen into a state of hibernation, mainly because, first, research was based on subjective aspects leading to self-referentiality. Second, the basic Gestalt rules explain perception only under very restricted conditions, and it remains to be shown how they combine. The same seems to be true in the studies reviewed here. Single Gestalt rules *per se* or simple combinations of them cannot explain crowding (see Rosen, Chakravarthi, & Pelli, 2011 for an attempt to explain crowding by Gestalt laws). For example, regularity seems to trump proximity (Bouma's law) in the letter crowding experiment in Figure 3C. The question remains, why? In addition, while proximity can easily be measured and defined this is not the case for the grouping cues of regularity and good Gestalt.

As a final point, grouping does by no means explain *why* performance deteriorates at all in crowding. Additional explanations are required.

Current models

Clearly, our results challenge most models of crowding. One reason is that models were made to capture the above characteristics of crowding, which are not as characteristic as previously thought.

First, basic pooling and limited attention models predict incorrectly that adding more flankers increases task irrelevant information and, hence, increases crowding strength. Substitution models predict that the more elements are presented, the more features can be confused. However, additional flankers can reduce crowding strength.

Second, more generally, we can rule out all crowding models that are in the spirit of hierarchical, feedforward processing as laid out in Figure 1 because (a) shape–shape interactions are crucial in crowding (Manassi et al., 2013). Since there are no feedback connections (feedforward processing), shape processing cannot influence vernier processing (hierarchical processing) and (b) vernier offset information is not irretrievably suppressed by the central square and, thus, "lost" during the first sweep of feedforward processing because vernier offset discrimination is good when many squares are presented (Figure
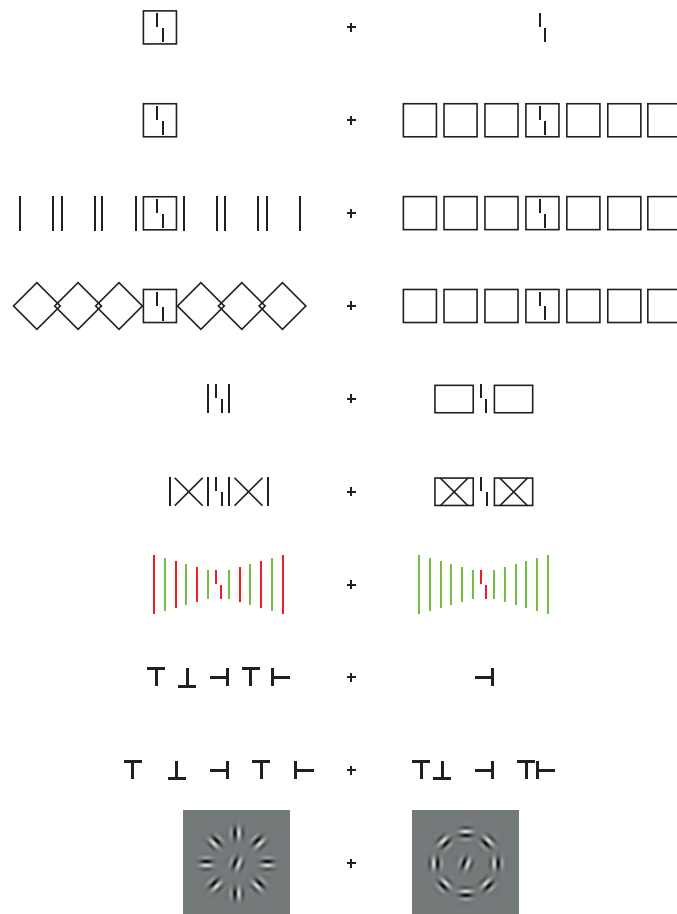
Figure 5. For illustrative purposes, we have plotted various stimuli for the studies. Fixate the central cross and compare stimuli on the right to those on the left hand side.

2B). Thus, one needs to give up one characteristic. One option is to give up the feedforward processing. The other option is that visual processing is not strictly hierarchical. For example in the reverse hierarchy model (Hochstein & Ahissar, 2002), when low-level features, such as lines, are attended, they may be represented on higher stages of vision, allowing for shape-line interactions. Likewise, shape may be processed at early stages (Altmann, Bülthoff, & Kourtzi, 2003) and can thus interact with the vernier. Future research will need to address these principled questions before embarking on detailed modeling.

Third, it may be argued that, for example, adding flankers increases the regularity of the stimulus configuration and, thus, simplifies the Fourier spectrum. However, we could not find evidence for such a model using standard Fourier analysis (Clarke, Herzog, & Francis, 2014).

Fourth, models based on eye movements (Nandy & Tjan, 2012) may explain why crowding occurs within Bouma's window and can account for anisotropies of crowding. However, as mentioned, Bouma's law does not always hold true. In addition, the models cannot explain the grouping and *un*grouping effects in foveal vision

(Malania et al., 2007; Sayim et al., 2008, 2010), where eye movements obviously play no role (see also for foveal crowding: Westheimer & Hauske, 1975; Levi et al., 1985; Huurneman, Boonstra, Cox, Cillessen, & Rens, 2012; Lev, Yehezkel, & Polat, 2014; Norgett & Siderov, 2014).

Future models

The results described in this review provide strong constraints and guidance for future modeling. We sketch briefly several principled avenues but there might be many more. As mentioned, basic pooling models of crowding (Wilkinson et al., 1997; Parkes et al., 2001; Greenwood et al., 2009, 2010; Van Den Berg et al., 2010) cannot explain our results, and the same is true for current pooling models based on summary statistics reproducing Bouma's law (Balas et al., 2009). However, summary statistics models may just give up the link to receptive field sizes or implement summary statistics on multiple levels (Freeman & Simoncelli, 2011; Whitney, Haberman, & Sweeny, 2014). Whether such or texture recognition models can explain uncrowding is an open, but important, question since

these models do not require the explicit computation of objects, such as the squares in Figure 2.

Another option is models where, indeed, object representations across large parts of the entire visual field are explicitly computed, and it is explicitly determined which elements group with each other. Interference occurs only within groups, for example, by averaging, pooling, confusion, or other mechanisms. One question for such a scenario is: Why should interference occur at all?

Yet, another option is that crowding occurs during recurrent processing where higher level feature processing interacts with lower in a time-consuming manner. Grouping and interference may occur dynamically in recurrent networks (Foley, Grossberg, & Mingolla, 2012), similar to models where perceptual grouping is crucial for depth perception (Cao & Grossberg, 2005; Francis, 2009), brightness perception (Grossberg & Kelly, 1999; Francis & Schoonveld, 2005), texture segmentation (Bhatt, Carpenter, & Grossberg, 2007), and metacontrast masking (Francis, 1997).

In the next subsection, we propose that, indeed, crowding reflects the outcome of recurrent processing and for the same reason, the best predictor for stimulus strength is how stimuli look, that is, appearance.

Recurrent processing and appearance

Vision is an ill-posed problem. For this reason, we suggest that crowding, as most other visual aspects, occurs during recurrent processing, when the human brain takes contextual information into account to solve the ill-posed problems of vision. For example, the light that arrives at the photoreceptors is the product of the light shining on the object (illuminance) and the material properties of the object (reflectance). To determine the reflectance, the brain needs to discount for the illuminance. For example, the brain tries to discount for shadows that may explain many brightness illusions (Adelson, 1993). Analysis of the shadow requires recurrent computations across the entire visual field. Where is the illuminance coming from? Where is the occluder?

We propose that crowding is related to the *outcome* of this processing. Crowding does not reflect interactions at the beginning of recurrent visual processing. As an illustration, in the tilt illusion, flankers can strongly bias perception by, for example, making a straight line *appearing* tilted. The whole determines the perception of its part, which seems to be a chicken–egg problem. We propose that first single elements are processed and their representations are "veridical" initially. Time-consuming recurrent processing of the overall configuration changes the representations of the single elements. The *output* of the processing is a tilted line and only this tilted line is

perceived. The processing *itself* remains unconscious (Scharnowski et al., 2009). Crowding strength is mainly related to this final processing stage and hence how things look (e.g., tilted). Whereas in the tilt illusion, neural normalization in V1 may or may not account for the results, we propose that crowding is determined by much more complex configural interactions. Crowding is determined by the perceptual organization of the entire stimulus (i.e., appearance). Appearance depends on the stimulus and the internal states of the observer. As mentioned, appearance currently cannot be predicted by Gestalt laws. Appearance can change without stimulus changes, as is evident in ambiguous figures. It was shown that when the appearance changes, crowding can change too (Dakin et al., 2011; Maus et al., 2011). In this sense, our results turn classical models of vision upside down. Appearance of the whole is first, appearance of the features is second (see also Hochstein & Ahissar, 2002). Particularly, it seems that the whole determines *performance* on its parts in analogy to the famous quote by Wertheimer, "the whole determines the *appearance* of its parts," i.e., not only appearance of the parts is determined by the whole but also fine-grained spatial discrimination of the parts (Wertheimer, 1922). Very similar considerations may hold true for other spatial aspects of vision.

Our imaging studies support the notion of recurrent processing (Chicherov et al., 2014). The P1 component in our crowding experiments reflects mainly the overall stimulus size of the stimuli being highest for the long flankers. It is not just before the N1 component, when neural activity corresponds to performance. It seems it takes about 50–80 ms to transform retinotopic encoding into an object-based representation, which correlates with crowding performance.

As mentioned, to the best of our knowledge, there is no model at the moment that can predict the appearance of the elements of a scene, particularly because there is no model that can predict good Gestalt. However, we may be able to determine appearance directly by subjective reports, that is, how stimuli look, in addition to the traditional objective measures, such as vernier acuity. As mentioned above, we asked observers to rate whether the target stands out from the flankers. These subjective ratings correlated more or less well with objective vernier thresholds (Malania et al., 2007; Saarela et al., 2009; Manassi et al., 2012; see also Wolford & Chambers, 1983).

Why are the target features invisible? As mentioned, for each retinal image, there are infinitely many possible stimuli in the external world; however, there can be only one object at a time. For this reason, the brain needs to suppress (or does not encode) infinitely many other interpretations (Herzog et al., 2014). In this sense, crowding is a purposeful process to see the forest rather than the trees when "intended" (Navon, 1977; Cavanagh, 2001). This hypothesis also explains why the

brain gives up excellent spatial resolution in certain but not other crowding conditions. Whereas it is good to have excellent resolution in psychophysical experiments, in real-life situations, it may be more important to have a mechanism available that quickly and automatically suppresses a plethora of irrelevant interpretations of a visual scene.

Object recognition

Our considerations have strong implications for object recognition and the philosophy of perception in general. In the model of Figure 1C, the visual input is subjected to a bank of filters, each with its fixed receptive field size and profile. In the next stage, the outputs of the filters are integrated (i.e., pooled) leading to more complex feature processing (and larger receptive fields). Object recognition occurs in a truly mechanistic and stereotyped fashion. In fact, the goal and beauty of these models is to explain perception by basic neural circuits, such as linking crowding to pooling, thus breaking down the complex problems of vision into simple, mathematically treatable computations. This stereotypical procedure is aimed to naturalize, (i.e., replace) the subjective aspects of vision, such as grouping and good Gestalt. However, our results challenge this view. It seems that a "flexible" grouping stage cannot be avoided (flexible in the sense that subtle changes in the spatial layout can strongly change grouping, and hence, crowding). For example, in Figure 3A-a, an orientation-sensitive neuron may be involved in crowding because the vernier and the flankers are in its receptive field. However, the responses of this neuron are the same when the flankers are part of the rectangles (Figure 3A-b). Hence, a mechanism is needed that "excludes" this neuron from contributing in the latter case. This argument seems to apply not only to feedforward, hierarchical models of the type shown in Figure 1C but also to any model, which does not compute grouping in an explicit or implicit way. However, hierarchical, feedforward models are particularly challenged because they aim to explain shape processing from lower level processing. Whereas these results may strongly challenge our current thinking and intuition about crowding and visual processing in general, they make very little constraints on modeling since the class of hierarchical, feedforward models is small compared to other classes of models, including models with all sorts of recurrent processing (Clarke et al., 2014).

We would like to mention that grouping plays a crucial role also in haptic (Overvliet & Sayim, 2013) and acoustical (Oberfeld & Stahn, 2012) crowding situations and in many other visual paradigms, such as pattern masking (Herzog & Koch, 2001; Herzog & Fahle, 2002), metacontrast masking (Duangudom, Francis, & Herzog, 2007; Sayim, Manassi, & Herzog, 2014), and surround suppression (Saarela & Herzog, 2009). Grouping is key for targets other than verniers, such as for letters (Saarela et al., 2010; Rosen & Pelli, in press) and Gabors (Saarela et al., 2009; Levi & Carney, 2009). Hence, our considerations seem to be crucial for information processing in general (Herzog & Manassi, 2015).

Summary

1. Crowding is usually implicitly or explicitly thought to be characterized by locally restricted (Bouma's law), feature-specific interactions, where adding elements can only deteriorate performance. Bouma's law is often seen as a definition of crowding. For this reason, almost all crowding research in the last 40 years has used single flankers that are close and similar to the target. However, we have reviewed ample evidence that these characteristics are less key in crowding than previously thought. Crowding does not depend only on the elements within Bouma's window but potentially on all elements of the visual scene. Remote elements can either increase or decrease crowding.
2. Crowding is not an inevitable bottleneck of low-level vision.
3. Most current models of crowding cannot explain crowding, particularly when they are made to explain the above characteristics of crowding. Uncrowding by adding elements (beyond Bouma's window) is hard to explain in basic pooling and substitution models. Also, models based on eye movements face the problem that Bouma's law is not always true, and, in addition, crowding can occur foveally.
4. Grouping seems to be the key in understanding crowding. When the target does not group with the flankers, crowding is weak. Only when the target groups with the flankers *can* crowding be strong. Hence, grouping is necessary but not sufficient for crowding.
5. Many grouping cues can lead to strong crowding and uncrowding, including similarity, regularity, contour grouping, and good Gestalt. However, at the moment, it is impossible to predict the overall grouping of elements (i.e., appearance).
6. Grouping is not a mechanism to explain why performance deteriorates in crowding. We propose that grouping is an intermediate step, which determines which elements are prone to mutual interference.
7. It remains an open question to which extent grouping is computed explicitly (e.g., the squares in Figure 2) or implicitly, e.g., as a byproduct of and during texture processing.

8. Our working hypothesis is that crowding occurs during recurrent processing, where high-level, figural processing interacts with low-level processing. Low-level information is not lost at the beginning of processing. Crowding is determined by the final states of processing and so is appearance. For this reason, crowding correlates with appearance.

9. Subjective terms, such as grouping, cannot be eliminated at the moment. For this reason it seems important to take subjective measures of appearance into account.

10. Our results challenge not only most existing models of crowding but also many classic feedforward and hierarchical models of object recognition where basic features (vernier offset) and shapes (squares) are processed at different levels. We propose that any successful model of object recognition cannot be based on stereotypical filtering. Models need to take a flexible grouping stage into account, with flexible meaning that small changes in the stimulus layout can lead to strong changes in perception.

Keywords: crowding, grouping, object recognition, appearance, vernier acuity

References

Adelson, E. H. (1993). Perceptual organization and the judgment of brightness. *Science*, *262*(5142), 2042–2044.

Altmann, C. F., Bülthoff, H. H., & Kourtzi, Z. (2003). Perceptual organization of local elements into global shapes in the human visual cortex. *Current Biology*, *13*(4), 342–349.

Anderson, E. J., Dakin, S. C., Schwarzkopf, D. S., Rees, G., & Greenwood, J. A. (2012). The neural correlates of crowding-induced changes in appearance. *Current Biology*, *22*(13), 1199–1206, doi:10.1016/j.cub.2012.04.063.

Andriessen, J., & Bouma, H. (1976). Eccentric vision: Adverse interactions between line segments. *Vision Research*, *16*(1), 71–78, doi:10.1016/0042-6989(76)90078-X.

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, http://www.journalofvision.org/content/9/12/13, doi:10.1167/9.12.13. [PubMed] [Article]

Banks, W. P., Larson, D. W., & Prinzmetal, W. (1979). Asymmetry of visual interference. *Perception & Psychophysics*, *25*(6), 447–456.

Banks, W. P., & White, H. (1984). Lateral interference and perceptual grouping in visual detection. *Perception & Psychophysics*, *36*(3), 285–295.

Bex, P. J., Dakin, S. C., & Simmers, A. J. (2003). The shape and size of crowding for moving targets. *Vision Research*, *43*(27), 2895–2904.

Bhatt, R., Carpenter, G. A., & Grossberg, S. (2007). Texture segregation by visual cortex: Perceptual grouping, attention, and learning. *Vision Research*, *47*(25), 3173–3211, doi:10.1016/j.visres.2007.07.013.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, *226*(5241), 177–178.

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

Cao, Y., & Grossberg, S. (2005). A laminar cortical model of stereopsis and 3d surface perception: Closure and Da Vinci stereopsis. *Spatial Vision*, *18*(5), 515–578.

Cavanagh, P. (2001). Seeing the forest but not the trees. *Nature Neuroscience*, *4*(7), 673–673.

Chakravarthi, R., & Pelli, D. G. (2011). The same binding in contour integration and crowding. *Journal of Vision*, *11*(8):10, 1–12, http://www.journalofvision.org/content/11/8/10, doi:10.1167/11.8.10. [PubMed] [Article]

Chanceaux, M., & Grainger, J. (2013). Constraints on letter-in-string identification in peripheral vision: Effects of number of flankers and deployment of attention. *Frontiers in Psychology, 4,* 119.

Chen, J., He, Y., Zhu, Z., Zhou, T., Peng, Y., Zhang, X., & Fang, F. (2014). Attention-dependent early cortical suppression contributes to crowding. *Journal of Neuroscience*, *34*(32), 10465–10474, doi:10.1523/JNEUROSCI.1140-14.2014.

Chicherov, V., Plomp, G., & Herzog, M. H. (2014).

Neural correlates of visual crowding. *NeuroImage*, *93*(Part 1), 23–31, doi:10.1016/j.neuroimage.2014. 02.021.

Chung, S. T., Levi, D. M., & Legge, G. E. (2001). Spatial-frequency and contrast properties of crowding. *Vision Research*, *41*(14), 1833–1850, http://linkinghub.elsevier.com/retrieve/pii/ S0042698901000712

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.

Dakin, S. C., Greenwood, J. A., Carlson, T. A., & Bex, P. J. (2011). Crowding is tuned for perceived (not physical) location. *Journal of Vision*, *11*(9):2, 1–13, http://www.journalofvision.org/content/11/9/2, doi:10.1167/11.9.2. [PubMed] [Article]

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research*, *36*(12), 1827–1837.

DiCarlo, J. J., Zoccolan, D., & Rust, N. C. (2012). How does the brain solve visual object recognition? *Neuron*, *73*(3), 415–434, doi:10.1016/j.neuron.2012. 01.010.

Duangudom, V., Francis, G., & Herzog, M. H. (2007). What is the strength of a mask in visual metacontrast masking? *Journal of Vision*, *7*(1):7, 1–10, http://www.journalofvision.org/content/7/1/7, doi: 10.1167/7.1.7. [PubMed] [Article]

Ester, E. F., Klee, D., & Awh, E. (2014). Visual crowding cannot be wholly explained by feature pooling. *Journal of Experimental Psychology: Human Perception & Performance*, *40*(3), 1022–1033, doi:10.1037/a0035377.

Ester, E. F., Zilber, E., & Serences, J. T. (2015). Substitution and pooling in visual crowding induced by similar and dissimilar distractors. *Journal of Vision*, *15*(1):4, 1–12, http://www.journalofvision. org/content/15/1/4, doi:10.1167/15.1.4. [PubMed] [Article]

Faivre, N., & Kouider, S. (2011). Multi-feature objects elicit nonconscious priming despite crowding. *Journal of Vision*, *11*(3):2, 1–10, http://www.journalofvision. org/content/11/3/2, doi:10.1167/11.3.2. [PubMed] [Article]

Farzin, F., Rivera, S. M., & Whitney, D. (2009). Holistic crowding of mooney faces. *Journal of Vision*, *9*(6):18, 1–15, http://www.journalofvision. org/content/9/6/18, doi:10.1167/9.6.18. [PubMed] [Article]

Felisberti, F. M., Solomon, J. A., & Morgan, M. J. (2005). The role of target salience in crowding. *Perception*, *34*(7), 823–833.

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.

Foley, N. C., Grossberg, S., & Mingolla, E. (2012). Neural dynamics of object-based multifocal visual spatial attention and priming: Object cueing, useful-field-of-view, and crowding. *Cognitive Psychology*, *65*(1), 77–117, doi:10.1016/j.cogpsych. 2012.02.001.

Francis, G. (1997). Cortical dynamics of lateral inhibition: metacontrast masking. *Psychological Review*, *104*(3), 572–594.

Francis, G. (2009). Cortical dynamics of figure-ground segmentation: Shine-through. *Vision Research*, *49*(1), 140–163, doi:10.1016/j.visres.2008.10.002.

Francis, G., & Schoonveld, W. (2005). Using afterimages for orientation and color to explore mechanisms of visual filling-in. *Perception & Psychophysics*, *67*(3), 383–397.

Freeman, J., Donner, T. H., & Heeger, D. J. (2011). Inter-area correlations in the ventral visual pathway reflect feature integration. *Journal of Vision*, *11*(4):15, 1–23, http://www.journalofvision.org/ content/11/4/15, doi:10.1167/11.4.15. [PubMed] [Article]

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

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, USA*, *106*(31), 13130–13135, doi:10.1073/ pnas.0901352106.

Greenwood, J. A., Bex, P. J., & Dakin, S. C. (2010). Crowding changes appearance. *Current Biology*, *20*(6), 496–501.

Grossberg, S., & Kelly, F. (1999). Neural dynamics of binocular brightness perception. *Vision Research*, *39*(22), 3796–3816.

Harp, T. D., Bressler, D. W., & Whitney, D. (2007). Position shifts following crowded second-order motion adaptation reveal processing of local and global motion without awareness. *Journal of Vision*, *7*(2):5, 1–13, http://www.journalofvision.org/ content/7/2/5, doi:10.1167/7.2.15. [PubMed] [Article]

Harrison, W. J., & Bex, P. J. (2014). Integrating retinotopic features in spatiotopic coordinates.

Journal of Neuroscience, 34(21), 7351–7360, doi:10.1523/JNEUROSCI.5252-13.2014.

Harrison, W. J., Mattingley, J. B., & Remington, R. W. (2013). Eye movement targets are released from visual crowding. *Journal of Neuroscience, 33*(7), 2927–2933, doi:10.1523/JNEUROSCI.4172-12.2013.

Harrison, W. J., Retell, J. D., Remington, R. W., & Mattingley, J. B. (2013). Visual crowding at a distance during predictive remapping. *Current Biology, 23*(9), 793–798, doi:10.1016/j.cub.2013.03.050.

He, S., Cavanagh, P., & Intriligator, J. (1996). Attentional resolution and the locus of visual awareness. *Nature, 383*(6598), 334–337.

Herzog, M. H., & Fahle, M. (2002). Effects of grouping in contextual modulation. *Nature, 415*(6870), 433–436.

Herzog, M. H., Hermens, F., & Öğmen, H. (2014). Invisibility and interpretation. *Frontiers in Psychology, 5*, 975.

Herzog, M. H., & Koch, C. (2001). Seeing properties of an invisible object: Feature inheritance and shine-through. *Proceedings of the National Academy of Sciences, USA, 98*(7), 4271–4275, doi:10.1073/pnas.071047498.

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.

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron, 36*(5), 791–804.

Huckauf, A. (2007). Task set determines the amount of crowding. *Psychological Research, 71*(6), 646–652, http://www.springerlink.com/index/w90204tr56615555.pdf, doi:10.1007/s00426-006-0054-6.

Huckauf, A., & Heller, D. (2002). Spatial selection in peripheral letter recognition: In search of boundary conditions. *Acta Psychologica, 111*(1), 101–123.

Hung, C. P., Kreiman, G., Poggio, T., & DiCarlo, J. J. (2005). Fast readout of object identity from macaque inferior temporal cortex. *Science, 310*(5749), 863–866.

Huurneman, B., Boonstra, F. N., Cox, R. F., Cillessen, A. H., & Rens, G. V. (2012). A systematic review on 'foveal crowding' in visually impaired children and perceptual learning as a method to reduce crowding. *BMC Ophthalmology, 12*(1), 27. Available from http://www.ncbi.nlm.nih.gov/pubmed/22824242, doi:10.1186/1471-2415-12-27.

Ikeda, H., Watanabe, K., & Cavanagh, P. (2013). Crowding of biological motion stimuli. *Journal of Vision, 13*(4):20, 1–6, http://www.journalofvision.org/content/13/4/20, doi:10.1167/13.4.20. [PubMed] [Article]

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

Joo, S. J., Boynton, G. M., & Murray, S. O. (2012). Long-range, pattern-dependent contextual effects in early human visual cortex. *Current Biology, 22*(9), 781–786.

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.

Kouider, S., Berthet, V., & Faivre, N. (2011). Preference is biased by crowded facial expressions. *Psychological Science, 22*(2), 184–189, http://pss.sagepub.com/content/22/2/184.short, doi:10.1177/0956797610396226.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35*(13), 1897–1916.

Krumhansl, C., & Thomas, E. (1977). Effect of level of confusability on reporting letters from briefly presented visual displays. *Perception & Psychophysics, 21*(3), 269–279, doi:10.3758/BF03214239.

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, doi:10.1152/jn.00476.2014.

Legge, G. E. (2007). *Psychophysics of reading in normal and low vision*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Lev, M., Yehezkel, O., & Polat, U. (2014). Uncovering foveal crowding? *Scientific Reports, 4*, 4067-4067, doi:10.1038/srep04067.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research, 48*(5), 635–654, doi: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.

Levi, D. M., Hariharan, S., & Klein, S. A. (2002). Suppressive and facilitatory spatial interactions in peripheral vision: Peripheral crowding is neither size invariant nor simple contrast masking. *Journal of Vision, 2*(2):3, 167–177, http://www.journalofvision.org/content/2/2/3, doi:10.1167/2.2.3. [PubMed] [Article]

Levi, D. M., Klein, S. A., & Aitsebaomo, A. (1985). Vernier acuity, crowding and cortical magnifica-

tion. *Vision Research*, 25(7), 963–977, doi:10.1016/0042-6989(85)90207-X.

Liu, T., Jiang, Y., Sun, X., & He, S. (2009). Reduction of the crowding effect in spatially adjacent but cortically remote visual stimuli. *Current Biology*, 19(2), 127–132.

Livne, T., & Sagi, D. (2007). Configuration influence on crowding. *Journal of Vision*, 7(2):4, 1–12, http://www.journalofvision.org/content/7/2/4, doi:10.1167/7.2.4. [PubMed] [Article]

Livne, T., & Sagi, D. (2010). How do flankers' relations affect crowding? *Journal of Vision*, 10(3):1, 1–14, http://www.journalofvision.org/content/10/3/1, doi:10.1167/10.3.1. [PubMed] [Article]

Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding: Selective interference between configural representations of faces in crowded scenes. *Journal of Vision*, 7(2):24, 1–11, http://www.journalofvision.org/content/7/2/24, doi:10.1167/7.2.24. [PubMed] [Article]

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

Manassi, M., Sayim, B., & Herzog, M. H. (2012). Grouping, pooling, and when bigger is better in visual crowding. *Journal of Vision, 12*(10):13, 114, http://www.journalofvision.org/content/12/10/114, doi:10.1167/12.10.114. [Abstract]

Manassi, M., Sayim, B., & Herzog, M. H. (2013). When crowding of crowding leads to uncrowding. *Journal of Vision, 13*(13):10, 1–10, http://www.journalofvision.org/content/13/13/10, doi:10.1167/13.13.10. [PubMed] [Article]

Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *Journal of Vision*, 5(1):6, 58–70, http://www.journalofvision.org/content/5/1/6, doi:10.1167/5.1.6. [PubMed] [Article]

Maus, G. W., Fischer, J., & Whitney, D. (2011). Perceived positions determine crowding. *PLoS One*, 6(5), e19796, doi:10.1371/journal.pone.0019796.

Millin, R., Arman, A. C., Chung, S. T., & Tjan, B. S. (2014). Visual crowding in V1. *Cerebral Cortex*, 24(12), 3107–3115.

Nandy, A. S., & Tjan, B. S. (2012). Saccade-confounded image statistics explain visual crowding. *Nature Neuroscience*, 15(3), 463–469, doi:10.1038/nn.3021.

Navon, D. (1977). Forest before trees: The precedence of global features in visual perception. *Cognitive Psychology*, 9(3), 353–383.

Nazir, T. A. (1992). Effects of lateral masking and spatial precuing on gap-resolution in central and peripheral vision. *Vision Research*, 32(4), 771–777, http://linkinghub.elsevier.com/retrieve/pii/004269899290192L.

Norgett, Y., & Siderov, J. (2014). Foveal crowding differs in children and adults. *Journal of Vision*, 14(12):23, 1–10, http://www.journalofvision.org/content/14/12/23, doi:10.1167/14.12.23. [PubMed] [Article]

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., & Sayim, B. (2013). Contextual modulation in haptic vernier offset discrimination. *Perception*, 42, 175.

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.

Pelli, D. G. (2008). Crowding: A cortical constraint on object recognition. *Current Opinions in Neurobiology*, 18(4), 445–451, doi:10.1016/j.conb.2008.09.008.

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, 1136–1169, http://www.journalofvision.org/content/4/12/12, doi:10.1167/4.12.12. [PubMed] [Article]

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, 11(10), 1129–1135.

Põder, E. (2006). Crowding, feature integration, and two kinds of "attention." *Journal of Vision*, 6(2):7, 163–169, http://www.journalofvision.org/content/6/2/7, doi:10.1167/6.2.7. [PubMed] [Article]

Põder, E. (2007). Effect of colour pop-out on the recognition of letters in crowding conditions. *Psychological Research*, 71(6), 641–645, doi:10.1007/s00426-006-0053-7.

Rashal, E., & Yeshurun, Y. (2014). Contrast dissimilarity effects on crowding are not simply another case of target saliency. *Journal of Vision*, 14(6):9, 1–12, http://www.journalofvision.org/content/14/6/9, doi:10.1167/14.6.9. [PubMed] [Article]

Riesenhuber, M., & Poggio, T. (1999). Hierarchical models of object recognition in cortex. *Nature Neuroscience*, 2(11), 1019–1025.

Rosen, S., Chakravarthi, R., & Pelli, D. G. (2011). Crowding reveals a third stage of object recognition. *Journal of Vision*, 11(11): 1142, http://www.

journalofvision.org/content/11/11/1142, doi:10. 1167/11.11.1142. [Abstract]

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, http://www.journalofvision.org/ content/14/6/10, doi:10.1167/14.6.10. [PubMed] [Article]

Rosen, S., & Pelli, D. (in press). Crowding by a pattern. *Journal of Vision,* in press.

Saarela, T. P., & Herzog, M. H. (2009). Size tuning and contextual modulation of backward contrast masking. *Journal of Vision, 9*(11):21, 1–12, http:// www.journalofvision.org/content/9/11/21, doi:10. 1167/9.11.21. [PubMed] [Article]

Saarela, T. P., Sayim, B., Westheimer, G., & Herzog, M. H. (2009). Global stimulus configuration modulates crowding. *Journal of Vision, 9*(2):5, 1–11, http://www.journalofvision.org/content/9/2/5, doi: 10.1167/9.2.5. [PubMed] [Article]

Saarela, T. P., Westheimer, G., & Herzog, M. H. (2010). The effect of spacing regularity on visual crowding. *Journal of Vision, 10*(10):17, 1–7, http:// www.journalofvision.org/content/10/10/17, doi:10. 1167/10.10.17. [PubMed] [Article]

Sayim, B., & Cavanagh, P. (2013). Grouping and crowding affect target appearance over different spatial scales. *PLoS One, 8*(8), e71188, doi:10.1371/ journal.pone.0071188.

Sayim, B., Greenwood, J. A., & Cavanagh, P. (2014). Foveal target repetitions reduce crowding. *Journal of Vision, 14*(6):4, 1–12, http://www.hubmed.org/ display.cgi?uids=25294741, doi:10.1167/14.6.4. [PubMed] [Article]

Sayim, B., Manassi, M., & Herzog, M. (2014). How color, regularity, and good gestalt determine backward masking. *Journal of Vision, 14*(7):8, 1–11, http://www.journalofvision.org/content/14/7/8, doi:10.1167/14.7.8. [PubMed] [Article]

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, 1–9, http://www. journalofvision.org/content/8/8/12, doi:10.1167/8. 8.12. [PubMed] [Article]

Sayim, B., Westheimer, G., & Herzog, M. H. (2010). Gestalt factors modulate basic spatial vision. *Psychological Science, 21*(5), 641–644.

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

Scharnowski, F., Ruter, J., Jolij, J., Hermens, F., Kammer, T., & Herzog, M. H. (2009). Long-lasting modulation of feature integration by transcranial magnetic stimulation. *Journal of Vision, 9*(6):1, 1– 10, http://www.journalofvision.org/content/9/6/1, doi:10.1167/9.6.1. [PubMed] [Article]

Scolari, M., Kohnen, A., Barton, B., & Awh, E. (2007). Spatial attention, preview, and popout: Which factors influence critical spacing in crowded displays? *Journal of Vision, 7*(2):7, 1–23, http://www. journalofvision.org/content/7/2/7, doi:10.1167/7.2. 7. [PubMed] [Article]

Serre, T., Oliva, A., & Poggio, T. (2007). A feedforward architecture accounts for rapid categorization. *Proceedings of the National Academy of Sciences, USA, 104*(15), 6424–6429.

Serre, T., Wolf, L., Bileschi, S., Riesenhuber, M., & Poggio, T. (2007). Robust object recognition with cortex-like mechanisms. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(3), 411–426.

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.

Tannazzo, T., Kurylo, D. D., & Bukhari, F. (2014). Perceptual grouping across eccentricity. *Vision Research, 103*, 101–108, doi:http://dx.doi.org/10. 1016/j.visres.2014.08.011.

Thorpe, S., Delorme, A., & Van Rullen, R. (2001). Spike-based strategies for rapid processing. *Neural Networks, 14*(6), 715–725.

Toet, A., & Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in the parafovea. *Vision Research, 32*(7), 1349–1357.

Tripathy, S. P., & Levi, D. M. (1994). Long-range dichoptic interactions in the human visual cortex in the region corresponding to the blind spot. *Vision Research, 34*(9), 1127–1138.

Van Den Berg, R., Roerdink, J. B., & Cornelissen, F. W., (2010). A neurophysiologically plausible population code model for feature integration explains visual crowding. *PLoS Computational Biology, 6*(1), el000646.

van Koningsbruggen, M. G., & Buonocore, A. (2013). Mechanisms behind perisaccadic increase of perception. *Journal of Neuroscience, 33*(28), 11327– 11328, doi:10.1523/JNEUROSCI.1567-13.2013.

Vickery, T. J., Shim, W. M., Chakravarthi, R., Jiang, Y. V., & Luedeman, R. (2009). Supercrowding: Weakly masking a target expands the range of crowding. *Journal of Vision, 9*(2):12, 1–15, http://

www.journalofvision.org/content/9/2/12, doi:10.1167/9.2.12. [PubMed] [Article]

Wallace, J. M., & Tjan, B. S. (2011). Object crowding. *Journal of Vision*, *11*(6):19, 1–17, http://www.journalofvision.org/content/11/6/19, doi:10.1167/11.6.19. [PubMed] [Article]

Wallis, T. S., & Bex, P. J., (2011). Visual crowding is correlated with awareness. *Current Biology*, *21*(3), 254–258, doi:10.1016/j.cub.20T1.01.011.

Wertheimer, M. (1922). Untersuchungen zur Lehre von der Gestalt. *Psychologische Forschung*, *1*(1), 47–58.

Wertheimer, M. (1923). Untersuchungen zur Lehre von der Gestalt. II. *Psychologische Forschung*, *4*(1), 301–350.

Westheimer, G., & Hauske, G. (1975). Temporal and spatial interference with vernier acuity. *Vision Research*, *15*, 1137–1141.

Whitney, D., Haberman, J., & Sweeny, T. D. (2014). From textures to crowds: multiple levels of summary statistical perception. In J. S. Werner & L. M. Chalupa (Eds.), *The new visual neurosciences* (pp. 685–710). Cambridge, MA: MIT Press.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Science*, *15*(4), 160–168.

Wilkinson, F., Wilson, H. R., & Ellemberg, D. (1997). Lateral interactions in peripherally viewed texture arrays. *Journal of the Optical Society of America*, *14*(9), 2057–2068.

Wolfe, B. A., & Whitney, D. (2014). Facilitating recognition of crowded faces with presaccadic attention. *Frontiers in Human Neuroscience, 8*, 103.

Wolford, G., & Chambers, L. (1983). Lateral masking as a function of spacing. *Perception & Psychophysics*, *33*(2), 129–138.

Yeh, S. L., He, S., & Cavanagh, P. (2012). Semantic priming from crowded words. *Psychological Science, 23*(6), 608–616.

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, doi: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, 1–12, http://www.journalofvision.org/content/10/10/16, doi:10.1167/10.10.16. [PubMed] [Article]

Zahabi, S., & Arguin, M., (2014), A crowdful of letters: Disentangling the role of similarity, eccentricity and spatial frequencies in letter crowding. *Vision Research*, *97*, 45–51, doi:10.1016/j.visres.2014.02.001.

Zhang, J.-Y., Zhang, G.-L., Liu, L., & Yu, C. (2012). Whole report uncovers correctly identified but incorrectly placed target information under visual crowding. *Journal of Vision*, *12*(7):5, 1–11, http://www.journalofvision.org/content/12/7/5, doi:10.1167/12.7.5. [PubMed] [Article]