Crowding results from optimal integration of visual targets with contextual information

Guido Marco Cicchini
Consiglio Nazionale delle Ricerche https://orcid.org/0000-0002-3303-0420
Giovanni D'Errico
CNR Neuroscience Institute https://orcid.org/0000-0002-0491-581X
David Burr (dave@in.cnr.it)
University of Florence https://orcid.org/0000-0003-1541-8832

Article

Keywords:

Posted Date: March 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1296243/v1

Version of Record: A version of this preprint was published at Nature Communications on September 30th, 2022. See the published version at https://doi.org/10.1038/s41467-022-33508-1.

Crowding results from optimal integration of visual targets with contextual information

Guido Marco Cicchini[1], Giovanni D'Errico[1] and David C. Burr[1,2]

1. Institute of Neuroscience, CNR, via Moruzzi, 1 , 56124 – Pisa (ITALY)
2. Department of Neurosciences, Psychology, Drug Research and Child Health, University of Florence, viale Pieraccini, 6 – 50139 Firenze (ITALY)

ABSTRACT

Crowding is the inability to recognize peripheral objects in clutter, usually considered a fundamental low-level bottleneck to object recognition. Here we advance and test an alternative hypothesis, that crowding, like "serial dependence", results from optimizing strategies that exploit redundancies in natural scenes. This notion leads to several testable predictions: (1) crowding should be greatest for unreliable targets and reliable flankers; (2) crowding-induced biases should be maximal when target and flankers have similar orientations, falling off for differences around 20°; (3) flanker interference should be associated with higher precision in orientation judgements, leading to lower overall error rate; (4) effects should be maximal when the orientation of the target is near that of the average of the flankers, rather than to that of individual flankers. All these effects were verified, and well simulated with ideal-observer models that maximize performance. The results suggest that while crowding can impact strongly on object recognition, it is best understood not as a processing bottleneck, but as a consequence of efficient exploitation of the spatial redundancies of the natural world.

INTRODUCTION

Crowding is the inability to recognize and identify objects in clutter, despite their being clearly visible, and recognizable when presented in isolation[1] (see examples in Figure 1A). It is particularly elevated in the periphery, scaling linearly with eccentricity, such that the minimal spacing between targets and flanking elements for uncrowded vision is equal to about half the eccentricity (Bouma's law[2]). Crowding impacts on many important daily tasks, such as face recognition and reading (for reviews see[3,4,5]), to the extent it has been considered a major bottleneck to object recognition.

Most popular current models of crowding involve some form of compulsory pooling (or substitution) of targets with flankers. For example, Parkes and colleagues[6] showed that while the orientation of a single line cannot be determined when embedded in flankers, it does influence the perceived orientation of the ensemble: hence it is merged with the flankers, rather than suppressed. This is reinforced by several studies showing that the targets can take on characteristics of the flanker stimuli[7-9]. Pelli and Tillmann[3] suggest that the compulsory integration occurs in higher cortical areas, such as V4, which have large receptive fields, appropriately sized to account for Bouma's law (see also [10]).

However, compulsory integration does not explain all the known facts about crowding. For example, flankers that are similar in size, colour or orientation cause more crowding than dissimilar ones[11-13]. More difficult to explain are the recent demonstrations of Herzog and colleagues[14] of "uncrowding", where the addition of extra flanking stimuli around the flankers can reduce drastically their crowding effect, particularly if the extra flankers group with the original flankers to form coherent objects. These data do not fit easily with compulsory integration, even with appropriate linear filtering, which could in principle account for other effects, such as orientation or size selectivity.

Crowding has been studied for decades, and usually considered to be a defect in the system, "an essential bottleneck to object perception"[15]. Certainly, it impacts heavily on object recognition in tasks like or face recognition: but is it possible that it may reflect processes that are in principle advantageous to perception? Perception is strongly affected by contextual information, particularly temporal context, where recent and longer term perceptual history has been shown to exert a major influence on current perception[16-19]. While the role of context and experience has been appreciated for some time[20,21], it has

become particularly topical in recent years within the framework of Bayesian analysis. This approach has revealed an interesting phenomenon termed “serial dependence”, where the appearance of many important attributes of a stimulus (including orientation, numerosity, facial identity, beauty etc) are biased towards previously viewed stimuli[17,18,22,23]. Counterintuitively, these consistent biases in perception have been shown to reflect an efficient perceptual strategy, exploiting temporal redundancies in natural viewing to reduce overall reproduction errors, despite the biases[24,17,25].

Could crowding also be a consequence of efficient integration processes that exploit spatial (rather than tempora)l redundancies to improve performance? We investigate this possibility by studying crowding with a similar paradigm used for serial dependence studies. If, like serial dependence, crowding is a by-product of efficient redundancy-reducing mechanisms, it should display several specific signature characteristics. One is that crowding-induced biases should be stronger for targets that are unreliably perceived, and for flankers that are reliably perceived. In addition, crowding should follow the signature pattern seen in serial dependence, highest when the orientations of target and flankers are similar, then steadily falling off. We verify these characteristics qualitatively and qualitatively, and show that crowding, while leading to biases, also improves overall performance. The results fit well with models simulating intelligent combination of signals from a small receptive field centred on the target with signals from a much larger integration region, following the same rules that govern serial dependence. On this view crowding should not be considered a defect, or bottleneck, in the system, but the unavoidable consequence of efficient exploitation of spatial redundancies of the natural world.

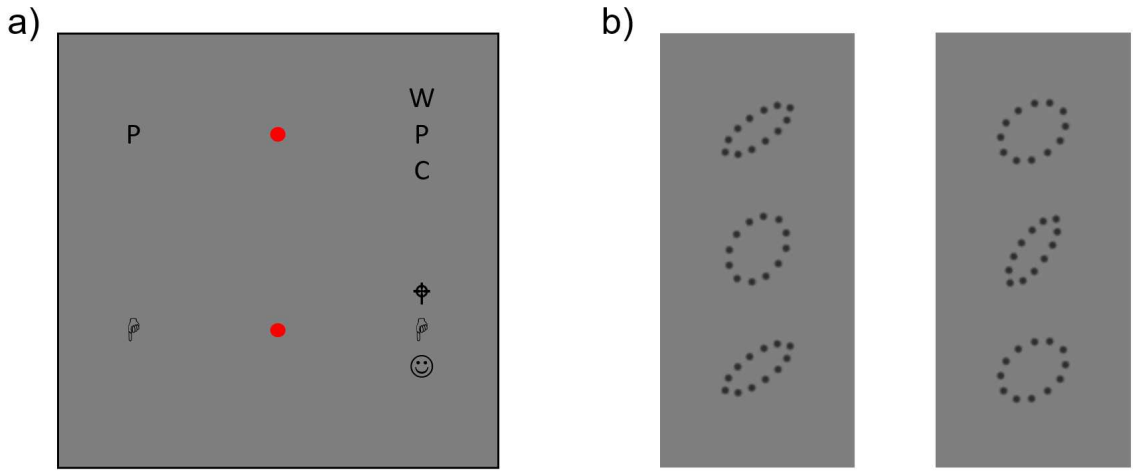

Figure 1

a) Crowding is a visual phenomenon where items that can be easily identified in isolation are not identifiable if surrounded by similar items. The P and hand symbol on the right are difficult to recognize, while fixating the central red dots.

b) Stimuli employed in this experiment. Observers judged the orientation of a peripheral target (the central oval), which was flanked above and below by oval stimuli. Two conditions were tested: a rounded target with elongated flankers (Low reliability target, high reliability flankers, at left) or an elongated target with rounded ovals (at right). In the main condition the centre-to-centre distance of flankers and targets was 5.5 deg, and eccentricity 26 degrees, leading to a Bouma ratio of 0.21.

RESULTS

To test if visual crowding follows the rules of optimal integration, which well describe serial dependence[18,25], we measured crowding with an orientation reproduction task. Participants reproduced the orientation of oval stimuli, which were either elongated (aspect ratio 1: 2.8) or rounded (aspect 1: 1.4). Targets were presented 26° to the right of fixation, and vertically flanked by similar oval stimuli, elongated if the target was rounded, and vice versa (see Fig. 1B). The orientation of the target was either 35° or 55° (at random). The orientations of the two flankers were yoked together, and varied randomly over a range of ±45° from target orientation. The clear prediction from the efficient integration model[24] (see Eqn 10) is that the effects of crowding will be far stronger for the rounded targets and elongated flankers than vice versa. The reasons are explained formally the modelling section, but the intuition is that the rounded stimuli have less reliable orientation signals and therefore benefit more from integration with contextual information, especially if it is reliable.

Figure 2A shows the bias in target reproduction as a function of difference in flanker orientation. Clearly, the rounded targets show the strongest contextual effects of crowding,

with peak biases varying by up to ±5.1°, compared with ±1.9° for the elongated targets. Furthermore, the pattern of bias follows closely that predicted and observed in serial dependence studies[25], varying non-linearly with difference between target and flanker orientation, increasing to a maximum around ±20°, then decreasing. These data are well fit by derivative of gaussian functions (eqn. 15, light-coloured lines), commonly used in serial dependence studies[18], and expected from a causal inference model (see modelling section[26]). The dark lines show the predictions of another Bayesian model (eqn. 10), which has also proven successful with serial dependence data[17,25]. While the models are detailed later, it is worth noting that they are almost entirely anchored by data, down to a simple scaling factor, suggesting that the data are consistent with ideal behaviour.

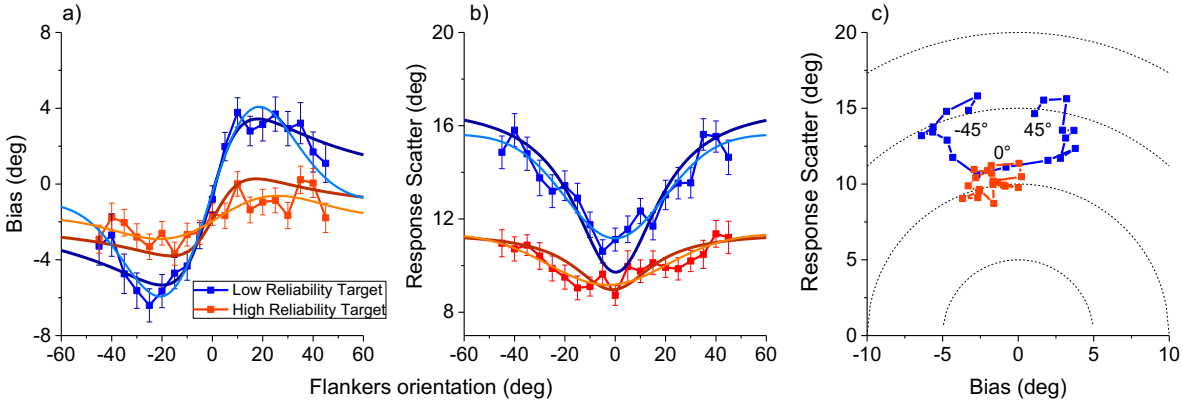

Figure 2

a) Average response bias (response minus target orientation) as a function of the orientation of two identical flankers. Low reliability (rounded) targets in blue, high reliability (elongated) in red. Error bars show ±1 SEM. Dark lines show predictions from an ideal-observer Bayesian model which scales the action of flankers according to their reliability and orientation difference (Eqn 10 of model section). Light blue and red curves show predictions for the causal inference model that doses flanker and target information according to their reliability and the probability of originating from a common cause (Eqn 15 of model section).

b) Response scatter as a function of the orientation of two identical flankers, together with model predictions. Colour coding as in A.

c) Response Scatter error plotted against bias errors for the two conditions. Dashed circles indicate regions with identical RMSE error (given by the Pythagorean sum of the two types of error). RMSE varies with orientation, and is least around 0°, when target and flankers coincide.

Another important prediction is that the contextual effects should improve performance. Figure 2B plots reproduction scatter (root-variance of reproduction trials) as a function of orientation difference. As expected, at all orientation differences, these are lower for the elongated than the rounded targets. However, for both targets, particularly the rounded

targets, the scatter decreased as the difference between target and flanker orientation decreased. Figure 2C plots scatter against bias, with points connected to follow the change in orientation. On this plot, total error (the Pythagorean sum of scatter and bias) is the radial distance from the origin. For the points with flanker orientation most distant from the target (near ±45°), the total error is around 15°. Between these extremes, total error falls off, despite the constant bias. When the flankers and targets have similar orientations, the error falls to around 11°, evidence that "crowding" improves overall performance.

If the effects shown in Figure 2 represent visual crowding, they should depend on critical spacing between target and flankers, and follow Bouma's law[1]. We therefore measured the effects as a function of target-flanker spacing, for 5 participants. Figure 3 shows the data for the rounded targets with elongated flankers (which show the strongest effects). For the two smallest spacings (5.5 and 7.5 deg), bias showed the characteristic S-shaped dependency on the orientation of the flankers. For the larger spacings (11.0 and 16.6 deg), however, the effect was much reduced and even inverted at 11 deg. As before, the curves are fit by a derivative of gaussian function (eqn 18), which is the product of a linear regression (illustrated by dashed line in Figure 3A) and a gaussian. The best fitting slope of this regression is an estimate of the weight given to the flankers when judging orientation. Figure 3B plots the fitted weight as a function of target-flanker spacing (lower abscissa), with the upper abscissa showing the Bouma constant, the distance between target and flanker centres divided by the eccentricity (26 deg). The weight drops from 0.5 to 0 for Bouma constants between 0.3 and 0.4, broadly in line with the literature, suggesting that the effects observed here relate to crowding.

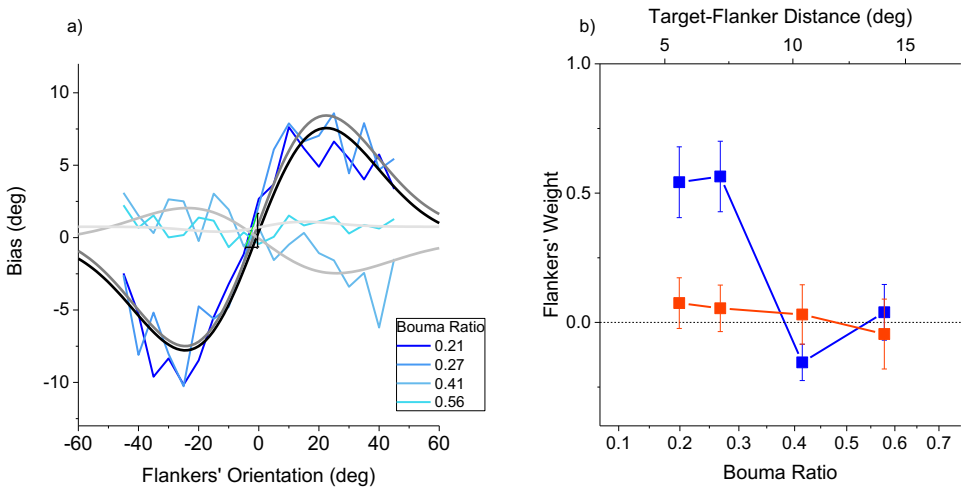

Figure 3

a) Response bias as function of flanker orientation for various target-flanker distances leading to four different Bouma ratios (distance between flanker and target centres divided by eccentricity. Data are fit with a derivative of gaussian function with free parameters (Eqn 18).

b) Weight of the flankers (maximal slope of the curves in panel A) as a function of the Bouma ratio (colour-code as before). Error bars show ±1 SEM.

The results so far show that integration is not obligatory, but depends on the reliability of both target and flankers, and on their orientation similarity. A remaining question is how the flankers integrate with the target: each separately, or after combination with each other. Figure 4 illustrates two possibilities (see also modelling section). One is feedforward model where the target integrates independently with low-level, high-resolution neural representations of each of the flankers. The other depicts integration with a broader representation including both flankers, potentially implemented through recurrent feedback.

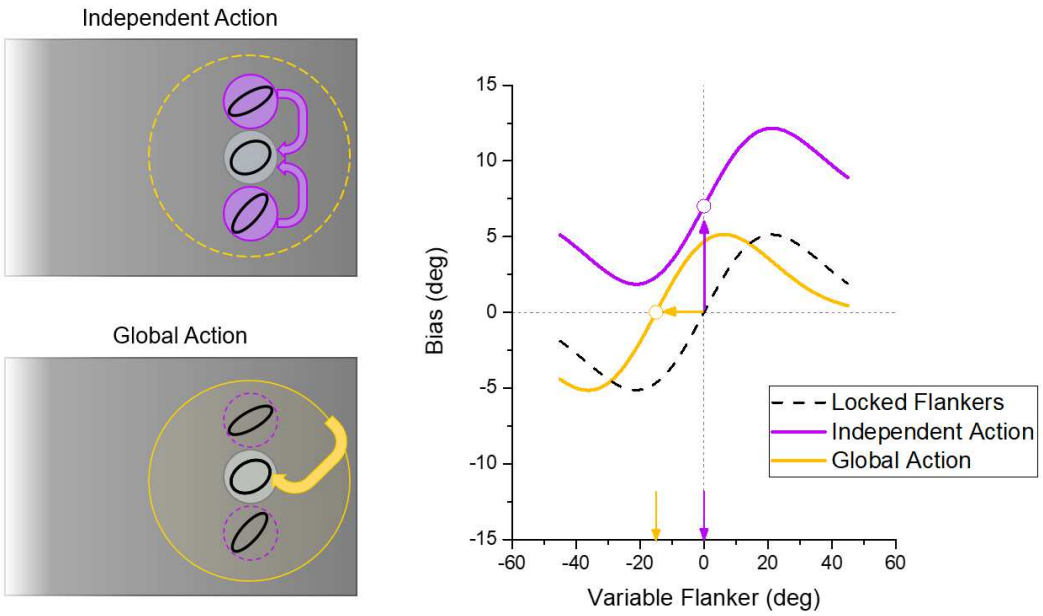

Figure 4

a) Rationale of Experiment 2. Flankers could either act independently on the target (as illustrated by purple arrows in top left panel), or first pooled into a larger RF, which in turn biases the target (illustrated by the large yellow circle and arrow in bottom left panel).

b) Predictions for the two hypotheses. If the flankers act independently, when one flanker is locked at +15° and the other free to vary, the pattern should be similar to that of the main experiment (centre close to 0°), but raised because of the action of the locked flanker (purple curve). If flankers are first integrated at a more global stage, maximal effect is expected when all the elements in the larger operator average 0°. Since one of the flankers is locked at +15°, this occurs when the other flanker is −15°, leading to a leftward shift of the curve of the main experiment (yellow curve)

To distinguish between these two plausible possibilities, we measured target bias with the orientation of the two flankers varying independently. Specifically, one flanker (randomly top or bottom) was always oriented +15° from the target, while the other varied randomly over the range. The logic is that the gaussian function windowing the contextual effect should be centred where the orientations of target and context coincide. If the integration occurs directly between the target and individual flankers, then the maximum effects should occur when the variable flanker coincides with the target; on the other hand, if the integration is with a broader representation including both flankers, maximum integration should occur when the flanker mean is zero, which occurs when the variable flanker is −15°. These predictions are illustrated in Figure 4B: note that the individual flanker effect also predicts the curve to be higher at all flanker orientations, as the fixed flanker will exert a constant effect at all orientations of the variable flanker.

The results for the rounded targets with elongated flankers are shown figure 5A. The biases clearly follow the signature pattern, well fit by a derivative of gaussian function. The centre of the function is −12.1°, closer to the 15° predicted by integration with the average orientation of the flankers, than 0° predicted by the individual flanker model. The mean height of the function is 0.5°, close to that observed in the previous experiment (−0.9°), while the individual-flanker integration model predicts a constant average bias 4.7°. Figure 5B shows the scatter for this experiment, which was reduced over the region of bias, well described by an inverted Gaussian.

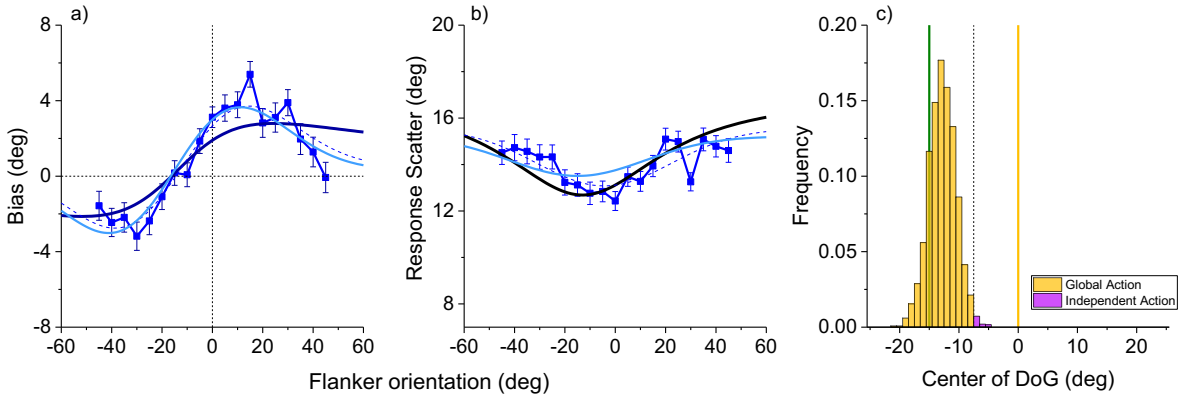

Figure 5.

a) Biasing errors as function of a single flanker orientation, while the other flanker was locked at +15°. Colours and conventions as for Figure 2. Thick dark lines refer to the ideal observer model (Eqn 10), thick light blue lines to the causal inference model (Eqn 15). Thin dashed lines show best-fitting derivative of gaussian, with all parameters free to vary (Eqn 18).

b) Response Scatter as a function of the variable flanker orientation. Conventions as in panel a.

c) Histogram of the centres of the gaussian derivative for 1000 bootstrap fits.

To test significance, we bootstrapped the data 1000 times (sampling with replacement) and measured the centre of the gaussian derivative on each iteration. The results plotted in the histogram of Figure 5C show that on only 16 out of 1000 iterations (1.6%) was the centre closer to 0° (individual flanker prediction) than to −15° (joint-flanker prediction). This leads to a likelihood ratio (Bayes factor) of 984/16 = 61.5, strong evidence in favour of the joint-flanker-integration model.

MODELLING

We propose two plausible models to explain the pattern of data. Both are motivated by principles of "optimal cue integration" commonly used in multi-sensory perception[27,28], which predict optimal combination of information from multiple sources after appropriate weighting to minimize overall root-mean-square error. The first is based on an ideal-observer model successfully used to model serial dependence[17], the second on a "causal-inference" model of multi-sensory integration[26]. Both models predict well the data.

Ideal Observer

Total RMS error (*E*) can be decomposed into bias (*B*) and precision (scatter standard deviation: *S*), whose squares sum to give total squared error:

$$E = \sqrt{B^2 + S^2} \qquad \text{[eq. 1]}$$

The ideal responses (*R*) in a pooling model can be expressed as a linear weighted combination of internal representation of target (*T*) and flankers (F_1 and F_2), each weighted by $w_1$and w_2.

$$R = w_1F_1 + w_2F_2 + (1 - w_1 - w_2)T \qquad \text{[eq. 2]}$$

As the two flankers of this study had the same aspect ratio they should be weighted equally, ($w_1 = w_2 = w$), so Eqn. 2 simplifies to:

$$R = wF_1 + wF_2 + (1 - 2w)T \qquad \text{[eq. 3]}$$

The mean of the responses (μ_R) is a simple linear combination of the means of flankers and target (μ_1, μ_2 and μ_T).

$$\mu_R = w\mu_1 + w\mu_2 + (1 - 2w)\mu_T = w(\mu_1 + \mu_2) + (1 - 2w)\mu_T \qquad \text{[eq. 4]}$$

Bias is the difference between the mean estimated response (μ_R) and real orientation, x_T; $B = \mu_R - x_T$. Using equation 4 and considering that the average target representation (μ_T) should be unbiased and coincide with target ($\mu_T = x_T$) it follows that:

$$B = \mu_R - x_T = w(\mu_1 + \mu_2) + \mu_T - 2w\mu_T - x_T = w(\mu_1 + \mu_2 - 2\mu_T) \qquad \text{[eq. 5]}$$

The term $\mu_1 + \mu_2 - 2\mu_T$ can be rearranged as $2(\ (\mu_1 + \mu_2)/2 - \mu_T)$ which is twice the distance between the average of the flanker representations, $(\mu_1 + \mu_2)/2$, and the target representation μ_T. For convenience we define:

$$d = (\mu_1 + \mu_2)/2 - \mu_T \qquad \text{[eq. 6]}$$

so that Eqn. 5 becomes:

$$B = w(\mu_1 + \mu_2 - 2\mu_T) = 2wd \qquad \text{[eq. 7]}$$

Variance of the linear combination of the flankers and target is itself a linear combination of the flanker and target variances (σ_F^2 and σ_T^2) with the squared coefficients

$$S^2 = w^2\sigma_F^2 + w^2\sigma_F^2 + (1 - 2w)^2\sigma_T^2 \qquad \text{[eq. 8]}$$

From Eqn 1, 7 and 8 it follows that $RMSE$ can be written as:

$$E = 4w^2d^2 + w^2\sigma_F^2 + w^2\sigma_F^2 + (1 - 2w)^2\sigma_T^2$$

$$= 4w^2d^2 + w^2\sigma_F^2 + w^2\sigma_F^2 + (1 - 4w + 4w^2)\sigma_T^2 \qquad \text{[eq. 9]}$$

Since RMSE is a function of second order of w, it is minimized when $w = \frac{-b}{2a}$, so the optimal weight (w_{opt}) is obtained at:

$$w_{opt} = -\frac{1}{2}\frac{-4\sigma_T^2}{4\sigma_T^2+2\sigma_F^2+4d^2} = \frac{\sigma_T^2}{2\sigma_T^2+\sigma_F^2+2d^2} \qquad \text{[eq. 10]}$$

This equation has much in common with that of all Bayesian-like integrations used in multi-sensory research and serial dependence: the weight depends directly on *target* variance σ_T^2, so *targets* of low reliability (inverse variance) benefit more from integration, resulting in higher weighting to the *flankers*. Increase in flanker variance (σ_F^2) has the opposite effect.

The term $2d^2$ is fundamental for the signature function, as the weight of the *flankers* will decrease with angular difference between target and average flanker orientation. This is reminiscent of serial dependence effects, and ensures that contextual cues are used only if they are plausibly similar to the target[24,17,25]. Importantly, the point that will ensure maximal weight of the flankers is when the target coincides with the average of the flankers (i.e. $d^2 = 0$).

Together the behaviour of Eqns 3 and 10 define the ideal observer behaviour. In order to accommodate suboptimal behaviour we introduce a scaling factor (α) which multiplies w_{opt} and sets the actual weight of the flankers:

$$R = \alpha w_{opt} F_1 + \alpha w_{opt} F_2 + (1 - 2\alpha w_{opt})T \qquad \text{[eq. 11]}$$

Causal Inference Model

An alternative model prescribes that the optimal blend of information can be obtained behaving as if the sources of information originated form one cause times the probability that two sources of information originate from the same cause[26]. Within this framework the maximal interaction between cues occurs when the two sources coincide, where the weight assigned of is the well known result known in sensory integration literature[27,28] (see also eq. 10):

$$w_A^{max} = \frac{\sigma_B^2}{\sigma_A^2 + \sigma_B^2} \qquad \text{[eq. 12]}$$

The probability of the two sources originating from a common cause can be calculated using Bayes' Theorem as demonstrated in[26]. Assuming gaussian probability distribution functions (with centres at μ_A and μ_B and variances σ_A^2 and σ_B^2), the solution is soluable analytically[26]:

$$p(A, B|C = 1) \propto \exp\left(-\frac{1}{2}\frac{(\mu_A - \mu_B)^2\sigma_P^2 + (\mu_A - \mu_P)^2\sigma_B^2 + (\mu_B - \mu_P)^2\sigma_A^2}{\sigma_A^2\sigma_B^2 + \sigma_A^2\sigma_P^2 + \sigma_B^2\sigma_P^2}\right) \qquad \text{[eq. 13]}$$

This is function of the variances of the two sources (σ_A^2 and σ_B^2), the centres of the representations (μ_A and μ_B) and their distance, and the a-prior likelihood of there being one cause (itself gaussian and characterized by mean and variance μ_P and σ_P^2). If no prior knowledge is available ($\sigma_P^2 \to \infty$) Eqn 13 simplifies to

$$p(A, B|C = 1) \propto \exp\left(-\frac{1}{2}\frac{(\mu_A - \mu_B)^2}{\sigma_A^2 + \sigma_B^2}\right) \qquad \text{[eq. 14]}$$

This is a gaussian peaking when the distribution of the two cues coincide ($\mu_A = \mu_B$) and falling off with a space constant related to the sum of their variances ($\sigma_A^2 + \sigma_B^2$).

In the specific case of our experiment we can map the two sources of information to the flanker compound (a gaussian with centre at $\mu_F = (\mu_1 + \mu_2)/2$, variance $\sigma_F^2/2$) and the

target (assumed gaussian with centre μ_T, and variance σ_T^2. Putting together Eqns 12 and 14, the bias (difference between the response and the target) is given by:

$$B = w_F^{max}\, p(F,T|C=1)(\mu_F - \mu_T) = \frac{\sigma_T^2}{\sigma_F^2+\sigma_T^2} \exp\left(-\frac{1}{2}\frac{(\mu_F-\mu_T)^2}{\sigma_F^2+\sigma_T^2}\right)(\mu_F - \mu_T) \qquad \text{[eq 15]}$$

Which is a derivative of gaussian as a function of flanker orientation μ_F

It also follows that response scatter is minimized only when the system considers a common cause likely (Eqn 14), predicting U-shaped (gaussian) plots of Figures 2B and 5B.

Again, to allow for suboptimal behaviour we introduced two free parameters that regulate the amplitude of the dependency on the flankers (β) and the breadth of the region of interaction (γ) so that the bias is:

$$B = \beta\, \frac{\sigma_T^2}{\sigma_F^2+\sigma_T^2} \exp\left(-\frac{1}{2}\frac{(\mu_F-\mu_T)^2}{\gamma^2(\sigma_F^2+\sigma_T^2)}\right)(\mu_F - \mu_T) \qquad \text{[eq 16]}$$

Interestingly, comparable behaviour is obtained if, instead of constructing a system which multiplies probabilities as in[26], one considers a system that measures the similarity between two distributions via their point-by-point product of the distributions and takes either the peak or area under the distribution.

The product of gaussians is itself a gaussian, is centred at $\left(\frac{\mu_B\,\sigma_A^2+\mu_A\,\sigma_B^2}{\sigma_A^2+\sigma_B^2}\right)$, has variance $\left(\frac{\sigma_A^2\sigma_B^2}{\sigma_A^2+\sigma_B^2}\right)$ and peak at:

$$\frac{1}{2\pi\sigma_A\sigma_B} \exp\left(-\frac{(\mu_A-\mu_B)^2}{2(\sigma_A^2+\sigma_B^2)}\right) \qquad \text{[eq. 17]}$$

So the peak embeds the same behaviour of Eqn. 14. It is easy to demonstrate that also the area under the curve follows the same gaussian dependency on the distance between cues as the area of a gaussian is equivalent to the peak (Eqn. 16) times the standard deviation of the curve $\left(\sqrt{\frac{\sigma_A^2\sigma_B^2}{\sigma_A^2+\sigma_B^2}}\right)$ and a constant factor $1/\sqrt{2\pi}$ all of which are constant once the distributions have known width and thus reduce to a scaling factor.

Model Fitting

The predictions of the two modelling approaches are overlayed on the data of Figures 2 and 5 with dark and light colours. To minimize degrees of freedom we derived the values of

sensory reliability from the data of Figure 2b, assuming that the extreme points (±30° and ±45°) give baseline data, not influenced by flanker integration: this is 17.1 for rounded targets (blue symbols), and 11.7 for elongated targets (red symbols).

We implemented the ideal observer model (Eqn. 11) with only a scaling constant (α), which allows for sub-optimal behaviour. These fits are particularly good for the rounded targets (with largest effects), with R^2 of 0.97 and 0.74 (for bias and scatter), and 0.24 and 0.60 for elongated targets) and come about assuming α = 0.57. One of the key features of the ideal observer model is that it reduces RMSE by leveraging on all available information. Thus it predicts the Global Integration of Figure 4, with centres of the Gaussian derivatives close to −15°. Besides capturing this key feature, the model also provides good quantitative fits to the data of Figure 5a with R^2 of 0.76 and average fits to those of Figure 5b 0.23 for bias and scatter respectively (α=0.32).

We used the same reliability values from Figure 2b to implement the "optimal causality gating model"[26], the derivative of gaussian function plotted with light colours in Figures 2 and 5. The sensory reliabilities fix both the maximal slope of the curve (see Eqn 12) and the width of the region of interaction (see Eqn 14). Assuming the same sensory precisions as above (17.1 and 11.7 for the two types of stimuli) maximal slopes should be 0.81 and 0.48 for the two conditions, larger than the real data. Also the widths (27.8 and 33.2) are larger than those predicted by Eqn 14 (19 and 16.8). For this reason we allowed two scaling factors, one enabling lower weighting of the context (β) and the other modulating the width (γ). Setting β=0.54 and γ=1.46 led to good fits with R^2 = 0.97 and 0.89 for the low reliability target (bias and scatter curves), and 0.67 and 0.79 for the high reliability target (β=0.26 and γ=1.97). As with the other model, the prediction in Experiment 2 is for large pooling of all available cues, thus the prediction is that of a centre at −15°. This model also provides good fits for response bias (R^2=0.89) and acceptable fits for response scatter (R^2=0.38, β=0.62 and γ=1.37).

DISCUSSION

The results of this study suggest a novel interpretation of visual crowding: that it is a by-product of efficient Bayesian processes, which lead to improved perceptual performance, minimizing production error. We tested and validated several key predictions of this idea. Firstly, crowding, measured as flanker-induced orientation bias, was greatest when targets had the weakest orientation signals (least reliability) and flankers had the strongest signals, as predicted from most models of optimal cue combination[27,28]. The magnitude of the bias varied with the difference of target and flanker orientation, following the predicted non-linear pattern, increasing to a maximum of around 15°, then falling off for larger orientation differences. Importantly, the interaction of the flankers and target was associated with a reduction in response scatter, which led to a reduction in total RMS error, an index of improved performance. Finally, the results suggest that the bias does not result from direct interactions with individual flankers, but from interaction with a representation of the average orientations of the two flankers. All these results were predicted by optimal feature combination principles, and quantitatively well modelled an ideal-observer model that minimizes reproduction errors.

These results are clearly difficult to reconcile with standard models of obligatory integration[6,29]. Passive integration systems may be tweaked to explain the stronger effects for more elongated flankers (such as having more Fourier energy at that orientation), but cannot explain the fall off in crowding effects when the difference exceeds 15°. Any basic integrator would necessarily combine orientation energy of all angles, not only similar angles. On the other hand, the flexible integrator models proposed here (Eqns 10 and 15) predict both the pattern and the magnitude of the results. Furthermore, the final experiment suggests that this intelligent orientation-dependent integration is unlikely to occur directly within a higher order cell itself, as the orientation-dependent integration function aligns with the average of two disparate flankers, rather than with each individual flanker. This suggests that the integration is between the target and a broad representation that includes both flankers. Mechanisms operating directly between target and individual flankers (such as the proposed "local association field"[30]) are not consistent with the results of Figure 5, which shows that flankers are first combined with each other before exerting their effects on the target.

Combination of target and a broad representation of both flankers could be implemented in several ways. One physiologically plausible mechanism would be feedback from mid-level areas, such as V4, which have large receptive fields, integrating over a wide area. These cells could contain information of both flankers (as well as the target), which could be fed back to low levels (eg V1) to integrate flexibly with finer representations of the target. Within this framework the fine-grain target information is not lost, but combined with broad contextual information in an optimal manner to improve performance. This is analogous to the process of serial dependence, where representations of perceptual history (often termed *Bayesian priors*) are generated at mid- to high-levels of analysis, but feed back onto fairly low processing levels[31]. Similar processes could evoke crowding, integrating over space rather than time.

The predictions of the crowding behaviour derive from theoretical minimization of total RMS errors, explained in detail in the modelling section, but readily understood intuitively. There are two orthogonal sources of error, bias (average accuracy) and response scatter (precision), which combine by Pythagorean sum to yield total error. Thus although the contextual effects do lead to inaccuracies (biases), these are more than offset by the decrease in response scatter (Fig. 2C). Clearly, if the effects were to increase continuously with orientation, then the bias would become large, and offset the reduction in scatter, leading to increased error: integration is therefore efficient only over a limited range. Note that the efficiency-driven ideal model gives good fits simultaneous to both bias and scatter data with only one free parameter, a scaling factor. This comes out at around 0.57 for the main data and probably reflects other processes in orientation judgements that we did not control for, such as regression to the mean[32,33].

The current experiment shows that under conditions of crowding, information about the target is not necessarily lost. This is consistent with a good deal of previous evidence (see reference[34] for review), including studies showing that it can affect the ensemble judgment[6], can cause adaptation[35] and that crowding induced biases may not affect grasping[36]. Even more dramatic are the demonstrations that increasing flanker length[37] or adding additional flankers[14] can decrease or eliminate crowding. Our study employed simple well controlled stimuli to allow quantitative prediction and measurement of crowding-effects, similar to the studies with serial dependence studies. Thus they do not readily relate

to the clever uncrowding studies of Herzog and colleagues. However, it is not difficult to envisage extensions to the model incorporating grouping principles within the rules of integration, in the spirit of the general principles of our model: flexible, "intelligent" combination of signals, rather than a rigid integration via "rectify and sum" or similar rules[10].

In summary, the current study suggests that crowding may be analogous to serial dependence, pointing to similar function and mechanisms. As serial dependence has been shown to exploit temporal redundancies to maximize performance, crowding may also reflect similar exploitation of redundancies over space. It is worth noting that while the rules governing crowding are flexible, leading to improved performance, crowding remains completely obligatory: no effort of will or deployment of attention can allow us to resolve the crowded objects, or to ignore the contextual effects of the orientated flankers. Indeed, while our proposed pooling process is flexible and "intelligent", it remains automatic, not subject to voluntary control. This is similar to many of the experience-driven perceptual illusions, such as the "hollow mask illusion"[21]: no effort of will can cause us to see the inside of a hollow mask as concave, we always see the convex face. However, while visual crowding remains an obligatory limitation to object recognition, we conclude that like the effects of temporal context and experience, it is best understood not as a defect or bottleneck of the system, but the consequence of efficient exploitation of spatial redundancies of the natural world.

METHODS

Participants

Fifteen healthy participants with normal or corrected-to-normal vision were recruited (aged 18-55 years, mean age = 36, 7 females). Experimental procedures are in line with the declaration of Helsinki and approved by the local ethics committee (*Commissione per l'Etica della Ricerca*, University of Florence, 7 July 2020). Written informed consent was obtained from each participant, which included consent to process, preserve and publish the data in anonymous form.

Stimuli

The stimuli, illustrated in Fig. 1A, were generated with Psychtoolbox for MATLAB (R2016b; MathWorks). They comprised an oval-shaped visual target flanked by oval-shaped upper and lower visual flankers, displayed 26 deg eccentric from the fixation point, with the target close to the horizontal meridian (vertical position was slightly varied from trial to trial to avoid pre-allocation of attention to the target) and flankers 5.5 deg away from the target. Both target and flankers were sketches of oval shapes, defined by 12 dark grey dots (diameter 0.3 deg, 1.4 deg inter-dots distant, 16.8 deg perimeter), presented against a uniform grey background. The target was orientated either +35° or +55° (clockwise) from the vertical, and flanker orientation randomly chosen in steps of 5° from −45° to +45° with respect to the target orientation. The two flankers were 5.5 deg from target, leading to a Bouma ratio of 0.2. We manipulated the reliability of orientation information of target and flanker stimuli by using two different aspect ratios, 2.8 (axes 3.48 and 1.23 deg) and 1.4 (axes 3.19 and 2.28 deg), illustrated in Fig. 1A. The more elongated target was always associated with more rounded flankers, and vice versa. In each experimental session of the three experiments, the two target-flanker combinations were shown both kinds of stimuli in random order.

Procedure

Stimuli were displayed on a linearized 22” LCD monitor (resolution 1920 x 1080 pixels, refresh rate 60 Hz). Observers were positioned 57 cm from the monitor, in a quiet room with dim lighting, and maintained fixation on a small (0.35 deg) black central dot. After a random delay from the observer initiating the trial, the stimulus was displayed for 167 ms. Then a thin rotatable white bar (0.05 x 5 deg with a gaussian profile) was presented at the fixation point with random orientation, and observers matched its orientation to that of the target by mouse control. In the first two experiments, the orientation of the two flankers was yoked, while in the third, one flanker was always +15° (clockwise) while the other varied from -45° to +45°. In the second experiment, the target-flanker distance varied, being 5.5, 7.5, 11.0 and 16.6 deg, leading to Bouma ratios of 0.21, 0.27, 0.4, 0.6.

Ten observers participated in the first experiment, five in the second, thirteen in the third. They contributed for a total of 10699 trials for the first experiment, 14377 for the second (spread across the four flanker-target distances) and 16574 for the last.

Data analysis

Responses occurred out from the range between 0.5 and 3 seconds after the stimulus offset were removed (for a total of 15.9% trials across the 3 experiments), as were responses with reproduction error greater than 35° (6.9% of trials).

For each target and relative orientation of the flanker, we calculated the average constant error (bias, positive meaning clockwise) and scatter. We then averaged the values for the two targets. Bias functions were fitted by a derivative of gaussian function, which can be considered to be a gaussian of width *s* multiplied by a straight line of slope a [or w], which can be considered the weighting given to the flankers: 1 means the flankers are weighted equally to the target. Bias is given by:

$$B = a \cdot (\theta - m) \exp\left(- \frac{(\theta - m)^2}{s^2} \right) + b \qquad \text{[eq. 18]}$$

Where θ is orientation difference, m the centre, and b the vertical offset of the function. a, b and m were free to vary.

Scatter (S) was defined as the average root variance in each condition. The variation with orientation a gaussian function in the form:

$$S = a \cdot \exp\left(- \frac{(\theta - m)^2}{s^2} \right) + b \qquad \text{[eq. 19]}$$

Where b is the baseline at high orientation differences and a is the amplitude of the Gaussian. As Bias and Scatter likely originate from the same process, we yoked the parameter s to best fit both curves.

REFERENCES

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

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

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

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

5 Whitney, D. & Levi, D. M. Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends Cogn Sci* **15**, 160-168, doi:10.1016/j.tics.2011.02.005 (2011).
6 Parkes, L., Lund, J., Angelucci, A., Solomon, J. A. & Morgan, M. Compulsory averaging of crowded orientation signals in human vision. *Nat Neurosci* **4**, 739-744, doi:10.1038/89532 (2001).
7 Balas, B., Nakano, L. & Rosenholtz, R. A summary-statistic representation in peripheral vision explains visual crowding. *J Vis* **9**, 13 11-18, doi:10.1167/9.12.13 (2009).
8 Levi, D. M. & Carney, T. Crowding in peripheral vision: why bigger is better. *Curr Biol* **19**, 1988-1993, doi:10.1016/j.cub.2009.09.056 (2009).
9 Greenwood, J. A., Bex, P. J. & Dakin, S. C. Crowding changes appearance. *Curr Biol* **20**, 496-501, doi:10.1016/j.cub.2010.01.023 (2010).
10 Freeman, J. & Simoncelli, E. P. Metamers of the ventral stream. *Nat Neurosci* **14**, 1195-1201, doi:10.1038/nn.2889 (2011).
11 Nazir, T. A. Effects of lateral masking and spatial precueing on gap-resolution in central and peripheral vision. *Vision Res* **32**, 771-777, doi:10.1016/0042-6989(92)90192-l (1992).
12 Kooi, F. L., Toet, A., Tripathy, S. P. & Levi, D. M. The effect of similarity and duration on spatial interaction in peripheral vision. *Spat Vis* **8**, 255-279, doi:10.1163/156856894x00350 (1994).
13 Kennedy, G. J. & Whitaker, D. The chromatic selectivity of visual crowding. *J Vis* **10**, 15, doi:10.1167/10.6.15 (2010).
14 Manassi, M., Sayim, B. & Herzog, M. H. When crowding of crowding leads to uncrowding. *J Vis* **13**, 10, doi:10.1167/13.13.10 (2013).
15 Levi, D. M. Crowding--an essential bottleneck for object recognition: a mini-review. *Vision Res* **48**, 635-654, doi:10.1016/j.visres.2007.12.009 (2008).
16 Chopin, A. & Mamassian, P. Predictive properties of visual adaptation. *Curr Biol* **22**, 622-626, doi:10.1016/j.cub.2012.02.021 (2012).
17 Cicchini, G. M., Anobile, G. & Burr, D. C. Compressive mapping of number to space reflects dynamic encoding mechanisms, not static logarithmic transform. *Proc Natl Acad Sci U S A* **111**, 7867-7872, doi:10.1073/pnas.1402785111 (2014).
18 Fischer, J. & Whitney, D. Serial dependence in visual perception. *Nat Neurosci* **17**, 738-743, doi:10.1038/nn.3689 (2014).
19 Pascucci, D. *et al.* Laws of concatenated perception: Vision goes for novelty, decisions for perseverance. *PLoS Biol* **17**, e3000144, doi:10.1371/journal.pbio.3000144 (2019).
20 Helmholtz, H. v. *Handbuch der physiologischen Optik*. (Voss, 1867).
21 Gregory, R. L. *Eye and brain; the psychology of seeing*. (McGraw-Hill, 1966).
22 Liberman, A., Fischer, J. & Whitney, D. Serial dependence in the perception of faces. *Curr Biol* **24**, 2569-2574, doi:10.1016/j.cub.2014.09.025 (2014).
23 Taubert, J., Van der Burg, E. & Alais, D. Love at second sight: Sequential dependence of facial attractiveness in an on-line dating paradigm. *Sci Rep* **6**, 22740, doi:10.1038/srep22740 (2016).
24 Burr, D. & Cicchini, G. M. Vision: efficient adaptive coding. *Curr Biol* **24**, R1096-1098, doi:10.1016/j.cub.2014.10.002 (2014).

25 Cicchini, G. M., Mikellidou, K. & Burr, D. C. The functional role of serial dependence. *Proc Biol Sci* **285**, doi:10.1098/rspb.2018.1722 (2018).
26 Kording, K. P. *et al.* Causal inference in multisensory perception. *PLoS One* **2**, e943, doi:10.1371/journal.pone.0000943 (2007).
27 Ernst, M. O. & Banks, M. S. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* **415**, 429-433, doi:10.1038/415429a (2002).
28 Alais, D. & Burr, D. The ventriloquist effect results from near-optimal bimodal integration. *Curr Biol* **14**, 257-262, doi:10.1016/j.cub.2004.01.029 (2004).
29 Rosenholtz, R., Yu, D. & Keshvari, S. Challenges to pooling models of crowding: Implications for visual mechanisms. *J Vis* **19**, 15, doi:10.1167/19.7.15 (2019).
30 Field, D. J., Hayes, A. & Hess, R. F. Contour integration by the human visual system: evidence for a local "association field". *Vision Res* **33**, 173-193, doi:10.1016/0042-6989(93)90156-q (1993).
31 Cicchini, G. M., Benedetto, A. & Burr, D. C. Perceptual history propagates down to early levels of sensory analysis. *Curr Biol* **31**, 1245-1250 e1242, doi:10.1016/j.cub.2020.12.004 (2021).
32 Jazayeri, M. & Shadlen, M. N. Temporal context calibrates interval timing. *Nat Neurosci* **13**, 1020-1026, doi:10.1038/nn.2590 (2010).
33 Cicchini, G. M., Arrighi, R., Cecchetti, L., Giusti, M. & Burr, D. C. Optimal encoding of interval timing in expert percussionists. *J Neurosci* **32**, 1056-1060, doi:10.1523/JNEUROSCI.3411-11.2012 (2012).
34 Manassi, M. & Whitney, D. Multi-level Crowding and the Paradox of Object Recognition in Clutter. *Curr Biol* **28**, R127-R133, doi:10.1016/j.cub.2017.12.051 (2018).
35 He, S., Cavanagh, P. & Intriligator, J. Attentional resolution and the locus of visual awareness. *Nature* **383**, 334-337, doi:10.1038/383334a0 (1996).
36 Bulakowski, P. F., Post, R. B. & Whitney, D. Visuomotor crowding: the resolution of grasping in cluttered scenes. *Front Behav Neurosci* **3**, 49, doi:10.3389/neuro.08.049.2009 (2009).
37 Banks, W. P., Larson, D. W. & Prinzmetal, W. Asymmetry of visual interference. *Percept Psychophys* **25**, 447-456, doi:10.3758/bf03213822 (1979).