The Bouma law accounts for crowding in fifty observers

Jan W. Kurzawski[1], Augustin Burchell[3], Darshan Thapa[2], Jonathan Winawer[1,2], Najib J. Majaj[2], Denis G. Pelli[1,2]

[1]Department of Psychology, New York University, New York, NY, USA
[2]Center for Neural Science, New York University, New York, NY, USA
[3]Cognitive Science & Computer Science, Swarthmore College

Corresponding Author: Denis G. Pelli, Department of Psychology, 6 Washington Place, New York University, NY, 10003, USA (denis.pelli@nyu.edu)

Conflict of interest: The authors declare no competing financial interests.

Keywords: crowding, critical spacing, crowding distance, Bouma law, object recognition, statistics of crowding, asymmetries around the visual field.

Funding: This research was funded by the US National Eye Institute R01-EY027964 to DGP and R01-EY031446 to NJM. It was also supported by an NIH core vision grant P30-EY013079.

Abstract

Crowding is the failure to recognize an object due to surrounding clutter. Our visual crowding survey measured 13 crowding distances (or "critical spacings") twice in each of 50 observers. The survey included three eccentricities (0, 5, and 10 deg), four cardinal meridians, two orientations (radial and tangential), and two fonts (Sloan and Pelli). The survey also tested foveal acuity, twice. Remarkably, fitting a two-parameter model, the well-known Bouma law — crowding distance grows linearly with eccentricity — explains 82% of the variance for all 13 × 50 measured log crowding distances, cross-validated. An enhanced Bouma law, with factors for meridian, crowding orientation, target kind, and observer, explains 94% of the variance, again cross-validated. These additional factors reveal several asymmetries, consistent with previous reports, which can be expressed as crowding-distance ratios: 0.62 horizontal:vertical, 0.79 lower:upper, 0.78 right:left, 0.55 tangential:radial, and 0.78 Sloan font:Pelli font. Across our observers, peripheral crowding is independent of foveal crowding and acuity. Evaluation of the Bouma factor *b* (the slope of the Bouma law) as a biomarker of visual health would be easier if there were a way to compare results across crowding studies that use different methods. We define a *standardized Bouma factor b'* that corrects for differences from Bouma's 25 choice alternatives, 75% threshold criterion, and linearly symmetric flanker placement. For radial crowding on the right meridian, the standardized Bouma factor *b'* is 0.24 for this study, 0.35 for Bouma (1970), and 0.30 for the geometric mean across five representative modern studies, including this one, showing good agreement across labs, including Bouma's. We found that guaranteeing fixation by gaze-contingent display halved the standard deviation across observers of the estimated log *b*. The reduction in standard deviation is explained by a "peeking" model in which the observer looked near an anticipated target location in 50% of *unmonitored*-fixation trials. Individual differences are robust, as evidenced by the much larger 0.08 SD of log *b* across observers than the 0.03 SD of test-retest within observers. Crowding's ease of measurement enhances its promise as a biomarker for dyslexia and visual health.

Introduction

Crowding is the failure to recognize an object due to surrounding clutter (Bouma, 1970, 1973; Pelli et al., 2004; Pelli & Tillman, 2008; Strasburger et al., 1991; Stuart & Burian, 1962). Crowding has been studied with several different tasks including letter identification (Bouma, 1970; Flom, Heath, et al., 1963; Strasburger et al., 1991), Landolt rings (Flom, Heath, et al., 1963; Flom, Weymouth, et al., 1963), Vernier acuity (Levi et al., 1985; Malania et al., 2007; Westheimer & Hauske, 1975), face recognition (Farzin et al., 2009; Louie et al., 2007; Martelli et al., 2005), and orientation discrimination (Andriessen & Bouma, 1976; Parkes et al., 2001; Toet & Levi, 1992; Westheimer et al., 1976). It is invariant with the size of target and flankers (Levi & Carney, 2009; Pelli et al., 2004; Pelli et al., 2007; Strasburger et al., 1991; Tripathy & Cavanagh, 2002). Crowding is usually measured by sandwiching the target between two similar flanking objects, or *flankers*, and is characterized by the *crowding distance* (or "critical spacing"), which is the center-to-center distance from target to flanker at which recognition attains a criterion level of performance. Crowding distance increases linearly with eccentricity (Bouma, 1970; Kooi et al., 1994; Levi & Carney, 2009; Pelli et al., 2004; Toet & Levi, 1992), and increases with target-flanker similarity (Andriessen & Bouma, 1976; Chastain, 1982; Kooi et al., 1994; Leat et al., 1999; Nazir, 1992; Pelli et al., 2004) as well as the number of distractors (Grainger et al., 2010; Strasburger et al., 1991). Crowding also occurs for moving stimuli (Bex & Dakin, 2005; Bex et al., 2003). For review of the crowding literature see (Herzog et al., 2015; Levi, 2008; Pelli & Tillman, 2008; Strasburger, 2020; Strasburger et al., 2011; Whitney & Levi, 2011). Among the normally sighted, crowding was first reported in the periphery, and, after some debate, has now been convincingly demonstrated in the fovea (Atkinson et al., 1986; Coates et al., 2018; Flom, Heath, et al., 1963 1963; Liu & Arditi, 2000; Malania et al., 2007; Pelli et al., 2016; Siderov et al., 2013; Toet & Levi, 1992).

We are interested in relating psychophysical measures of crowding to brain physiology, especially cortical magnification measured by fMRI (functional Magnetic Resonance Imaging) in areas V1, V2, V3, and hV4. For this purpose, we tested crowding in 50 observers to characterize the statistics of crowding within and across individuals. The comparison with fMRI will be reported separately (Kurzawski et al., 2021). Here we report only the psychophysics. We tested with letters, which, after little or no training, provide the many possible targets that are needed for quick testing (Pelli & Robson, 1991). Long term, we are interested in testing crowding in children (Waugh et al., 2018) as an early biomarker for susceptibility to visual problems like dyslexia. Crowding distance is highly conserved across object kind (Kooi et al., 1994; Pelli & Tillman, 2008), which suggests that letters, vernier, and Gabors might have similar crowding distances, but Grainger et al. (2010) reported different crowding distances for letters and symbols.

Crowding exhibits several striking asymmetries. Crowding distance measured radially — along a line passing through the foveal center — is roughly twice that measured tangentially — the orthogonal orientation (Greenwood et al., 2017; Kwon et al., 2014; Pelli, 2008; Petrov & Meleshkevich, 2011; Toet & Levi, 1992). Crowding distance has often been reported to be smaller in the lower than upper visual field (Fortenbaugh et al., 2015; Greenwood et al., 2017; He et al., 1996; Petrov & Meleshkevich, 2011) and on the horizontal than vertical

midline (Chung, 2013; Coates et al., 2021; Liu et al., 2009; Petrov & Meleshkevich, 2011; Toet & Levi, 1992; Wallis & Bex, 2012).

Crowding distance is a potentially valuable biomarker for several reasons. Crowding severely limits what we see, how fast we read, and is associated with dyslexia. There are large individual differences in crowding distance and correspondingly large physiological differences in the sizes of relevant areas of visual cortex, which invite analysis by correlation (Kurzawski et al., 2021). Here we measured crowding in 50 observers. Previous in-person crowding surveys (Grainger et al., 2010; Greenwood et al., 2017; Petrov & Meleshkevich, 2011; Toet & Levi, 1992) included at most 27 observers. The only remote crowding survey tested 793 observers but did not report any asymmetries (Li et al., 2020). The above cited works used various kinds of stimuli, including letters of various fonts. The original reports of the crowding phenomenon were mostly letter-based (Anstis, 1974; Bouma, 1970, 1973; Ehlers, 1936; Ehlers, 1953; Korte, 1923; Stuart & Burian, 1962). Historical review of crowding is described elsewhere (Levi, 2008; Pelli et al., 2004; Strasburger, 2020; Strasburger et al., 2011). Here, we too use letters, because they do not require training, and provide a large number of stimulus alternatives, which speeds threshold estimation in laboratory and clinical testing (Pelli et al., 1988).

Whether crowding can be explained by the neural computations in any particular cortical location remains unknown, but several candidate areas have been suggested: V1 (Millin et al., 2014), V2 (Freeman & Simoncelli, 2011; He et al., 1996), V3 (Bi et al., 2009; Tyler & Likova, 2007), hV4 (Burchell et al., 2019; Liu et al., 2009; Motter, 2006; Zhou et al., 2017) and higher-order areas (Aghdaee, 2005; Louie et al., 2007). The magnitude of the BOLD signal in V1 is lower in the presence of crowding (Millin et al., 2014). Crowding distance is different for stimuli tuned to stimulate either the parvo- or magno-cellular pathway (Atilgan et al., 2020). Although both crowding and acuity increase linearly with eccentricity, which might suggest a common physiological origin, the two lines have very different intercepts with the eccentricity axis, i.e. the E_2 value for acuity is more than 5 times larger than the E_2 value for crowding [$E_2 = 2.72$ for acuity and $E_2 = 0.45$ for crowding] (Latham & Whitaker, 1996; Petrov & Meleshkevich, 2011; Rosenholtz, 2016; Song et al., 2014; Strasburger, 2020). This seems inconsistent with a common cause. Here, we use our data from 50 observers to study the relationship between acuity and crowding in the fovea.

In our 50 participants, we measured 13 crowding distances at three eccentricities (0, 5, and 10 deg), on all four cardinal meridians, in two crowding orientations (radial and tangential), using two fonts (Sloan and Pelli). We also measured acuity in the fovea. We used the Sloan and the Pelli fonts. As far as we know, the Pelli font is still the only letter font skinny enough to measure crowding distance in the fovea. Apart from letters, foveal crowding can be measured with vernier targets (Malania et al., 2007). We also assessed crowding's variation along the four cardinal meridians, in two crowding orientations, and across individuals. Crowding varies two-fold across meridians, producing several asymmetries.

METHODS

Measuring crowding

We measured thresholds of many participants, collecting two datasets with similar methods except for one important difference. Details about Quest and stimulus presentation were similar for both and are described in the sections below. Here we focus on the differences.

Method 1: Unmonitored fixation. Threshold was measured without gaze tracking. Viewing distance was measured before each session, and no chin rest or forehead support was provided. The participant identified the target by pressing that letter in the keyboard. Participants were naive to the task and received no advance training. In each block, crowding distance was measured at two randomly interleaved target locations, which were horizontally or vertically symmetric about the fixation cross. This *unmonitored*-fixation dataset includes one radial crowding distance on each of the four cardinal meridians with the Sloan font for 100 participants.

Method 2: Awaited fixation. Each trial began only once the participant had continuously fixated within ±1.5 deg of the crosshair for 250 ms, and we only saved trials in which gaze remained within ±1.5 deg of the crosshair center until stimulus offset. The experimenter was present during data acquisition. Viewing distance was measured at the beginning of each session and maintained by use of a chinrest with forehead support. Participants identified the target by using a mouse to click on one of the letters displayed on the response screen. For each participant, data collection began after a total of 10 correct trials. Crowding distance was measured at four randomly interleaved target locations symmetric about the fixation cross, one on each cardinal meridian. We acquired two thresholds at each location to estimate test-retest reliability. This *awaited*-fixation dataset includes two radial crowding distances on each of the four cardinal meridians with the Sloan font for 50 participants.

Comparing results obtained with the two methods at ±5 deg on the horizontal midline, reveals large differences in the mean and distribution of the Bouma factor (**Fig. 1**). [In this paper, "log" is the logarithm base 10.] Using *unmonitored* fixation, the geometric mean *b* was 0.12, with a 0.31 SD of log *b*. Using *awaited* fixation, the geometric mean *b* was higher, 0.20, with a lower SD of log *b*, 0.18. The awaited fixation histogram (red) is compact. The unmonitored fixation histogram (green) is much broader, extending to much lower values of *b*. Our interpretation of the broader histogram and lower geometric-mean *b* in *unmonitored* fixation is that observers occasionally "peek", that is fixate near an anticipated location of the target instead of the fixation cross as instructed. Indeed, at the end of Results, we present a quantitative peeking model showing that peeking reduces the geometric mean *b* and broadens its distribution, consistent with the observed results.

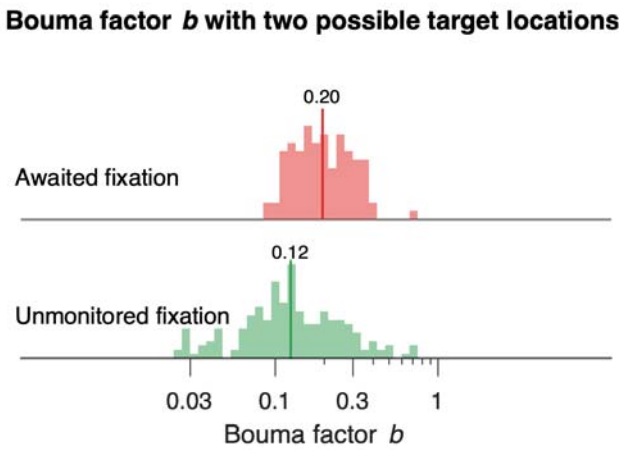

Figure 1. Histograms of the Bouma factor estimated by two methods. Each histogram shows the Bouma factor *b* at ±5 deg eccentricity along the horizontal midline. For *awaited* fixation we only used data from first session of the experiment. The geometric mean *b* is indicated by a dark vertical line capped by a number. *Unmonitored* fixation gave a 0.12 geometric mean *b* with 0.31 SD of log *b*. *Awaited* fixation gave a higher 0.20 geometric mean *b* with a lower 0.18 SD of log *b* (see Table 3). Each histogram includes *b* estimates made at both +5 and -5 deg eccentricity on the horizontal midline. The Results section below reports a 0.78:1 right:left advantage. Note: Mixing data from the two locations (-5 deg left and +5 deg right) makes the combined histogram slightly broader than that for either location. The 0.78:1 *b* ratio corresponds to a -0.107 log *b* difference. If we suppose that mixing log *b* estimates from the two locations is equivalent to taking all the data from the right location and adding +0.107 to a random half of the log *b* estimates, then mixing the two locations increases the variance by +0.0025, which is only 8% of the measured variance of log *b* for awaited fixation, and only 3% for unmonitored fixation.

This paper focusses on the *awaited*-fixation data, which can be downloaded from OSF (https://osf.io/83p6u/).

Crowding dataset

Data were acquired with the *CriticalSpacing* software (Pelli et al., 2016) using Quest (Watson & Pelli, 1983), allowing for reliable and relatively fast measurement of crowding distances. Our crowding database consists of measurements of crowding distance with the Sloan font (with radial and tangential flankers) and with the Pelli font (radial flankers) in 50 observers. With the Sloan font, we measured crowding at 8 different peripheral locations in the visual field: 2 eccentricities (5 and 10 deg) along the four cardinal meridians (upper, lower, left, and right). Sloan tangential crowding was measured only at ±5 deg eccentricity on the horizontal midline. With the Pelli font, we measured crowding at the fovea and at ±5 deg on the horizontal midline. The Sloan font acuity size is too big to allow measuring foveal crowding distance in adults. The Pelli font was specially designed for measuring foveal crowding distance (Pelli et al., 2016). We also measured acuity in the fovea. A spatial map of the testing is shown below, in **Figure 4** in Results. We also tested 10 observers at 20 and 30 deg eccentricity with radial flankers (only 1 session) and plotted the results in **Figure 9**. To estimate test-retest reliability of our measurements we used two sessions to measure each threshold twice. Sessions were scheduled at least a day apart over a maximum of 5 days apart. We report our results as the Bouma factor *b* (slope of crowding distance vs. eccentricity) estimated from **Eq. 10**, to minimize error in fitting log $\hat{s}$. Here, crowding

distance $\hat{s}$ is the required center-to-center spacing (in deg) for 70% correct report of the middle letter in a triplet.

Participants

Table 1 describes our main dataset, and **Figure 4** (in Results) plots its spatial coverage of the visual field. The study tested 50 observers (mean age = 23), mostly New York University undergraduate students. Each observer had normal or corrected-to-normal vision. All experiments were conducted in accordance with the Declaration of Helsinki and were approved by New York University's ethics committee on activities involving human observers. In all analyses, except the test-retest section, we average the first- and second-session thresholds.

The peeking-model section in Results refers to these "awaited-fixation" peripheral measurements on 50 participants and compares them to separate "unmonitored fixation" peripheral measurements on 100 participants.

N	Measure	Font	Crowding orientation	Radial eccentricity (deg)	Cardinal meridians	Thresholds per observer (each measured twice)	Gaze tracking
50	Crowding	Sloan	radial	5, 10	all	8	Yes
	Crowding	Sloan	tangential	5	right, left	2	Yes
	Crowding	Pelli	radial	5	right, left	2	Yes
	Crowding	Pelli	horizontal	0	–	2	No
	Acuity	Sloan	–	0	–	2	No

Table 1. Data summary. In the periphery, we measured crowding distance radially and tangentially with the Sloan font. With the Pelli font, we measured crowding both in the fovea and periphery. We also measured foveal acuity with the Sloan font. Each threshold was measured once in two sessions separated by at least 24 hours. For peripheral thresholds, we used gaze tracking to guarantee fixation within ±1.5 deg of the crosshair center. Foveal crowding required long viewing distance which made gaze tracking impractical, so participants were merely instructed to fixate the center of the crosshair. We suppose good fixation of the central crosshair because the participants expected a foveal target.

Apparatus

Each testing session was completed on an Apple iMac 27" with an external monitor. The observer viewed the LG 27" 5K monitor 27MD5KL-B, with a screen resolution of 5120 x 2880 and a white background with luminance 275 cd/m2. The white background never changed throughout the experiment; the black crosshair and letters were drawn on it. The observer viewed the screen binocularly at one of several different viewing distances. The software required a special keypress by the experimenter at the beginning of every block with a new observer or a new viewing distance, to affirm that the new viewing distance (eye to screen) was correct as measured with a tape measure, and that the screen center was orthogonal to the observer's line of sight. To measure crowding and acuity in the fovea, the viewing

distance was 200 cm. For ±5 and ±10 deg eccentricity the distance was 40 cm, and for ±20 and ±30 deg it was 20 cm. The long viewing distance gives good rendering of small stimuli; the short viewing distance results in a wide angular subtense of the display, to allow presentation of peripheral targets on either side of a central fixation. Stimuli were rendered using *CriticalSpacing.m* software (Pelli et al., 2016) implemented in MATLAB 2021 using the Psychtoolbox (Brainard, 1997; Pelli, 1997). Every Sloan letter was at least 8 pixels wide and every Pelli digit was at least 4 pixels wide.

Stimuli and procedure

To measure acuity, we show one letter. To measure crowding we show a trigram of three letters or digits. For each trial, the three letters or digits are drawn randomly, without replacement, from the 9 letters (DHKNORSVZ) or digits (123456789) available. Letters and digits are rendered as black in the Sloan or Pelli font, presented on a uniform white background (Pelli et al., 2016; Sloan et al., 1952). We omit the C in the Sloan font because it's too easily confused with the O (Elliott et al., 1990). For crowding, each trigram was arranged either radially or tangentially. Each testing session included several blocks and was about an hour long. Most blocks measured 4 thresholds, interleaved, usually four crowding thresholds on the 4 cardinal meridians at the same radial eccentricity. For the Pelli font and tangential crowding we measured two thresholds at symmetric locations about fixation along the horizontal midline. To minimize the temptation to look away from fixation toward an expected target location, we randomly interleaved conditions measuring threshold at the same radial eccentricity at 2 or 4 symmetric locations around fixation. A sample stimulus sequence appears in **Figure 2A**. A central crosshair (the fixation mark) is displayed until the observer presses a key to initiate the trial. Then, after 250 ms of correct fixation, the letter trigram appears on the screen for 150 ms and the computer waits for the observer to identify the middle stimulus letter by using a mouse to click on a letter in a row of all the possible letters on the response screen. Observers are instructed to return their eyes to fixation before clicking their response. A correct response is acknowledged with a brief beep. Then the computer waits indefinitely for the observer to fixate within 1.5 deg of the crosshair for 250 ms, and then immediately presents the stimulus for the next trial. If the observer fails to fixate for 250 ms within a 10 second window, the software asks for recalibration of the gaze tracker.

Measuring crowding distance

Crowding distance was estimated using the Pelli et al. (2016) procedure. Letter spacing is controlled by QUEST. Spacing scales with letter size, maintaining a fixed ratio of 1.4:1. We set the Weibull function guessing rate parameter γ to the reciprocal of the number of characters in the test alphabet for that font, usually 9. We set the "finger error" probability δ to 1%, to help QUEST cope with an occasional reporting mistake. We set the Weibull function steepness parameter β to 2.3, based on fits to two observers' psychometric data for radial crowding. At the end of each block, QUEST estimates threshold (crowding distance in deg, **Fig. 2B**). To measure acuity, we followed a similar procedure, except that the target was presented without flankers. Threshold was defined as the letter spacing (crowding distance) in deg or letter size (acuity) in deg that achieved 70% correct identification, using Quest to control the stimulus parameters trial by trial, and make the final estimate.

Each threshold measurement was based on 35 trials (one condition). A block consists of all the trials in however many conditions are randomly interleaved, e.g., 4 × 35 = 140 trials to measure threshold at 4 meridians. Each condition measures threshold for one meridian. The interleaving keeps the observer uncertain as to which location is tested on each trial. We do this to minimize the urge to "peek" away from fixation. On a crowding trial, until the target and flankers appear there is nothing about the display that distinguishes which of the interleaved conditions this trial belongs to.

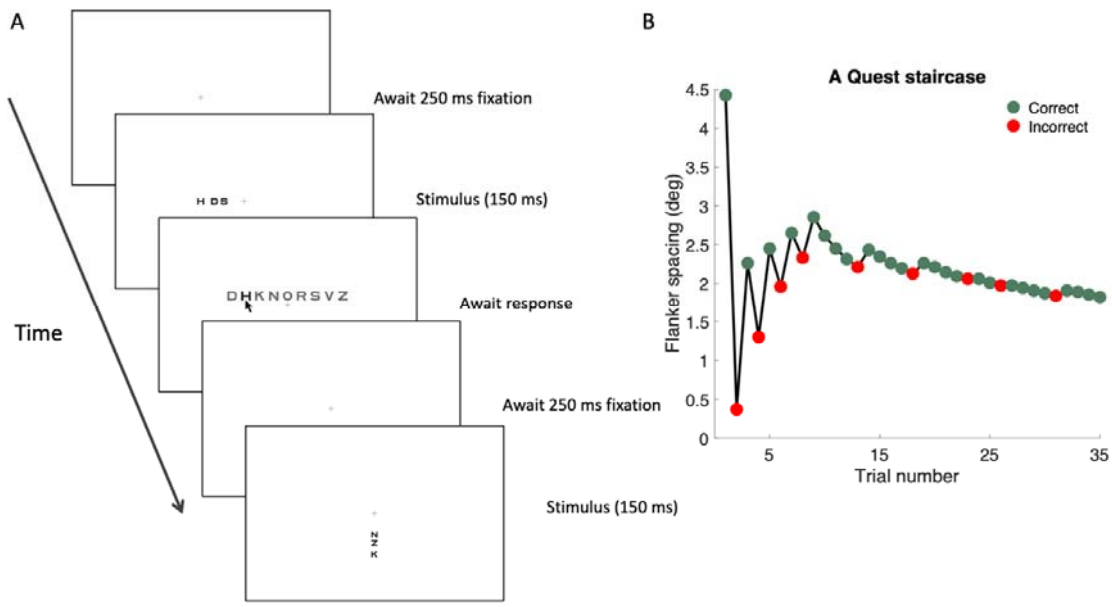

Figure 2. Stimulus and procedure. A) The display sequence for a peripheral trial and part of the next. While gazing at the crosshair, which is always present, the observer presses a space bar key, which begins the first trial. The target is presented only if the observer was continuously fixating for 250 ms. Stimulus presentation is accompanied by a low-pitched purr. Then the observer identifies the target by using a mouse to click on one letter out of all possible letters that appear above the fixation on the response screen. If the response was correct, the observer hears a brief beep acknowledging correctness and silence otherwise. Then the computer again waits for 250 ms of fixation within 1.5 deg of the crosshair center. The four conditions (one for each meridian) are randomly interleaved, so the observer does not know which location comes next. **B)** The staircase sequence of spacings tested on 35 successive trials of one condition (+5 deg eccentricity), under control of QUEST. On each trial, the letter size was a fraction 1/1.4 of the spacing. QUEST picks the most informative spacing to test on each trial to minimize variance of its final threshold estimate. Finally, after 35 trials, QUEST estimates the crowding distance, i.e., spacing to achieve 70% correct. Notice that testing quickly homes in on threshold.

Gaze tracking

We used gaze-contingent display to guarantee fixation while measuring all peripheral thresholds. We used an EyeLink 1000 eye tracker (SR Research, Ottawa, Ontario, Canada) with a 1000 Hz sampling rate. To allow short viewing distance (40 cm) we used the EyeLink Tower mount with a 25 mm lens mounted on the EyeLink camera. Each trial presented the stimulus once gaze had been within 1.5 deg of the crosshair center (the fixation mark) for 250 ms. If, during the stimulus presentation, gaze deviated more than 1.5 deg from the crosshair center, then the trial was not saved, the fixation cross turned red (to alert the

participant), and the trial was repeated with a fresh letter trigram. Thus, each threshold estimate was based on 35 trials with fixation within 1.5 deg of the crosshair center. The foveal thresholds demanded a long 200 cm viewing distance that was incompatible with our gaze tracking set up, so they were measured without gaze tracking, but fixation is generally good when the participant knows that the target is foveal.

Model fitting

It is generally found that SD of repeated measurement of threshold spacing *s* is roughly proportional to mean spacing, but the SD of log spacing *S* is independent of mean spacing. Therefore, our fitting minimizes the RMS error in log spacing $S = \log_{10} s$. The fitting is nonlinear (using the MATLAB `fmincon` function) because we minimize error in *S*, whereas each model is linear in *s*, not *S*. We estimate the participant, meridional, crowding orientation, and font factors by solving several models (see model equations in Table 4).

Our fitting minimizes the RMS error in predicting the log crowding distances, which is equivalent to minimizing the summed square error, *SSE*:

$$SSE = \sum_i \left(S_i - \hat{S}_i\right)^2 \tag{1}$$

where S_i is the *i*-th log crowding distance, and $\hat{S}_i$ is the *i*-th predicted log crowding distance. The variance explained by each model is:

$$R^2 = 1 - \sum_i \left(S_i - \hat{S}_i\right)^2 \Big/ \sum_i \left(S_i - \bar{S}\right)^2 \tag{2}$$

where $\bar{S} = \text{mean}_i S_i$ is the mean log crowding distance.

Model comparison

An F test was used for pairwise model comparison. The model with fewer parameters is called "simple", and the model with more parameters is called "full". After calculating the sum of square errors SSE_{simple} and SSE_{full} for each model (**Eq. 1**), we calculate the F-statistic:

$$F = \frac{(SSE_{\text{simple}} - SSE_{\text{full}})/(n_{\text{full}} - n_{\text{simple}})}{SSE_{\text{full}}/(N - n_{\text{full}})} \tag{3}$$

Where n_{full} is the number of parameters in the full model, n_{simple} is the number of parameters in the simple model, and *N* is the number of observations. We estimate the p-value using the F-distribution. A p-value less than 0.05 indicates that the model with more parameters provides a significantly better explanation of the data.

Cross validation

First, we divide the thresholds into 6 random subsets of equal size. In each cross-validation step, one subset of data is retained as the validation set for testing the model, while the remaining subsets are used as training data. Each subset is chosen only once for testing. We

repeat leave-one-out testing 6 times to obtain the full dataset. Variance explained R^2 is calculated by **Eq. 2.**

Standardized Bouma factor can be compared across studies

To facilitate comparison across studies and the cooperative evaluation of the Bouma factor as a biomarker of visual health, we define the *standardized* Bouma factor *b*′ as the slope of crowding distance vs radial eccentricity multiplied by a correction factor that accounts for methodological differences from Bouma's number of choices (25), threshold criterion (75% correct), and linear spacing (vs. log).

Table 2 computes the correction factors needed to compare the Bouma factor *b* across studies that used various numbers of response choices (e.g., 9 Sloan letters or 2 orientations of a tumbling T), various threshold criteria (e.g., 70% or 75% correct), and linear or log flanker spacing. Including the present one, we know of five studies that compare crowding distance across meridians. We take Bouma (1970) as the standard for this standardized way of reporting the strength of crowding.

Study	**Target**	**Ecc. (deg)**	**Choices *n***	**Guessing rate γ**	**Threshold criterion proportion correct *P***	**Threshold criterion "true" proportion correct *P****	**Log threshold shift $\Delta S = S - S_{Bouma}$**	**Correction factor to standardize Bouma factor for number of choices and the threshold criterion**	**Correction factor to standardize Bouma factor for log vs. linear flanker spacing**
[Calculation]			n	$\gamma = 1/n$	P	Eq. 5	Eq. 9	10^–ΔS	1.18 for log 1 for linear
Bouma 1970	Courier 10, letter	1-7	25	1/25	0.75	0.74	0.00	1.00	1
Our data	Sloan or Pelli, letter	5,10	9	1/9	0.70	0.66	-0.04	1.10	1.18
Greenwood et al., 2017	Tumbling clock	4,8	4	1/4	0.80	0.73	0.00	1.01	1
Toet & Levi, 1992	Tumbling T	2.5, 5 ,10	2	1/2	0.75	0.50	-0.13	1.33	1
Grainger et al., 2010	Courier New, letter or symbol	3	9	1/9	0.75	0.72	-0.01	1.03	1
Coates et al., 2021	Tumbling T	9	4	1/4	0.625	0.50	-0.13	1.33	1

Table 2. Correction factor to standardize each study's Bouma factor to compare studies. The resulting standardized Bouma factors are reported in Table 6. Each study's difference from Bouma's 25 choice alternatives and 75% correct threshold criterion offset the threshold log spacing by ΔS relative to Bouma's. The correction factor is 10^–ΔS. We provide a brief Excel spreadsheet to calculate the correction factor and a MATLAB script that produces plots like Fig. 3. The correction factor to account for linear vs. log flanker spacing is estimated in the Supplementary material. Note that crowding measured with tangential flankers does not require correction for log vs. linear spacing symmetry. Tangential spacing was always linear.

The log-threshold shift ΔS is illustrated in **Figure 3B**. Using the Coates et al. (2021) reanalysis of Bouma's (1970) data, we estimated the Bouma factor for a 75% threshold criterion applied to Bouma's (1970) results (see the Bouma factor paragraph in Discussion).

Proportion correct. To account for the different number of choices and the threshold criterion, we assume that, with accurate fixation, the proportion correct *P* is a Weibull function of the log spacing *S*,

$$P(S) = \gamma + (1-\gamma)[1 - \exp(-10^{\beta(S-\Sigma)})] \quad (4)$$

with a threshold parameter Σ, where γ is the guessing rate (the reciprocal of the number of target choices, which is 1/25=0.04 in Bouma's 1970 results), and β is the steepness parameter, which we set to 2.3, based on fitting psychometric functions to hundreds of trials at several spacings by two experienced observers. **Figure 3A** shows this psychometric function for six studies, taking the guessing rate γ to be the reciprocal of the number of choices *n*, and using our own estimate of the steepness parameter β=2.3.

"True" proportion correct. To accommodate various numbers of choices *n*, and thus guessing rates $\gamma = 1/n$, we "correct for guessing,"

$$P^* = (P - \gamma)/(1 - \gamma) \quad (5)$$

This is a popular transformation of psychometric data, usually justified by assuming that the guessing rate can be modelled as an independent process. Because it discounts false alarms, the corrected hit rate is called the "true hit rate". That makes sense for a yes-no task, but not for an identification task. Here we proceed regardless, and compute the "true" proportion correct, because, with this Weibull function (**Eq. 4**), correction for guessing (**Eq. 5**) removes all dependence on γ. Applying correction for guessing to any given threshold criterion *P* gives us the corresponding "true" proportion correct criterion *P** to apply after correction for guessing. Similarly applying correction for guessing to the psychometric function (**Eq. 4**) gives us the "true" proportion correct,

$$P^*(S) = 1 - \exp(-10^{\beta(S-\Sigma)}) \quad (6)$$

The inverse of **Eq. 6** is

$$S = \mathrm{invP}^*(P^*) = \Sigma + \log[-\ln(1 - P^*)]/\beta \quad (7)$$

Figure 3B plots "true" proportion correct (**Eq. 6** with β=2.3 and Σ=0), the same function for all studies, and, for each study, a vertical line reads off the log threshold spacing *S* at its threshold criterion *P**.

Relative to the Bouma standard. Thus, a study's number of choices and threshold criterion increase its log threshold by ΔS relative to the Bouma standard:

$$\Delta S = \mathrm{invP}^*(P^*) - \mathrm{invP}^*(P^*_{\mathrm{Bouma}}) \quad (8)$$

$$= \log\left(-\ln(1-P^*)\right)/\beta \; - \log\left(-\ln(1-P^*{}_{\text{Bouma}})\right)/\beta \qquad (9)$$

where β=2.3, and P^* and $P^*{}_{\text{Bouma}}$ are the "true" proportion correct threshold criteria computed by **Eq. 5** from the study's criterion P and Bouma's P_{Bouma}=0.75 (Andriessen & Bouma, 1976)

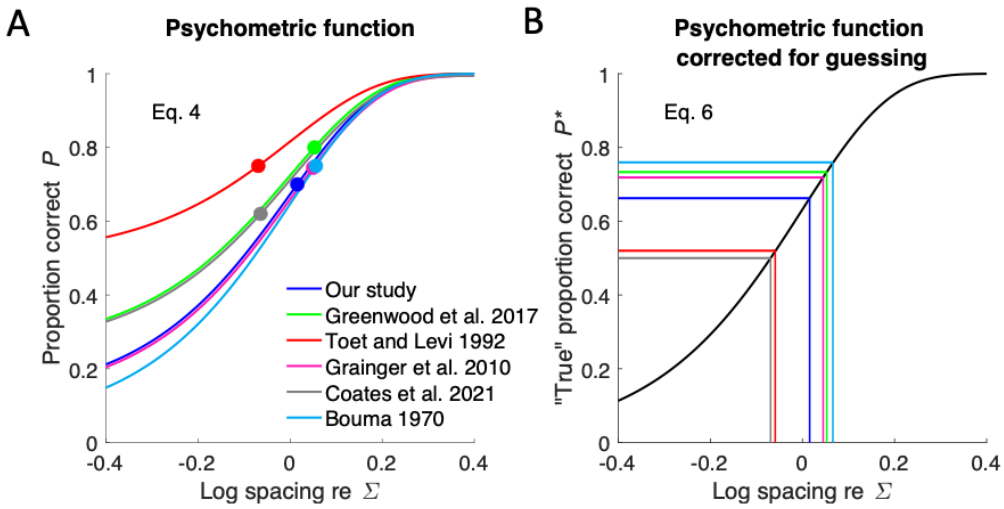

Figure 3. Effect of guessing rate and criterion on threshold log crowding distance *S*. A) Proportion correct (Eq. 4) of the six studies, with threshold parameter Σ set to zero, showing the effects of the number of response choices n (sets lower asymptote $\gamma=1/n$) and the threshold criterion (height of each colored dot). **B) "True" proportion correct (Eq. 6)** for the same studies. Each study's threshold criterion P^* is represented by a horizontal line. For each study, a vertical line reads off the log threshold spacing S at its threshold criterion P^*. Results from Bouma (1970) are used with the Andriessen and Bouma (1976) 75% threshold criterion. **Table 2** computes the difference between each study's threshold and Bouma's. To avoid occlusion, the Toet and Levi and Bouma lines in this panel were offset by +0.02 horizontally and vertically.

Log-symmetric spacing of flankers

Since (Bouma, 1970), most crowding studies measure crowding distance as the center-to-center spacing between the target and each of two flankers on opposite sides of the target that yields a criterion level of performance. When crowding is measured in the radial orientation, the Bouma law tells us that crowding distance increases linearly with eccentricity. Several studies have documented that, when flankers are arranged symmetrically about the target on a radial line from fixation, the outer flanker has much more effect (Banks et al., 1977; Bex & Dakin, 2005; Estes et al., 1976; Krumhansl, 1977). This is to be expected since crowding distance grows with eccentricity and the outer flanker is more eccentric. In fact, crowding distance on the cortical surface (in mm)—the product of crowding distance in deg and cortical magnification in mm/deg—is conserved across eccentricity (for eccentricities above 5 deg) because psychophysical crowding distance scales with eccentricity (Bouma, 1970; Kooi et al., 1994; Levi & Carney, 2009; Pelli et al., 2004; Toet & Levi, 1992). Given the logarithmic cortical mapping of the visual field (Fischer, 1973), when measuring radial crowding we space the trigram so that the log eccentricity of the target is midway between the log eccentricities of the flankers and report the inner spacing. This raises the question of how to compare crowding distances between

experiments that spaced the flankers linearly vs logarithmically. Given the Bouma law (**Eq. 8** below), supposing that crowding distance depends primarily on the flanker-to-flanker distance and only negligibly on the target position between them, we show in the Supplement ("Effect of symmetric placement of flankers with regard to either linear or log eccentricity") that the crowding distance is expected to be 1.18 times larger when measured with linearly-spaced flankers than with log-spaced flankers. To ease comparison across studies, the correction factors in Table 2 include this effect of log vs linear spacing on the estimated Bouma factor.

Might attention help explain differences in the reported Bouma factor?

Attention reduces many perceptual thresholds (Carrasco, 2011). Many researchers have assessed the effects of attention on crowding, but they have yet to reach a consensus. Several found an attentional benefit in crowding tasks (Bacigalupo & Luck, 2015; Kewan-Khalayly et al., 2022) including reduction of crowding distance (Yeshurun & Rashal, 2010), but others did not find such effects (Scolari et al., 2007; Strasburger, 2005; Strasburger & Malania, 2013). All our peripheral crowding thresholds are measured with either two-fold or four-fold uncertainty about target location, and we suppose that attention was distributed among the possible target locations. It is possible that attentional bias contributes to some of the Bouma factor asymmetries.

Data from other studies

Data were extracted from Figure 6 of Toet and Levi (1992) using WebPlotDigitizer (Rohatgi, 2020). Data were extracted from Figure 7 of Grainger et al. (2010). Data from of Greenwood et al. (2017) Supplementary Fig. 1 and Coates et al. (2021) Figure 10 were received as personal communications from the authors. Data from Bouma (1970) were used by means of the recent reanalysis by Coates et al. (2021).

Statistical analysis

Statistics of the log Bouma factor $B = \log b$ were assessed by an ANOVA with B as the dependent variable. Two sample comparisons are made with the Wilcoxon rank sum test. We report Pearson's r correlation coefficient for test-retest reliability and correlations of crowding distance.

RESULTS

Crowding and acuity

As shown in the map of testing (**Fig. 4**), radial crowding thresholds were measured in 50 adults at 9 visual field locations. Using the Sloan font, radial crowding thresholds were measured at the four cardinal meridians at 5 and 10 deg eccentricity, and tangential crowding thresholds were measured on the left and right meridians at 5 deg eccentricity. Using the Pelli font, the horizontal crowding threshold was measured, in the fovea and on the right and left meridians at 5 deg eccentricity. Foveal acuity was also measured with the Sloan font. Each threshold was measured twice.

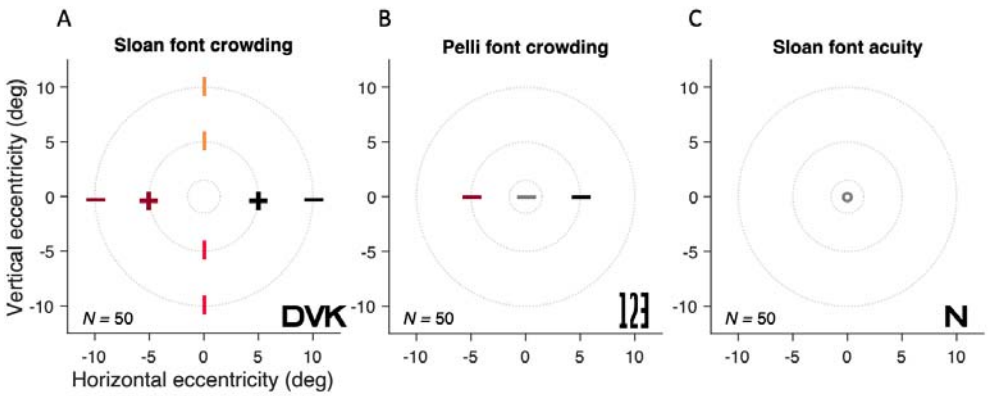

Figure 4. Map of testing. Each panel title indicates the font and threshold task. The number of observers tested is indicated in the lower left. The +, –, and o symbols indicate testing of crowding with radial (—) or radial and tangential (+) flankers, and testing of acuity (o). Typical stimuli appear in the lower right of each panel. Beyond the main dataset described here, Fig. 9 shows additional results from 10 observers at 20 and 30 deg eccentricity.

Test-retest reliability of visual threshold

Measurement reliability was assessed by measuring each threshold twice, at least one and no more than five days apart. Crowding thresholds are converted to Bouma factors *b* (see **Eq. 10** below). Foveal crowding and acuity are presented as crowding distance (deg) and acuity as letter size (deg). **Figure 5** plots a scatter diagram of estimates from first vs second session for each combination of font and task. For the Pelli font (**Fig. 5B**), we found a clear improvement of measured crowding distance in the second session (ratio of geometric mean retest:retest = 0.88). This training benefit was much smaller for the Sloan font (0.95), presumably because Sloan is more similar (than Pelli) to familiar fonts. In general, each threshold is derived from a Quest staircase with 35 trials, which takes about 3.5 minutes, and has very good reproducibility. The analyses performed in the following sections are based on the geometric average threshold across both sessions.

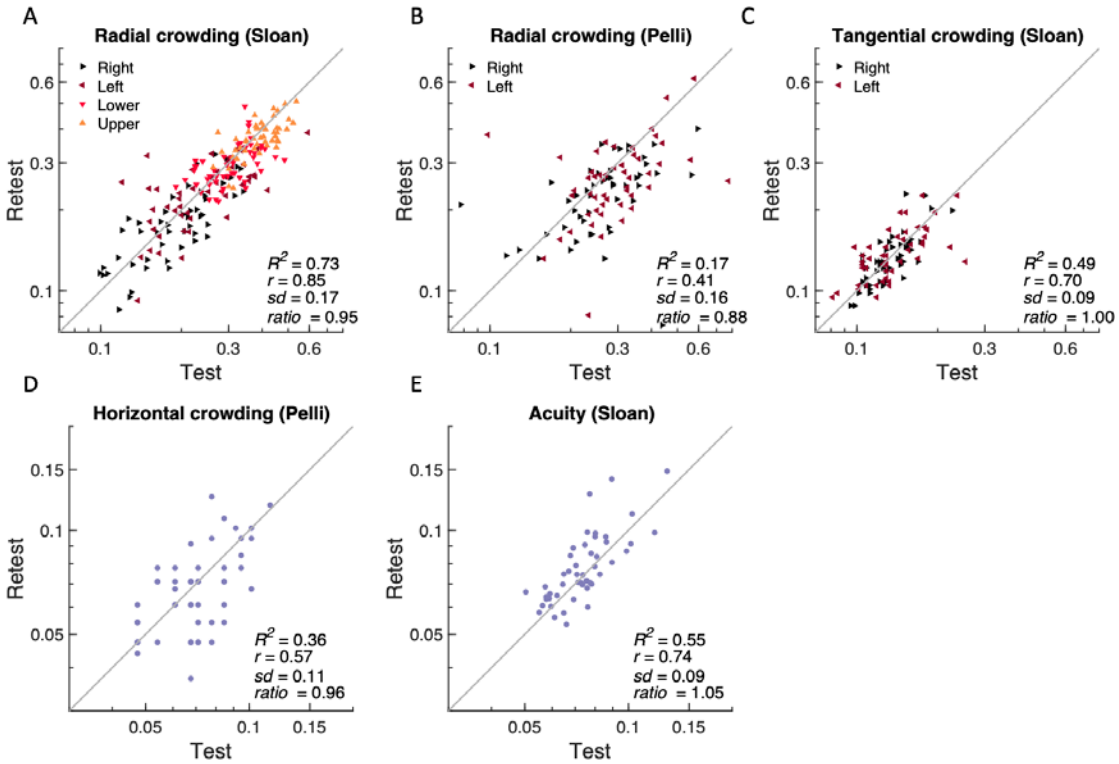

Figure 5. Test-retest reliability of threshold estimates. Estimates of Pearson's *r* correlation coefficient, standard deviation, retest:test ratio, and R^2 are based on the log of the font-task threshold, in degrees of visual angle, named at the top of each panel. For peripheral crowding (**A-C**) each measurement is represented by a triangle pointing towards the tested meridian. The gray line equality.

Analysis of variance

Table 3 presents an ANOVA analysis of the radial Sloan Bouma factors (also plotted in Figure 5A). The 0.18 SD for radial Sloan with 2 meridians (with awaited fixation) in **Figure 1** corresponds to the 0.17 total SD with 4 meridians (also with awaited fixation) in **Table 3**. The 0.08 SD for Sloan radial crowding in **Figure 8** corresponds to the 0.08 SD across observers in **Table 3**. Meridian contributes the most variance.

Factor	d.f.	Variance	SD	p-value
meridian	3	0.0143	0.1195	<0.01
observer	1	0.0059	0.0770	<0.01
test-retest	59	0.0001	0.0109	0.016
error	346	0.0071	0.0843	
total	399	0.0274	0.1656	

Table 3 – Analysis of variance. We computed the contribution of each parameter to overall variance. Meridian contributes the most (SD = 0.12) and test-retest contributes the least (SD = 0.01). Degrees of freedom d.f. is number of parameters minus 1. Error is the remaining variance not accounted for by a linear combination of

meridian, observer, and test-retest. There were no significant pairwise interactions among meridian, test-retest, and observer (all p>0.5).

82% of variance explained by Bouma law

Bouma (1970) discovered the linear relationship between crowding distance and eccentricity. He initially reported a slope of 0.5, which he later revised to 0.4 (Andriessen & Bouma, 1976). The Bouma law is:

$$\hat{s} = (\varphi_0 + \varphi)\, b \tag{10}$$

where $\hat{s}$ is crowding distance in deg, φ is radial eccentricity in deg, and φ_0 (in deg) and b (dimensionless) are positive fitted constants (Bouma, 1970; Rosen et al., 2014). The dimensionless slope b is the Bouma factor. The horizontal intercept is -φ_0, and the vertical intercept is $\varphi_0 b$ (Liu & Arditi, 2000; Strasburger et al., 2011; Toet & Levi, 1992).

Crowding is one of several tasks whose threshold increases linearly with radial eccentricity, and such a task can be summarized by an E_2 value that is the eccentricity at which threshold reaches twice its foveal value (Levi et al., 1985). In the Bouma law (**Eq. 10)**, $E_2 = \varphi_0$.

Our large database of visual crowding thresholds (**Table 1**) is very well fit (R^2 = 82.45%) by the two-parameter linear Bouma law (**Eq. 10**), showing that most of crowding's variation in our data is explained by eccentricity. Just two degrees of freedom, b and φ_0, suffice to fit all 650 data points (13 thresholds measured in each of 50 observers). The estimated slope b was 0.23, just over half of Bouma's 0.4. Our database consists of measurements at five locations with radial and tangential flankers, and two fonts. To capture the effect of these parameters on the Bouma factor we propose an extended version of the Bouma law.

94% of variance is explained by extended Bouma law

Crowding depends on more than just eccentricity. Crowding varies substantially across meridians (right, left, up, or down), crowding orientation (radial or tangential), target kind (e.g., letters or symbols) and across individuals. Here, we enhance the Bouma law by including these other variables. One by one, the extensions add model parameters for meridian, crowding orientation, target kind, and observer. The models and the variance that they account for are summarized in **Table 4**.

Meridian. Factor b_θ, which allows b to depend on the meridian θ (right, left, up, or down)

$$\hat{s} = \varphi_0 b + \varphi b_\theta \tag{11}$$

where b from **Eq. 10** now represents the *geometric mean* of b_θ, b = 10^(mean(log$_{10}$(b_θ))). Note that the meridian is undefined at the fovea.

Crowding orientation. Factor f_{dir} depends on crowding orientation (radial or tangential).

$$\hat{s} = (\varphi_0 b + \varphi b_\theta) f_{\text{dir}} \tag{12}$$

*Target kind***.** Factor t_{kind} depends on target kind (e.g. Pelli or Sloan font).

$$\hat{s} = (\varphi_0 b + \varphi b_\theta) f_{\text{dir}} t_{\text{kind}} \tag{13}$$

Observer. Finally, factor o_i depends on the observer.

$$\hat{s} = (\varphi_0 b + \varphi b_\theta) f_{\text{dir}} t_{\text{kind}} o_i \tag{14}$$

where $\prod_i \; o_i = 1$. Adding factors to the original Bouma law accounts for more variance. Going from the simplest to the most enhanced model (**Eqs. 10 and 14**) increases explained variance from 82% to 94% (**Table 4**). Model performance is improved by adding the meridian factor (R^2 = 89%) and crowding orientation (R^2 = 93%). Adding the target-kind factor explains hardly any more variance, increased from 92.54% to 92.63%. Finally, the most enhanced model, with an observer factor, explains 94% of variance. The models are all cross validated, so the additional variance explained is not a necessary consequence of the increase in parameters. If the additional parameters were overfitting the training data, then we would find less variance accounted for in the left-out test data.

Model	**Equation**	R^2	R	**RMSE**	**No. of parameters**
Bouma law	$s\square = (\varphi_0 + \varphi)\, b$ (10)	82.45%	0.90	4.80	2
× **meridional factor**	$s\square = \varphi_0\, b + \varphi\, b_\theta$ (11)	88.96%	0.94	3.80	5
× **crowding orientation**	$s\square = (\varphi_0\, b + \varphi\, b_\theta)\, f_{\text{dir.}}$ (12)	92.54%	0.96	3.13	7
× **target-kind factor**	$s\square = (\varphi_0\, b + \varphi\, b_\theta)\, f_{\text{dir}}\, t_{\text{kind}}$ (13)	92.63%	0.96	3.11	9
× **observer factor**	$s\square = (\varphi_0\, b + \varphi\, b_\theta)\, f_{\text{dir}}\, t_{\text{kind}}\, o_i$ (14)	93.86%	0.97	2.84	59

Table 4. How well the Bouma law and its extensions predict crowding distance. We begin with the Bouma law (**Eq. 10**). Successive models try to account for more variance by adding factors that depend on the meridian, crowding orientation, font, and observer. Each row gives a significantly better fit than the row above (assessed with F-test using **Eq. 3**). The R^2 (**Eq. 2**) column shows cross-validated variance accounted for in predicting log crowding distance over the whole visual field (13 thresholds per observer). Pearson's *R* shows the correlation between acquired and predicted data, and RMSE is root mean square error.

Since model parameters can be added in any order to form the final model (**Eq. 14**), we asked how much each parameter contributes to the total explained variance. For each parameter, we begin with the full model, remove that parameter, calculate the explained variance for the reduced model, and assess the drop in explained variance. We find that, after eccentricity, meridian contributes the most (4.5%) and target kind contributes the least (0.2%). (Note that most of our data were collected with one font; we expect target

kind to explain more variance in studies that emphasize comparison of fonts, or other target kinds.) Results are shown in **Table 5**.

Removed factor	Equation	Decrease in R^2
meridian	$= (\ _0 + \ ) b \, f_{dir} \, t_{kind} \, o_i$	4.5%
crowding orientation	$= (\ _0 b + \ b \ ) \, t_{kind} \, o_i$	3.1%
observer	$= (\ _0 b + \ b \ ) \, f_{dir} \, t_{kind}$	1.4%
target kind	$= (\ _0 b + \ b \ ) \, f_{dir} \, o_i$	0.2%

Table 5 – Parameter contribution to the most extended Bouma law. We measured the contribution of each parameter to the full model.

Since the enhanced model accounts for more variance than the original Bouma law, we looked in the data for systematic effects of these parameters. Figure 6 plots each set of model parameters after normalizing by the geometric mean of that set, where a set is the four meridians, two crowding orientations, two fonts, or fifty observers. Asymmetry within each factor is discussed in the next section. Except for target kind, each of the factors (meridian, crowding orientation, and observer) accounts for a roughly two-fold variation in the Bouma factor (dashed horizontal lines in **Fig. 6**).

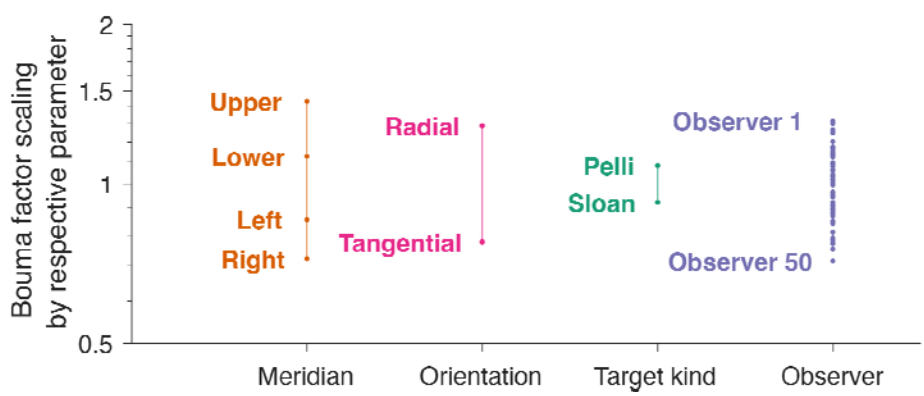

Figure 6. How several parameters scale the Bouma factor. To reveal the effect of each parameter (horizontal axis) each set of model parameters was normalized by the geometric mean of that set. The vertical axis plots the model's estimates of that parameter. Because the model is multiplicative, the final Bouma factor is proportional to the product of all the parameters. We find a similar variation of Bouma factor with meridian, crowding orientation, and observer. Target kind has the least effect, but that is partly because nearly all of our data are with one font.

Radial Bouma factor varies twofold across meridians

Most of the thresholds in our dataset are for the Sloan font with radial flankers. Using these data, we explore the variation of Bouma factor across meridians. We estimate the Bouma factor by fitting **Eq. 10** for each participant and meridian independently (**Fig. 7A**). The Bouma factor is smallest along the right meridian and highest along the upper meridian (**Fig.**

7B). Bouma factor is 0.184 right, 0.237 left, 0.300 lower, and 0.381 upper meridian; overall geometric mean = 0.27).

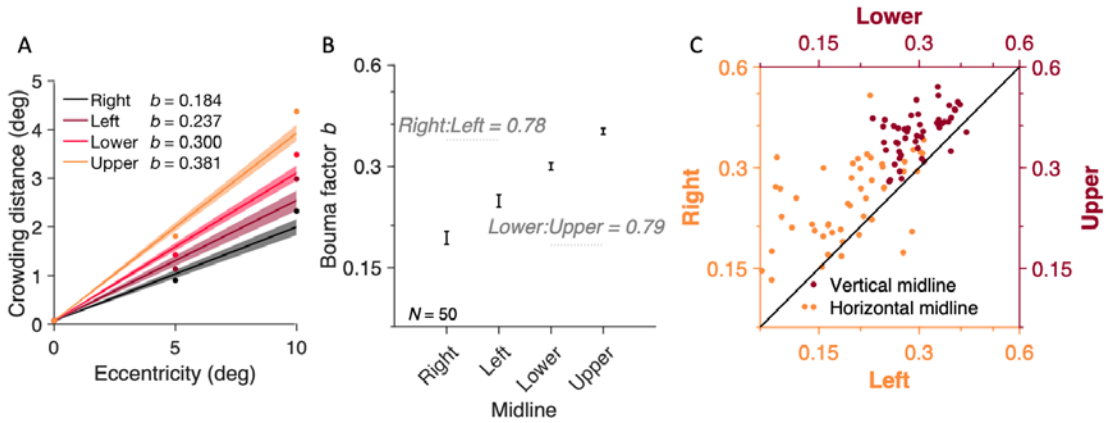

Figure 7. Bouma factor vs. meridian. A) Bouma law estimates for radial crowding with the Sloan font estimated using **Eq. 10**. Each point represents the mean across participants and error bars represent 95% confidence intervals. **B)** Bouma factor vs. meridian. **C)** Individual-participant data plotted for the vertical (dark red) and horizontal midline (orange).

Meridional asymmetries. We find three asymmetries. (As a reporting convention, we refer to the "advantage" of a smaller Bouma factor.) 1) Along the vertical midline, there is a 0.79 lower:upper advantage. 2) Along the horizontal midline, there is a 0.78 right:left advantage. 3) Finally, there is a 0.62 horizontal:vertical advantage (based on geometric mean of the Bouma factors from right and left meridian vs upper and lower meridian). All reported asymmetries are highly consistent across participants (**Fig. 7C**). ANOVA reveals that there is a significant effect of meridian $F(3,196) = 92.76$, $p < 0.001$ and post-hoc analysis shows that Bouma estimates at each meridian are significantly different from each other (all $p < 0.001$; corrected for multiple comparisons). For each meridian we also estimate the eccentricity φ_0 at which the crowding distance reaches twice its foveal value. φ_0 was 0.37 deg ± 0.02 for right, 0.29 deg ± 0.02 for left, 0.22 deg ± 0.01 for lower and 0.17 deg ± 0.01 for upper meridian.

Tangential Bouma factor is roughly half of radial

Unlike radial crowding, tangential crowding is the same in left and right meridians according to the Wilcoxon rank sum test ($z = -0.73$, $p = 0.49$). The standardized (see section on corrected Bouma factor below) tangential Bouma factor is much smaller than radial: 0.13 on the right and 0.14 on the left meridian. The tangential:radial ratio is 0.60 in the right meridian and 0.50 in the left.

Bouma factor varies with target kind

Pelli and Tillman (2008) highlighted the remarkable degree to which crowding distance is conserved across stimulus kind, but later work shows that crowding distance does differ substantially between some target kinds (e.g. letters vs. symbols; Grainger et al., 2010). Along the horizontal midline, standardized Bouma factor for Sloan font was 0.239 on the right and 0.308 on the left, and slightly higher for the Pelli font: 0.325 on the right and 0.377 on the left. Overall, there is a 0.78 Sloan:Pelli ratio of standardized Bouma factors

(geometric mean of the ratio taken at each meridian) and the difference between fonts was statistically significant ($z = 3.58$, $p < 0.001$). The model performance is slightly improved by adding the target-kind factor (**Eq. 13**). This factor contributed little to the overall variance explained by the model because most of the data came from trials with the same target kind (Sloan letters). So even though excluding target kind as a factor from the model caused inaccurate predictions for the Pelli font, the reduction in variance explained is negligible because nearly all of the dataset is based on one font. We anticipate that the target-kind factor will account for more variance in datasets that focus on comparing target kinds.

Bouma factor varies twofold across observers

Bouma factor varies with meridian, crowding orientation, and target kind. Here, in this section, we quantify differences between observers. First, we estimated how well the Bouma law fits individual-participant data. Fitting **Eq. 10** to the right meridian data for each participant results, on average, in 97% explained variance, confirming that individual crowding data are well described by the linear model. Next, for each observer we fit the whole model to estimate the observer's overall Bouma factor (**Fig. 8**). We also report individual differences in acuity. Individual differences are characterized by the standard deviation of the log of the threshold. Radial Bouma factor for the Sloan font varies approximately two-fold across observers (SD of log b = 0.08). This variation is unchanged for tangential flankers (SD of log b = 0.08) and nearly doubles for the Pelli font (SD of log b = 0.11). Foveal acuity a and foveal crowding distance s also vary two-fold. For crowding, the φ_0 values also vary two-fold and range between 0.17 and 0.37 (Song et al., 2014). We also report the SD of the retest minus test difference for the Bouma factor estimated with radial flankers and Sloan font (**Fig. 8B**). For each observer we fit one Bouma factor for the test session and one Bouma factor for the retest. Differences across observers are much larger than those of test-retest. The 0.08 SD of the log Bouma factor across observers is 3 times larger than the 0.03 SD of test and retest, showing that one measurement is enough to distinguish individual differences.

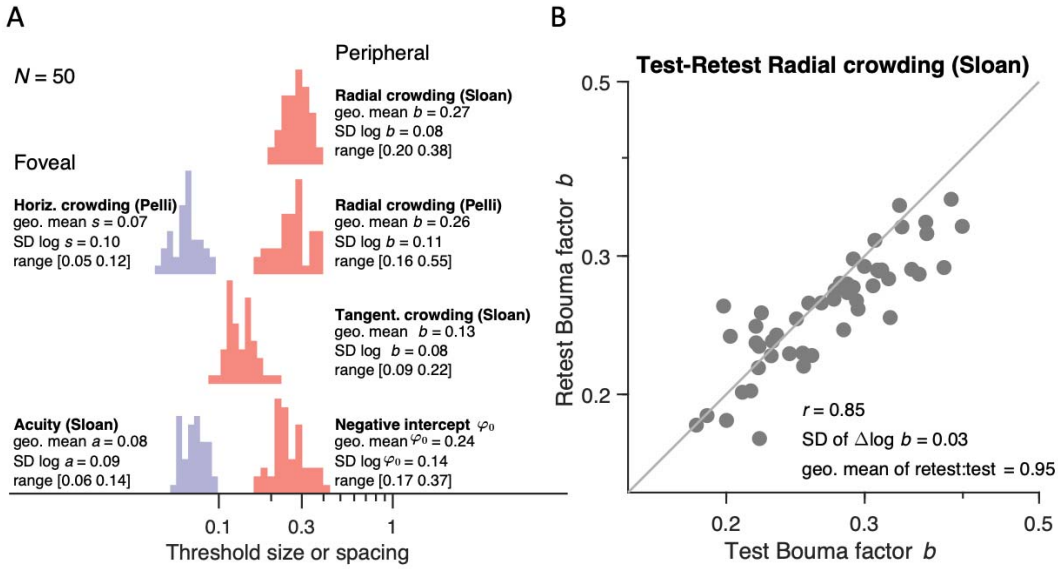

Figure 8. Histograms of crowding and acuity. Histograms of **A)** radial Sloan, radial Pelli, and tangential Sloan Bouma factor (dimensionless), and negative intercept φ_0 (in deg), horizontal crowding distance s (in deg), and

acuity a (in deg). To estimate individual differences, we used all acquired data (e.g., 16 thresholds for radial crowding with Sloan font or 4 thresholds for radial crowding with Pelli font). **B)** Retest vs test of Bouma factor b for radial Sloan.

Supralinearity: Bouma factor increases with eccentricity

Bouma discovered the linear increase of crowding distance with eccentricity. We have seen that this linear equation fits our data well. However, seeing that we have 50 participants and data at 0 to 10 deg, a reviewer suggested that we examine how well the Bouma factor is conserved across eccentricity. To estimate the Bouma factor we fit **Eq. 10** for Sloan font with radial flankers to our data at 0 deg plus either 5 or 10 deg (**Fig. 9A**). On average Bouma factor is 1.4 higher at 10 than 5 deg eccentricity. This effect was statistically significant ($F(1,398) = 42.3$ $p < 0.001$), and there was no interaction between eccentricity and meridian ($F(3,392) = 0.513$ $p = 0.674$). This shows that the growth of crowding distance with eccentricity is actually more than linear. Indeed, the Coates et al. reanalysis of Bouma 1970 shows a similar supralinearity. Motivated by this finding, we invited 10 observers already in the main dataset (0, 5, 10 deg) to also measure crowding distance at 20 and 30 deg eccentricity.

When fitting data, there is a long tradition of using the shortest polynomial that fits adequately. Bouma (1970) initially suggested proportionality, with one degree of freedom. Measurements of nonzero crowding distance at 0 deg eccentricity led to a linear equation, with two degrees of freedom. Seeing curvature in our data from 10 observers from 0 to 30 deg, we enhanced the Bouma law from linear to quadratic (three degrees of freedom) to fit the data. **Eq. 15** adds a quadratic term to **Eq. 7** to allow the slope to grow with eccentricity. Replacing **Eq. 10** by **15** increases the degrees of freedom from 2 to 3, and increases the explained variance from 90% to 95% (**Fig. 9B-C**),

$$\hat{s} = (\varphi_0 + \varphi + c\varphi^2)\, b \tag{15}$$

where $\hat{s}$ is predicted crowding distance (in deg), φ is radial eccentricity (in deg), and φ_0 (in deg), b, and c are degrees of freedom.

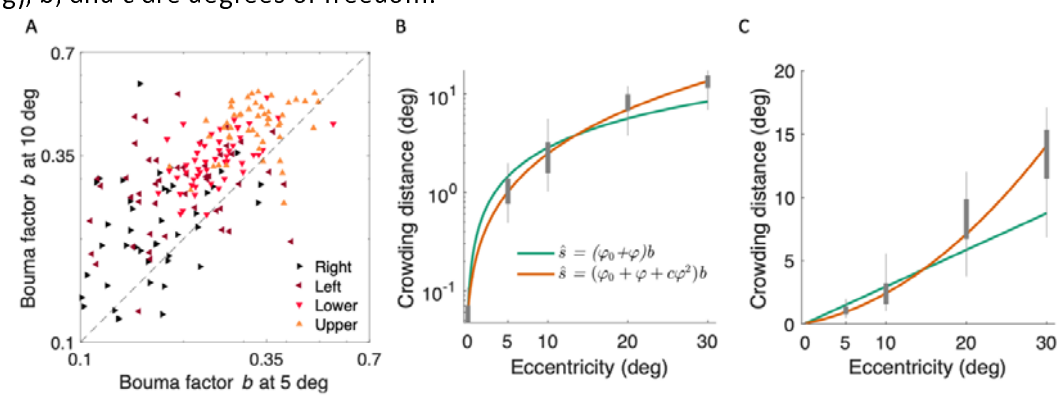

Figure 9. Supralinearity: Crowding distance slope grows with eccentricity. A) Radial Bouma factor at 5 vs. 10 deg eccentricity for Sloan font. Each color shows data on a different meridian. Data are plotted for all 50 participants included in the main study. **B)** Log crowding distance for 10 participants is plotted against eccentricity out to 30 deg. The linear Bouma law is green, and the quadratic Bouma law is orange. We use log coordinates because the fit minimizes the error in log coordinates. Enhancing the Bouma law from linear to quadratic increases the explained variance from 90% to 95%. **C)** Same fits replotted in linear coordinates. The

nonlinear growth of crowding distance with eccentricity is not an artifact of perspective transformation: The computation of target angular size and eccentricity was done correctly using the arc tangent function. Eq. 10 fit with RMSE 0.20 and b = 0.30, and φ_0 = 0.20. **Eq. 15** fit with RMSE 0.16 and b = 0.15, φ_0 = 0.43, and c = 0.06.

Correlation of log crowding distance across visual field, crowding orientation, and target kind

We explored the pattern of correlations of log crowding distance for visual field locations, crowding orientations, and target kinds. These correlations are shown in **Figure 10A** where each cell shows Pearson's r between two measurements. Rows are sorted so that the average correlation decreases from top to bottom. We find that log crowding distance measured on the right meridian at 10 deg eccentricity with Sloan font and radial flankers yields the highest average correlation with other log crowding distances (r = 0.39 with all, and r = 0.41 when fovea is excluded). Foveal log crowding distance measured with Pelli font yields the smallest average correlation with the rest of the log crowding distances.

To summarize how correlation depends on stimulus properties we estimated the average correlation across measurements when 1, 2 or 3 stimulus properties (eccentricity, meridian, target kind, crowding orientation) are modified (**Fig. 10B**). The test-retest correlation of 0.54 is plotted at zero changes. The average correlation drops to 0.30 with 1 change, to 0.25 with 2 changes, and to 0.18 with 3 changes. We also estimated what is the average correlation at the same stimulus location, and we only vary font and crowding orientation (right or left meridian at 5 deg). We find an average correlation (across 2 changes) of r = 0.54. On the other hand, when we change the location only and keep stimulus properties (e.g., radial flankers, Sloan font, 5 deg) we get a much lower correlation of r = 0.32. This indicates that when correlating crowding distances, location matters more than any other stimulus property.

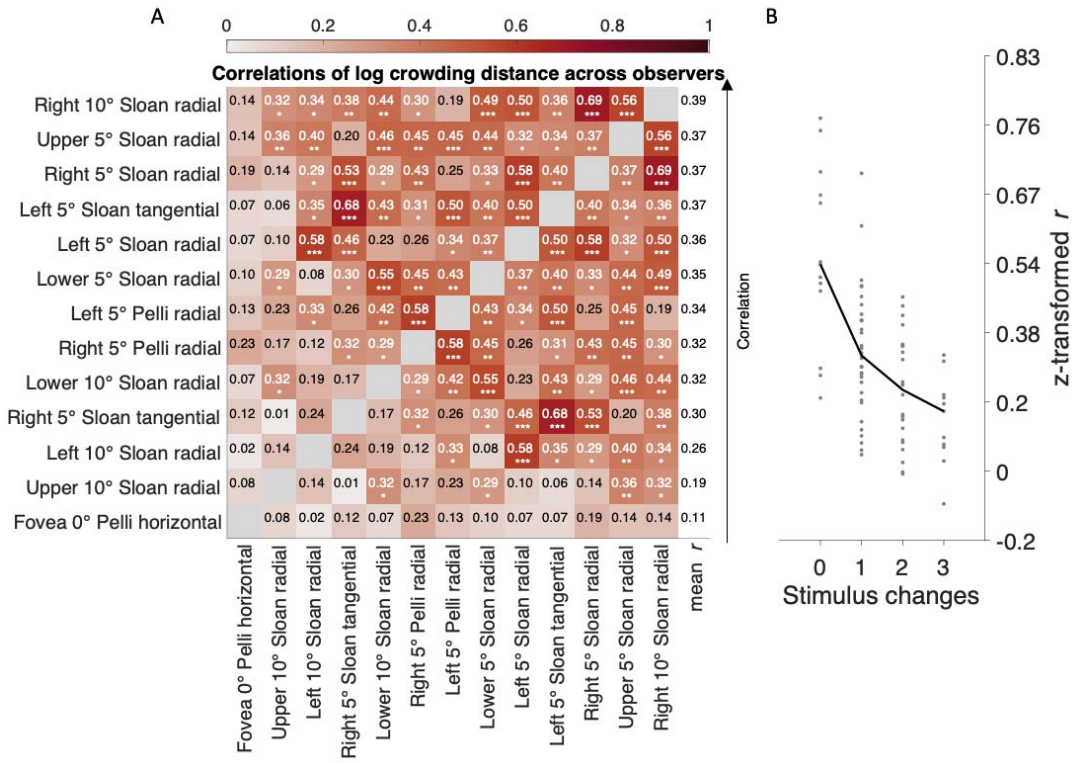

Figure 10. Correlations of log crowding distance. A) Pair-wise correlations between crowding distances for various conditions (Table 1). Rows are sorted so that the average correlation decreases from top to bottom. **B)** Average correlation as a function of the number of stimulus property differences in: radial eccentricity, meridian, crowding orientation, and target kind. For example, comparing two eccentricities is a 1-parameter difference.

Correlation of Bouma factor *b* and intercept φ_0

The Bouma law has two degrees of freedom, φ_0 and b, which are anticorrelated, r = -0.51 (geometric mean across meridians).

Standarized Bouma factor and its asymmetries across different studies

Visual field asymmetries can help identify the neural origin of perceptual phenomena (Afraz et al., 2010; Himmelberg et al., 2023). We compare our estimates of the Bouma factor to all the previous studies that measured crowding asymmetry.

Estimating slope from just one point. The Bouma law has two degrees of freedom, the slope b and the negative intercept φ_0. Estimating two parameters requires two measurements but many crowding studies report crowding distance at only one eccentricity. In the complete case, we have thresholds S_0 and S at eccentricities 0 and φ, and we use the definition of the Bouma factor b as the slope $b = (s - s_0)/(\varphi - 0)$. In the incomplete case, we have only threshold S at eccentricity φ. One might try to estimate the missing foveal threshold s_0 or negative intercept φ_0, but the simplest thing to do is to neglect φ_0 (pretend it's zero), and estimate $\hat{b} = s/\varphi$. The estimate has fractional error $\epsilon = (\hat{b} - b)/b = \varphi_0/\varphi$. Thus neglecting φ_0, possibly because the foveal threshold is unknown, leads to a fractional error φ_0/φ. The studies in Table 2 used eccentricities $\varphi \geq 2$ deg. Our measurements estimate

φ_0 = 0.24 deg. Thus at 2 deg or beyond, the fractional error in estimated Bouma factor will be at most 0.24/2 = 12%. The fractional error drops to 5% at 5 deg and 2% at 10 deg.

Each Bouma factor was multiplied by a correction factor to account for criterion differences and log vs. linear flanker spacing (**Table 2**). Correction provides a *standardized* Bouma factor *b'* for each study (**Table 6**). **Figure 11** compares crowding across studies by plotting the standardized Bouma factor vs. meridian.

Row	Study	N	Gaze tracking	Target	Crowding orientation	mean standardized Bouma factor (*b'*) [95% CI]			
						Right	Left	Lower	Upper
1	Our data	50	Yes	Sloan font	radial	0.239 [0.22 0.25]	0.308 [0.30 0.32]	0.390 [0.38 0.40]	0.495 [0.48 0.49]
2	Our data	50	Yes	Sloan font	tangential	0.143 [0.14 0.15]	0.154 [0.14 0.15]	– –	– –
3	Our data	50	Yes	Pelli font	radial	0.325 [0.31 0.34]	0.377 [0.36 0.39]	– –	– –
4	Greenwood et al., 2017	12	Yes	tumbling clock	radial	0.313 [0.28 0.33]	0.343 [0.30 0.38]	0.464 [0.42 0.51]	0.636 [0.59 0.70]
5	Greenwood et al., 2017	12	Yes	tumbling clock	tangential	0.172 [0.16 0.19]	0.152 [0.13 0.17]	0.232 [0.20 0.25]	0.293 [0.26 0.33]
6	Toet & Levi, 1992	6	No	tumbling T	radial	– –	0.426 [0.36 0.51]	0.638 [0.57 0.72]	– –
7	Toet & Levi, 1992	6	No	tumbling T	tangential	– –	0.213 [0.20 0.23]	0.293 [0.27 0.32]	– –
8	Grainger et al., 2010	27	No	Courier New letter	radial	0.227 [0.19 0.28]	0.340 [0.27 0.37]	– –	– –
9	Grainger et al., 2010	27	No	Courier New symbol	radial	0.340 [0.29 0.39]	0.433 [0.38 0.48]	– –	– –
10	Coates et al., 2021	4	No	tumbling T	radial	0.386 [0.37 0.40]	– –	0.559 [0.55 0.57]	– –
B	Bouma 1970	25	No	Courier 10	radial	0.35	–	–	–

Table 6 – Standardized Bouma factor vs. meridian across studies. We compare our results with those of 4 studies that estimated crowding distance on at least two of the four cardinal meridians and report the Bouma factor by dividing crowding distance by the target's eccentricity. We standardized each Bouma factor by correcting for differences in number of target choices and threshold criterion (see last column of **Table 2**). Data from Greenwood et al. (2017) and Coates et al. (2021) were obtained through personal communication. Data from Toet and Levi (1992) and Grainger et al. (2010) were digitized from figures in their papers. Data from Bouma 1970 were replotted from Coates et al. (2021). We estimated the standard error of Bouma factors in the Grainger et al. (2010) study based on the p-value they reported for meridian differences.

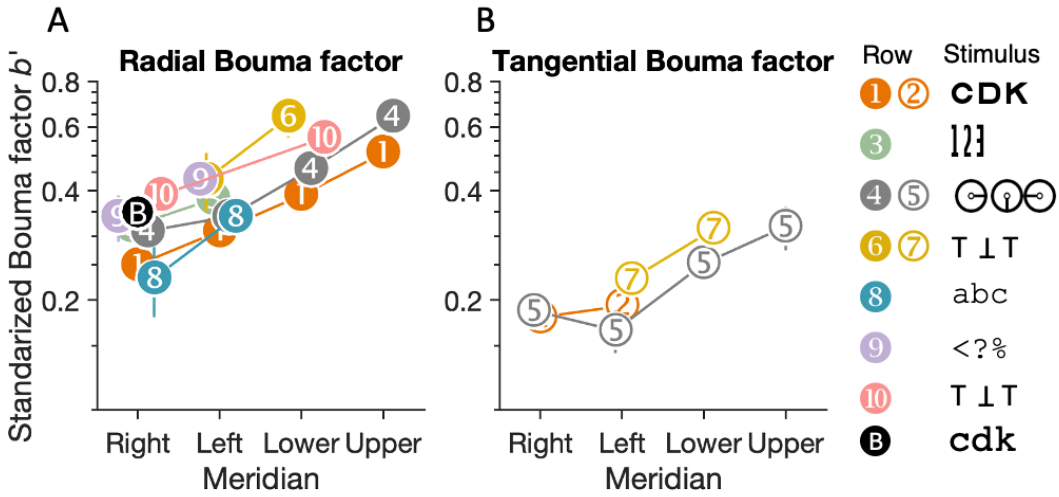

Figure 11. Standardized Bouma factor vs. meridian for various studies and target kinds. Each numbered point corresponds to a numbered row of data in Table 6. **A)** Comparison of radial-crowding studies and **B)** tangential-crowding studies. Both panels plot standardized Bouma factor vs. meridian. The legend shows the crowding stimuli. The white-on-black *B* symbol is the standardized Bouma factor estimated from Bouma (1970) data at 4 deg eccentricity with a 75% threshold criterion, with help from the reanalysis in Coates et al. (2021).

Effects of meridian and target kind. Two rows in **Table 6** — 1 (Sloan letter) and 4 (Tumbling clock) — report the radial standardized Bouma factor for all four cardinal meridians. **Figure 11A** shows that although the standardized Bouma factor is higher for the clock than for Sloan, by a factor of 1.3 (ratio of means for the right meridian 0.313/0.239 = 1.3), the two curves are otherwise similar, showing the same dependence on meridian. The 1.3:1 difference is not an artifact of number of choices or threshold criterion (**Table 2**). Both studies used gaze tracking to exclude fixation errors, so the difference is not a consequence of bad fixation. Thus, this seems to be a real 1.3:1 difference in standardized Bouma factor between target kinds, precisely what the target-kind factor t_{kind} is meant to account for in **Eq. 13**. The tumbling clocks may be more like each other than the 9 Sloan letters are and therefore produce larger crowding distance. Almost all other studies (Rows 3, 6, 9, 10) cluster above the Sloan font and show the same dependence on meridian. In general, we find that Courier New letters (Row 8) produce the smallest radial standardized Bouma factor (0.23 on the right meridian) and Tumbling Ts (Row 10) produce the largest radial standardized Bouma factor (0.39 on the right meridian).

Tangential crowding. We also compared standardized Bouma factors estimated with tangential flankers across studies (**Fig. 11B**). The tangential Bouma factor did not vary as much as radial, especially in the left meridian (**Fig. 11B**, Rows 2,5,7). Radial crowding estimates, even with the same stimuli, showed more variation in the standardized Bouma factor (**Fig. 11A**, Rows 1, 4, 6). Although our data did not show any difference between right and left meridians (Row 2), data extracted from Greenwood et al. (2017) do show a slight right:left advantage (Row 5).

Meridional asymmetries. **Table 7** and **Figure 12** report three Bouma-factor asymmetries (horizontal:vertical, right:left, and, lower:upper Bouma-factor ratios) for our and four selected studies. On average, the Bouma factor asymmetry is larger radially than tangentially. Radially, there is an advantage of horizontal over vertical meridian, right over

left meridian, and lower over upper meridian in every study. The horizontal:vertical advantage seems to be insensitive to object kind as the estimates are clustered around ratio of 0.6-0.7. Similarly, lower visual field advantage is close to 0.8 for both studies that tested at this location (Rows 1 and 4). The right:left asymmetry is the most variable. The right:left ratio is smallest for Courier New letters (Row 8) and largest for clocks (Row 4).

Row	Study	*N*	Gaze tracking	Target	Crowding orientation	Ratio of Bouma factors [95% CI]		
						Horizontal :vertical	Right:left	Lower:upper
1	Our data	50	Yes	Sloan font	radial	0.62 [0.60 0.64]	0.78 [0.78 0.91]	0.79 [0.76 0.82]
2	Our data	50	Yes	Sloan font	tangential	– –	0.93 [0.93 0.96]	– –
3	Our data	50	Yes	Pelli font	radial	– –	0.86 [0.82 0.90]	– –
4	Greenwood et al., 2017	12	Yes	tumbling clock	radial	0.59 [0.57 0.64]	0.91 [0.88 1.05]	0.73 [0.70 0.84]
5	Greenwood et al., 2017	12	Yes	tumbling clock	tangential	0.68 [0.6 0.75]	1.13 [0.97 1.40]	0.79 [0.73 0.91]
6	Toet & Levi, 1992	6	–	tumbling T	radial	0.67 [0.61 0.78]	– –	– –
7	Toet & Levi, 1992	6	–	tumbling T	tangential	0.73 [0.67 0.84]	– –	– –
8	Grainger et al., 2010	27	–	Courier New letter	radial	– –	0.67 [0.56 0.76]	– –
9	Grainger et al., 2010	27	–	Courier New symbol	radial	– –	0.79 [0.70 0.90]	– –
10	Coates et al., 2021	4	–	tumbling T	radial	0.69 [0.66 0.70]	– –	– –

Table 7 – Bouma factor and its asymmetries, compared with the literature. As in **Table 6**, we compare our results with those of studies that measured crowding distance on more than one of the four cardinal meridians**.** Each row lists the target type, crowding orientation, and mean ratio with standard-error intervals across participants.

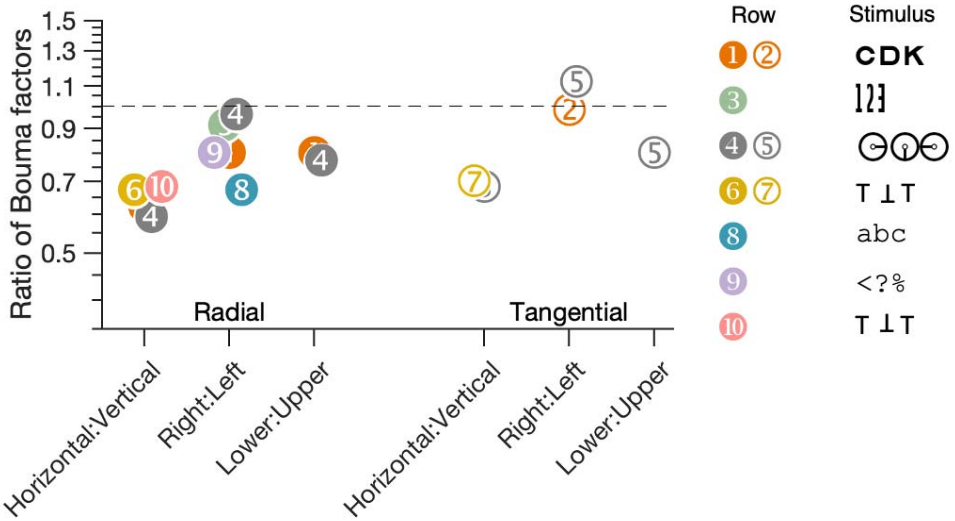

Figure 12. Plot of the three asymmetries, expressed as the ratio of Bouma factors. Each numbered point corresponds to a numbered row of data in Table 7. The horizontal dashed line at 1 represents no asymmetry. Each point is the ratio between Bouma factors. As in **Figure 11**, the legend shows examples of the objects that were used to measure crowding distance.

What does peripheral crowding distance add to foveal acuity?

Not predicted by acuity. Any evaluation of the usefulness of crowding distance as a biomarker must assess what crowding tells us about the observer over and above what can be gleaned from foveal acuity, which is routinely measured in all optometric and ophthalmic exams. For our 50 observers, foveal acuity failed to predict peripheral crowding, with an insignificant average correlation of 0.04 (**Fig. 13;** grey peripheral circles). More generally both acuity and crowding measured in the fovea fail to predict peripheral crowding (average foveal-peripheral crowding correlation is an insignificant correlation of 0.15 – red circles). Within the fovea, we do find a significant correlation between acuity and crowding (r = 0.64). Thus, foveal acuity predicts foveal crowding but not peripheral crowding. If peripheral crowding is of interest, e.g., as a possible limit to reading speed, then it should be measured, since it is not predicted by foveal acuity.

Foveal acuity and crowding. Our procedure measures threshold by covarying size and spacing. Since we found that foveal acuity and crowding are correlated, one might ask whether the correlation is due to the measured crowding threshold being contaminated by acuity limits. For five observers, using the same CriticalSpacing.m software, Pelli et al. (2016) measured foveal spacing threshold with the Pelli font with several spacing:size ratios and, for each observer, confirmed that all the measured thresholds correspond to one spacing at different sizes. For those five normally-sighted observers, this showed that the procedure measured a crowding threshold, not acuity.

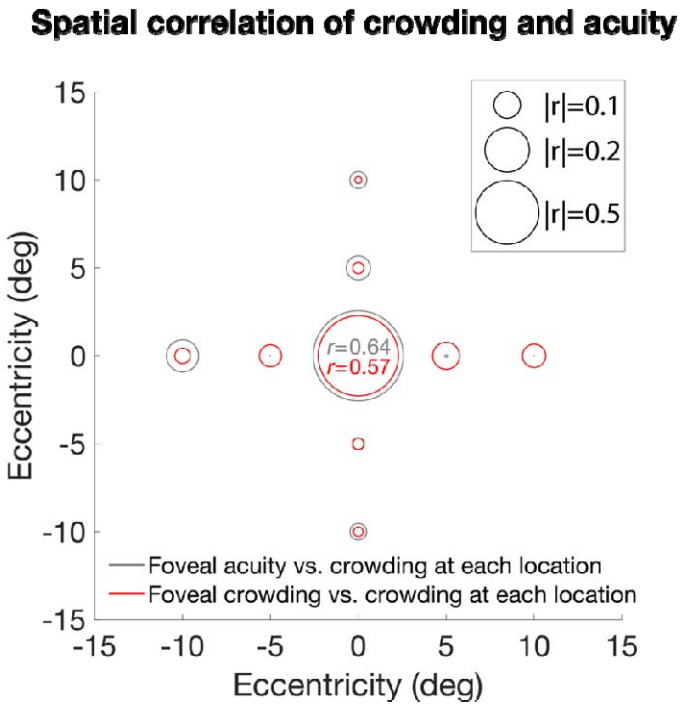

Figure 13. Foveal acuity and crowding fail to predict peripheral crowding. Correlation of foveal acuity (grey) and foveal crowding (red) with crowding everywhere. The central circles are test-retest for acuity (black) and crowding (red).

The peeking-observer model

Without gaze tracker. We wondered how the Bouma factor estimate depends on fixation accuracy and we wondered if fixation accuracy might explain the difference between the two Bouma factor histograms in **Figure 1**. Peripheral identification is hard, so, in ordinary life, we typically first foveate a peripheral target that we need to identify. Despite instructing observers to fixate on the central cross, we know that gaze could be elsewhere during the target presentation. The observer is torn between the desire to follow the instruction to fixate the cross and the natural impulse to fixate an anticipated peripheral target location. When the target location is randomly one of several peripheral locations, the participant's anticipation of location is often wrong. We model the participant's gaze position by two distributions, one without and one with peeking. First, for the no-peeking *awaited-fixation* distribution, we used the measured eye position in the *awaited-fixation* dataset, in which the participant's gaze was within 1.5 deg of the crosshair center for 250 ms immediately before target presentation, and we discarded trials in which gaze was more than 1.5 deg from the crosshair center during target presentation. The awaited-fixation gaze-position distribution is compact and roughly centered on the fixation crosshair. Second, we consider peeking toward a possible target location. We suppose that the participant peeks on a fraction p of the trials, and that the peeking eye movement travels only a fraction k of the distance from the crosshair to the possible target location, with a gaussian error (0.5 deg SD in x and in y). The *peeking distribution* has a mode corresponding to each possible target location, but at a fraction k of the possible target eccentricity. Gaze position is randomly sampled from the *peeking* distribution on a proportion p of trials and otherwise from the *awaited*-fixation distribution.

In the spirit of the Bouma law, our peeking-observer model assumes that the probability of identifying the target is given by a psychometric function

$$P(r) = 1 - 0.5\exp(-(\frac{r}{b_{\text{true}}})^{\beta}) \tag{16}$$

that depends solely on the ratio *r* of target-flanker spacing to actual target eccentricity, where b_{true} is the true Bouma factor and the steepness β is 2.3. For simplicity, the model omits threshold criterion and finger-error probability delta. Bouma factor *b* is estimated by 35 trials of Quest, assuming the true psychometric function, with a prior guess=0.11 of *r* and an assumed SD=2 of log *r.*

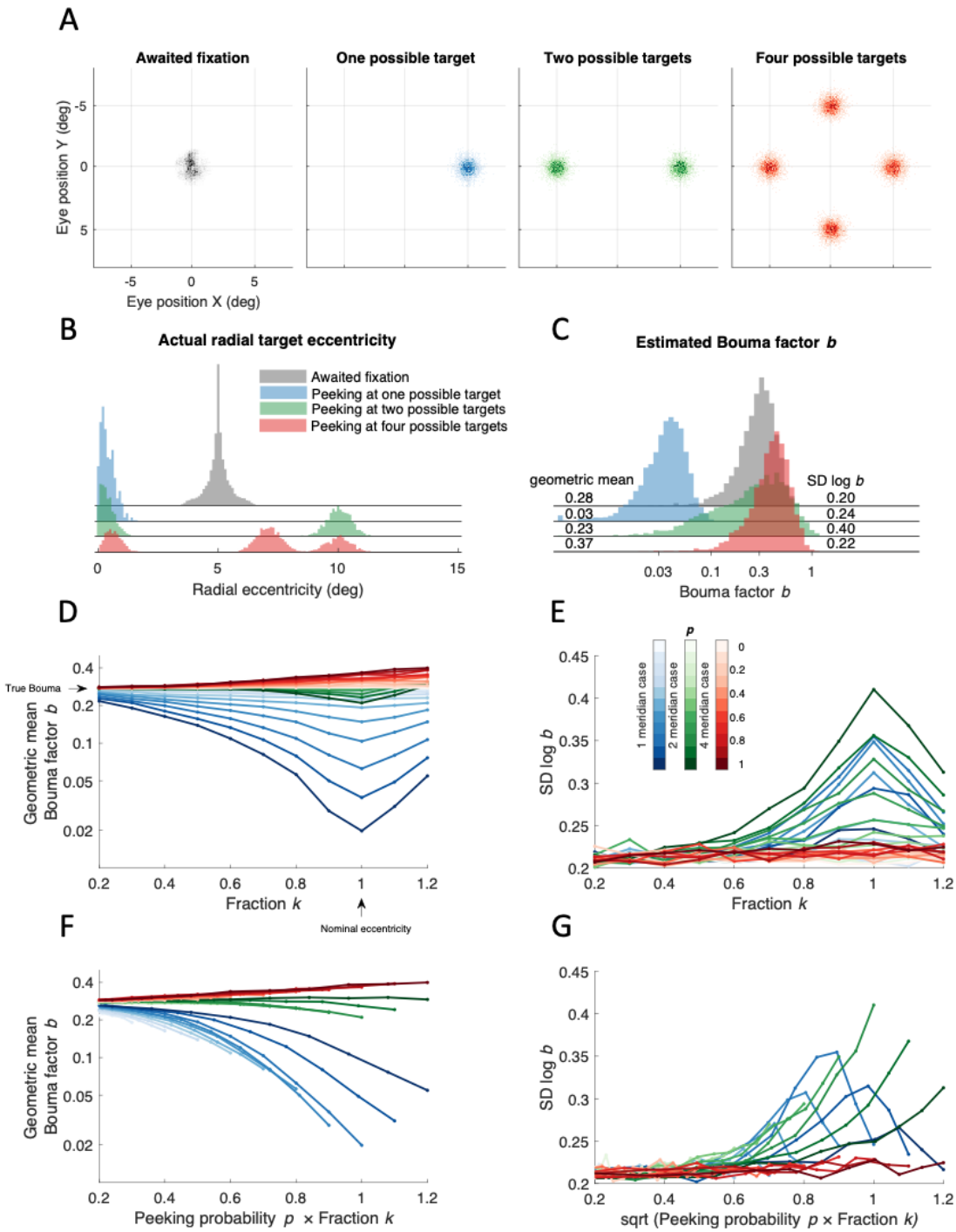

Figure 14. The peeking model. A) Scatter diagrams of the distribution of gaze position for each number of possible target locations. The gray histogram was measured by the EyeLink eye tracker in the *awaited-fixation* dataset. The green, blue, and red distributions of gaze position were synthesized assuming a full peek fraction *k*=1, and a gaussian (SD = 0.5 deg in x and 0.5 deg in y) centered at each possible target location **B)** Four histograms showing "actual" radial eccentricity of the target based on the distributions in panel A. ("actual" indicates that target eccentricity is computed relative to gaze position rather than the crosshair that the observer was asked to fixate.) **C)** 5000 Bouma factors estimated using Quest (35 trials) where each trial used the actual eccentricity. **D)** Geometric mean of the Bouma factor for each simulated case plotted vs. Fraction *k*.

Arrow on the vertical axis indicates true value of the Bouma that we assumed (0.3) and arrow on the horizontal axis shows nominal eccentricity (5 deg). **E)** SD of Bouma factor vs. fraction *k*. Color saturation indicates the percentage of trials in which observers' peek (see inset colorbar). **F-G)** Same as Panel D and E, however, the geometric mean of Bouma factor and its SD are plotted vs combination of peeking probability *p* with fraction *k*.

Awaited-fixation distribution. The *awaited*-fixation distribution was 3500 actual gaze positions at stimulus onset (35 trials x 50 participants x 2 sessions) measured with our EyeLink eye tracker in our *awaited*-fixation dataset. Recall that the stimulus was presented only once the gaze had been within 1.5 deg of the crosshair center for 250 ms.

Peeking distribution. We considered 1, 2, and 4 possible target locations (**Fig. 14A**). First, the target was always presented at one location (right meridian at 5 deg). Second, the target was randomly presented at ±5 deg on the horizontal midline. Third, the target was at 5 deg radial eccentricity on a random one of the four cardinal meridians (right, left, upper, lower). When participants "peek" a possible target location, depending on the number of possible locations, they have a 100%, 50%, or 25% chance of selecting the target location. Target and gaze position together define the actual target eccentricity.

Nominal eccentricity of the target is relative to the crosshair. *Actual* eccentricity of the target is relative to gaze position when the target is presented. Peeking near the target position will reduce the actual eccentricity to practically zero, while peeking another target location could result in an actual eccentricity greater than nominal. **Figure 14B** shows the actual target eccentricities calculated based on the four distributions of gaze position in **Figure 14A**.

The nominal radial eccentricity of the target is always 5 deg, so we simulate an observer with a threshold spacing of 1.5 deg by using a psychometric function provided by Quest (beta = 2.30, delta = 0.01, gamma = 0.11). Thus, our model assumes a Bouma factor of 1.5/5 = 0.3. This is the parameter we try to estimate. The point of this exercise is to evaluate how various methods estimate the true Bouma factor. We simulate a block of 35 trials using Quest to estimate the 70%-correct spacing threshold from which we calculate the Bouma factor. We repeat this many times to get a histogram of estimated Bouma factor (**Fig**. **14C**) for *awaited fixation* (gray) and *peeking* (colored) distributions.

The modelling shows that the geometric mean estimated Bouma factor *b* is lowest (0.03) for peeking with one possible location and highest with four (0.37), given a true Bouma factor *b* = 0.30. With no peeking (*p*=0), gaze position is from the *awaited*-fixation distribution, and the model estimate of Bouma factor is 0.28, very close to the true value of 0.30. The standard deviation of log Bouma factor *b* is highest (0.40) with two possible locations and lowest (0.22) with four.

The error (deviation from the assumed Bouma factor *b* = 0.3, see arrow on the vertical axis in **Fig. 14D**) in estimating the Bouma factor grows with the proportion *p* and fraction *k* of peeking (**Fig. 14D**). Observing that the error of estimated Bouma factor grows proportionally with *k* and that its SD grows proportionally with k^2 (note the parabolic shape in **Fig. 14E**), we produced new figures (**Fig. 14F-G**) showing that geometric mean *b* is roughly linear with $p \times k$ and SD log *b* with $\sqrt{p \times k}$.

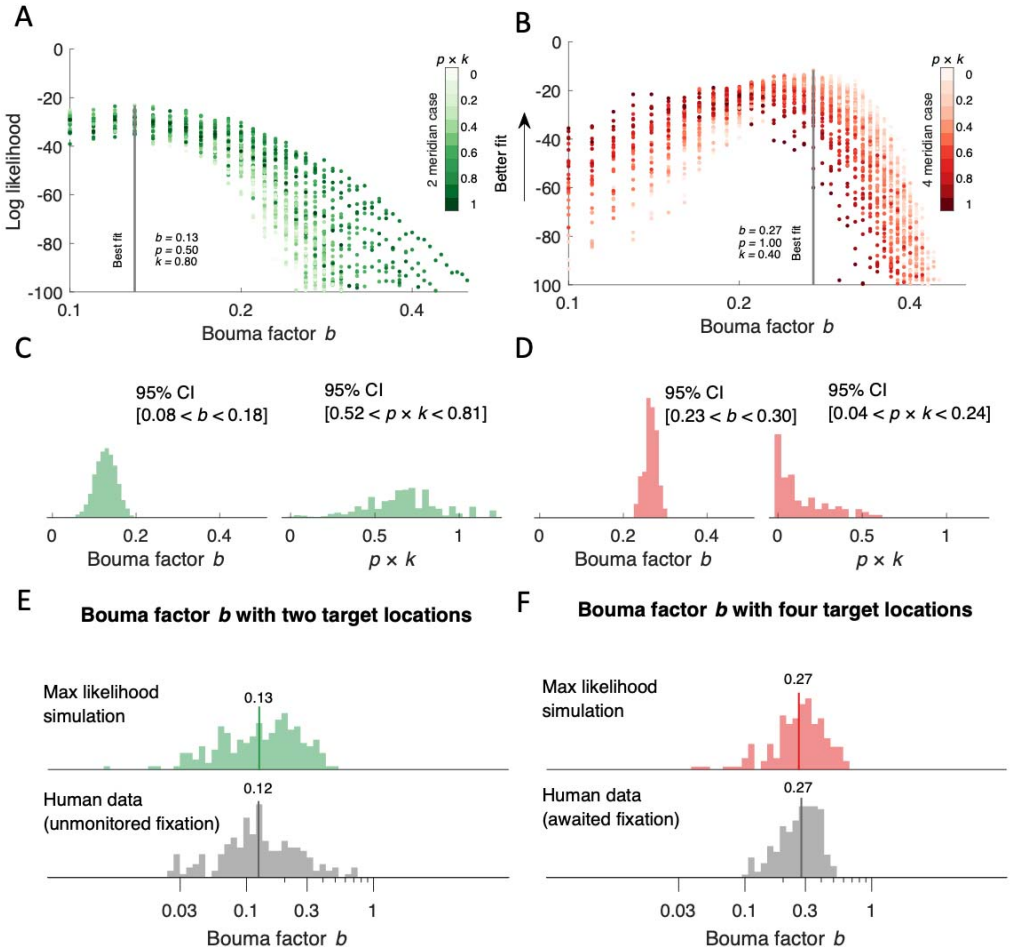

Figure 15. Comparing the peeking model with data. We used log-likelihood estimation to find the model parameter values that best fit our data. The higher the likelihood the better the fit. **A-B)** Log-likelihood is plotted vs. estimated Bouma factor *b*. Grey lines indicate best fit for which we plot model parameters. Brightness indicates the product $p \times k$. **C-D**) Bootstrapped model estimates. For each model we bootstrapped the fit (n=100) by randomly removing 25% of data at each iteration and fitting the model on the remaining data. Each histogram contains 100 best fitted parameters from each iteration. The $p \times k$ confidence intervals were calculated after bootstrapping the $p \times k$ parameter (n=1000) and averaging 10 random samples at each iteration. This creates a normal distribution and allows the calculation of confidence intervals. **E-F)** Comparison of the histograms of best-fitting simulated and acquired data. The solid vertical line indicates the geometric mean. For *unmonitored* fixation, the geometric mean of Bouma factor *b* was 0.13 for the simulated data and 0.12 for the human data. The SD of log Bouma factor was 0.32 for simulated data and 0.31 for human data. For the *awaited* fixation geometric mean of Bouma factor was 0.27 for the simulated data and 0.27 for the human data. The SD of log Bouma factor was 0.21 for simulated data and 0.19 for human data. Note that the red distribution in panel D is different from the red distribution in **Fig. 1**. Here we plot data from all four meridians. Data plotted in **Fig. 1** are extracted only from the right and left meridians.

Given *p*, *k,* and the true Bouma factor, our peeking-observer model predicts the estimated Bouma factor *b*. We estimate *p* and *k* by fitting the model using maximum likelihood optimization. The higher the log-likelihood the better the fit. We wanted to know how often (*p*) and how far (*k*) the observer peeks and what is the estimated Bouma factor *b.* In **Figure 15A-B**, the scatter and breadth of the log-likelihood distribution result in a broad confidence interval for the product $p \times k$. To estimate the error of the fit, we bootstrapped it by removing 25% of data at each iteration (n = 100). For *unmonitored* fixation, **Figure 15A**, bootstrapped parameters are consistent with high peeking ($0.5 < p \times k \leq 1$) and reject no

peeking ($p \times k = 0$). For *awaited* fixation, **Figure 15B**, bootstrapped parameters are consistent with low peeking ($p \times k < 0.5$) and reject high peeking ($p \times k = 1$). For each dataset, **Figure 15E-F** show that the human data are well matched by the simulated histogram. The two geometric means of b match, as do the standard deviations of log b. Our peeking model of eye position and crowding predicts the estimated Bouma factor in both cases, showing that the *unmonitored* fixation results are well fit by high peeking and the *awaited* fixation results are well fit by low peeking. One simple model of eye position and crowding fits all our data.

It is our impression that as observers gain experience with peripheral viewing, they peek less. Based on **Figure 15**, Bouma's (1970) peeking rate must have been practically zero.

DISCUSSION

Levi (2008), Pelli and Tillman (2008), Herzog et al. (2015), Strasburger (2020), and Coates et al. (2021) have reviewed the crowding literature. Most recently, Coates et al. (2021) provided a compact summary of the effects on the Bouma factor of contrast, size, target-flanker similarity and visual field location. This summary includes reanalysis of old data and shows a weak effect of stimulus duration. They also measured new data with two durations and two meridians confirming the effect of duration on the Bouma factor. Most of the data in their paper were acquired on fewer than 5 observers. We measured the effect of meridian, eccentricity, crowding orientation, and font with 50 observers. We did not measure effects of contrast, duration, or target-flanker similarity, but otherwise we confirm all the effects that they reported. We provide an equation predicting how crowding distance depends on meridian, target kind, and crowding orientation for each observer. We also show that crowding is reliable across days.

The Bouma law and factor

Bouma law. The Bouma law describes the linear increase of crowding distance with eccentricity (Bouma, 1970, 1973; Levi, 2008; Pelli & Tillman, 2008; Rosen et al., 2014). The Bouma factor is the slope of that line (Rosen et al., 2014). The Bouma law is robust when fit to individual observer's data (Pelli et al., 2004; Rosen et al., 2014; Strasburger, 2020; Strasburger et al., 1991). In this study, for the first time, we fit the Bouma law to data that include measurements from 50 observers tested with two crowding orientations at 9 locations of the visual field. The Bouma law is an excellent fit to our data and explains 82.5% of the variance despite being just a straight line with two degrees of freedom. We tried adding terms to the Bouma law to account for known factors: crowding orientation (Greenwood et al., 2017; Kwon et al., 2014; Petrov & Meleshkevich, 2011; Toet & Levi, 1992), meridional location of the stimulus (Fortenbaugh et al., 2015; Greenwood et al., 2017; He et al., 1996), target-kind (Coates et al., 2021; Grainger et al., 2010) and individual differences (Petrov & Meleshkevich, 2011; Veríssimo et al., 2021). We find that the enhanced model explains a bit more variance (increased from 82.5% to 94%). Eccentricity remains the dominant factor, accounting for 82.5% of the variance.

Standardized Bouma factor. We define the *standardized* Bouma factor *b*' as the reported Bouma factor *b* (ratio of crowding distance to radial eccentricity) multiplied by a correction factor that account for differences in task from Bouma's 25 choice alternatives, 75% threshold criterion, and linear flanker symmetry. Bouma reported a "roughly" 0.5 slope for radial letter crowding vs. eccentricity (Bouma, 1970). Andriessen and Bouma (1976) later reported a slope of 0.4 for crowding of lines. Coates et al. reanalyzed Bouma's original data with various threshold criteria so we interpolated between the 70% and 80% thresholds to estimate the 75% threshold. Estimating the Bouma factor from Bouma's original data using this criterion yields a Bouma factor of 0.35, in line with modern estimates of 0.3 (**Table 6** and Supplementary **Table 2**). **Figure 11A** shows that the corrected Bouma factor *b*' ranges from 0.23 for Courier New letters to 0.39 for Tumbling T measured with radial flankers on the right meridian. That residual difference may be due to target kind (Coates et al., 2021; Grainger et al., 2010). This is further supported by our finding that the Bouma factor was 0.78 lower for the Sloan font than for the Pelli font.

Supralinearity and the Bouma law

The linearity of the Bouma law implies that the Bouma factor is independent of eccentricity. The Coates et al. (2021) reanalysis of Bouma's 1970 data found a twofold increase of the Bouma factor with eccentricity (from 1 to 7 deg), with a log-log slope of 0.35. Coates et al. speculated that this eccentricity dependence might be due to Bouma's use of constant size stimuli at all eccentricities. Both acuity and crowding can limit measured thresholds for size and spacing across the visual field (Pelli et al., 2016; Song et al., 2014). If the threshold is independent of size, it is a crowding threshold and if the threshold is independent of spacing it is an acuity threshold. In 10 participants, we measured crowding distance at eccentricities of 0, 5, 10, 20, and 30 deg, scaling letter size with spacing as is now usual (see Methods). In our results with proportional letter size and controlled eye position, we find a similar twofold increase of the Bouma factor with eccentricity (from 5 to 30 deg), with a log-log slope of 0.38. Enhancing the Bouma law to allow a nonlinear dependence on eccentricity improves the fit to 10 observers' data slightly, increasing the variance accounted for from 90% to 95%. This effect is small but detectable in data from 0, 5, and 10 deg (**Fig. 9A**) and becomes pronounced at eccentricities of 20 and 30 deg (**Fig. 9B**). To our knowledge, only a few past studies measured crowding beyond 10 deg eccentricity (Bouma, 1970; Kalpadakis-Smith et al., 2022; Kwon & Liu, 2019; Pelli et al., 2004) and all these datasets show supralinear growth with eccentricity. From the perspective of mathematical modeling Bouma initially suggested a simple proportionality with one term, which later was extended to linearity with two terms, and the evidence for supralinearity justifies a three-term quadratic polynomial. Biologically, it seems possible that the increase of the Bouma factor at high eccentricity reflects a compression of eccentric visual field in higher order areas. Indeed, hV4 has a reduced peripheral representation when compared to earlier visual areas, V1, V2, and V3 (Arcaro et al., 2009; Goddard et al., 2011; Kolster et al., 2010; Winawer & Witthoft, 2015). This parallels the idea that the ventral visual stream, specialized in object recognition, emphasizes the central visual field (Levy et al., 2001; Ungerleider & Haxby, 1994).

Crowding asymmetries

At any given eccentricity, the Bouma factor varies with polar angle. The Bouma factor is lower along the horizontal than vertical meridian (Greenwood et al., 2017; Petrov & Meleshkevich, 2011; Toet & Levi, 1992), is higher in the upper than lower meridian (Fortenbaugh et al., 2015; Greenwood et al., 2017; He et al., 1996; Toet & Levi, 1992), tends to be lower in the right than left meridian (Grainger et al., 2010; White et al., 2020) and approximately halves with tangential flankers (Greenwood et al., 2017; Kwon et al., 2014). In this work, we replicated all these asymmetries (**Fig. 12** and **Table 7**). The horizontal vs. vertical advantage and better performance in the lower vs. upper visual field is found for many visual tasks (Himmelberg et al., 2023), and these asymmetries parallel those found in population receptive field size, cortical magnification, retinal ganglion cell density, and the BOLD signal magnitude (Benson et al., 2020; Himmelberg et al., 2021; Kupers et al., 2022; Kupers et al., 2019; Kurzawski et al., 2022; Kwon & Liu, 2019; Liu et al., 2006; Silva et al., 2018). The right:left asymmetry seems to be least described and does not generalize across all tasks. Beyond crowding, right visual field advantages have been reported: For native readers of left-to-right written languages, like English, the right meridian outperforms left in word recognition (Mishkin & Gorgays, 1952). Worrall and Coles (1976) examined letter recognition across the visual field and found a significant right hemifield advantage only along the right horizontal midline. The similarities in asymmetry suggest a common mechanism, and the differences may be useful hints toward the cortical substrate of crowding.

Standard deviation of measured acuity and crowding

To estimate the reliability of our measurements we acquired each threshold twice. Previous work showed improved performance in crowding tasks for repeated measurements (Chung, 2007; Malania et al., 2020). From their figures we estimate the second-block benefit to be 13% for Malania et al. and 20% for Chung. [Chung shows thresholds before and after 60 100-trial blocks. She shows percent correct for each block. By eye, we estimate that the benefit from first to second block is about a third that provided by the 60 blocks of training. Thus her 62% advantage (see average data from Table 1 – Chung 2007) after 60 blocks corresponds to the 20% advantage after the first block.] We find a modest second-threshold improvement for crowding thresholds measured with Sloan font at all tested locations and with the Pelli font in the fovea (less than 10%). Thresholds measured with Pelli font in the periphery yielded the highest improvement (23%). In our data the improvement is likely not due to acquiring familiarity with the task, since all observers participated in a training session, which consisted of repeated trials until 10 answers are correct. We find no improvement in acuity.

Overall, we find very good reproducibility of crowding and acuity thresholds (**Fig. 5**). The standard deviation of log Bouma factor *b* measured with Sloan font and radial flankers for test-retest is much lower than the standard deviation of log Bouma factor across observers (0.03 vs 0.08).

Individual differences

Estimating individual differences requires data from many observers. In this paper we measured crowding in 50 observers, which is the biggest dataset of crowding measurements

to date. Previous in-person crowding surveys included at most 27 observers (Grainger et al., 2010; Greenwood et al., 2017; Petrov & Meleshkevich, 2011; Toet & Levi, 1992). [An online crowding study tested 793 observers, but did not report individual differences (Liu et al., 2009).] To capture individual differences in the Bouma factor we included an observer factor o_i in the enhanced model. Adding the observer factor improved the explained variance from 92.6% to 94% (**Table 4**). Although this effect may seem negligible at first, we find that the Bouma factor varies twofold across observers, ranging from 0.20 to 0.38 (**Fig. 8A**). A similar, two-fold variation is observed for all other thresholds that we estimated (**Fig. 8A**).

Bouma factor as a biomarker. Large individual differences enhance crowding's potential as a biomarker for studying cortical health and development. Specifically, crowding varies across children too (Kalpadakis-Smith et al., 2022) and predicts RSVP reading speed (Pelli et al., 2007). Foveal crowding distance drops threefold from age 3 to 8 (Waugh et al., 2018). If crowding correlates with reading speed of beginning readers, then preliterate measures of crowding might help identify the children who need extra help before they learn to read. Measuring crowding distance across individuals in several diverse populations might expose any limit that crowding imposes on reading, yielding a norm for the development of crowding. Huge public interventions seek to help dyslexic children read faster and to identify them sooner. A virtue of crowding distance as a potential biomarker for dyslexia and cortical health is that it can be measured in 3.5 minutes.

Crowding correlations

This paper reports 13 crowding thresholds for each of 50 observers. Such a comprehensive dataset allows for a correlation analysis to assess how well each crowding threshold predicts the others. We find a moderate correlation of crowding between peripheral locations (r = 0.39 averaged across all peripheral locations) and hardly any between fovea and periphery (r = 0.11). We also find that crowding measured with radial flankers correlates highly with crowding measured with tangential flankers at the same location (r = 0.53 for the right meridian, r = 0.50 for the left meridian). The threshold measurement that best predicts all other peripheral thresholds (excluding the fovea), with a correlation r = 0.41, is radial Sloan crowding at 10 deg in the right meridian.

Effect of stimulus configuration vs. location. We find higher correlation (r = 0.54) when the location is the same and the stimulus configuration is changed, than (r = 0.32) when stimulus configuration is the same and location is changed. Correlation of crowding distance depends more on location than configuration. Paralleling our result, Poggel and Strasburger (2004) found only a weak correlation across meridians for visual reaction times. Surprisingly little is known about spatial correlation of basic measures like acuity and contrast sensitivity.

The peeking-observer model

We always asked the observer to fixate on the crosshair during each trial. We acquired data with two methods: *unmonitored fixation*, without gaze tracking, and *awaited fixation,* in which the stimulus was only presented when gaze was near the fixation cross. Both methods are described in detail in the Methods section. The two methods yield different Bouma factor distributions. Upgrading from *unmonitored* to *awaited* fixation increased the Bouma factor mean b from 0.12 to 0.20 and nearly halved the standard deviation of log b

from 0.31 to 0.18. Histograms are shown in **Figure 1**. This peeking-observer model assumes firstly that performance on each trial depends solely on target eccentricity (relative to gaze position), secondly that the observer peeks on a fraction *p* of the trials and fixates near the crosshair on the rest of the trials, and thirdly that the location of the peek is a fraction *k* of the distance from the fixation mark to the anticipated target location.

In *unmonitored* fixation, the observer peeks with probability *p*. In *awaited* fixation, peeking is prevented by using gaze-contingent display and discarding any trials where gaze left the fixation cross while the target was present. Suppose there are two possible target locations. The Bouma factor distribution is unimodal for low values of *p* and becomes bimodal for high values for *p*. Our *unmonitored b* histogram is bimodal is best fit with a peeking probability of 50%. *Our awaited* fixation b histogram is unimodal and best fit by peeking restricted to the 1.5 deg from the crosshair allowed by the gaze tracker. Upgrading from the bimodal to the unimodal *b* distribution raised the mean *b* from 0.12 to 0.27 and nearly halved the standard deviation of log *b* from 0.31 to 0.19.

The peeking model does not account for the reduction of a target's crowding distance that occurs in anticipation of a saccade to the target (Harrison et al., 2013). It is conceivable that on some *awaited*-fixation trials the observer was planning an eye movement to the correct target location and that this reduced crowding before the eye moved.

Effects of duration and peeking. Coates et al. (2021) reanalyzed crowding data from 16 studies and presented a scatter diagram of Bouma factor versus stimulus duration. The plot of Bouma factor vs. log stimulus duration had a semi-log slope of -0.16 describing how the Bouma factor drops with duration. Their analysis included many studies, with various threshold criteria, from various meridians, which introduced differences in the Bouma factor. To avoid these confounds, Coates et al. collected new data using a consistent threshold criterion and consistent locations. In their new results, increasing the duration from 67 to 500 ms decreased the Bouma factor by a factor of 1/1.6. However, none of these studies monitored fixation. Our **Figure 1** shows that, relative to controlled fixation, peeking can reduce the Bouma factor by a factor of 1.6 which is the size of the decrease with duration reported by Coates et al. If the probability of peeking grows with duration, then peeking might explain their drop in Bouma factor with duration.

Why measure crowding?

Peripheral crowding provides additional information about visual health. Acuity is the threshold size of a target for recognition, while crowding is a spacing threshold. Clinical assessment routinely includes foveal acuity and not crowding. Both limit recognition of everyday objects. Our results show that peripheral crowding is independent of foveal acuity and might be a useful biomarker of visual health. Specifically, peripheral crowding might predict dyslexia (Bouma & Legein, 1977; Martelli et al., 2009; O'Brien et al., 2005). There are hints that crowding tends to be worse in dyslexia (Pelli et al., 2007). If crowding correlates with reading speed of beginning readers, then preliterate measures of peripheral crowding might help identify the children who need extra help before they learn to read.

What about foveal crowding? In healthy individuals, foveal crowding correlates with foveal acuity but there are some conditions in which the two are dissociated. Strabismic amblyopia makes crowding worse in the fovea, but not in the periphery (Song et al., 2014). This suggests that the fovea might be the most sensitive place to detect the increase in crowding associated with amblyopia. Traditional tests for crowding are mostly peripheral and use a fixation mark and a brief peripheral target, which are poorly suited for testing children and dementia patients whose attention may wander. Such participants will fixate much more reliably on a foveal target.

We hope there will be clinical studies to assess the diagnostic benefit of measuring crowding, which takes 3.5 minutes.

CONCLUSIONS

1. The well-known Bouma law — crowding distance depends linearly on radial eccentricity — explains 82% of the variance of log crowding distance, cross-validated. Our enhanced Bouma law, with factors for observer, meridian, and target kind, explains 94% of the variance, cross-validated. The very good fit states the central accomplishment of the paper and shows how well the linear Bouma law fits human data.
2. The Bouma factor varies twofold across meridians, and radial vs. tangential crowding orientations.
3. Consistent with past reports, five asymmetries each confer an advantage expressed as a ratio of Bouma factors: 0.62 horizontal:vertical, 0.79 lower:upper, 0.78 right:left, 0.55 tangential:radial, and 0.78 Sloan font:Pelli font.
4. The Bouma factor varies twofold across observers. Differences across observers are much larger than those of test-retest. The 0.08 SD of log Bouma factor across observers is triple the 0.03 SD of test-retest, so one 3.5-minute threshold is enough to capture individual differences.
5. The growth of crowding distance with eccentricity is supralinear, but a linear fit is nearly as good, unless the range of eccentricities is huge.
6. Crowding distance measured at 10 deg eccentricity along the right meridian is the best predictor of average crowding distance elsewhere ($r = 0.39$).
7. Peripheral crowding is independent of foveal crowding and foveal acuity.
8. Peeking can be avoided by use of a gaze-contingent display. Peeking nearly halves the geometric mean Bouma factor b, and nearly doubles the standard deviation of log b, from 0.18 to 0.31.

ACKNOWLEDGEMENTS

We thank reviewer Hans Strasburger for many helpful suggestions which led to significant improvements, including the standardized Bouma factor. Thanks to Nina Hanning for helping with EyeLink hardware and software and for providing code to extract eye position at each trial. Thanks to Larry Maloney for suggesting bootstrap analysis of log-likelihood. Thanks to Barbara Finlay for helpful discussion of how psychophysical crowding correlations might relate to retinal developmental. Thanks to Maria Pombo for ANOVA help. Thanks to Ashley Feng and Laura Suciu for comments on the manuscript. The authors contributed as follows: Jan Kurzawski collected data with *awaited* fixation, analyzed data with *awaited* and *unmonitored* fixation and wrote the paper; Augustin Burchell performed initial analysis of *unmonitored fixation* data. Darshan Thapa supervised testing for data with *unmonitored fixation*; Najib Majaj discussed, designed, analyzed, and wrote; Jonathan Winawer discussed, and helped design, analyze, and write; Denis Pelli directed the project, including design, software, data collection, analysis, and writing. Jan, Najib, and Denis met most days for a year to do the revisions.

SUPPLEMENTARY MATERIALS

Effect of equating either the linear or log spacing of two radial flankers

When measuring radial crowding, the target lies between two flankers on a radial line from fixation. Bouma spaced the flankers equally, and most investigators have followed suit. However, we spaced the flankers symmetrically on a logarithmic rather than linear scale. This raises the question of how to compare crowding distances between experiments that spaced the flankers linearly vs logarithmically. Given the Bouma law (**Eq. 10**) and assuming that crowding distance depends primarily on the flanker-to-flanker distance, and negligibly on the target position between them, we show here that the crowding distance is expected to be 1.18 times larger when measured with linearly-spaced flankers than with log-spaced flankers.

Specifically, for a target at φ we choose the two flanker eccentricities φ_{in} and φ_{out} so that φ is located between φ_{in} and φ_{out}.

$$\log(\varphi_0+\varphi) = [\log(\varphi_0+\varphi_{in}) + \log(\varphi_0+\varphi_{out})] / 2, \quad \text{(S1)}$$

where φ_0=0.15 deg, and we report the inner spacing $s = \varphi-\varphi_{in}$. We can rearrange

$$\log(\varphi_0+\varphi_{out}) = 2 \log(\varphi_0+\varphi) - \log(\varphi_0+\varphi_{in}) \quad \text{(S2)}$$

Thus, the two flankers are at different (linear) distances from the target, but we suppose that they are equally effective in crowding the target. We report the center-to-center spacing from the inner flanker to the target as the “spacing” *s*.

We now estimate the relation of crowding distances measured in these two ways. We suppose that degree of crowding is determined by the separation between flankers on the log scale, as expected from the Bouma law:

$$\Delta\Phi = \log(\varphi_0+\varphi_{out}) - \log(\varphi_0+\varphi_{in}) \quad \text{(S3)}$$

If we are in the periphery, i.e. $\varphi_{in} > 1$, then φ_0=0.15 is negligible, and we can simplify,

$$\Delta\Phi \approx \log(\varphi_{out} / \varphi_{in}) \quad \text{(S4)}$$

Of course, increasing spacing alleviates crowding, so the degree of crowding will drop as log spacing $\Delta\Phi$ grows. Note that this model is at best an approximation, as it neglects position of the target. We are using it solely to compare crowding for two different ways of centering the target between flankers so the two target positions won't differ by much.

Linear flanker spacing: With flankers spaced symmetrically about the target on a linear scale, both at distance *s* from the target:

$$\Delta\Phi \approx \log (\varphi+s)/(\varphi-s) \quad \text{(S5)}$$

Log flanker spacing: With flankers spaced symmetrically on a log scale, the log flanker-to-flanker spacing is twice the log target-to-flanker spacing:

$$\Delta\Phi \approx 2 \log \varphi/(\varphi-s') \quad \text{(S6)}$$

Now we equate the two log flanker spacings, one with linearly symmetric spacing *s*, the other with log-symmetric spacing *s'*.

$$\log (\varphi+s)/(\varphi-s) = 2 \log \varphi/(\varphi-s') \quad \text{(S7)}$$

Solve for *s'*,

$$s' = \varphi - \varphi\, [(\varphi-s)/(\varphi+s)]^{0.5} \quad \text{(S8)}$$

Now substitute $b=s/\varphi$ and $b'=s'/\varphi$,

$$b' = 1 - [(1-b)/(1+b)]^{0.5} \quad \text{(S9)}$$

Figure S1 shows that, to a good approximation, this is a proportionality, with error of at most 0.014 over the relevant range $0 \le b \le 0.9$,

$$b' \approx 0.845\, b \quad \text{(S10)}$$

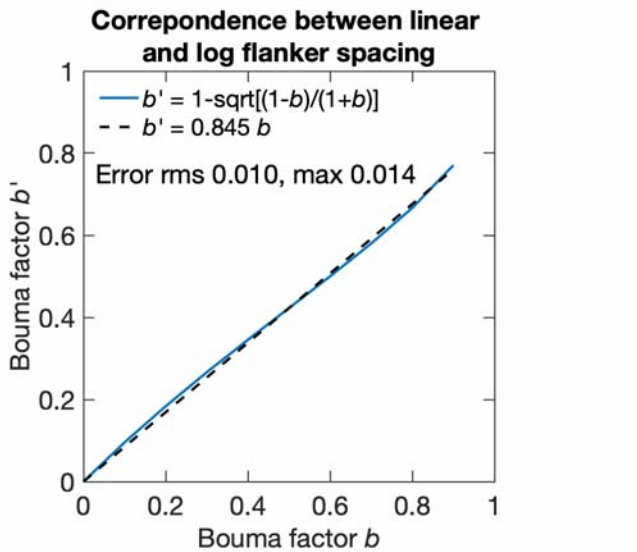

Figure S1. Log-symmetric spacing $s'=\varphi b'$ for same crowding effect as each linearly symmetric spacing $s = \varphi b$.

Thus, our log-symmetric spacing $s'=\varphi b'$ is approximately 0.845 times the linearly-symmetric spacing $s=\varphi b$ that is traditionally reported.

Log-log versions of the Bouma models

The negative intercept φ_0 is small and negligible at large eccentricity. (Zeroing it in the Bouma law produces less than 5% error in predicted crowding distance at eccentricities beyond 4.8 deg.) If we consider only peripheral results (>4.8 deg eccentric) we can set $\varphi_0=0$, and express the Bouma models in log coordinates (Table S2). The multiplicative combination rule becomes additive in the new coordinates. As with the linear models, the cross-validated variance explained R^2 increases with more parameters. At large eccentricity, these models are equivalent to the linear models presented in the main text, but fitting is quicker because the fitting error can be minimized by linear regression.

Model	Equation	R^2 (%)	Pearson's R	No. of parameters
Bouma law	$S \approx \Phi + B$	53.53	0.78	1
× meridional factor	$S \approx \Phi + B_\theta$	71.38	0.85	4
× crowding orientation	$S \approx \Phi + B_\theta + F_d$	80.47	0.89	6
× target-kind factor	$S \approx \Phi + B_\theta + F_d + T_{kind}$	81.13	0.90	8
× observer factor	$S \approx \Phi + B_\theta + F_d + T_{kind} + O_i$	84.32	0.92	58

Table S1 – Fitting the log-log version of the Bouma law (setting $\varphi_0=0$ and modeling only peripheral data $\varphi>4.8$ deg). Uppercase variables are the log10 of corresponding lowercase variables. R^2 represents variance explained after model cross-validation. All our fitting minimizes error in log crowding distance so fitting the log-log version can be fit using linear regression.

Correcting the Bouma factor

Comparison of crowding measured with different threshold criterion, number of choices, and log vs. linear spacing of flankers is facilitated by calculating the Standardized Bouma factor *b*′, which corrects for these factors. Each row number in Table S2 corresponds to a row in Table 6. The correction factors come from Table 2.

Row	**Correction factor**	Bouma factor **Standardized Bouma factor**			
		Right	**Left**	**Lower**	**Upper**
1	**× 1.30** **(1.10×1.18)**	0.184 **0.239**	0.237 **0.308**	0.300 **0.390**	0.381 **0.495**
3	**× 1.30** **(1.10×1.18)**	0.25 **0.325**	0.29 **0.377**		
4	**× 1.01**	0.31 **0.313**	0.34 **0.343**	0.46 **0.464**	0.63 **0.636**
6	**× 1.33**		0.32 **0.426**	0.48 **0.638**	
8	**× 1.03**	0.22 **0.227**	0.33 **0.340**		
9	**× 1.03**	0.33 **0.340**	0.42 **0.433**		
10	**× 1.33**	0.29 **0.386**		0.42 **0.559**	
Geometric mean		0.26 **0.30**	0.32 **0.37**	0.40 **0.50**	0.49 **0.56**

Table S2 – Radial Bouma factor before and after the correction. Row numbers correspond to numbers in Table 6. Each cell contains two numbers. The upper number is the Bouma factor before accounting for measurement differences and the lower one is the Standardized Bouma factor (already multiplied by the correction factor). Correction factors are calculated in Table 2. The Standardized Bouma factor is overall higher (e.g., 0.26 to 0.32 on the right meridian).

Afraz, A., Pashkam, M. V., & Cavanagh, P. (2010). Spatial heterogeneity in the perception of face and form attributes. *Curr Biol*, *20*(23), 2112-2116. https://doi.org/10.1016/j.cub.2010.11.017

Aghdaee, S. M. (2005). Adaptation to spiral motion in crowding condition. *Perception*, *34*(2), 155-162. https://doi.org/10.1068/p5298

Andriessen, J. J., & Bouma, H. (1976). Eccentric vision: adverse interactions between line segments. *Vision Res*, *16*(1), 71-78. https://doi.org/10.1016/0042-6989(76)90078-x

Anstis, S. (1974). Letter: A chart demonstrating variations in acuity with retinal position. *Vision Research*, *14 7*, 589-592.

Arcaro, M. J., McMains, S. A., Singer, B. D., & Kastner, S. (2009). Retinotopic organization of human ventral visual cortex. *J Neurosci*, *29*(34), 10638-10652. https://doi.org/10.1523/JNEUROSCI.2807-09.2009

Atilgan, N., Yu, S. M., & He, S. (2020). Visual crowding effect in the parvocellular and magnocellular visual pathways. *J Vis*, *20*(8), 6. https://doi.org/10.1167/jov.20.8.6

Atkinson, J., Pimm-Smith, E., Evans, C., Harding, G., & Braddick, O. (1986). Visual Crowding in Young Children. In B. Jay (Ed.), *Detection and Measurement of Visual Impairment in Pre-Verbal Children: Proceedings of a workshop held at the Institute of Ophthalmology, London on April 1–3, 1985, sponsored by the Commission of the European Communities as advised by the Committed on Medical Research* (pp. 201-213). Springer Netherlands. https://doi.org/10.1007/978-94-009-4263-9_27

Bacigalupo, F., & Luck, S. J. (2015). The allocation of attention and working memory in visual crowding. *J Cogn Neurosci*, *27*(6), 1180-1193. https://doi.org/10.1162/jocn_a_00771

Banks, W. P., Bachrach, K. M., & Larson, D. W. (1977). The asymmetry of lateral interference in visual letter identification. *Perception & Psychophysics*, *22*(3), 232-240. https://doi.org/10.3758/BF03199684

Benson, N. C., Kupers, E. R., Barbot, A., Carrasco, M., & Winawer, J. (2020). Cortical Magnification in Human Visual Cortex Parallels Task Performance around the Visual Field. *bioRxiv*, 2020.2008.2026.268383. https://doi.org/10.1101/2020.08.26.268383

Bex, P. J., & Dakin, S. C. (2005). Spatial interference among moving targets. *Vision Res*, *45*(11), 1385-1398. https://doi.org/10.1016/j.visres.2004.12.001

Bex, P. J., Dakin, S. C., & Simmers, A. J. (2003). The shape and size of crowding for moving targets. *Vision Res*, *43*(27), 2895-2904. https://doi.org/10.1016/s0042-6989(03)00460-7

Bi, T., Cai, P., Zhou, T., & Fang, F. (2009). The effect of crowding on orientation-selective adaptation in human early visual cortex. *J Vis*, *9*(11), 13 11-10. https://doi.org/10.1167/9.11.13

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

Bouma, H. (1973). Visual interference in the parafoveal recognition of initial and final letters of words. *Vision Res*, *13*(4), 767-782. https://doi.org/10.1016/0042-6989(73)90041-2

Bouma, H., & Legein, C. P. (1977). Foveal and parafoveal recognition of letters and words by dyslexics and by average readers. *Neuropsychologia*, *15*(1), 69-80. https://doi.org/10.1016/0028-3932(77)90116-6

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spat Vis*, *10*(4), 433-436. https://www.ncbi.nlm.nih.gov/pubmed/9176952

Burchell, A., Benson, N. C., Zhou, J. Y., Winawer, J. A., & Pelli, D. G. (2019). Using fMRI to link crowding to hV4. *Journal of Vision*, *19*(10), 14a-14a. https://doi.org/10.1167/19.10.14a

Carrasco, M. (2011). Visual attention: the past 25 years. *Vision Res*, *51*(13), 1484-1525. https://doi.org/10.1016/j.visres.2011.04.012

Chastain, G. (1982). Confusability and interference between members of parafoveal letter pairs. *Perception & Psychophysics*, *32*(6), 576-580. https://doi.org/10.3758/BF03204213

Chung, S. T. (2007). Learning to identify crowded letters: does it improve reading speed? *Vision Res*, *47*(25), 3150-3159. https://doi.org/10.1016/j.visres.2007.08.017

Chung, S. T. (2013). Cortical reorganization after long-term adaptation to retinal lesions in humans. *J Neurosci*, *33*(46), 18080-18086. https://doi.org/10.1523/JNEUROSCI.2764-13.2013

Coates, D. R., Levi, D. M., Touch, P., & Sabesan, R. (2018). Foveal Crowding Resolved. *Sci Rep*, *8*(1), 9177. https://doi.org/10.1038/s41598-018-27480-4

Coates, D. R., Ludowici, C. J. H., & Chung, S. T. L. (2021). The generality of the critical spacing for crowded optotypes: From Bouma to the 21st century. *J Vis*, *21*(11), 18. https://doi.org/10.1167/jov.21.11.18

Ehlers, H. (1936). V: THE MOVEMENTS OF THE EYES DURING READING [https://doi.org/10.1111/j.1755-3768.1936.tb07306.x]. *Acta Ophthalmologica*, *14*(1-2), 56-63. https://doi.org/https://doi.org/10.1111/j.1755-3768.1936.tb07306.x

Ehlers, H. (1953). Clinical testing of visual acuity. *AMA Arch Ophthalmol*, *49*(4), 431-434. https://doi.org/10.1001/archopht.1953.00920020441007

Elliott, D. B., Whitaker, D., & Bonette, L. (1990). Differences in the legibility of letters at contrast threshold using the Pelli-Robson chart [https://doi.org/10.1111/j.1475-1313.1990.tb00877.x]. *Ophthalmic and Physiological Optics*, *10*(4), 323-326. https://doi.org/https://doi.org/10.1111/j.1475-1313.1990.tb00877.x

Estes, W. K., Allmeyer, D. H., & Reder, S. M. (1976). Serial position functions for letter identification at brief and extended exposure durations. *Perception & Psychophysics*, *19*(1), 1-15. https://doi.org/10.3758/BF03199379

Farzin, F., Rivera, S. M., & Whitney, D. (2009). Holistic crowding of Mooney faces. *J Vis*, *9*(6), 18 11-15. https://doi.org/10.1167/9.6.18

Fischer, B. (1973). Overlap of receptive field centers and representation of the visual field in the cat's optic tract. *Vision Res*, *13*(11), 2113-2120. https://doi.org/10.1016/0042-6989(73)90188-0

Flom, M. C., Heath, G. G., & Takahashi, E. (1963). Contour Interaction and Visual Resolution: Contralateral Effects. *Science*, *142*(3594), 979-980. https://doi.org/10.1126/science.142.3594.979

Flom, M. C., Weymouth, F. W., & Kahneman, D. (1963). Visual Resolution and Contour Interaction. *J Opt Soc Am*, *53*, 1026-1032. https://doi.org/10.1364/josa.53.001026

Fortenbaugh, F. C., Silver, M. A., & Robertson, L. C. (2015). Individual differences in visual field shape modulate the effects of attention on the lower visual field advantage in crowding. *J Vis*, *15*(2). https://doi.org/10.1167/15.2.19

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

Goddard, E., Mannion, D. J., McDonald, J. S., Solomon, S. G., & Clifford, C. W. (2011). Color responsiveness argues against a dorsal component of human V4. *J Vis*, *11*(4). https://doi.org/10.1167/11.4.3

Grainger, J., Tydgat, I., & Issele, J. (2010). Crowding affects letters and symbols differently. *J Exp Psychol Hum Percept Perform*, *36*(3), 673-688. https://doi.org/10.1037/a0016888

Greenwood, J. A., Szinte, M., Sayim, B., & Cavanagh, P. (2017). Variations in crowding, saccadic precision, and spatial localization reveal the shared topology of spatial vision. *Proc Natl Acad Sci U S A*, *114*(17), E3573-E3582. https://doi.org/10.1073/pnas.1615504114

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

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

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

Himmelberg, M. M., Kurzawski, J. W., Benson, N. C., Pelli, D. G., Carrasco, M., & Winawer, J. (2021). Cross-dataset reproducibility of population receptive field (pRF) estimates and cortical magnification asymmetries. *bioRxiv*, 2021.2004.2012.439348. https://doi.org/10.1101/2021.04.12.439348

Himmelberg, M. M., Winawer, J., & Carrasco, M. (2023). Polar angle asymmetries in visual perception and neural architecture. *Trends Neurosci*. https://doi.org/10.1016/j.tins.2023.03.006

Kalpadakis-Smith, A. V., Tailor, V. K., Dahlmann-Noor, A. H., & Greenwood, J. A. (2022). Crowding changes appearance systematically in peripheral, amblyopic, and developing vision. *J Vis*, *22*(6), 3. https://doi.org/10.1167/jov.22.6.3

Kewan-Khalayly, B., Migo, M., & Yashar, A. (2022). Transient attention equally reduces visual crowding in radial and tangential axes. *J Vis*, *22*(9), 3. https://doi.org/10.1167/jov.22.9.3

Kolster, H., Peeters, R., & Orban, G. A. (2010). The retinotopic organization of the human middle temporal area MT/V5 and its cortical neighbors. *J Neurosci*, *30*(29), 9801-9820. https://doi.org/10.1523/JNEUROSCI.2069-10.2010

Kooi, F. L., Toet, A., Tripathy, S. P., & Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spat Vis*, *8*(2), 255-279. https://doi.org/10.1163/156856894x00350

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

Krumhansl, C. L. (1977). Naming and locating simultaneously and sequentially presented letters. *Perception & Psychophysics*, *22*(3), 293-302. https://doi.org/10.3758/BF03199693

Kupers, E. R., Benson, N. C., Carrasco, M., & Winawer, J. (2022). Asymmetries around the visual field: From retina to cortex to behavior. *PLoS Comput Biol*, *18*(1), e1009771. https://doi.org/10.1371/journal.pcbi.1009771

Kupers, E. R., Carrasco, M., & Winawer, J. (2019). Modeling visual performance differences 'around' the visual field: A computational observer approach. *PLoS Comput Biol*, *15*(5), e1007063. https://doi.org/10.1371/journal.pcbi.1007063

Kurzawski, J. W., Gulban, O. F., Jamison, K., Winawer, J., & Kay, K. (2022). Non-neural factors influencing BOLD response magnitudes within individual subjects. *J Neurosci*, *42*(38), 7256-7266. https://doi.org/10.1523/JNEUROSCI.2532-21.2022

Kurzawski, J. W., Pelli, D. G., & Winawer, J. A. (2021). Conservation across individuals of cortical crowding distance in human V4. *Journal of Vision*, *21*(9), 2675-2675. https://doi.org/10.1167/jov.21.9.2675

Kwon, M., Bao, P., Millin, R., & Tjan, B. S. (2014). Radial-tangential anisotropy of crowding in the early visual areas. *J Neurophysiol*, *112*(10), 2413-2422. https://doi.org/10.1152/jn.00476.2014

Kwon, M., & Liu, R. (2019). Linkage between retinal ganglion cell density and the nonuniform spatial integration across the visual field. *Proc Natl Acad Sci U S A*, *116*(9), 3827-3836. https://doi.org/10.1073/pnas.1817076116

Latham, K., & Whitaker, D. (1996). Relative roles of resolution and spatial interference in foveal and peripheral vision. *Ophthalmic and Physiological Optics*, *16*(1), 49-57. https://doi.org/https://doi.org/10.1016/0275-5408(95)00124-7

Leat, S. J., Li, W., & Epp, K. (1999). Crowding in central and eccentric vision: the effects of contour interaction and attention. *Invest Ophthalmol Vis Sci*, *40*(2), 504-512. https://www.ncbi.nlm.nih.gov/pubmed/9950611

Levi, D. M. (2008). Crowding--an essential bottleneck for object recognition: a mini-review. *Vision Res*, *48*(5), 635-654. https://doi.org/10.1016/j.visres.2007.12.009

Levi, D. M., & Carney, T. (2009). Crowding in peripheral vision: why bigger is better. *Curr Biol*, *19*(23), 1988-1993. https://doi.org/10.1016/j.cub.2009.09.056

Levi, D. M., Klein, S. A., & Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Res*, *25*(7), 963-977. https://doi.org/10.1016/0042-6989(85)90207-x

Levy, I., Hasson, U., Avidan, G., Hendler, T., & Malach, R. (2001). Center-periphery organization of human object areas. *Nat Neurosci*, *4*(5), 533-539. https://doi.org/10.1038/87490

Li, Q., Joo, S. J., Yeatman, J. D., & Reinecke, K. (2020). Controlling for Participants' Viewing Distance in Large-Scale, Psychophysical Online Experiments Using a Virtual Chinrest. *Sci Rep*, *10*(1), 904. https://doi.org/10.1038/s41598-019-57204-1

Liu, L., & Arditi, A. (2000). Apparent string shortening concomitant with letter crowding. *Vision Res*, *40*(9), 1059-1067. https://doi.org/10.1016/s0042-6989(99)00247-3

Liu, T., Heeger, D. J., & Carrasco, M. (2006). Neural correlates of the visual vertical meridian asymmetry. *J Vis*, *6*(11), 1294-1306. https://doi.org/10.1167/6.11.12

Liu, T., Jiang, Y., Sun, X., & He, S. (2009). Reduction of the crowding effect in spatially adjacent but cortically remote visual stimuli. *Curr Biol*, *19*(2), 127-132. https://doi.org/10.1016/j.cub.2008.11.065

Louie, E. G., Bressler, D. W., & Whitney, D. (2007). Holistic crowding: selective interference between configural representations of faces in crowded scenes. *J Vis*, *7*(2), 24 21-11. https://doi.org/10.1167/7.2.24

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

Malania, M., Pawellek, M., Plank, T., & Greenlee, M. W. (2020). Training-Induced Changes in Radial-Tangential Anisotropy of Visual Crowding. *Transl Vis Sci Technol*, *9*(9), 25. https://doi.org/10.1167/tvst.9.9.25

Martelli, M., Di Filippo, G., Spinelli, D., & Zoccolotti, P. (2009). Crowding, reading, and developmental dyslexia. *J Vis*, *9*(4), 14 11-18. https://doi.org/10.1167/9.4.14

Martelli, M., Majaj, N. J., & Pelli, D. G. (2005). Are faces processed like words? A diagnostic test for recognition by parts. *J Vis*, *5*(1), 58-70. https://doi.org/10.1167/5.1.6

Millin, R., Arman, A. C., Chung, S. T., & Tjan, B. S. (2014). Visual crowding in V1. *Cereb Cortex*, *24*(12), 3107-3115. https://doi.org/10.1093/cercor/bht159

Mishkin, M., & Gorgays, D. G. (1952). Word recognition as a function of retinal locus. *J Exp Psychol*, *43*(1), 43-48. https://doi.org/10.1037/h0061361

Motter, B. C. (2006). Modulation of transient and sustained response components of V4 neurons by temporal crowding in flashed stimulus sequences. *J Neurosci*, *26*(38), 9683-9694. https://doi.org/10.1523/JNEUROSCI.5495-05.2006

Nazir, T. A. (1992). Effects of lateral masking and spatial precueing on gap-resolution in central and peripheral vision. *Vision Res*, *32*(4), 771-777. https://doi.org/10.1016/0042-6989(92)90192-l

O'Brien, B. A., Mansfield, J. S., & Legge, G. E. (2005). The effect of print size on reading speed in dyslexia. *J Res Read*, *28*(3), 332-349. https://doi.org/10.1111/j.1467-9817.2005.00273.x

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nat Neurosci*, *4*(7), 739-744. https://doi.org/10.1038/89532

Pelli, D., & Robson, J. (1991). Are letters better than gratings? *Clin Vis Sci*, *6*, 409-411.

Pelli, D., Waugh, S., Martelli, M., Crutch, S., Primativo, S., Yong, K., Rhodes, M., Yee, K., Wu, X., Famira, H., & Yiltiz, H. (2016). A clinical test for visual crowding. *F1000Research*, *5*. https://doi.org/10.12688/f1000research.7835.1

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming numbers into movies. *Spat Vis*, *10*(4), 437-442. https://www.ncbi.nlm.nih.gov/pubmed/9176953

Pelli, D. G. (2008). Crowding: a cortical constraint on object recognition. *Curr Opin Neurobiol*, *18*(4), 445-451. https://doi.org/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. *J Vis*, *4*(12), 1136-1169. https://doi.org/10.1167/4.12.12

Pelli, D. G., Robson, J. G., & Wilkins, A. J. (1988). The design of a new letter chart for measuring contrast sensitivity [Article]. *Vision Research*, *2*(3), 187-199. http://www.scopus.com/inward/record.url?scp=0023948250&partnerID=8YFLogxK http://www.scopus.com/inward/citedby.url?scp=0023948250&partnerID=8YFLogxK

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nat Neurosci*, *11*(10), 1129-1135. https://doi.org/10.1038/nn.2187

Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *J Vis*, *7*(2), 20 21-36. https://doi.org/10.1167/7.2.20

Petrov, Y., & Meleshkevich, O. (2011). Asymmetries and idiosyncratic hot spots in crowding. *Vision Res*, *51*(10), 1117-1123. https://doi.org/10.1016/j.visres.2011.03.001

Poggel, D. A., & Strasburger, H. (2004). Visual perception in space and time--mapping the visual field of temporal resolution. *Acta Neurobiol Exp (Wars)*, *64*(3), 427-437. https://www.ncbi.nlm.nih.gov/pubmed/15283484

Rohatgi, A. (2020). Webplotdigitizer: Version 4.4. In.

Rosen, S., Chakravarthi, R., & Pelli, D. (2014). The Bouma law of crowding, revised: Critical spacing is equal across parts, not objects. *Journal of Vision*, *14*. https://doi.org/10.1167/14.6.10

Rosenholtz, R. (2016). Capabilities and Limitations of Peripheral Vision. *Annu Rev Vis Sci*, *2*, 437-457. https://doi.org/10.1146/annurev-vision-082114-035733

Scolari, M., Kohnen, A., Barton, B., & Awh, E. (2007). Spatial attention, preview, and popout: which factors influence critical spacing in crowded displays? *J Vis*, *7*(2), 7 1-23. https://doi.org/10.1167/7.2.7

Siderov, J., Waugh, S. J., & Bedell, H. E. (2013). Foveal contour interaction for low contrast acuity targets. *Vision Res*, *77*, 10-13. https://doi.org/10.1016/j.visres.2012.11.008

Silva, M. F., Brascamp, J. W., Ferreira, S., Castelo-Branco, M., Dumoulin, S. O., & Harvey, B. M. (2018). Radial asymmetries in population receptive field size and cortical magnification factor in early visual cortex. *Neuroimage*, *167*, 41-52. https://doi.org/10.1016/j.neuroimage.2017.11.021

Sloan, L. L., Rowland, W. M., & Altman, A. (1952). Comparison of three types of test target for the measurement of visual acuity. *Q Rev Ophthalmol*, *8*(1), 4-16.

Song, S., Levi, D. M., & Pelli, D. G. (2014). A double dissociation of the acuity and crowding limits to letter identification, and the promise of improved visual screening. *J Vis*, *14*(5), 3. https://doi.org/10.1167/14.5.3

Strasburger, H. (2005). Unfocused spatial attention underlies the crowding effect in indirect form vision. *J Vis*, *5*(11), 1024-1037. https://doi.org/10.1167/5.11.8

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

Strasburger, H., Harvey, L. O., Jr., & Rentschler, I. (1991). Contrast thresholds for identification of numeric characters in direct and eccentric view. *Percept Psychophys*, *49*(6), 495-508. https://doi.org/10.3758/bf03212183

Strasburger, H., & Malania, M. (2013). Source confusion is a major cause of crowding. *J Vis*, *13*(1), 24. https://doi.org/10.1167/13.1.24

Strasburger, H., Rentschler, I., & Juttner, M. (2011). Peripheral vision and pattern recognition: a review. *J Vis*, *11*(5), 13. https://doi.org/10.1167/11.5.13

Stuart, J. A., & Burian, H. M. (1962). A study of separation difficulty. Its relationship to visual acuity in normal and amblyopic eyes. *Am J Ophthalmol*, *53*, 471-477. https://www.ncbi.nlm.nih.gov/pubmed/13917936

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

Tripathy, S. P., & Cavanagh, P. (2002). The extent of crowding in peripheral vision does not scale with target size. *Vision Res*, *42*(20), 2357-2369. https://doi.org/10.1016/s0042-6989(02)00197-9

Tyler, C. W., & Likova, L. T. (2007). Crowding: a neuroanalytic approach. *J Vis*, *7*(2), 16 11-19. https://doi.org/10.1167/7.2.16

Ungerleider, L. G., & Haxby, J. V. (1994). 'What' and 'where' in the human brain. *Curr Opin Neurobiol*, *4*(2), 157-165. https://doi.org/10.1016/0959-4388(94)90066-3

Veríssimo, I. S., Hölsken, S., & Olivers, C. N. L. (2021). Individual differences in crowding predict visual search performance. *Journal of Vision*, *21*(5), 29-29. https://doi.org/10.1167/jov.21.5.29

Wallis, T. S., & Bex, P. J. (2012). Image correlates of crowding in natural scenes. *J Vis*, *12*(7). https://doi.org/10.1167/12.7.6

Watson, A. B., & Pelli, D. G. (1983). QUEST: a Bayesian adaptive psychometric method. *Percept Psychophys*, *33*(2), 113-120. https://doi.org/10.3758/bf03202828

Waugh, S., Pelli, D., Álvaro, L., & Formankiewicz, M. (2018). Crowding distance in healthy children. *Journal of Vision*, *18*(10), 855-855. https://doi.org/10.1167/18.10.855

Westheimer, G., & Hauske, G. (1975). Temporal and spatial interference with vernier acuity. *Vision Res*, *15*, 1137-1141. https://doi.org/10.1016/0042-6989(75)90012-7

Westheimer, G., Shimamura, K., & McKee, S. P. (1976). Interference with line-orientation sensitivity. *J Opt Soc Am*, *66*(4), 332-338. https://doi.org/10.1364/josa.66.000332

White, A., Tang, K., & Yeatman, J. (2020). The field of view for word recognition: crowding and hemifield asymmetries. *Journal of Vision*, *20*(11), 911-911. https://doi.org/10.1167/jov.20.11.911

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

Winawer, J., & Witthoft, N. (2015). Human V4 and ventral occipital retinotopic maps. *Vis Neurosci*, *32*, E020. https://doi.org/10.1017/S0952523815000176

Worrall, N., & Coles, P. (1976). Visual field differences in recognizing letters. *Perception & Psychophysics*, *20*(1), 21-24. https://doi.org/10.3758/BF03198698

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

Zhou, J., Benson, N., Pelli, D., & Winawer, J. (2017). Conservation of crowding distance in human V4. *Journal of Vision*, *17*(15), 19-19. https://doi.org/10.1167/17.15.19