Crowding follows the binding of relative position and orientation

John A. Greenwood

UCL Institute of Ophthalmology, University College London, London, UK, Laboratoire Psychologie de la Perception, Université Paris Descartes, Paris, France, & CNRS UMR 8158, Paris, France

Peter J. Bex

Schepens Eye Research Institute, Harvard Medical School, Boston, MA, USA

Steven C. Dakin

UCL Institute of Ophthalmology, University College London, London, UK

Crowding—the deleterious influence of clutter on object recognition—disrupts the identification of visual features as diverse as orientation, motion, and color. It is unclear whether this occurs via independent feature-specific crowding processes (preceding the *feature binding* process) or via a singular (late) mechanism tuned for combined features. To examine the relationship between feature binding and crowding, we measured interactions between the crowding of relative position and orientation. Stimuli were a target cross and two flanker crosses (each composed of two near-orthogonal lines), 15 degrees in the periphery. Observers judged either the orientation (clockwise/counterclockwise) of the near-horizontal target line, its position (up/down relative to the stimulus center), or both. For single-feature judgments, crowding affected position and orientation similarly: thresholds were elevated and responses biased in a manner suggesting that the target appeared more like the flankers. These effects were tuned for orientation, with near-orthogonal elements producing little crowding. This tuning allowed us to separate the predictions of independent (feature specific) and combined (singular) models: for an independent model, reduced crowding for one feature has no effect on crowding for other features, whereas a combined process affects either all features or none. When observers made conjoint judgments, a reduction of orientation crowding (by increasing target–flanker orientation differences) increased the rate of correct responses for both position and orientation, as predicted by our combined model. In contrast, our independent model incorrectly predicted a high rate of position errors, since the probability of positional crowding would be unaffected by changes in orientation. Thus, at least for these features, crowding is a singular process that affects bound position and orientation values in an all-or-none fashion.

Keywords: crowding, orientation, position, feature binding, peripheral visual field

Introduction

Our recognition of complex visual objects and scenes requires the encoding of values along a number of dimensions—color, orientation, and spatial frequency, for instance—and their accurate combination. We refer to these values as the *features* of an object: variation along these dimensions recruits distinct populations of feature detectors (or "channels") and alters the appearance of the object in question (Braddick, Campbell, & Atkinson, 1978; Graham, 1989; Pelli, Burns, Farell, & Moore-Page, 2006). Given the specialization of these feature detectors (Felleman & Van Essen, 1991; Lennie, 1998; Ungerleider & Mishkin, 1982), distinct features from the same object must be correctly co-localized in space and time, a process known as *feature binding* (Treisman, 1996; Treisman &

Gelade, 1980). Failures of this process can lead to objects with incorrect feature conjunctions (e.g., misperceiving a red X as being green; Treisman & Schmidt, 1982).

Within each feature dimension, accurate recognition can be severely limited by *crowding*, a disruptive interaction among adjacent objects that are otherwise visible in isolation (Bouma, 1970; Flom, Weymouth, & Kahneman, 1963). Crowding occurs when clutter falls within a region of space surrounding the target object, known as the *interference zone*, which increases in size with retinal eccentricity (Bouma, 1970; Toet & Levi, 1992). Errors made under crowded conditions correlate strongly with the features present within flanking objects (Dakin, Cass, Greenwood, & Bex, 2010; Huckauf & Heller, 2002; Strasburger, Harvey, & Rentschler, 1991), likely because crowded target features change to more closely resemble those of the flankers (Greenwood, Bex, & Dakin, 2010).

A range of theories has been proposed to account for these effects (reviewed by Levi, 2008), though a weighted averaging process that combines target and flanker features has been arguably the most successful, with clear application for the crowding of orientation (Parkes, Lund, Angelucci, Solomon, & Morgan, 2001) and position (Dakin et al., 2010; Greenwood, Bex, & Dakin, 2009). The net effect is that the visual scene becomes simplified toward texture (Freeman & Simoncelli, 2011).

Crowding has a pervasive effect on feature identification, with documented effects on features including orientation and spatial frequency (Wilkinson, Wilson, & Ellemberg, 1997), color and size (van den Berg, Roerdink, & Cornelissen, 2007), position (Greenwood et al., 2009), and motion (Bex & Dakin, 2005). However, these wide ranging effects raise the question of whether crowding reflects a singular mechanism or a group of independent feature-specific processes. Put another way, the relation between crowding and the feature binding process is unclear. Were crowding to occur after (or even during) feature binding, it is conceivable that it would be tuned for feature conjunctions so that when crowding occurs it would affect all relevant features at the same time. Alternatively, if it were to precede feature binding, crowding in one feature domain would not necessarily be accompanied by the crowding of other features. If crowding were independent for each feature, it is also conceivable that its operation could vary in each domain. This issue is therefore central to our understanding of crowding.

Perhaps the strongest test of the "singularity of crowding" is whether a change in the probability of crowding for one feature affects the probability of crowding for other features. A series of independent processes would allow crowding to occur for one feature without affecting others within the same object. In contrast, a single (late) crowding process is necessarily all or none for the features present within an object. Consistent with the latter, when target and flanker elements differ in color, contrast polarity, or binocular disparity, crowding is reduced and the identity of both letter (Hess, Dakin, Kapoor, & Tewfik, 2000; Kooi, Toet, Tripathy, & Levi, 1994) and vernier acuity targets (Butler & Westheimer, 1978; Sayim, Westheimer, & Herzog, 2008) is easier to discern. That is, the features required for these tasks (e.g., orientation and position) are less crowded when crowding is reduced in other domains. However, a singular crowding process also predicts that the occurrence of errors for one feature type should correlate with errors for other features (i.e., that when crowding occurs it gives errors for all features present). This was not found when observers made conjoint judgments of color, spatial frequency, and orientation (Põder & Wagemans, 2007). Instead, a mix of single-feature errors and occasional conjunction errors was observed—a pattern that is more consistent with several independent processes than a single combined mechanism.

There are two ways to reconcile these prior findings. The first is that crowding could involve a series of independent processes that are nonetheless able to interact. That is, independent processes could give uncorrelated errors in each domain, but a release in crowding for one feature domain might reduce the probability of crowding in other domains through these interactions. The second is that crowding is a singular process, but we may expect correlated errors only when crowding is strong. The large feature differences used by Põder and Wagemans (2007)—red/green, horizontal/vertical, and low/high spatial frequencies—might have produced a release from crowding on some trials and not others since crowding is tuned for color and orientation differences (Kooi et al., 1994; Levi & Carney, 2009; Wilkinson et al., 1997). It is possible that the conditions that break crowding for one feature break it for others, which would have led to a mix of correlated and uncorrelated errors.

Our approach

From the above, it is clear that a full consideration of the unity of crowding requires that we examine the nature of the errors that are made under both strong and weak crowding. This was the aim of the present study.

We examined whether crowding operates on independent or combined visual features with two features that are critical for letter recognition: relative position and orientation. There are several reasons to select these features. First, the identity of letters is strongly determined by the relative position and orientation of the constituent strokes (Chastain, 1981; Watt & Dakin, 2010; Wolford, 1975). A shift in stroke position can change X into V; rotation can change W into M. Second, the effect of crowding on these features is relatively well understood, with computational models available for both orientation (Parkes et al., 2001; Solomon, Felisberti, & Morgan, 2004) and position (Dakin et al., 2010; Greenwood et al., 2009). Finally, crowding is tuned for orientation, with less crowding for large target–flanker orientation differences than for small differences (Levi & Carney, 2009; Wilkinson et al., 1997), allowing us to modulate the strength of crowding for one feature and examine the effects on another.

However, are these features representative of those used by the visual system? At the outset, we defined features as image components for which there is selectivity within the visual system. The role of orientation and position in determining letter identity (as above) is a likely indication of their importance here, but a wealth of physiological evidence also demonstrates that neurons in the primate visual system are tuned for both orientation (Hubel & Wiesel, 1968) and spatial phase (Hamilton, Albrecht, & Geisler, 1989). Psychophysical evidence similarly supports the existence of a range of orientation-selective

channels (Blakemore & Nachmias, 1971; Campbell & Kulikowski, 1966) and at least two phase-selective channels (Burr, Morrone, & Spinelli, 1989; Huang, Kingdom, & Hess, 2006). Without knowing the precise means by which letter features are encoded, we suggest that these orientation- and phase-selective channels must be involved in some fashion. Both position and orientation thus satisfy our criteria to be candidate features.

Besides our mechanistic definition, others have sought to define the feature dimensions of vision using more behavioral criteria. A popular argument from Feature Integration Theory (FIT) is that basic features must be processed pre-attentively, allowing pop-out in visual search tasks (Treisman & Gelade, 1980) and effortless texture segmentation (Julesz, 1981). Though this is clear for orientation in both visual search (Sagi & Julesz, 1985) and texture segmentation (Beck, 1966), position holds a special place in FIT as it is the attentive co-localization of feature conjunctions that allows binding to occur. Here though, we specifically examine the relative position of features within object boundaries, as opposed to their gross position in the visual field. These small-scale position differences (often included under the umbrella terms "shape" and "form" in the visual search literature) determine *what* rather than *where* and are themselves susceptible to crowding (Dakin et al., 2010; Greenwood et al., 2009) as opposed to large-scale position shifts that typically relieve crowding (Bouma, 1970; Toet & Levi, 1992). Along these lines, pop-out has been observed with differences in phase (Heathcote & Mewhort, 1993) and the relative position of letter strokes (Duncan & Humphreys, 1989). Texture segmentation has similarly been observed with phase differences (Hansen & Hess, 2006) and shifts in letter-stroke position (Bergen & Julesz, 1983). By all of the above criteria then, orientation and position are reasonable features to select for our analysis.

Returning to our aims, if crowding were a single, combined process, then it should be "all or none" when it arises. That is, if the probability of crowding is reduced for one attribute, it should be reduced for all attributes at the same time. Strong crowding should produce the combination of position and orientation errors, while weak crowding should produce neither. Conversely, if crowding were independent for each feature, then a release from crowding in one domain (e.g., orientation) would have no effect on crowding in the other (e.g., position). Here, strong crowding should produce largely uncorrelated errors, while weak crowding for orientation should have no effect on the rate of errors for position judgments.

Because we wish to model the crowding of multiple features, we first measured the crowding of each feature in isolation (Experiments 1 and 2) allowing us to consider whether the mechanism of crowding is similar in both feature domains. We then measured the effect of orientation on uncrowded position judgments to ensure that the manipulation of orientation—outside of its effect on crowding—would not unduly influence our results

(Experiment 3). Finally, we examined the effects of crowding on combined judgments of position and orientation (Experiment 4) and compare the predictions of independent and combined models of this process.

General methods

Observers

There were four observers for the whole set of experiments: two were authors (JAG and SCD) and two were naive. All had normal or corrected-to-normal visual acuity and were experienced psychophysical observers.

Apparatus

Experiments were programmed in MATLAB (MathWorks) and run on a Power Macintosh G5 computer with Psych-Toolbox software (Brainard, 1997; Pelli, 1997). Stimuli were presented on a CRT monitor (LaCie Electron Blue 22), with 1152×870 pixel resolution and 75-Hz refresh rate. The monitor was calibrated with a Minolta photometer and linearized in software, giving a mean and maximum luminance of 50 and 100 cd/m^2, respectively. Stimuli were viewed monocularly with the dominant eye from 57 cm, with responses made on a keypad. No feedback was provided.

Stimuli

Target and flankers were white "cross-like" elements consisting of two near-orthogonal lines (see Figure 1A), as used previously (Dakin et al., 2010; Greenwood et al., 2009). Stimulus size was 1.8 deg, approximately twice the size–acuity thresholds for each observer (see below), and each line was 0.36 deg wide (one-fifth the stimulus length, as with Sloan letters; Sloan, 1959). Stimuli were presented at 50% Weber contrast above the mean luminance. Judgments of line position (above/below the stimulus midpoint) and/or orientation (tilted clockwise/counterclockwise) were made regarding the horizontal (or near-horizontal) line.

The target cross was presented at an eccentricity of 15 deg in the upper visual field. Under crowded conditions, one flanker was presented to the left and one to the right of the target, with a center-to-center separation of 2.5 deg for three observers. This separation is $0.17\times$ the target eccentricity, well within the standard interference zone (Bouma, 1970; Toet & Levi, 1992). One naive observer (MST) required a separation of 3.75 deg to perform the task reliably, which should, nonetheless, give robust crowding at $0.25\times$ the target eccentricity. Stimuli were

A. Example stimuli
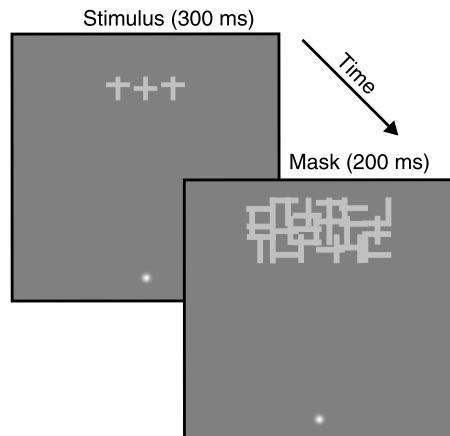

B. Example time course

Figure 1. (A) Sample cross-like stimuli. In uncrowded conditions, only the middle "target" element was presented; two flanker elements were present to the left and right in crowded conditions. In all experiments, observers made judgments about the near-horizontal line of the target: either its position relative to the stimulus midpoint (up/down, as depicted here), its tilt relative to horizontal (clockwise/counterclockwise), or both. (B) Sample time course. Stimuli were presented at 15 deg in the upper visual field for 300 ms. A mask was then presented for 200 ms before responses were made.

presented for 300 ms, followed by a dense 7.5×3.5 deg masking array of cross stimuli with identical size and contrast, randomized feature positions, and a matched range of orientations (centered on 15 deg eccentricity; Figure 1B). The mask was presented for 200 ms before being replaced by a blank interval until observers responded. A white Gaussian blob with a standard deviation of 0.1 deg was present near the bottom of the monitor for fixation during the trial.

Acuity measurement

Prior to the main experiments, size–acuity thresholds were measured for each observer by requiring judgments of the tilt (clockwise/counterclockwise) of the near-horizontal bar of an uncrowded target. The tilt was presented at ±24°, the maximum tilt used in Experiment 1. Stimulus size was determined using QUEST (Watson & Pelli, 1983), which converged on 75% correct identification. Stimuli were presented for 300 ms and post-masked as above. This procedure was repeated three times for

each observer and gave mean thresholds of 0.6 deg (EJA and SCD) and 0.7 deg (JAG and MST). Stimuli were subsequently presented at a size of 1.8 deg, 2.5–3 times these thresholds, to ensure that all elements would be clearly discriminable when presented in isolation.

Experiment 1: Orientation crowding

We first examined the effect of crowding on orientation judgments, with two aims in mind. The first was to measure the effects of crowding on perceived orientation with our stimulus configuration, given that the biases induced by crowding can vary with both the flanker orientations (Solomon et al., 2004) and the target eccentricity (Mareschal, Morgan, & Solomon, 2010). The second was to examine whether the magnitude of crowding varies with target–flanker similarity. It has been shown previously, using Gabor elements, that flankers oriented orthogonally to the target produce less crowding than similarly oriented flankers (Levi & Carney, 2009; Wilkinson et al., 1997). Given that we intend to modulate crowding during combined position and orientation judgments (Experiment 4), we wished to ensure that we could produce similar results with our stimuli.

Methods

Observers judged the orientation (clockwise/counterclockwise relative to horizontal) of the near-horizontal line in the target element (see Figure 2A). For both crowded and uncrowded targets, the horizontal line was presented at orientations between −24° (clockwise of horizontal) and +24° (counterclockwise) in 4° steps using the method of constant stimuli. When crowded, two identical flanker crosses were present, with their horizontal lines at one of eight orientations relative to horizontal: 0°, ±20°, ±40°, ±60°, and 90° (note that with the horizontal line at 90°, elements collapse to a single vertical line). In all cases, the vertical line was fixed. Flanker orientations were tested in two blocks: 0°, ±40°, and 90° in one and ±20° and ±60° in the other. With 8 trials per target orientation, there were 416 trials per block, which each observer repeated 3 times. Uncrowded trials were blocked separately. Responses were scored as the proportion of trials on which observers indicated that the near-horizontal line of the target was counterclockwise of horizontal. Psychometric functions were fit to these data (Wichmann & Hill, 2001), from which we extracted the midpoint (where observers were equally likely to say CW or CCW, indicative of bias) and the threshold for orientation discrimination (the difference in tilt required to shift from 50% to 75% performance).
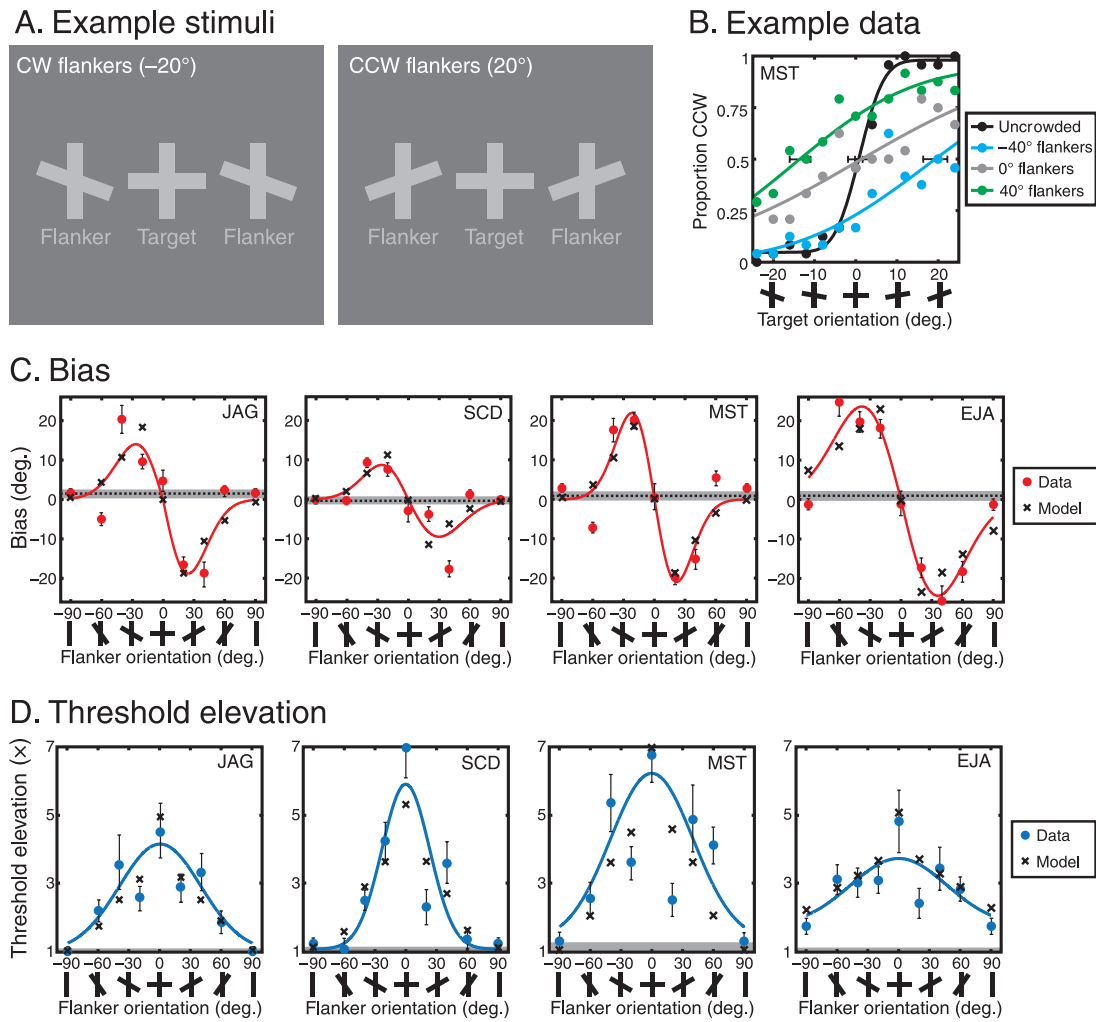
Figure 2. (A) Two sample stimuli from Experiment 1. When flankers were present, the near-horizontal flanker line could be tilted clockwise (left panel, 20° depicted) or counterclockwise of horizontal (right panel). (B) Sample data and psychometric functions for observer MST. The proportion of "counterclockwise" responses is plotted as a function of the orientation of the near-horizontal target line (shown schematically below the *x*-axis). Data are presented for isolated targets (black) and crowded conditions where the near-horizontal line of flankers was oriented at 0° (gray), −40° (cyan), or +40° (green). Error bars depict 95% confidence intervals (CIs) derived from bootstrapping. (C) Midpoints of the psychometric functions for each observer (red points) plotted as a function of the orientation of the near-horizontal flanker lines. Positive flanker orientations produce negative shifts in bias, indicating assimilation. Data are fit with the first derivative of a Gaussian function. Uncrowded biases are shown as a dotted line, while error bars and the gray region depict 95% CIs. (D) As in (C), for threshold elevation (relative to uncrowded baseline). The greatest threshold elevation occurs with near-horizontal flanker elements, with a Gaussian decline for larger orientation differences. In (C) and (D), black crosses show the best-fitting simulations of a weighted averaging model.

Results and discussion

Data from one observer are displayed in Figure 2B, where the proportion of CCW responses is plotted as a function of the target line orientation (illustrated schematically on the abscissa). For uncrowded orientation judgments (black points), the psychometric function is symmetrically distributed around horizontal (0°) with a steep slope (thresholds ~3–5°). In the presence of flankers with an untilted horizontal line (gray points), performance remains unbiased (centered on horizontal), but now larger tilts are required to reliably report the orientation of the target line.

Similarly shallow psychometric functions are evident in conditions with tilted flanking features (blue points: +40°, green points: −40°), with an additional shift of the psychometric function. Counterclockwise flankers (+40°), for instance, shift the entire function leftward to give a negative midpoint value (indicating more "counterclockwise" responses), while clockwise flankers give a positive midpoint value.

The midpoint/bias values for all four observers are plotted in Figure 2C as a function of the orientation of the near-horizontal flanker lines (shown schematically on the abscissa; note that 90° data are repeated as −90° for

symmetry). For all observers, flankers with clockwise orientations produce predominantly positive shifts in bias, indicating an increase in "clockwise" responses. This pattern reverses when flanker lines are tilted counterclockwise. We refer to this as *assimilation*, because the target is biased under crowding to resemble the flankers more closely. Assimilation increases with increasing tilt, up to around ±20–40° where it peaks and then declines until, by ±60°, it is largely abolished. For two of the observers, ±60° flankers induce some repulsion (e.g., counterclockwise flankers increase "clockwise" responses) although the other two observers show either no repulsion or continued assimilation for these flanker orientations. In all cases, the data are well described by the first derivative of a Gaussian function:

$$y = -\gamma x \cdot e^{\left(\frac{-(x-\mu)^2}{2\sigma^2}\right)}. \tag{1}$$

Here, μ is the center of the underlying Gaussian (constrained to be 0°) and γ and σ were two free parameters that gave the scale and variance of the Gaussian, respectively. Curves were fit independently to each half of the data.

Threshold elevation values were obtained by dividing orientation discrimination thresholds in the crowded conditions by uncrowded thresholds. These are presented in Figure 2D, again plotted as a function of the flanker line orientation. For all observers, threshold elevation peaks with untilted flankers (0°) and decreases with increasing tilt of the flankers such that thresholds return to uncrowded levels with flanker line tilts around 60–90°. This pattern is well described by a three-parameter Gaussian function (fitting the variance, baseline, and peak values). Note also that the observers with broader Gaussian functions (especially EJA) are those that tend to show broader tuning for bias in Figure 2C.

In Figures 2C and 2D, black crosses show the simulations of a weighted averaging model of orientation crowding (see Appendix A for details) used previously to predict the crowding of position (Dakin et al., 2010; Greenwood et al., 2009). Briefly, the model consists of four stages. In the first, the veridical orientation values are corrupted by Gaussian noise. The second stage determines the probability of crowding—this probability is tuned for target–flanker dissimilarity in a Gaussian fashion, peaking at matched orientations and declining on either side. If crowding occurs, the model then takes a combination of the target and flanker orientations and produces a weighted average. The final stage rounds the crowded orientation value to a 2AFC decision. We determine the best-fitting parameters using a least-squares fit. The resulting simulations, shown in Figures 2C and 2D, successfully reproduce both the observed pattern of assimilative biases and the Gaussian-shaped pattern of threshold elevation.

In summary, we observe both elevated orientation discrimination thresholds and a systematic bias such that the perceived orientation of the target horizontal line shifts predominantly toward that of the flankers. This pattern differs slightly from the more complex pattern of bias and threshold elevation observed by Solomon et al. (2004), who reported that assimilation occurred only for the smallest target–flanker orientation differences and that larger differences produced repulsion of the target orientation from that of the flankers. However, although this kind of target–flanker repulsion (as in the tilt illusion) is dominant within central and parafoveal vision, recent work demonstrates that more peripheral target presentations strengthen the assimilation of target orientations toward the flankers and minimize repulsion (Mareschal et al., 2010). Our results, collected at 15 deg eccentricity, are thus consistent with crowding being largely characterized by assimilation. Our model further demonstrates that a weighted averaging process provides an excellent simulation of this pattern.

We also demonstrate a clear selectivity of crowding for the orientation difference between target and flanking elements: sensitivity losses are greatest for similar target–flanker orientations and smallest for dissimilar (i.e., near-orthogonal) orientations, as in prior results (Levi & Carney, 2009; Wilkinson et al., 1997). For present purposes, we can thus modulate the strength of crowding by varying the orientation difference between the target and flanker elements—small orientation differences produce a large degree of induced assimilation and threshold elevation, both of which lessen considerably as the orientation difference approaches 90°.

Experiment 2: Position crowding

We next consider the effects of crowding on the perceived position of the lines forming these cross-like elements. Our previous work demonstrates that crowding induces both assimilative bias and decreased sensitivity for such position judgments, in a manner well described by weighted averaging (Greenwood et al., 2009). Because that work used a limited range of position offsets, here we sought to determine if larger positional shifts produce the same degree of bias and threshold elevation (i.e., whether there is tuning for feature positions as observed in Experiment 1 for orientation).

Methods

Observers judged whether the position of the horizontal target line was above or below the stimulus midpoint (see Figure 3A). The horizontal target line was presented
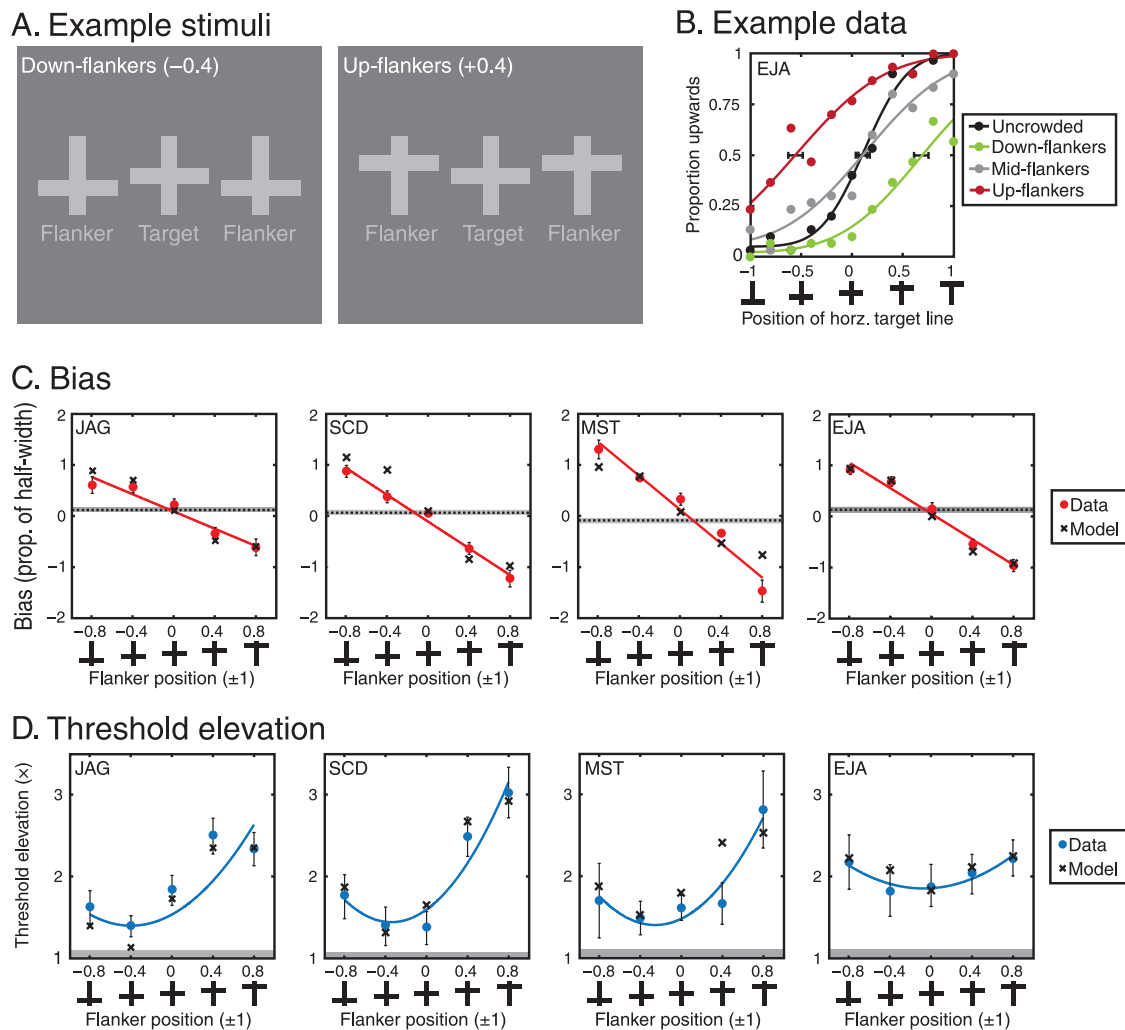
Figure 3. (A) Sample stimuli from Experiment 2. With flankers present, their horizontal line could be displaced downward (left panel, −0.4 shift depicted) or upward (right panel). (B) Sample data for observer EJA, with the proportion of "upward" responses plotted as a function of the position of the horizontal target line (shown on the abscissa). Data are shown for uncrowded targets (black) and crowded conditions with the horizontal flanker line positioned at the midpoint (gray), downward by −0.4 (green), or upward by +0.4 the stimulus half-width (red). Error bars depict 95% CIs derived from bootstrapping. (C) Midpoints of the psychometric functions (bias) for the 4 observers plotted as a function of the position of the horizontal flanker lines (as on the abscissa). Positive shifts in flanker position give negative shifts in bias and vice versa, indicating assimilation. Data are fit with a straight line. Uncrowded biases are shown as a dotted line, while error bars and the gray region depict ±1 *SEM*. (D) As in (C), for threshold elevation relative to the uncrowded baseline. Data are fit with a shifted parabolic function. In (C) and (D), black crosses show the best-fitting simulations of a weighted averaging model.

according to the method of constant stimuli at 11 positions between ±1× the stimulus half-height in steps of 0.18 deg (where −1 is the lowest position and gives an inverted "T"; +1 gives an upright "T"). Two flanker crosses were presented left and right of the target in crowded trials, with their horizontal lines either positioned at the midpoint or displaced above or below the midpoint by ±0.4 or ±0.8 of the stimulus half-width (0.72 deg and 1.44 deg, respectively). The vertical line was fixed in all cases. All conditions were tested in the same block (one uncrowded and five crowded conditions), with 10 trials per target line position to give 660 trials per block. Each observer repeated this three times. Responses were scored as the proportion of trials

that observers indicated "upward" displacements, with psychometric functions fit as before.

Results and discussion

Sample data from observer EJA are displayed in Figure 3B as a function of the target line position shown on the abscissa. Uncrowded data (black points) are symmetrically distributed around the stimulus midpoint with a steep slope (thresholds were 0.16–0.22 of the stimulus half-width for each observer). With flankers present, larger positional offsets were required to accurately report the target line

position (i.e., thresholds were elevated). As before, there is also a shift in the midpoint of the crowded data that follows the position of the flanker lines: upward-shifted flankers (+0.4 stimulus half-width; red points) produce an increase in "upward" responses, shifting the function leftward to give a negative midpoint value, with the opposite for downward-shifted flankers (−0.4 half-width; green points).

The resulting midpoint/bias values are plotted in Figure 3C as a function of the position of the flanker horizontal lines (illustrated under the abscissa). For all observers, flankers with horizontal lines positioned below the stimulus midpoint produce positive shifts in bias, indicating increased "downward" responses. This assimilation also occurs for upward-displaced flankers to give negative bias values. Unlike the effects of crowding on orientation, bias increases with larger flanker line positions (i.e., there is no tuning evident) and the data are well described by a simple straight-line fit.

Threshold elevation values (crowded divided by uncrowded thresholds) are shown in Figure 3D. All observers show U-shaped functions (as predicted in our earlier study; Dakin et al., 2010), with the least threshold elevation for flankers with a horizontal line located between the stimulus midpoint and the next most downward position (−0.4 half-width units). Displacements away from the midpoint produce more threshold elevation, with upward (and more peripheral) positions tending to give more threshold elevation than downward positions. This pattern is well described by a three-parameter function combining a straight line with a parabola:

$$y = m(x - \mu)^2 + b. \qquad (2)$$

Here, m is the slope, μ is the inflection point, and b is the offset value. The slope of this function simulates the increase in positional noise for more eccentric positions, consistent with the general decline of positional sensitivity in the periphery (Morrone, Burr, & Spinelli, 1989; Rentschler & Treutwein, 1985).

These effects are consistent with our prior results using these stimuli, which were well described by a weighted averaging model (Dakin et al., 2010; Greenwood et al., 2009). Accordingly, black crosses in Figures 3C and 3D depict simulations with the best-fitting parameters of a four-stage weighted averaging model for position crowding (see Appendix A for details). The general form of this model is similar to that described above for orientation. Both the pattern of increasing bias and the U-shaped functions for threshold elevation are well captured with these simple computations. A probabilistic weighted averaging approach can thus simulate crowding for both position and orientation, despite differences in the precise pattern of bias and threshold elevation. The differences arise largely through the narrower tuning for orientation differences than for position. Broad tuning for position

differences (relative to stimulus dimensions) means that crowding does not significantly decline with large positional offsets, giving a linear pattern of positional bias. The U-shaped pattern of threshold elevation arises when these large flanker position offsets are incorporated into the average, causing responses to approach floor/ceiling values and the psychometric functions to flatten out, giving an increase in threshold.

This could suggest that positional crowding is less "tuned" than orientation crowding, though it is unlikely that there would never be a release from crowding with larger positional offsets. In extreme cases, feature positions outside the interference zone would give a release from crowding, though at 15 deg eccentricity this would extend to approximately ±7.5 deg Elongation of our stimuli on the vertical axis would make it possible to test this hypothesis, though this would also introduce uncertainty about the stimulus midpoint that would degrade performance substantially. For our present purposes, it is sufficient that crowding is clearly tuned for orientation, and we use this property to modulate the strength of crowding in Experiment 4.

Experiment 3: Position judgments with tilted lines

The results of Experiments 1 and 2 demonstrate that crowding changes the perception of both orientation and position in letter-like elements. When crowded, targets appear more similar to the flankers. To examine the effect of crowding on conjoint judgments of position and orientation, it is thus important to ensure that changes in orientation do not produce significant changes in position judgments. We thus performed a control experiment to examine position judgments for a single, uncrowded target with a range of tilts applied to the near-horizontal feature.

Methods

On each trial, a single cross-like stimulus was presented at 15 deg in the upper periphery. The target line was either fixed at horizontal (0° tilt) or rotated counterclockwise by 20°, 40°, or 60°. For each tilt, the target line was presented at one of 11 positions along the fixed vertical line, with the veridical position considered to lie at the intersection of the two lines, regardless of tilt. Observers judged whether the midpoint of the target line was above or below the stimulus midpoint (2AFC). Each position was shown 10 times in a block and 3 blocks were run according to the method of constant stimuli.
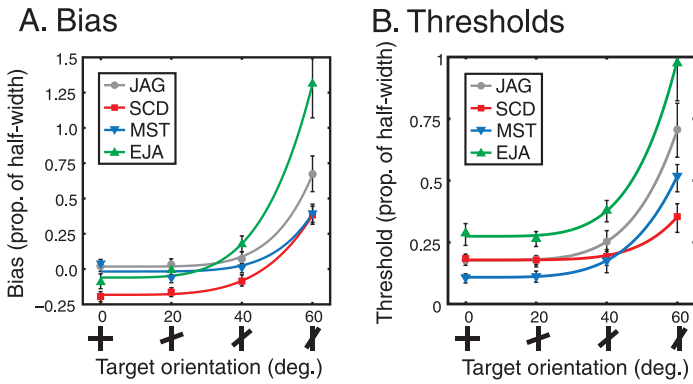
Figure 4. Uncrowded position discrimination as a function of target line orientation. (A) Midpoints of the psychometric functions (bias) for 4 observers plotted as a function of the orientation of the near-horizontal target line in the flankers (illustrated on the abscissa). Data are fit with a three-parameter power function; error bars show 95% CIs. (B) As in (A), for position identification thresholds. In both cases, performance is stable until tilts exceed ±40° when thresholds and bias rise steeply.

Results and discussion

Psychometric functions were fit as before, and the midpoint and threshold values are plotted in Figures 4A and 4B, respectively. Considering thresholds first, with untilted crosses observers required displacements between 0.1 and 0.3 the stimulus half-width (i.e., spatial displacements between 0.09 and 0.27 deg). Thresholds remained largely unchanged for line orientations between 0 and 40°, with a slight increase at 40° and a sharper increase at 60°. Performance at this largest orientation is such that, for some observers, lines need to be positioned almost at the extreme ends of the crosses to be discriminable. A similar pattern occurred for the midpoint/bias values in Figure 4A—observers show a slight, idiosyncratic, bias for untilted lines that is largely constant for the first 0–40° of tilt. A large degree of bias becomes evident with tilts of 60°, which in all cases is positive, indicating that observers were predisposed to indicate "downward" positions with these elements, regardless of their actual position. This was not related to the direction of the orientation, as both clockwise and counterclockwise rotations (examined during pilot testing) gave identical patterns of bias. For both midpoints and thresholds, the data are well described by a three-parameter power function:

$$y = \alpha x^\gamma + b. \tag{3}$$

We suspect that this pattern arises because the separation of the two stimulus lines becomes difficult to discern at the largest tilts, making the position of the near-horizontal element difficult to estimate. It is curious that this led all four observers to increasingly respond "downward,"

though regardless of the origin of this effect, the important point is that when the horizontal line is tilted less than ±40°, perception of position is not affected by orientation. This defines the range for reliable measurement of the joint perception of orientation and position. As we discuss shortly, these threshold estimates also provide an important input to our models of crowding.

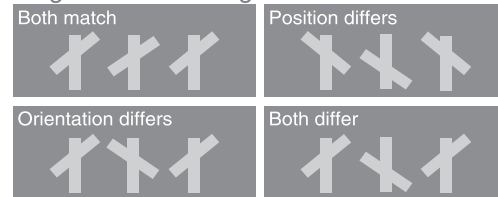# Experiment 4: Conjoint position and orientation crowding

Our final experiment examined whether the effects of crowding on position and orientation operate through a single combined mechanism or multiple independent processes. The results of Experiment 1 demonstrate that cross-elements with similar orientations (at 15 deg eccentricity) produce large degrees of assimilation and threshold elevation for orientation judgments, while those with near-orthogonal orientations produce little to no effect. We can thus modulate the strength of crowding using orientation and examine the concomitant effects on position judgments. However, because Experiment 3 indicates that large tilts can decrease the reliability of positional judgments, line tilts need to be kept below ±40°.

For this task, observers were required to make conjoint judgments of the orientation and position of the near-horizontal target line in the presence of two flankers. These judgments were measured under four conditions, depicted in Figure 5A. In the *both match* condition, the orientation of target and flanker near-horizontal lines was both CW (or CCW) and the positions both upward (or downward), though the precise value of each cue varied. In the other three conditions, either the orientation of the target and flankers differed (i.e., one CW and one CCW but with matched position cues), the position differed (but the orientation cue was matched), or both differed. Across these four conditions, the orientation cue could be small (±10° for flankers and ±5° for targets) to cause *strong crowding* or large (±40° for flankers and ±35° for targets) to cause *weak crowding*. Note that these orientations never exceeded ±40° to maintain the reliability of position judgments.

With this 4 × 2 design, we can separate the predictions of independent and combined models of crowding. We know from Experiments 1 and 2 that crowding causes both threshold elevation and assimilation in each feature domain. The key difference between the models is the probability with which these effects occur: in the independent model, crowding can occur for one feature and not the other, while the combined model is all or none. For the first three conditions, we do not expect a difference between the models. In the *both match* condition, both

A. Stimulus conditions

Target-flanker configurations
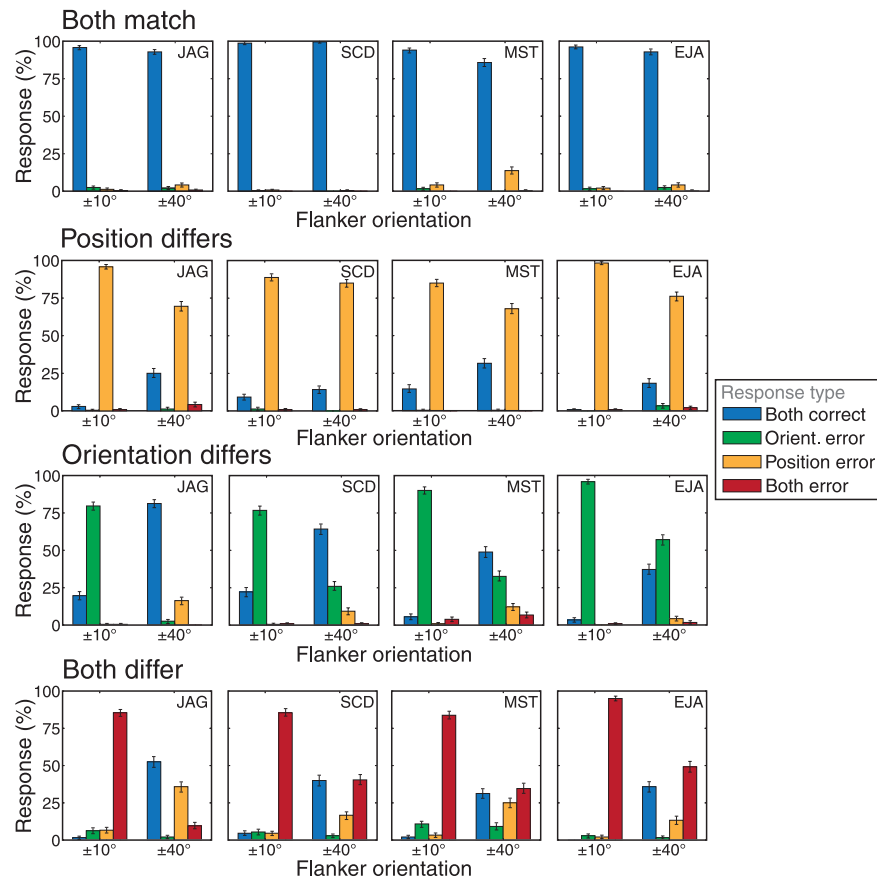
Flanker orientations
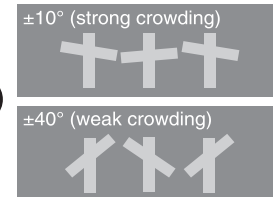
B. Results

Figure 5. (A) Examples of the stimulus conditions in Experiment 4. As well as four target–flanker configurations (each with four combinations of position and orientation values), elements had one of two orientation levels: small tilts that give strong crowding or large tilts that give weak crowding. (B) Results of the conjoint experiment, where responses are characterized as "both correct" (blue bars) or an error in orientation (green), position (yellow), or both features (red). Responses are plotted separately for the condition with small tilt (±10° flanker orientations) and large tilt (±40°). Error bars show 95% CIs derived from bootstrapping. The four target–flanker configurations are plotted in distinct rows: the target and flankers either *both match* (top row), their *position differs* (second row), the *orientation differs* (third row), or *both differ* (bottom row).

models predict mostly correct responses—without crowding, the target should be perceived veridically, and even when crowding does occur, biases in each feature domain will shift the perceived target position and orientation in the correct response direction. In the *position differs* condition, both models predict a high proportion of position errors, but because the tilt cue is always matched, the effect of any orientation crowding is obscured. Both models also predict a high proportion of orientation errors in the *orientation differs* condition with small tilt and a

release from crowding with larger tilts. Results from these conditions are nonetheless essential for constraining our computational models.

The key condition that separates predictions for the two models is the *both differ* condition. With strong orientation crowding (i.e., small tilt), the combined model predicts a high proportion of errors in both feature domains. When orientation crowding is weak (i.e., large tilt), the combined model predicts a low proportion of errors in both feature domains, because its operation is

"all or none." The independent model predictions depend on the probabilities of crowding for each feature—errors in both domains are the outcome of two independent processes. Nonetheless, when orientation crowding is released with large tilts, the independent model predicts no effect on the crowding of position, making position errors the predominant response type.

Methods

Here, the near-horizontal lines of the target and flankers varied in both position and orientation. The near-horizontal line was positioned above or below the stimulus midpoint by ±0.4 of the target half-width and ±0.8 for the flankers. In the *strong crowding* conditions, the orientation of the near-horizontal target line was presented at ±5° for each position, while flanker elements were tilted by ±10° from horizontal. For the *weak crowding* conditions, the target was oriented at ±35° and flankers at ±40°. In all cases, the vertical lines were fixed in position and orientation.

Notice the asymmetry in feature values above: flanker positional offsets and orientations were always more extreme than the target. This ensured that any averaging of position or orientation would not produce an outcome centered on the decision boundary (i.e., the stimulus midpoint for position or horizontal for orientation). Were this the case, errors would be predominantly determined by noise. By utilizing the assimilation caused by crowding and biasing the outcome of this process away from the decision boundary, we can ensure that crowding-based errors are more diagnostic.

In total, 32 combinations of target and flanker elements were tested, which can be grouped into 8 conditions: for each of the *strong* and *weak* crowding conditions, there are four target–flanker configurations as depicted in Figure 5A (*both match*, *position differs*, *orientation differs*, or *both differ*). Observers indicated both the position and orientation of the near-horizontal element of targets, making a 4AFC response using the keyboard (up/CCW, up/CW, down/CCW, or down/CW). *Strong* and *weak* crowding conditions were run in separate blocks, with 20 trials of each target–flanker configuration at a time, to give 320 trials per block. Observers completed three runs of each.

Results and discussion

Responses from the 32 target–flanker configurations were sorted into the 8 conditions depicted in Figure 5A and scored as the percentage of total responses that were either: correct for both features, an error in the sign of the target position (e.g., an "upward" response to a downward target), an error in orientation sign, or an error for both. Results are presented in Figure 5B for ±10° flanker tilt (strong crowding) and ±40° flanker tilt (weak crowding), separately for each observer. In the *both match* condition

(top row), observers were correct in indicating both the target position and orientation 86–99% of the time with both small and large tilts. This is consistent with the assimilative nature of crowding; even if the targets were strongly crowded, the perceived orientation and position would still be shifted toward the correct answer. Of the errors that that did occur, position errors were most common, particularly in the "weak crowding" condition, where tilts were larger.

In the *position differs* condition (second row), position errors were dominant with both levels of flanker tilt. That is, with flankers that differ in the sign of their positions (e.g., "upward" flankers and a "downward" target), observers most commonly reported an offset direction that matched the flankers. This is consistent with the strong assimilation shown here (Experiment 2) and elsewhere (Greenwood et al., 2009). Correct responses (for both features) were the second most common, indicating that crowding did not occur on all trials. Note that even with large tilt in this condition, we do not expect any release from crowding in the ±40° flanker orientation condition since the orientation sign of target and flankers is always matched. We suspect that the reduced rate of position errors with large tilts is due to uncertainty regarding the position of tilted lines (as in Experiment 3) rather than a reduction in crowding strength, as we consider in the modeling section.

For the *orientation differs* condition (third row) with *strong crowding* (±10° flankers), the dominant response is an orientation error, reflecting the strong assimilative crowding that occurs under these conditions in Experiment 1 and elsewhere (Mareschal et al., 2010; Parkes et al., 2001). With weak crowding (±40° flankers), where stimuli differ in their orientation but not the sign of their position, the rate of orientation errors is reduced considerably. For three observers, this produces a greater rate of "both correct" responses, while for EJA there remains a majority of orientation errors despite their overall rate being reduced. This is further evidence for the reduction of crowding with large orientation differences, as shown here (Experiment 1) and elsewhere (Levi & Carney, 2009; Wilkinson et al., 1997).

The key condition is when *both differ* (bottom row). With *strong crowding*, errors are most often for both features at once ("both errors") rather than for either feature in isolation. In the *weak crowding* condition, the probability of both errors was strongly reduced for all observers. For three observers, "both errors" remained the predominant response type, with "both correct" responses second most common, while for JAG correct responses became most common. We also observe an increase in position errors in this condition, although this is never the most common response type.

Results in the *both differ* condition with strong crowding (±10° flankers) are inconsistent with the simplest independent model of crowding. Independent, low probabilities of crowding in either domain should have produced a greater proportion of position and orientation

errors in this condition. It is nonetheless still consistent with an independent crowding model in which *both errors* are simply the most likely outcome of a crowding process (i.e., when the probability of crowding is high for each feature). The results from the weak crowding condition are harder to explain with independent crowding processes. The release from crowding in the orientation domain should have had no effect on positional crowding, resulting in a preponderance of position errors. Although all observers do show a slight increase in position errors in this condition, there is not a simple replacement of *both errors* with *position errors* as the independent model predicts. Rather, in all cases, correct responses become the next most likely response, as the combined model predicts. We propose that the positional errors in this condition arise from the positional uncertainty with tilted lines, as in Experiment 3, and suggest that our results support a combined model of crowding—a notion we now set out to test explicitly.

Computational modeling

To test the predictions of independent and combined models, we developed two computational models implementing these processes (see Appendix A for details). Each is a four-stage process, depicted schematically in Figure 6A. The major difference is whether there are independent probabilities of crowding for position and orientation or a single combined probability. Both share an identical first stage and begin with noisy estimates of the position and orientation of the near-horizontal line in target and flanker elements. Because positional error depends on orientation (Experiment 3), the magnitude of positional noise was determined by a power function dependent on orientation.

The second stage is a probabilistic determination of whether crowding occurs or not. This is similar to the processes employed in our recent model, where the target–flanker separation set the probability of whether crowding occurred (Dakin et al., 2010). Here, it is the orientation difference between target and flanker elements that sets this probability. The combined and independent models differ at this point. For the independent model, the probability of orientation crowding is set by a Gaussian function that peaks at matched target–flanker orientations and declines on either side. Because position crowding was untuned in Experiment 2, the probability of position crowding was set with a single free parameter. In contrast, for the combined model, the probability of crowding was set only by the orientation difference between the elements (though a combined and perhaps multiplicative feature difference would most likely be used here, the probability of position crowding is factored out by being unchanged in our manipulations, as in Experiment 2). In this way, with sufficiently large orientation differences, crowding would not occur for either position or orientation.

The effects of crowding were then applied in the third stage of each model. Following the simulations of Experiments 1 and 2, both models applied a weighted average of target and flanker elements for both features. The only difference between the models in this stage was that the prior "gating stage" could allow crowding for one feature and not the other in the independent model, whereas the combined model was "all or none." Note that although we treat these computations as a distinct stage, we do not propose that this need be physiologically distinct from the "gate" in stage two that determines whether crowding occurs or not. Finally, these estimates of position and orientation were converted to a binary decision regarding each feature (up/down or CW/CCW), which allowed a 4AFC decision regarding the target identity. Note that while our prior model used "reference repulsion" to push responses away from the decision boundary (Greenwood et al., 2009), this was not required here as we do not simulate the actual perceived values of position and orientation.

The best-fitting parameters were selected by minimizing the least-squares error between simulated responses and those of either individual observers or the average. Fits were to the whole eight-condition data set in each case. Simulated responses for the averaged data are displayed in Figure 6B for both models (*independent*: closed triangles, *combined*: open circles) and overlaid on the averaged responses of our observers (bars, colored as in Figure 5). Both models capture performance in the *both match* condition, since here, even if crowding occurs, it will shift perceived target position and orientation in the correct direction. Similarly, the *position differs* condition is similar for both models because the matched target–flanker orientations ensure a high probability of crowding. Notice the elevation in position errors in the large tilt (±40°) condition, which arises through the power function for positional noise. In the *orientation differs* condition, both models again replicate the observed errors since the flanker positions match the target and either the probability of orientation crowding (for the independent model) or the total probability (combined model) is modulated by orientation.

What separates the models is the *both differ* condition. With *strong crowding* (±10° flankers), the combined model necessarily produces conjunction errors because of its all-or-none operation. The independent model can similarly produce these errors when the probability of crowding in both domains is sufficiently high (best-fitting probabilities here were 0.97 for position crowding and 0.95 for orientation). Because crowding was so strong for both features, conjunction errors were simply the most likely outcome. This relationship breaks down in the *both differ* condition with *weak crowding* (±40° flankers). Here, the independent model is able to predict the decreased conjunction errors because the probability of orientation crowding is reduced (to 0.31 at this orientation difference). However, it incorrectly predicts that the dominant response will be position errors because the probability of position crowding is unchanged by the change in
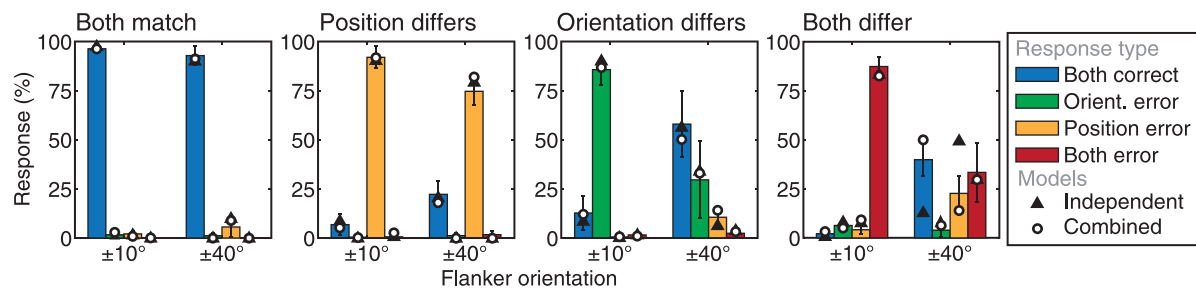
A. Model schematics
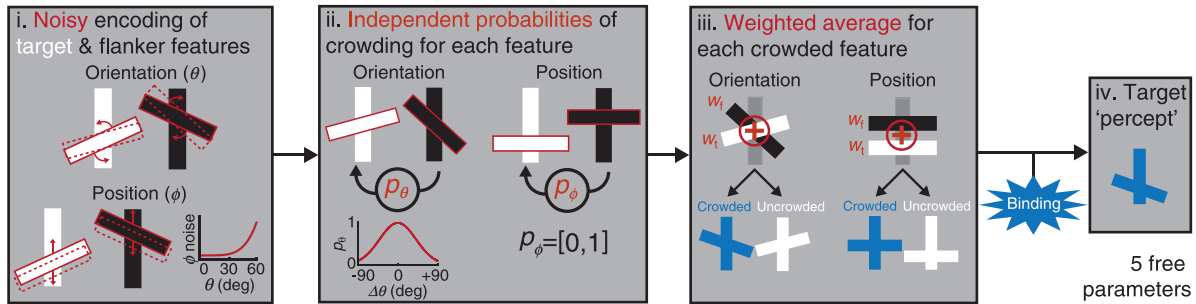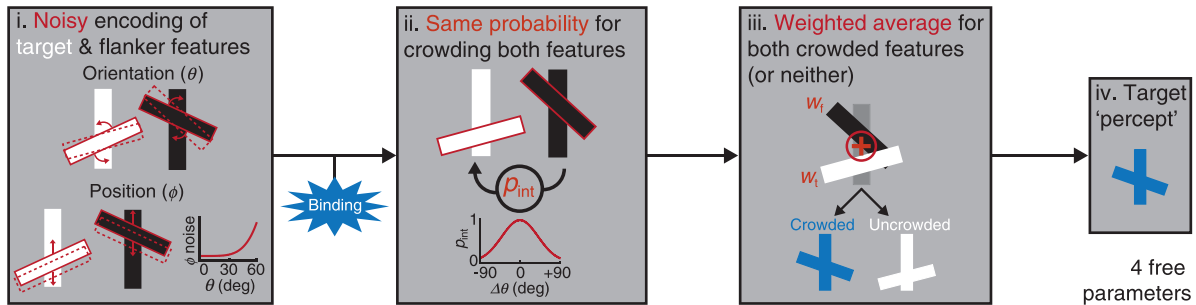
B. Averaged data & simulations

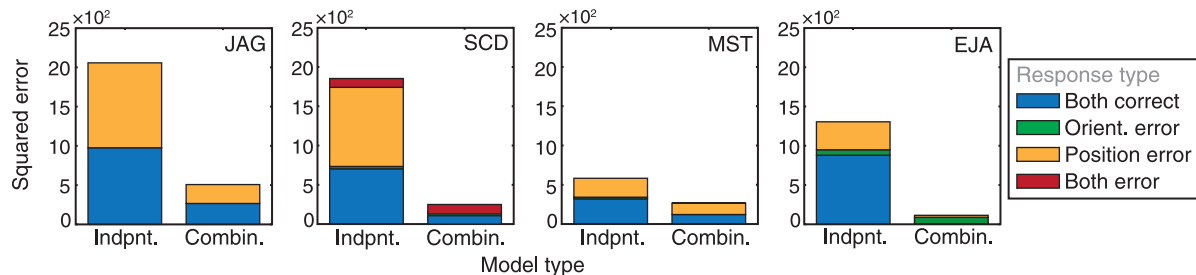## C. Error in individual fits (both differ, ±40° flankers)

Figure 6. (A) Schematic of the two crowding models tested. (B) Simulations from the two models, overlaid on averaged data plotted as in Figure 5. Independent model responses are shown as closed black triangles and combined model responses as open white circles. Both models perform equivalently for all conditions except the *both differ* condition with *weak crowding* (±40° flankers) where the combined model more closely matches performance. (C) A stacked bar plot showing squared error in the individual fits to the *both differ* condition with *weak crowding* (values are divided by 100 for clarity) for the two models. The types of errors made by the models are color-coded, showing the total proportion of errors made in each response category. For all observers, the independent model produces more error (i.e., performs worse) than the combined model as it predicts both too few correct responses and too many positional errors.

orientation. The proportion of correct responses is under-predicted as a consequence. In contrast, the combined model successfully predicts that the decrease in conjunction errors is accompanied by an increase in correct responses because it is all or none. Notice that although both models predict the increase in position errors, this is due to the positional noise induced by tilted elements and not an increase in position crowding.

The above describes fits of the models to averaged data, but the same is true for individual fits. Table A2 of Appendix A displays the squared error from the fits to both individual and averaged data (i.e., the squared difference between simulated and observed responses). In all cases, the combined model better characterizes the observed responses and generates less error than the independent model. For illustration, Figure 6C plots the squared error from individual fits in the *both differ* condition with *weak crowding*, the source of the most error. It is clear that the independent model performs worse than the combined model. These stacked bar plots show the reason for this: most of the independent model failures occur because the model under-predicts correct responses and over-predicts position errors. Note that the independent model could actually simulate responses in this latter condition if the probability of positional crowding were lower than that of orientation crowding. However, this would then under-predict the rate of positional crowding in the *position differs* condition as well as the rate of conjunction errors in the *both differ* condition with *strong crowding*. It is through fitting the model to all of the conditions simultaneously that we can separate the models. Together then, our data unambiguously reject the possibility of an independent set of crowding processes and support instead a process that operates on combined features.

Discussion

Our aim was to determine whether the crowding of multiple features (here, orientation and position) occurs independently or via a unitary mechanism that operates after feature binding. Our results and modeling demonstrate that crowding has a single probability of occurrence that affects position and orientation in an all-or-none fashion and that the perceptual effects of crowding are similar in both of these feature domains.

We first demonstrated that crowding operates similarly for position and orientation. In both cases, crowding elevates discrimination thresholds and introduces a largely assimilative bias that caused the target to resemble the flankers. This is consistent with prior work demonstrating assimilation in both orientation (Greenwood et al., 2010; Mareschal et al., 2010; Parkes et al., 2001) and position (Dakin et al., 2010; Greenwood et al., 2009). Assimilation is thus the dominant mode of crowding for eccentricities of 10–15 deg and above (as in the present study). At closer eccentricities, repulsion of the target orientation by flankers has also been reported under crowded conditions (Mareschal et al., 2010; Solomon et al., 2004). Our proposal of a unitary crowding mechanism therefore makes the prediction that similar repulsion would occur

for the crowding of position at these eccentricities. Repulsive effects on position have indeed been found previously (Levi, Li, & Klein, 2003), making this likely.

The strongest support for a unitary crowding mechanism comes from our conjoint position and orientation experiment (Experiment 4). We demonstrate that strong crowding of both features induces a high likelihood of conjunction errors, a pattern that can be simulated both with independently determined errors and with combined all-or-none errors. However, when the strength of crowding was reduced for orientation, this also reduced the probability of crowding for position, causing an increase in correct responses (as the combined model predicts) rather than an increase in position errors (as the independent model predicts). The clear failure of the independent model under these conditions is contrasted with the markedly better performance of the combined model.

One potential shortcoming is that we may have selected two features that are inextricably linked, where others could be more independent (e.g., Fujisaki & Nishida, 2010). The linkage between orientation and position can potentially be seen in observations such as the oblique effect for vernier judgments (Leibowitz, 1955; Saarinen & Levi, 1995), though these thresholds are not a pure measure of positional acuity (Carney & Klein, 1999). Likewise, our own results suggest a potential linkage between the two features (Experiment 3), though this is likely an issue of resolution rather than interaction. Nonetheless, at the outset, we did not require complete independence between our candidate features, merely that the visual system is tuned selectively for their dimensions. There are many inter-relations between more clearly distinct variables, such orientation and size (Finger & Spelt, 1947) or motion and stereoscopic depth (Edwards & Badcock, 2003), which (we think) do not rule out their candidacy as visual "features." As outlined in the Introduction section, the role of position and orientation in determining letter identity, the selectivity of the visual system along these dimensions, and their production of both pop-out and texture segmentation suggests that relative position and orientation are basic features in visual processing.

The generality of our results is also bolstered by a range of prior work demonstrating a release from crowding when target and flankers differ in color, contrast polarity, or binocular disparity (Butler & Westheimer, 1978; Hess et al., 2000; Kooi et al., 1994; Sayim et al., 2008). The combined nature of crowding is thus unlikely to be restricted to position and orientation. Our results differ from those of Põder and Wagemans (2007) however, who found a mix of independent and conjunction errors on a task requiring judgments of color, spatial frequency, and orientation. We suggest that this is due to their use of large stimulus differences in each of the three feature domains, which would reduce crowding (Kooi et al., 1994; Levi & Carney, 2009; Wilkinson et al., 1997). A mix of independent and conjunction errors would be

expected if the stimuli were crowded on some trials and not others.

An alternative explanation of our findings is that distinct features are crowded via independent processes that are linked in some fashion. For instance, a release from crowding in one feature domain could cue a release from crowding at the same location within other feature maps through spatially precise interactions between stimulus features. For this to occur however, the binding problem must have already been solved—interactions between the precise locations of distinct features would indicate a lack of feature uncertainty that would avoid both crowding and the binding problem altogether. Rather, mislocalizations between these feature maps may be a significant source of errors in feature binding (Neri & Levi, 2006). Another alternative is that top-down connections could release crowding in other feature domains when one of the features is correctly perceived. This type of architecture is seen in the Guided Search model, where a top-down system can determine the location of objects with a particular tilt when this is known in advance (Wolfe, 1994). However, observers did not know in advance which orientation or position would be present on each trial, making this kind of top-down guidance unlikely.

Rather, we suggest that position and orientation estimates for an object are combined prior to crowding. That is, crowding occurs after (or during) the binding process that takes visual features and combines them into objects. This places crowding as a relatively late-stage process within the visual processing hierarchy. Accordingly, we know from prior studies that crowding must occur at least at the level of binocular cells in V1, since it is undiminished by dichoptic presentation of the target and flanker elements (Flom, Heath, & Takahashi, 1963). Recent work also demonstrates that crowding depends not on the physical position of target and flanking elements but on their perceived positions (Dakin, Greenwood, Carlson, & Bex, 2011; Maus, Fischer, & Whitney, 2011) and that the magnitude of crowding depends on an awareness of the flanker elements (Wallis & Bex, 2011), both properties to be expected of a higher order process. Together, these results support a minimally two-stage model of crowding, where features are first detected and subsequently pooled (He, Cavanagh, & Intriligator, 1996; Pelli, Palomares, & Majaj, 2004). We argue that this later combinatorial stage of crowding originates at, or beyond, the site of feature binding.

Though it was first suggested that crowding did not strongly affect the initial stages of feature detection, with no effect on orientation-selective adaptation (He et al., 1996) or contrast detection thresholds (Levi, Hariharan, & Klein, 2002; Pelli et al., 2004), subsequent studies demonstrate that modulations of the strength of crowding can reveal these effects (Blake, Tadin, Sobel, Raissian, & Chong, 2006; Põder, 2008). Indeed, recent work suggests that the effects of crowding may operate at multiple levels of the visual hierarchy (Whitney & Levi, 2011). In our model, the stage determining the probability of crowding is separate from both feature detection and the application of the effects of crowding on feature appearance. We do not suggest that these latter two stages need necessarily be separate however. It is possible, for instance, that the alterations of feature appearance could be achieved via feedback to the initial feature detection stage (thus occasionally modulating adaptation and contrast detection effects), though it is equally likely that these effects occur at a higher level stage that is more "object-based" in operation (producing both the occurrence of crowding and the associated changes in appearance). Our present consideration of the "singularity" of crowding thus applies only to the application of crowding to different feature dimensions—our results do not speak to the singularity of the mechanisms within the visual hierarchy. However, at the very least, our results demonstrate that the "switch" that determines the probability of crowding is a singular process that affects all of the features within an object.

As with crowding, the process of feature binding is also likely to have at least two stages: one in which features are encoded in distinct retinotopic maps and another where they are combined to form objects (Treisman & Gelade, 1980). In fact, these two processes share several similarities. When feature binding fails, we see illusory conjunctions of features that belong to distinct objects (Prinzmetal, Henderson, & Ivry, 1995; Treisman & Schmidt, 1982). The rate of these illusory conjunctions increases with both target eccentricity and inter-object similarity (Ivry & Prinzmetal, 1991; Prinzmetal et al., 1995), as it does for crowding (Bouma, 1970; Kooi et al., 1994). Impaired feature binding is also seen in the fovea of those with strabismic amblyopia (Neri & Levi, 2006), where crowding is similarly elevated (Flom, Weymouth et al., 1963; Levi & Klein, 1985). These similarities have led some to suggest that crowding and feature binding may be one and the same (Pelli & Tillman, 2008). That is, illusory conjunctions could reflect excessive crowding that links features from adjacent objects. Our results are consistent with this possibility: the same process that ensures the correct features are assigned to an object would likely also give a release from crowding for all features when there is a clear difference in one feature domain.

Finally, our results suggest that crowding takes *objects*, rather than elementary features, as its basic unit of organization. This is consistent with prior work demonstrating that crowding affects letters in their entirety rather than individual strokes (Martelli, Majaj, & Pelli, 2005), as well as the observation that tuning for the crowding of faces is determined holistically (Farzin, Rivera, & Whitney, 2009; Louie, Bressler, & Whitney, 2007). We demonstrate here that the features of letter-like stimuli influence the probability of crowding in an ensemble fashion—if crowding does not occur in one feature domain, it will not occur in the other. As the net effect

of crowding appears to be a simplification of the peripheral visual field (Freeman & Simoncelli, 2011; Greenwood et al., 2010), these object-based computations could maintain the general structure of adjacent objects while also increasing their similarity. The end result would be a more structured simplification of the peripheral visual field that is more efficiently encoded by the limited neural resources afforded to these regions.

Appendix A: Computational models

Single-feature crowding models

To simulate the crowding of position and orientation in isolation, two 4-stage models were developed. For orientation, the first stage involved noisy estimates of the orientation of the near-horizontal line in target and flanking elements. For each of the element orientations, this estimate $\theta*$ is calculated as

$$\theta* = \theta + n_o\sigma, \tag{A1}$$

where θ represents the veridical target or flanker feature orientation, σ represents Gaussian error, and n_o is a free parameter that sets the magnitude of this error. The second stage is a probabilistic determination of whether crowding occurs or not. Here, the probability of orientation crowding (p_θ) is set between 0 and 1 by a Gaussian function:

$$p_\theta = \alpha e^{\frac{(-\delta\theta-\mu)^2}{2\sigma_w^2}}, \tag{A2}$$

where $\delta\theta$ represents the orientation difference between target and flanker elements, σ_w sets the width of the tuning function (the second free parameter), α sets the peak of the tuning function (parameter three), and μ was centered on 0°. If crowding occurs, the effects are applied

in the third stage with a weighted average of the orientations of each element:

$$t_c = \frac{t_v w_t + f_1 w_f + f_2 w_f}{w_t + 2w_f}. \tag{A3}$$

Here, t_c is the crowded orientation of the target, t_v is the veridical target orientation (corrupted by noise in stage one), f_1 and f_2 are the flanker orientations, w_t is the weight (0–1) of the target value in the average (parameter four), and w_f is the flanker weight, equal to $1 - w_t$. These estimates of orientation were then converted to a 2AFC response and simulated 1024 times per target orientation. The best-fitting parameters were selected as those that minimized the least-squares error between the predicted midpoint and threshold values and those plotted in Figure 2.

The general structure of the model for position crowding was identical to that for orientation. The first stage involved noisy estimates of the position of the near-horizontal line in target and flanking elements. Because we expect positional noise to rise with eccentricity, we utilize a straight line with two free parameters. These noisy feature positions were clipped between ±1, corresponding to the upper and lower extremes of the stimuli. The probability of positional crowding was then determined using a Gaussian function, again with two free parameters as in Equation A2, substituting positional differences for the orientation difference used previously. If crowding occurred on a given trial, a weighted average was employed as in Equation A3. The final values were then converted to a 2AFC response and best-fitting parameters again determined using the least-squares fit to the data, with the final result shown in Figure 3.

Multi-feature crowding models

Two models were developed to test the predictions of independent and combined models, each with four stages (shown in Figure 6A). Both were identical in the first stage, with noisy estimates of the position and orientation

	Independent model					Combined model			
	n_o	n_p	σ_w	p_ϕ	w_t	n_o	n_p	σ_w	w_t
JAG	5.04	0.45	34.96	0.95	0.28	5.78	0.39	35.62	0.33
SCD	4.46	0.40	53.98	0.96	0.30	5.45	0.36	52.46	0.30
MST	4.35	0.41	60.00	0.84	0.30	4.90	0.41	59.18	0.32
EJA	3.98	0.32	76.24	1.00	0.29	4.88	0.38	73.93	0.23
Average	4.66	0.40	53.65	0.97	0.28	5.85	0.37	52.48	0.27

Table A1. Best-fitting parameters for the independent and combined crowding models for each observer and the averaged data set. The independent model (left table) has five free parameters; the combined model (right) has four. Of these, n_o gives the magnitude of orientation noise, n_p gives the magnitude of position noise, σ_w determines the width of the orientation tuning function for the independent model and the width of the combined feature tuning for the combined model, p_ϕ is the probability of positional crowding (not present in the combined model), and w_t is the weight of the target in the crowded average.

	JAG	SCD	MST	EJA	Average
Independent model	2729.54	2729.07	965.66	1452.51	1659.96
Combined model	1166.92	961.47	791.41	386.08	441.62

Table A2. Squared error values for the final simulations of each model (fit to either individual or averaged data) using the best-fitting parameters from Table A1. In each case, there is more error in the independent model than the combined model.

of the near-horizontal line in target and flanking elements. For orientation, this was calculated using Equation A1. Because positional error depends on orientation (Experiment 3), the magnitude of positional noise was here determined by a power function dependent on orientation:

$$\phi^* = \phi + (n_p \theta^\gamma)\sigma, \qquad (A4)$$

where ϕ is the veridical target or flanker feature position, σ represents Gaussian positional error, and the bracketed function is a power function. The magnitude of positional error is thus set by an interaction between the element orientation (θ) and a free parameter n_p that sets the overall magnitude. The γ parameter was determined from the power function fit to the averaged data of Experiment 3 (fixed at 5.54).

As in the single-feature models, the second stage is a probabilistic determination of whether crowding occurs or not. The combined and independent models differ here: for the independent model, the probability of orientation crowding was set by a one-parameter Gaussian function (Equation A2, with σ_w as a free parameter and α set to 1), while the probability of positional crowding was set with a single free parameter p_ϕ, which could vary between 0 and 1. For the combined model, the probability of crowding was set only by the relative orientation of the elements using a single-parameter Gaussian function (Equation A2). In this way, if the orientation difference were sufficiently large, crowding would not occur for either position or orientation. The effects of crowding were applied in the third stage of each model using the weighted averaging in Equation A3. The only free parameter in this stage is the target weight, which was the same for both position and orientation. The two models differed here—for the independent model, the "gating" stage could allow crowding for one feature and not the other, whereas the combined model was "all or none."

These estimates were converted to a binary value regarding each feature (up/down or CW/CCW) to give a 4AFC decision. The best-fitting parameters were selected as those that minimized the least-squares error between the predicted responses and those in Figure 5. Parameters were fit for each observer separately as well as to the averaged responses but were fit to the entirety of the data set in each case. This is important because allowing parameters to vary across conditions would allow the models to alter their tuning properties when one feature differs and when

both features differ. Because we do not expect these tuning properties to vary with stimulus conditions, we explicitly fit both models to the entire data set. The final output (fit to the averaged data set), generated using 1024 iterations of these best-fitting parameters, is displayed in Figure 6B. Best-fitting parameters for both models are displayed in Table A1 for all four observers as well as for the averaged data, and the resultant squared error values (the sum of squared differences between observed and simulated responses) are shown in Table A2. The combined model gives a better fit, with less error, in all cases.

References

Finger, F. W., & Spelt, D. K. (1947). The illustration of the horizontal–vertical illusion. *Journal of Experimental Psychology, 37,* 243–250.

Leibowitz, H. (1955). Some factors influencing the variability of vernier adjustments. *The American Journal of Psychology, 68,* 266–273.

Sloan, L. L. (1959). New test charts for the measurement of visual acuity at far and near distances. *American Journal of Ophthalmology, 48,* 807–813.

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

Flom, M. C., Weymouth, F. W., & Kahneman, D. (1963). Visual resolution and contour interaction. *Journal of the Optical Society of America, 53,* 1026–1032.

Beck, J. (1966). Perceptual grouping produced by changes in orientation and shape. *Science, 154,* 538–540.

Campbell, F. W., & Kulikowski, J. J. (1966). Orientational selectivity of the human visual system. *The Journal of Physiology, 187,* 437–445.

Hubel, D. H., & Wiesel, T. N. (1968). Receptive fields and functional architecture of monkey striate cortex. *The Journal of Physiology, 195,* 215–243.

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

Blakemore, C., & Nachmias, J. (1971). The orientation specificity of two visual after-effects. *The Journal of Physiology, 213,* 157–174.

Wolford, G. (1975). Perturbation model for letter identification. *Psychological Review, 82,* 184–199.

Braddick, O. J., Campbell, F. W., & Atkinson, J. (1978). Channels in vision: Basic aspects. In R. Held, H. W. Leibowitz, & H. L. Teuber (Eds.), *Handbook of sensory physiology* (vol. 8, pp. 3–38). Heidelberg, Germany: Springer.

Butler, T. W., & Westheimer, G. (1978). Interference with stereoscopic acuity: Spatial, temporal, and disparity tuning. *Vision Research, 18,* 387–392.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12,* 97–136.

Chastain, G. (1981). Inhibition of feature extraction with multiple instances of the target feature in different orientations. *Psychological Research, 43,* 45–56.

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

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology, 14,* 107–141.

Ungerleider, L. G., & Mishkin, M. (1982). Two cortical visual systems. In D. J. Ingle, M. A. Goodale, & R. J. W. Mansfield (Eds.), *Analysis of visual behavior* (pp. 549–586). Cambridge, MA: MIT Press.

Bergen, J. R., & Julesz, B. (1983). Parallel versus serial processing in rapid pattern discrimination. *Nature, 303,* 696–698.

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

Levi, D. M., & Klein, S. A. (1985). Vernier acuity, crowding and amblyopia. *Vision Research, 25,* 979–991.

Rentschler, I., & Treutwein, B. (1985). Loss of spatial phase relationships in extrafoveal vision. *Nature, 313,* 308–310.

Sagi, D., & Julesz, B. (1985). "Where" and "what" in vision. *Science, 228,* 1217–1219.

Burr, D. C., Morrone, M. C., & Spinelli, D. (1989). Evidence for edge and bar detectors in human vision. *Vision Research, 29,* 419–431.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review, 96,* 433–458.

Graham, N. V. S. (1989). *Visual pattern analyzers.* New York: Oxford University Press.

Hamilton, D. B., Albrecht, D. G., & Geisler, W. S. (1989). Visual cortical receptive fields in monkey and cat: Spatial and temporal phase transfer function. *Vision Research, 29,* 1285–1308.

Morrone, M. C., Burr, D., & Spinelli, D. (1989). Discrimination of spatial phase in central and peripheral vision. *Vision Research, 29,* 433–445.

Felleman, D. J., & Van Essen, D. C. (1991). Distributed hierarchical processing in the primate cerebral cortex. *Cerebral Cortex, 1,* 1–47.

Ivry, R. B., & Prinzmetal, W. (1991). Effect of feature similarity on illusory conjunctions. *Perception & Psychophysics, 49,* 105–116.

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

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

Heathcote, A., & Mewhort, D. J. K. (1993). Representation and selection of relative position. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 488–516.

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,* 255–279.

Wolfe, J. M. (1994). Guided Search 2.0 A revised model of visual search. *Psychonomic Bulletin & Review, 1,* 202–238.

Prinzmetal, W., Henderson, D., & Ivry, R. (1995). Loosening the constraints on illusory conjunctions: Assessing the roles of exposure duration and attention. *Journal of Experimental Psychology: Human Perception and Performance, 21,* 1362–1375.

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

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

Treisman, A. M. (1996). The binding problem. *Current Opinion in Neurobiology, 6,* 171–178.

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

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

Wilkinson, F., Wilson, H. R., & 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.

Lennie, P. (1998). Single units and visual cortical organization. *Perception, 27,* 889–935.

Carney, T., & Klein, S. A. (1999). Optimal spatial localization is limited by contrast sensitivity. *Vision Research, 39,* 503–511.

Hess, R. F., Dakin, S. C., Kapoor, N., & Tewfik, M. (2000). Contour interaction in fovea and periphery. *Journal of the Optical Society of America A, Optics, Image Science, and Vision, 17,* 1516–1524.

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

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

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

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]

Edwards, M., & Badcock, D. R. (2003). Motion in depth affects perceived stereoscopic depth. *Vision Research, 43,* 1799–1804.

Levi, D. M., Li, R. W., & Klein, S. A. (2003). "Phase capture" in the perception of interpolated shape: Cue combination and the influence function. *Vision Research, 43,* 2233–2243.

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]

Solomon, J. A., Felisberti, F. M., & Morgan, M. J. (2004). Crowding and the tilt illusion: Toward a unified account. *Journal of Vision, 4*(6):9, 500–508, http://www.journalofvision.org/content/4/6/9, doi:10.1167/4.6.9. [PubMed] [Article]

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

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]

Blake, R., Tadin, D., Sobel, K. V., Raissian, T. A., & Chong, S. C. (2006). Strength of early visual adaptation depends on visual awareness. *Proceedings of the National Academy of Sciences of the United States of America, 103,* 4783–4788.

Hansen, B. C., & Hess, R. F. (2006). The role of spatial phase in texture segmentation and contour integration. *Journal of Vision, 6*(5):5, 594–615, http://www.journalofvision.org/content/6/5/5, doi:10.1167/6.5.5. [PubMed] [Article]

Huang, P.-C., Kingdom, F. A. A., & Hess, R. F. (2006). Only two phase mechanisms, ±cosine, in human vision. *Vision Research, 46,* 2069–2081.

Neri, P., & Levi, D. M. (2006). Spatial resolution for feature binding is impaired in peripheral and amblyopic vision. *Journal of Neurophysiology, 96,* 142–153.

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research, 46,* 4646–4674.

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]

Põder, E., & Wagemans, J. (2007). Crowding with conjunctions of simple features. *Journal of Vision, 7*(2):23, 1–12, http://www.journalofvision.org/content/7/2/23, doi:10.1167/7.2.23. [PubMed] [Article]

van den Berg, R., Roerdink, J. B. T. M., & Cornelissen, F. W. (2007). On the generality of crowding: Visual crowding in size, saturation, and hue compared to orientation. *Journal of Vision, 7*(2):14, 1–11, http://www.journalofvision.org/content/7/2/14, doi:10.1167/7.2.14. [PubMed] [Article]

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

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

Põder, E. (2008). Crowding with detection and coarse discrimination of simple visual features. *Journal of Vision, 8*(4):24, 1–6, http://www.journalofvision.org/content/8/4/24, doi:10.1167/8.4.24. [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]

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]

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 of the United States of America, 106,* 13130–13135.

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

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

Fujisaki, W., & Nishida, S. (2010). A common perceptual temporal limit of binding synchronous inputs across different sensory attributes and modalities. *Proceedings of the Royal Society of London B: Biological Sciences, 277,* 2281–2290.

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

Mareschal, I., Morgan, M. J., & Solomon, J. A. (2010). Cortical distance determines whether flankers cause crowding or the tilt illusion. *Journal of Vision, 10*(8):13, 1–14, http://www.journalofvision.org/content/10/8/13, doi:10.1167/10.8.13. [PubMed] [Article]

Watt, R. J., & Dakin, S. C. (2010). The utility of image descriptions in the initial stages of vision: A case study of printed text. *British Journal of Psychology, 101,* 1–26.

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]

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

Maus, G. W., Fischer, J., & Whitney, D. (2011). Perceived positions determine crowding. *PLoS ONE, 6,* e19796.

Wallis, T. S. A., & Bex, P. J. (2011). Visual crowding is correlated with awareness. *Current Biology, 21,* 254–258.

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