



# Unraveling brain interactions in vision: The example of crowding

Maya A. Jastrzębowska<sup>a,b,\*</sup>, Vitaly Chicherov<sup>a</sup>, Bogdan Draganski<sup>b,c</sup>, Michael H. Herzog<sup>a</sup>

<sup>a</sup> Laboratory of Psychophysics, Brain Mind Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

<sup>b</sup> Laboratory for Research in Neuroimaging, Department of Clinical Neuroscience, Lausanne University Hospital and University of Lausanne, Lausanne, Switzerland

<sup>c</sup> Neurology Department, Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

## ARTICLE INFO

### Keywords:

Spatial contextual effects  
Crowding  
Brain imaging  
Effective connectivity  
Visual cortex

## ABSTRACT

Crowding, the impairment of target discrimination in clutter, is the standard situation in vision. Traditionally, crowding is explained with (feedforward) models, in which only neighboring elements interact, leading to a “bottleneck” at the earliest stages of vision. It is with this implicit prior that most functional magnetic resonance imaging (fMRI) studies approach the identification of the “neural locus” of crowding, searching for the earliest visual area in which the blood-oxygenation-level-dependent (BOLD) signal is suppressed under crowded conditions. Using this classic approach, we replicated previous findings of crowding-related BOLD suppression starting in V2 and increasing up the visual hierarchy. Surprisingly, under conditions of *uncrowding*, in which adding flankers improves performance, the BOLD signal was further suppressed. This suggests an important role for top-down connections, which is in line with global models of crowding. To discriminate between various possible models, we used dynamic causal modeling (DCM). We show that recurrent interactions between all visual areas, including higher-level areas like V4 and the lateral occipital complex (LOC), are crucial in crowding and uncrowding. Our results explain the discrepancies in previous findings: in a recurrent visual hierarchy, the crowding effect can theoretically be detected at any stage. Beyond crowding, we demonstrate the need for models like DCM to understand the complex recurrent processing which most likely underlies human perception in general.

## 1. Introduction

Crowding, the degradation of object discrimination due to the presence of flanking objects, has traditionally been explained by pooling models, where features of the target and flankers are averaged or otherwise combined (Pelli et al., 2004; Pelli and Tillman, 2008; Greenwood et al., 2009). Pooling is often thought to occur at the earliest stages of vision, i.e., in primary visual cortices, and the loss of feature discriminability is considered to be irrecoverable (e.g., Strasburger et al., 1991; Strasburger, 2005; Levi and Carney, 2009; Pelli and Tillman, 2008; Balas et al., 2009; Rosenholtz et al., 2019). While neuroimaging studies of crowding have yielded equivocal results regarding the neural locus of crowding, they have all pointed to an early locus, i.e., either V1 (Anderson et al., 2012; Millin et al., 2014) or V2

(Fang and He, 2008; Bi et al., 2009; He et al., 2019), which is consistent with the view of crowding as a “bottleneck” at the earliest stages of vision (Levi, 2008; Whitney and Levi, 2011). However, most models of crowding cannot explain *uncrowding* since they are insensitive to global aspects of the stimulus, which is more determinant of crowding than any other aspect (Banks et al., 1979; Malania et al., 2007; Manassi et al., 2012, 2013, 2015; Herzog and Manassi, 2015a; Herzog et al., 2015b, 2016a, 2016b). For example, performance of a Vernier discrimination task deteriorates when the Vernier is surrounded by a square, but this deterioration can be undone by adding further squares, a fact which cannot be explained by most models of crowding in which more flankers can only deteriorate performance (Manassi et al., 2013). Moreover, the exact configuration of the flankers matters (Manassi et al., 2016), which suggests that perceptual grouping plays an important role in crowding

**Abbreviations:** AIC, Akaike's information criterion; ANOVA, analysis of variance; BMA, Bayesian model averaging; BF, Bayes factor; BOLD, blood-oxygenation-level-dependent; CW/CCW, clockwise/counterclockwise; DCM, dynamic causal modeling; DDM, drift diffusion model; DIC, deviance information criterion; EEG, electroencephalography; EPI, echo-planar imaging; fMRI, functional magnetic resonance imaging; GLM, general linear model; HDDM, hierarchical drift diffusion model; LOC, lateral occipital complex; PEB, parametric empirical Bayes; Pp, posterior probability; pRF, population receptive field; PSC, percent BOLD signal change; RF, radio-frequency; ROI, region of interest; SEM, standard error of the mean; SOA, stimulus onset asynchrony; SPM, statistical parametric mapping; T1w, T1-weighted MRI; TE, echo time; TR, repetition time; VOI, volume of interest.

\* Corresponding author at: Laboratory of Psychophysics, Brain Mind Institute, School of Life Sciences, École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland.

<sup>1</sup> Present address: Department of Education and Psychology, Freie Universität Berlin, Habelschwerdter Allee 45, 14195 Berlin, Germany.

E-mail address: [maya.jastrzebowska@gmail.com](mailto:maya.jastrzebowska@gmail.com) (M.A. Jastrzębowska).

<https://doi.org/10.1016/j.neuroimage.2021.118390>

Received 26 January 2021; Received in revised form 9 July 2021; Accepted 12 July 2021

Available online 13 July 2021.

1053-8119/© 2021 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

(Manassi et al., 2012; Hermens and Bell, 2014; Herzog et al., 2015b; Manassi et al., 2015). Similar results have been obtained with Gabors (Livne and Sagi, 2007; Levi and Carney, 2009; Saarela et al., 2009; Livne and Sagi, 2010; Chakravarthi and Pelli, 2011; Yeotikar et al., 2011; Joo et al., 2012; Robol et al., 2012) and other visual stimuli (Wolford and Chambers, 1983; Banks and White, 1984; Pöder, 2006; Louie et al., 2007), as well as in other perceptual modalities, including haptics (Overvliet and Sayim, 2016) and audition (Oberfeld et al., 2014).

Uncrowding demonstrates the inadequacy of local and feedforward models. However, rejecting purely feedforward models puts the problem in the much more complex universe of feedback models (Clarke et al., 2014; e.g., Wilson and Cowan, 1973; Zhaoping, 2003; Cao and Grossberg, 2005; Chaney et al., 2014; Francis et al., 2017; Doerig et al., 2020). These models range from lateral interaction models within one area, e.g., Zhaoping's V1 recurrent model (Zhaoping, 2003), through more complex models such as the LAMINART, a two-stage model which includes a basic grouping mechanism but no explicit object recognition (Francis et al., 2017), to even more complex models, e.g., Capsule Neural Networks, which require full object representations and recurrent connections from object-related areas to lower visual areas (Doerig et al., 2020). While some models perform better than others in their ability to replicate behavioral results of (un)crowding (Doerig et al., 2019), we cannot definitively discriminate between them using only behavior. This is where the need for neuroimaging comes in.

The methodology adopted by the majority of imaging studies of crowding consists in estimating the percent blood-oxygenation-level-dependent (BOLD) signal change (PSC) within different visual areas and identifying the regions in which the PSC is significantly suppressed in crowded conditions as compared to conditions without crowding. Recent advances in the modeling of the interactions between brain regions (Granger, 1980; Geweke, 1982; McIntosh and Gonzalez-Lima, 1994; Friston et al., 2003; Ding et al., 2006) now enable a comprehensive approach, which goes beyond the localization of within-region BOLD level attenuation. For example, dynamic causal modeling (DCM) allows us to unearth the complex interplay between brain areas underlying the functional magnetic resonance imaging (fMRI) signal observed during the performance of an experimental task, i.e., the effective connectivity. Here, we used DCM to probe the global vs. local and the feedforward vs. feedback nature of visual processing in (un)crowding. Besides identifying the specific architecture of connectivity in crowding, DCM also made it possible for us to examine the excitatory versus inhibitory nature of the connectivity.

## 2. Methods

### 2.1. Participants

We recruited 12 participants (six females, 11 right-handed, age range: 18 to 30 years) to take part in the experiment. All participants had normal or corrected-to-normal vision as assessed with the Freiburg Visual Acuity test, i.e., acuity values above 1.0 (Bach, 1996). Participants gave written informed consent and were informed that they could discontinue the experiment at any time. All experimental procedures complied with the Declaration of Helsinki except for pre-registration (§35) and were approved by the local ethics committee.

### 2.2. Stimuli and apparatus

The experiment was conducted at the MRI-scanning facilities of the Centre Hospitalier Universitaire Vaudois in Lausanne. Participants lay in the MRI scanner and looked at a screen, placed inside the 60 cm scanner bore, through a mirror (viewing distance: 70 cm). A Sony VPL-FH31 projector (size of projected image: 53 × 30 cm, chosen pixel resolution:

1920 × 1080 pixels, refresh rate: 60 Hz) was used to back-project images on the screen. The mean background luminance of the screen was 50 cd/m<sup>2</sup>, achieved through an adjustable neutral density filter (reduction approximately 10:1) made of a pair of polarization filters that were put on the far end of the waveguide. Participants held a button box and used the index and middle fingers of their right hands to push one of two buttons to respond to the stimuli. Stimulus generation and response collection were done in MATLAB 9.1 (The MathWorks, Inc., Natick, MA) using the Psychtoolbox (ver. 3.0; Brainard, 1997).

Subjects were asked to look at a black central fixation dot at all times. Eight circular sine-wave target gratings (spatial frequency: 2.5 cpd, diameter: 1.4 deg, contrast: 50%; mean luminance: 50 cd/m<sup>2</sup>) were presented at an eccentricity of 4° from the dot (radian angle locations from 22.5° to 337.5° with regular intervals of 45°). The gratings were identical to each other and were rotated either counterclockwise (CCW) or clockwise (CW) from the vertical. We used eight identical target gratings in order to elicit a spatially extended fMRI response. Participants were instructed to push the first button (index finger) for CCW gratings and the second button (middle finger) for CW gratings. We conducted a practice session before the experiment proper, during which the tilt magnitude of the target gratings was adjusted for each observer individually so as to achieve intermediate task difficulty. This observer-specific tilt was used in the fMRI experiment. The spatial phase of the gratings was either 0 or  $\pi$ . All combinations of orientations and phases were counter-balanced and randomized.

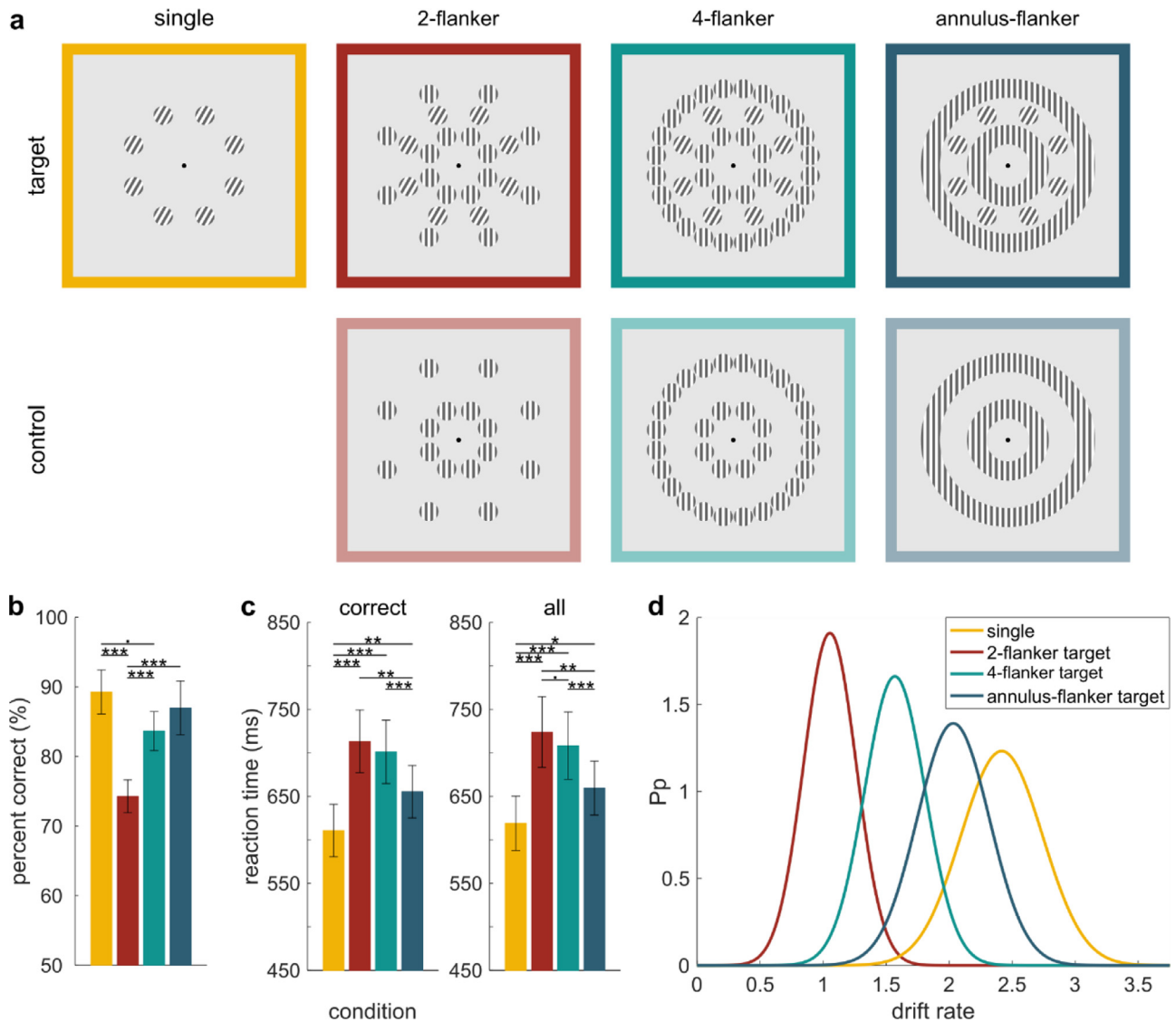
The experiment comprised seven experimental conditions (Fig. 1A), including four target conditions and three corresponding control conditions, which differed between each other in terms of the configuration of gratings. In the single-target condition (1), only the eight circular target gratings were presented. In the 2-flanker condition (2), two additional vertically oriented circular gratings were presented radially on either side of each target ("inner" and "outer" flanker: diameter 1.4 and 1.3 deg, eccentricity 2.3 deg and 5.7 deg, respectively, the same spatial frequency and contrast as the targets). In the 4-flanker condition (3), in addition to the "inner" and "outer" flankers, two vertically oriented circular gratings were presented on either side of each "outer" flanker in such a way that the "outer" flankers of all eight target gratings formed a circle. In the annulus-flanker condition (4), "inner" and "outer" flankers were connected into annuli of the same widths as the diameters of flankers in condition (2) and located at the same eccentricities. In addition, three control conditions (5–7) were presented that contained the same flankers as in the target conditions (2–4) but did not contain the target gratings. Participants were asked to push a randomly chosen button when they saw one of the control conditions.

We assumed that in cortical regions corresponding to the target locations the PSC in a given crowded target condition would correspond to the PSC in the single target condition plus the PSC in the corresponding control condition plus the crowding effect:

$$PSC_{\text{crowded target}} = PSC_{\text{single target}} + PSC_{\text{control}} + \text{crowding effect} \quad (1)$$

Thus, the purpose behind the control conditions was to isolate the component of the BOLD response corresponding to the crowding effect, independent of the configuration of the flankers. Our assumption of additive target-flanker BOLD effects was in keeping with the standard approach of previous fMRI studies of crowding (Anderson et al., 2012; Millin et al., 2014; Bi et al., 2009).

Each participant completed five sets of 88 trials (20 trials/condition/orientation/spatial phase in the target conditions and 20 trials/condition/spatial phase in the control conditions). In each trial, the stimulus appeared for 150 ms. The instructions to the participants were to respond as accurately as possible but within 1.5 s after stimulus onset. The next stimulus appeared at a random stimulus onset asynchrony (SOA)  $3 \pm 1$  s. If no response was made within 1.5 s after stimulus onset, the trial was considered invalid and repeated later in the set.



**Fig. 1.** (A) The four target conditions (single target, 2-flanker target, 4-flanker target, annulus-flanker target) and three corresponding control conditions with targets absent. (B) Accuracy of responses in each of the 4 target conditions. (C) Reaction times in correct trials (left) and all trials (right) in each of the 4 target conditions. Error bars correspond to the standard error of the mean (SEM). Post hoc comparisons between conditions: ‘,’  $BF_{10} \in [1,3)$  – anecdotal evidence in favor of the alternative hypothesis, ‘\*’  $BF_{10} \in [3,10)$  – substantial evidence, ‘\*\*’  $BF_{10} \in [10,30)$  – strong evidence, ‘\*\*\*’  $BF_{10} > 30$  – very strong evidence (Jeffreys, 1961; Lee and Wagenmakers, 2014). (D) Posterior distributions of the drift rate  $v$ . Bayesian hypothesis testing shows that all conditions differ strongly from each other ( $p[v_{2-flanker} < v_{4-flanker}] = 1.00$ ;  $p[v_{4-flanker} < v_{annulus-flanker}] = 0.98$ ;  $p[v_{annulus-flanker} < v_{single}] = 0.90$ ). The color coding in all panels is the same (see frames around stimulus displays in panel A and legend in panel D).

### 2.3. MRI data acquisition

MRI data were acquired using a 3 T whole-body MRI system (Magnetom Prisma, Siemens Medical Systems, Germany), using a 64-channel radio-frequency (RF) receive head coil and RF body coil for transmission. fMRI data for the main experiment and functional localizers consisted of partial 2D multi-echo echo-planar imaging (EPI) volumes ( $3.0 \times 3.0 \times 2.5$  mm; 20% distance factor; 1.98 s repetition time (TR); 30 ms echo time (TE);  $90^\circ$  flip angle). A  $64 \times 64$  base resolution was used, with 30 axial slices approximately parallel to the AC/PC line covering the occipital and parietal cortices, with a 192 mm field of view. A whole-brain EPI volume was also acquired and used in an intermediate step in the spatial registration of the partial functional image with the anatomical image. Slices for all functional images were acquired in ascending order. A standard 1.0 mm isotropic T1-weighted (T1w)

MPRAGE sequence image was acquired and used as high-resolution anatomical data for fMRI data pre-processing.

#### 2.3.1. LOC localizer

To localize the object-sensitive lateral occipital complex (LOC), we showed participants intact and scrambled images of objects. A subset of 120 images of objects ( $280 \times 200$  pixels) was selected from the SVLO database (Rossion and Pourtois, 2004) and converted to grayscale. We then scrambled these images by randomly shuffling square pieces of  $20 \times 20$  pixels (only those that contained parts of an object). On top of the intact and scrambled images, we superimposed a black grid (1 px wide).

Six blocks of each type of images were presented in succession (intact, scrambled, intact, etc.). In each 16 s block, a series of 20 images appeared separated by an empty screen with a fixation dot (200 ms stim-

ulus and 600 ms fixation per image). Participants were asked to look at the fixation dot.

### 2.3.2. Target ROI localizer

Ultimately, we sought to carry out analyses on voxels corresponding to the target grating locations. To localize the target regions of interest (ROIs), eight contrast-reversing flickering checkerboards were presented at the same locations as the target gratings in the main experiment (contrast 100%, spatial frequency 2.5 cpd, temporal frequency 6 Hz). Six blocks with flickering checkerboards (16 s) were interleaved with six blocks of empty screen (16 s).

Participants performed a fixation task. A small digit '9' appeared at the fixation. Sometimes it turned to '6' for 170 ms and turned back to '9' (SOA  $5 \pm 3$  s). Participants were asked to push the button with their index finger within 1.5 s after digit '6' onset. The task was designed to keep the participants' gaze at the fixation point.

### 2.3.3. Retinotopic mapping stimulus

We used retinotopic mapping to localize areas V1, V2, V3 and V4. A contrast-reversing flickering checkerboard wedge rotated counterclockwise eight times with a cycle time of 30 s (contrast 100%, temporal frequency of flicker 6 Hz). The wedge spanned 22.5 deg of radial angle and was restricted to eccentricities [0.5 ÷ 12] deg. Checks had approximately equal dimensions in the radial and tangential directions and were scaled with eccentricity. In the initial position, the wedge pointed upwards (12 o'clock position). The task was the same as in the target ROI localizer experiment.

## 2.4. Behavioral data analysis

### 2.4.1. Accuracy and reaction times

We first analyzed the behavioral data from the task that participants performed while in the scanner, to corroborate that our experimental conditions were successful in inducing (un)crowding. We used the open-source JASP software (Jeffreys's Amazing Statistics Program; version 0.13.1; <https://jasp-stats.org>; Love et al., 2019) for a Bayesian repeated-measures analysis of variance (ANOVA) of the percent correct (PC), reaction times in all trials (RT all) and reaction times in correct trials (RT correct) in three separate ANOVAs. For each of the ANOVAs, we compared a model which included a condition factor of four levels – corresponding to the single, 2-flanker, 4-flanker and annulus-flanker conditions – to a null model which only included a subject factor. We compared models using the Bayes factor ( $BF_{10}$ ), i.e., the evidence for the alternative hypothesis relative to the null hypothesis. Model comparison showed that the model which included the condition factor was far superior to the null model in all three behavioral measure models (PC:  $BF_{10} = 8189.29$ ; RT all:  $BF_{10} = 1.59e+6$ ; RT correct:  $BF_{10} = 6.61e+6$ ). We then conducted post hoc tests on the winning model, comparing the behavioral measures between conditions pairwise (Fig. 1B and C).

### 2.4.2. Drift-diffusion model

In order to obtain a single measure of stimulus difficulty that would combine accuracy and reaction times, we estimated a drift-diffusion model (DDM). In the DDM, decision-making is seen as a noisy process, with information accumulating over time from a given starting point  $z$ , corresponding to decision bias, towards one of two response boundaries, 0 and  $a$ . The distance  $a$  between the two boundaries is referred to as the threshold and corresponds to the speed/accuracy trade-off. Information accumulates at a drift rate  $\nu$ , a proxy for stimulus difficulty or the quality of evidence present in the stimulus. The model also includes a variable corresponding to the non-decision time  $t$ , i.e., the time before information starts accumulating, related to perception and movement initiation and execution. Please see [Ratcliff and McKoon, 2008] for details.

We used the Python-based HDDM (hierarchical drift diffusion model) toolbox for a hierarchical Bayesian estimation of the joint posterior distribution of model parameters (Wiecki et al., 2013). HDDM allows for a

hierarchical Bayesian estimation of the posterior distributions of model parameters based on the observed single-trial behavior, i.e., reaction times and responses (correct vs. incorrect). The hierarchical nature of this approach rests in the simultaneous estimation of group and subject parameters, with subject parameters assumed to be drawn from the group distribution. Such an approach necessitates a smaller number of trials per subject and condition than do non-hierarchical methods. The use of a Bayesian framework naturally lends itself to hierarchical modeling. Moreover, the Bayesian approach enables the estimation of the full posterior, giving information about the uncertainty of parameter estimates in addition to their most likely values. HDDM approximates the posterior distributions of DDM parameters using Markov Chain Monte Carlo sampling. We generated 10,000 samples and discarded 1,000 samples as burn-in. In order to verify that the models had converged, we inspected the traces of model parameters, their autocorrelation and calculated the R-hat (Gelman-Rubin) statistics.

We first assessed the validity of our model's assumptions (drift rate and threshold both modulated by the crowding condition – model  $\nu a$ ), comparing it against two nested models: one in which only drift rate was modulated by the experimental condition – model  $\nu$  – and the other, in which only threshold was modulated by condition – model  $a$ . We conducted model comparison of these three models using the deviance information criterion (DIC). The DIC is a measure which is used to compare complex hierarchical Bayesian models (Spiegelhalter et al., 2002). Like in the case of Akaike's information criterion (AIC), the lower the DIC the better the model, taking into account both model fit and complexity. Generally, a difference in DIC values between 5 and 10 corresponds to a substantial difference between two models, while a difference of more than 10 is decisive evidence in the favor of the model with the lower DIC.

The DIC values for models  $\nu a$ ,  $\nu$  and  $a$  were  $-414.97$ ,  $-371.34$  and  $-164.84$ , respectively. Thus, the model containing condition-dependent modulation of both drift rate and threshold was found to be the winning model. In order to ascertain whether DDM parameter estimates differed between conditions, we compared the posterior distributions corresponding to each pair of conditions using Bayesian hypothesis testing. Given that condition-specific estimates can only be interpreted in conjunction with the estimates of the intercept, we compared the parameter's overall intercept ( $a_{single}/\nu_{single}$ ) to the sums of  $a_{single}/\nu_{single}$  and the corresponding condition's intercept ( $a_{2-flanker}/\nu_{2-flanker}$ ,  $a_{4-flanker}/\nu_{4-flanker}$  and  $a_{annulus-flanker}/\nu_{annulus-flanker}$  for the 2-flanker, 4-flanker and annulus-flanker target conditions, respectively). Since we were interested in the parameter representing stimulus difficulty, we only show the drift rate results in the main text (Fig. 1D). The threshold results can be found in the Supplemental Information (Supplemental Fig. S1).

## 2.5. MRI data analysis

### 2.5.1. Pre-processing and univariate analysis

Imaging data pre-processing, statistical analysis and dynamic causal modeling were done using the statistical parametric mapping (SPM) software package (SPM12, Wellcome Trust Centre for Neuroimaging, London, UK, <http://www.fil.ion.ucl.ac.uk>) in MATLAB 9.1 (The MathWorks, Inc., Natick, MA). The fMRI data consisted of eight experimental sessions, i.e., (1–5) the five sessions of the main experiment, (6) the LOC localizer session, (7) the target ROI localizer session and (8) the retinotopic mapping session. The first five volumes of each of the experimental sessions were discarded to control for magnetic saturation effects. The remaining functional images were spatially realigned to the mean of the whole time-series using rigid-body transformations to correct for head motion. A B0 field map image, acquired during the session, was used to correct for EPI geometric distortions. We then performed slice timing correction, using the first acquired slice as the reference, and intensity bias correction. The anatomical (T1w) image was co-registered first to the whole-brain fMRI in an intermediary step and then to the mean fMRI volume using mutual information. Finally, the images from the LOC and

target localizer sessions were smoothed with a 4 mm full-width-at-half-maximum Gaussian kernel. We did not normalize the images to standard MNI space, but instead kept the images in each subject's native space.

First-level analysis consisted of three separate general linear models (GLMs), corresponding to (1) the main experiment, (2) the LOC localizer and (3) the target localizer. In the main experiment GLM, we included a separate regressor for each of the four target conditions – single target, 2-flanker target, annulus-flanker target, 4-flanker target – and each of the three control conditions – 2-flanker control, annulus-flanker control, 4-flanker control. The corresponding seven regressors modelled conditions as events (0 s duration) and timed them according to the presentation of visual cues. We specified a separate T-contrast for each of the seven experimental conditions, as well as two F-contrasts – one for the four target conditions and another for the three control conditions.

In the LOC localizer GLM, we included two regressors – one for intact and one for scrambled images. The regressors modelled the two conditions as 200 ms duration blocks, timed according to the image presentation. We specified a T-contrast for intact compared to scrambled images.

Finally, in the target localizer GLM, we included a single regressor which modelled the time of presentation of the eight flickering checkerboard target gratings. The regressor consisted of six blocks of 16 s duration. We specified a T-contrast for the presence of target gratings.

In each of the three GLMs, the time-series in each voxel was high-pass filtered at 1/128 Hz to remove low-frequency drifts, and regressors were convolved with a canonical hemodynamic response function. The GLM parameters were estimated using the classic SPM approach of Restricted Maximum Likelihood. The main experiment GLM results were further used for the percent signal change analysis and dynamic causal modeling. The LOC and target localizer GLM results were used in the volume of interest (VOI) definition procedure.

### 2.5.2. Retinotopic mapping

Cortical reconstruction was performed on the co-registered T1w image using FreeSurfer's recon-all function (FreeSurfer software package version 6.0, <http://surfer.nmr.mgh.harvard.edu/>, Dale et al., 1999). Retinotopic mapping was done using the SPM6 toolbox for population receptive field (pRF) modeling (Schwarzkopf et al., 2019). The procedure included projecting the pre-processed retinotopic mapping fMRI data onto the cortical surface and then fitting a standard Gaussian 2d pRF model to the data. While our wedge-only stimulus did not allow for an estimation of eccentricities or pRF size, we were only interested in the polar angle maps. Based on reversals in the polar angle map, we manually delineated our ROIs: V1, ventral and dorsal V2 (V2v/d, respectively), V3v/d and V4. ROI labels, defined on the cortical surface, were then converted back to volume space as binary masks. Masks from right and left hemispheres as well as from ventral and dorsal regions were combined. This is the only part of the analysis which was carried out in Linux (all other analyses were done in Windows).

### 2.5.3. VOI definition procedure

We created the target ROIs by applying a statistical threshold to the SPM at the output of the contrast defined on the target localizer GLM. Statistical values in SPMs can vary considerably between subjects – due to different levels of noise, different values of BOLD signal amplitudes, etc. – and we wanted to ensure specificity in our identification of target-responsive voxels. Thus, we used a subject-specific statistical thresholding procedure, basing the thresholds on the expected size of the targets' cortical representations.

For visual ROIs V1 to V4, we defined target-location ROIs based on the target localizer GLM. Instead of setting a standard statistical threshold for all participants, we instead ranked the voxels according to the corresponding T-statistic and retained a set number. We set this voxel cutoff number according to the expected size of the cortical representation of the eight target gratings in V1. At the target eccentricity of 4°, the cortical magnification factor is expected to be between

3 and 4 mm/degree (Duncan and Boynton, 2003). Given the stimulus diameter of 1.4°, an average cortical thickness of 2.5 mm and the  $3.0 \times 3.0 \times 2.5$  mm voxel resolution of the fMRI images, the cortical representation of the eight target gratings should amount to between approximately 10 and 20 voxels. Since the striate cortex can vary in size by as much as a factor of two between individuals (Dougherty et al., 2003), we set the voxel cutoff number to between 10 and 50. Thus, for each subject, the procedure consisted of the following: (1) start with a statistical threshold of  $p = 0.001$ , (2) find the intersection between the thresholded target localizer SPM, the V1 binary mask found through retinotopic mapping and the negation of the thresholded main experiment SPM corresponding to the F-contrast over control conditions ( $p = 0.01$  uncorrected, cluster extent-based threshold = 10), (3) count the number of voxels at the intersection described in (2); if less than 10, increase the  $p$ -value threshold; if greater than 50, decrease the  $p$ -value threshold. The final  $p$ -value threshold identified on V1 was then used for the given subject's remaining ROIs (V2 – V4). For two participants, this procedure yielded an empty target V4 ROI. In order not to exclude them from the analysis, we set the threshold to  $p = 0.05$  in these two instances. Across all participants, this resulted in  $44.08 \pm 3.21$  (mean  $\pm$  standard error of the mean [SEM]) voxels in V1,  $74.67 \pm 8.91$  voxels in V2,  $82.08 \pm 12.51$  voxels in V3 and  $44.33 \pm 7.39$  voxels in V4, with  $p$ -values  $\in [2.00e-14, 5.00e-2]$ .

In the case of the LOC, our VOI was based solely on the LOC localizer described in the MRI data acquisition section above. We did not refine the VOI further to target-specific voxels because the LOC is known to be a largely non-retinotopic area (Grill-Spector et al., 2001). In order to create the LOC VOI, we applied a subject-specific statistical threshold to the SPM at the output of the LOC localizer GLM so that the number of voxels surviving the threshold lay between 50 and 200 (mean  $\pm$  SEM:  $157.33 \pm 17.90$ ,  $p \in [2.00e-12, 4.20e-3]$ ).

## 2.6. Percent signal change analysis

We first sought to replicate findings from previous studies as a point of reference. Previous fMRI studies of crowding have shown that crowding attenuates the BOLD response (see Eq. (1); Anderson et al., 2012; Millin et al., 2014; Bi et al., 2009). Thus, we extracted the percent signal change (PSC) in the seven experimental conditions from ROIs corresponding to the locations of the target gratings in the retinotopically-identified visual regions V1, V2, V3 and V4 and from the non-retinotopic LOC.

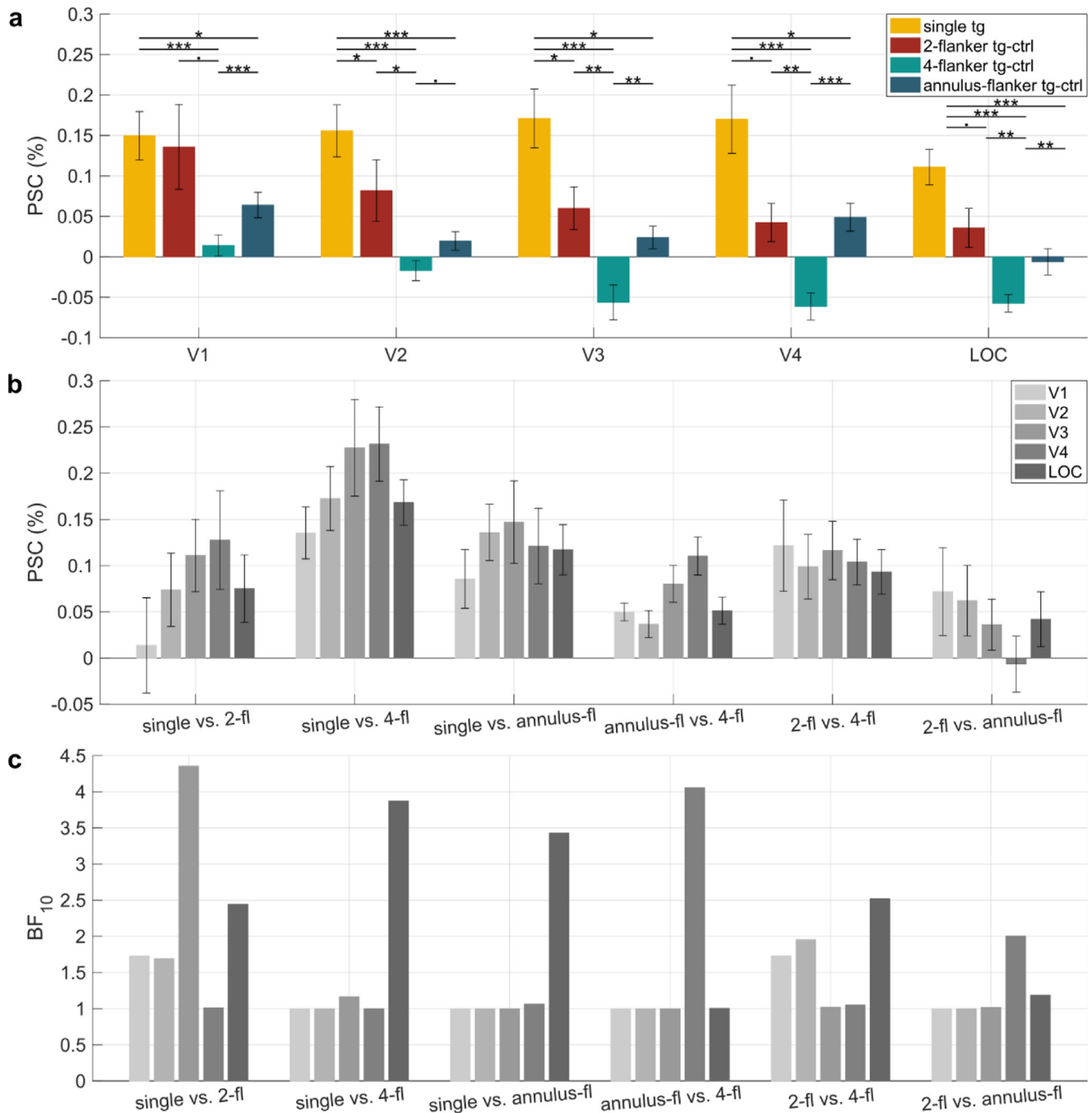
### 2.6.1. Group average analysis

We used the main experiment GLM to quantify activation magnitudes in PSC in each experimental condition using the Marsbar toolbox (<http://marsbar.sourceforge.net/>; Brett et al., 2002).

For each subject, each ROI and each experimental condition, we estimated the maximum of the absolute value of the first eigenvariate of the BOLD response ('abs max' option in Marsbar). The raw PSC results can be found in the Supplemental Information (Supplemental Fig. S2). We then subtracted the PSC in the control conditions from the corresponding target conditions, in accordance with Eq. (1) (Fig. 2A).

We analyzed the PSC in the conditions of interest using a Bayesian repeated-measures ANOVA with two factors, condition and ROI, in JASP. The condition factor had four levels, one for each experimental condition, and the ROI factor had five levels, one for each ROI (V1 to V4 and LOC). We compared a series of four models with different combinations of main effects and interaction of the two factors to a null model which only included a subject factor. Model comparison showed that the model which included main effects of condition and ROI was the winning model and was far superior to the null model in all three behavioral measure models ( $BF_{10} = 1.355e+24$ ).

In order to determine the between-condition differences between conditions within each ROI, we followed up the above analysis with separate repeated-measures ANOVAs for each ROI separately. Here, we



**Fig. 2.** (A) PSC estimates in the conditions-of-interest, i.e., the single target condition and the differences in PSC between the 2-flanker, 4-flanker and annulus-flanker target conditions and their respective control conditions. Post hoc comparisons between conditions: ‘.’  $BF_{10} \in [1,3)$ , ‘\*’  $BF_{10} \in [3,10)$ , ‘\*\*’  $BF_{10} \in [10,30)$ , ‘\*\*\*’  $BF_{10} > 30$ . (B) Pairwise differences in PSC between conditions of interest from panel (A). The first comparison (single vs. 2-flanker) corresponds to the effect of crowding w.r.t. no crowding. The second and third comparisons (single vs. 4-flanker and single vs. annulus-flanker) correspond to the effect of uncrowding w.r.t. no crowding. The fourth comparison (annulus-flanker vs. 4-flanker) corresponds to the effect of the connecting flankers into annuli. The final two comparisons (2-flanker vs. 4-flanker and 2-flanker vs. annulus-flanker) correspond to the effect of uncrowding w.r.t. crowding. We observe a clear hierarchical pattern, with increasing pairwise differences between conditions in the first four comparisons and decreasing pairwise differences in the last two comparisons, as one ascends the visual hierarchy. (C) The pairwise between-condition PSC differences from the different ROIs were used to predict the drift rate. We compared the ROI-based models against a null model, which only included a group intercept, using the Bayes factor (BF). Across almost all pairwise comparisons, either the V4 or LOC model performs best and substantially better than the null model. Error bars correspond to the *SEM*. The color coding in panels B and C is the same (see legend in panel B).

included a single between-subject factor: condition. For all five ROIs, the model that included the condition factor was far superior to the null model which only included a subject effect. We conducted pairwise between-condition post hoc tests for each ROI model (Fig. 2A).

In a complementary analysis, we subtracted the PSC in each of the conditions of interest from that of every other condition, pairwise. Our

goal with this analysis was to explore the patterns of BOLD attenuation due to (un)crowding across ROIs (Fig. 2B).

#### 2.6.2. Within-subject analysis

To draw a direct link between neural and behavioral measures, we conducted a model comparison to determine which ROI’s PSC could

best predict the DDM parameter of drift rate. In JASP, we estimated a series of Bayesian linear regression models (Castillo et al., 2015), one for each pair of conditions and each ROI separately using local empirical Bayesian (EB-local) priors (Hansen and Yu, 2001). The response variable consisted of each participant's difference in drift rates between the two conditions, while the dependent variable consisted of the differences in PSC between the two conditions for each participant in the given ROI. For each of these models, we also tested a corresponding null model that did not contain any information about the PSC. We compared the non-null models to the corresponding null model using the Bayes factor (Fig. 2C).

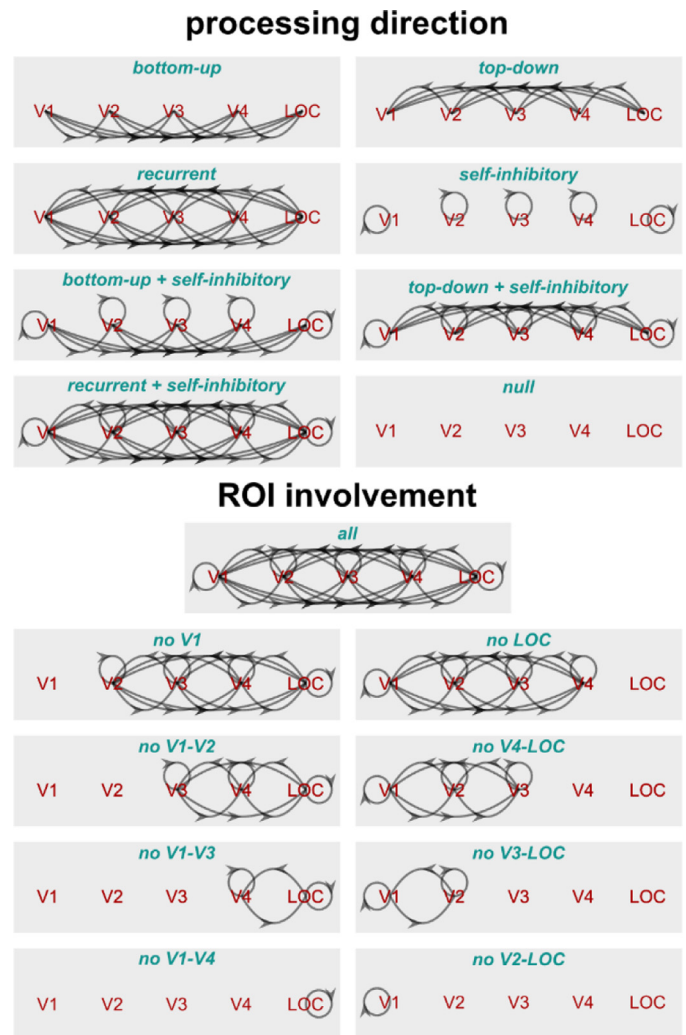
## 2.7. Dynamic causal modeling

We used DCM to determine how the five visual ROIs – V1 to V4 and LOC – interact with each other during our (un)crowding task. Dynamic causal modeling estimates the effective connectivity between regions in a brain network (Friston et al., 2003). A generative Bayesian model links the hidden neural states within the nodes – i.e., regions – of the network and their interactions with the signal observed with fMRI. Inverting this model gives us the parameter estimates corresponding to the connectivity between regions, including (1) the latent connectivity between the nodes of the network, i.e., the intrinsic connections among regions in the absence of input (the DCM.A matrix); (2) task-dependent modulation of connectivity between nodes (DCM.B); (3) the driving input (DCM.C). We constructed a series of DCMs including our five visual ROIs. We defined the latent connectivity between the ROIs to be a fully connected network, based on the known anatomical connectivity in macaques (Felleman and Van Essen, 1991). The direct driving effects of the input were set to target V1.

As in most DCM studies, our hypothesis pertained to the architecture of the task-dependent modulation of connectivity (DCM.B). For each of the four experimental conditions, we sought to distinguish between models with architectures that could be characterized as exhibiting local (within ROI) versus global (across ROI) processing and only bottom-up (feedforward) versus only top-down (feedback) versus recurrent (both feedforward and feedback) connectivity. Notably, within-ROI or self-inhibitory connections determine the excitatory/inhibitory balance within brain areas and provide a biologically plausible mechanism for changes in a region's activity, through the interplay of pyramidal cells and interneurons (Bastos et al., 2012). We also wanted to determine which ROIs were most important in each condition. Thus, we defined a model space consisting of 82 models, whose architectures varied according to two factors: processing direction – with 8 levels – and ROI involvement – with 10 levels (see Fig. 3 for an illustration of the two factors). The levels of the 'processing direction' and 'ROI involvement' factors were fully crossed (see Supplemental Fig. S3 for an illustration of the full model space). Model 1 corresponds to the full model with all possible connections between the five ROIs and model 82 corresponds to the null model, with no input-dependent modulation of connectivity (see the two panels in the lower right of Supplemental Fig. S3).

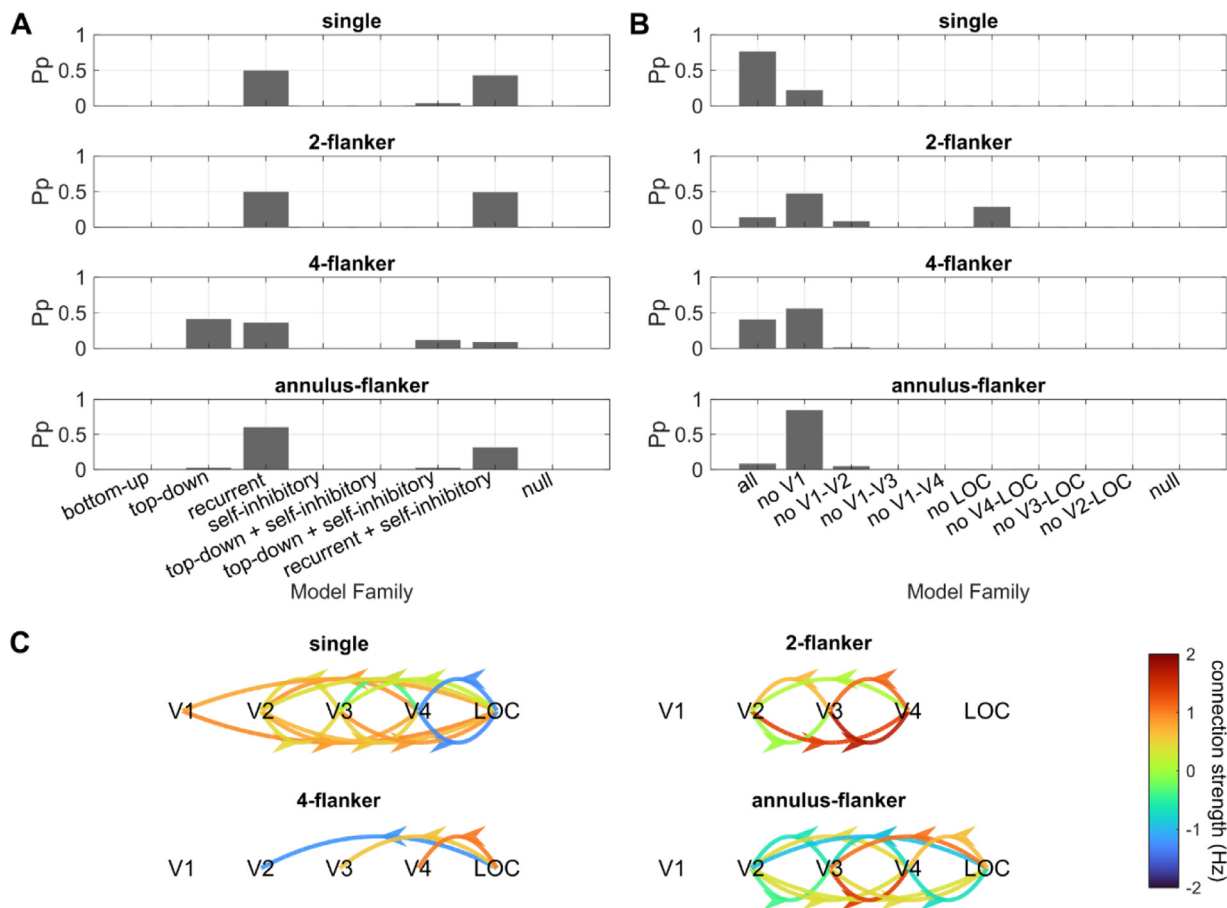
We used the parametric empirical Bayes (PEB) framework for DCM, which consists of a hierarchical Bayesian linear model over connectivity parameters (Friston et al., 2015; Litvak et al., 2015; Friston et al., 2016). We first estimated the full DCM model for all subjects, alternating between estimating individual DCMs and estimating group effects. The group effects are then used as (empirical) priors in the following iteration to constrain subject's connectivity estimates, thus drawing subject-level estimates out of local optima towards the group average. We then specified a GLM design matrix in order to estimate a group-level PEB model over connectivity parameter estimates. In this case, we were only interested in the group mean, so our design matrix consisted of an array of ones. We fit the model to the DCM.B connectivity parameters.

We then used Bayesian model reduction to estimate the remaining models (2–82) in our model space. To determine which models best ex-



**Fig. 3.** The two factors of the model space for the DCM analysis: 'processing direction' and 'ROI involvement', describing the possible architectures of condition-dependent modulation of connectivity (corresponding to the DCM.B matrix) between the five ROIs (V1 to V4 and LOC). The 'processing direction' factor consisted of 8 levels, defined by the directionality of connections. We considered the following directionalities of connections: bottom-up, top-down and self-inhibitory (i.e., local connections), as well as different combinations of the three. The 'ROI involvement' factor consisted of 10 levels, defined by the presence and absence of specified ROIs from the input-dependent modulation. The 'processing direction' and 'ROI involvement' factors were fully crossed, giving a total of 82 models (see Supplemental Fig. S3 for a complete illustration).

plain the data, we used Bayesian model selection (see Supplemental Fig. S4). Notably, Bayesian model comparison takes into account log model evidence, a trade-off between accuracy and complexity. Thus, in accordance with the principle of parsimony, the simplest model will be favored. We also conducted Bayesian model family selection over families of PEB models. More specifically, we conducted two model family selections, one for the 'processing direction' factor (Fig. 4A) and the other for the 'ROI involvement' factor (Fig. 4B), comparing among the PEB models which corresponded to the different levels of these factors. Posterior probabilities (Pp) of the models or model families were used to determine the winning model(s). Finally, we used Bayesian model averaging (BMA) to obtain the weighted average of parameter estimates across all 82 models. The weights correspond to the posterior probabilities of the corresponding models (Hoeting et al., 1999; Penny et al., 2006).



**Fig. 4.** DCM model selection results for each of the four conditions. (A) Posterior probabilities (Pp) of the 8 ‘processing direction’ model families. The recurrent model family was the winner in all conditions except the 4-flanker condition, where the top-down family was slightly superior. (B) Posterior probabilities of the 10 ‘ROI involvement’ model families. The winning model family in all but the single condition was the family which included all ROIs except for V1. In the single target condition, the winning family included all ROIs. (C) Condition-dependent modulations of connection strengths, as estimated through Bayesian model averaging (BMA) over the 82 models in the model space. BMA averages the parameters from the different models and weights them by the models’ posterior probabilities. Only those connections with a posterior probability greater than 0.95 are shown. The values of the coupling parameters correspond to connection strengths, i.e., how quickly and with what strength a response is elicited in the target region and are given in units of Hz (Friston et al., 2003). Positive/negative connections can be interpreted as excitatory/inhibitory, respectively.

### 3. Results

#### 3.1. Behavioral results

In the crowded condition (2-flanker target), the percentage of correct responses (PC) was much lower (mean  $\pm$  SEM:  $74.27\% \pm 2.36\%$ ) and reaction times (RT) were much higher ( $723.94 \text{ ms} \pm 40.68 \text{ ms}$ ) than in the single target condition (PC:  $89.27\% \pm 3.16\%$ ; RT:  $618.96 \text{ ms} \pm 31.23 \text{ ms}$ ). The presence of additional flanker gratings in the 4-flanker target condition lead to *uncrowding*, i.e., increased PC ( $83.65\% \pm 2.80\%$ ) and decreased RT ( $708.22 \text{ ms} \pm 38.80 \text{ ms}$ ). Connecting the flankers into annuli further improved performance (PC:  $86.98\% \pm 3.87\%$ ; RT:  $659.48 \text{ ms} \pm 31.02 \text{ ms}$ ). These results (Fig. 1B and C) are consistent with previous findings that global aspects of a stimulus are crucial in determining crowding magnitude (Malania et al., 2007; Manassi et al., 2012, 2013, 2015; Herzog and Manassi, 2015a; Herzog et al., 2015b, 2016a, 2016b).

The above results show that response accuracies and reaction times go together, i.e., when PC is higher, RT is lower. In order to obtain a single measure of stimulus difficulty, we estimated a drift diffusion model (DDM; Ratcliff and McKoon, 2008) using the Python-based HDDM toolbox (Wiecki et al., 2013). We built a model which included condition-dependent modulations of the threshold  $a$  – which accounts for the

speed-accuracy trade-off – and drift rate  $\nu$  – corresponding to stimulus difficulty. Since we were specifically interested in stimulus difficulty, we focus on the drift-rate parameter here (see Supplemental Information for details on the threshold parameter  $a$ ). Results are shown in Fig. 1D.

Similarly to the performance in terms of correct responses, we found the highest drift rate in the single target condition ( $2.42 \pm 0.32$ ), followed, in order, by the annulus-flanker ( $2.03 \pm 0.29$ ), 4-flanker ( $1.57 \pm 0.24$ ) and 2-flanker ( $1.05 \pm 0.21$ ) target conditions. All conditions differed strongly from each other. This finding was expected, as drift rate is known to quantify difficulty and can likewise be interpreted as an index for the signal-to-noise ratio of the information processing system.

#### 3.2. Neuroimaging results: crowding attenuates the BOLD response but uncrowding does more

Before conducting the DCM analysis, we first sought to replicate findings from previous studies as a point of reference. We extracted the PSC in the seven experimental conditions from ROIs corresponding to the locations of the target gratings in the retinotopically-identified visual regions V1, V2, V3 and V4 and from the non-retinotopic LOC.

In accordance with Eq. (1), we subtracted the PSC in each of the control conditions from that of the corresponding crowded target condition. Previous studies have indicated that crowding attenuates the BOLD re-



sponse, with the implicit expectation that low BOLD corresponds to low performance and high BOLD to good performance. Although we replicated previous results of crowding-related PSC attenuation starting in V2 and following a positive gradient from early to late retinotopic visual areas, we found that uncrowding attenuates PSC even more (Fig. 2A). Thus, it appears that the PSC is not (monotonically) related to the level of crowding, contrary to assumptions.

In order to investigate the patterns in crowding-related BOLD attenuation across ROIs, rather than simply comparing crowding and control conditions within ROIs, we subtracted the PSC in one condition from the other for all pairs of conditions (Fig. 2B). The higher the difference in PSC, the greater the level of BOLD amplitude modulation by crowding with respect to no crowding and uncrowding with respect to crowding, respectively (Anderson et al., 2012). Interestingly, a hierarchical pattern emerged in each of the pairwise crowding level indices. Across the first four comparisons (i.e., single vs. 2-flanker, single vs. 4-flanker, single vs. annulus-flanker and annulus-flanker vs. 4-flanker), V4 reflected the crowded percept better than lower visual areas; for uncrowding (i.e., 2-flanker vs. 4-flanker and 2-flanker vs. annulus-flanker), this trend was reversed. While the level of crowding-related attenuation in the LOC was higher than that in V1 – indicating that the LOC reflects the crowded percept better – the level of attenuation was the same or lower than that of the remaining ROIs. This is potentially due to the LOC being a largely non-retinotopic area (Grill-Spector et al., 2001) and, relatedly, the fact that we used the entire LOC for this analysis instead of just the target locations as in the other ROIs, making the PSC an average over target- and flanker-related areas.

One of the weaknesses of the PSC analysis approach of most studies – which we replicated above – is that it only considers the *group averaged* differences between the crowding conditions. Aiming to unearth the causes behind this result and to draw a direct link between neural and behavioral measures, in a second analysis we used a within-subject approach. We conducted a model comparison to determine which ROI could best explain the behavioral performance estimated by the drift rate. For each pair of conditions and each ROI, we estimated a regression model in which the difference in drift rates between the two conditions was the dependent variable and the difference in PSC between the two conditions was the independent variable. We compared the models for each ROI to a null model which did not include any PSC information using the Bayes factor, i.e., the evidence for the alternative hypothesis relative to the null hypothesis.

We found that across almost all pairwise between-condition comparisons, the model which best explained the drift rate data was the one that includes the PSC from one of the higher visual areas, i.e., V4 or LOC (Fig. 2C). The one exception was the single vs. 2-flanker comparison, in which the V3 model was the winning model, though the LOC model was the second best. In the two comparisons between the crowding condition and the uncrowding conditions, the BF of the winning models was below 3; thus, the winning models could not be considered substantially better than the null model. Hence, early visual areas do not contain substantial discriminatory information regarding (un)crowding. In fact, in all but one of the pairwise condition comparisons (i.e., single vs. 2-flanker), the V1 model performed at the same level as the null model.

While this last analysis linking PSC to drift rate offers a more complete picture than the standard group-level PSC approach, it is still limited by the fact that ROIs are considered independently. The visual hierarchy is dynamic and interactive. In order to investigate the potential recurrent connections between visual areas, we used DCM, a state-of-the-art technique for modeling the neural interactions underlying the signals observed through neuroimaging (Friston et al., 2003).

### 3.3. (Un)crowding is mediated by global, recurrent processing

Here, we used DCM to determine how the five visual ROIs – V1 to V4 and LOC – interact with each other during our (un)crowding task.

We sought to distinguish between models that could be characterized as exhibiting local (within ROI) versus global (across ROI) processing and only bottom-up (feedforward) versus only top-down (feedback) versus recurrent (both feedforward and feedback) connectivity – which we call the ‘processing direction’ factor. Moreover, we wanted to determine which ROIs were most important in each condition – which we call the ‘ROI involvement’ factor. Our hypotheses were expressed through a model space consisting of 82 models varying according to 8 levels of ‘processing direction’ and 10 levels of ‘ROI involvement’ (see Fig. 3 for an illustration of the two factors and Supplemental Fig. S3 for an illustration of all 82 models). As the final outcome measure of the model comparison, we computed the posterior probability (Pp) for the model families corresponding to the 8 levels of the ‘processing direction’ (Fig. 4A) and the 10 levels of ‘ROI involvement’ (Fig. 4B).

In all but the 4-flanker condition, the winning ‘processing direction’ model family was either the recurrent processing family or the recurrent and self-inhibitory processing family. In the 4-flanker condition, the winning family was the top-down processing family (Pp = 0.42), though it was closely followed by the recurrent processing family (Pp = 0.37). This finding is particularly interesting given the potential role of perceptual grouping in uncrowding. The fact that the top-down processing family was not more prominent in the annulus-flanker condition (Pp = 0.03) could be related to the fact that the flankers in this condition are connected into annuli and are thus more readily separable from the target even in lower regions without additional top-down processing as compared to the no crowding condition (single target). In the single target condition, the winning ‘ROI involvement’ model family was the family that contains all ROIs (Pp = 0.77). In all three crowding conditions, the winning model family was consistently the family that omits V1. The posterior probability of this model family increased successively as uncrowding increased, i.e., it was lowest in the 2-flanker condition (Pp = 0.48), higher in the 4-flanker condition (Pp = 0.57) and highest in the annulus-flanker condition (Pp = 0.85). While in the uncrowding conditions, the model family that includes all ROIs was the runner-up (4-flanker: Pp = 0.41; annulus-flanker: Pp = 0.09), in the 2-flanker condition, the runner-up was the model family that includes all regions but the LOC (Pp = 0.29). For detailed model comparison results (winning models in each of the four conditions, as opposed to winning model families), please see Supplemental Fig. S4.

We averaged the parameters from the different models, weighting them by the models’ posterior probabilities, a procedure known as Bayesian model averaging (Hoeting et al., 1999; Penny et al., 2006). Fig. 4C shows the Bayesian model average of connectivity parameters, at a threshold of Pp > 0.95. The connectivity parameters correspond to the strength of connections between regions, i.e., how quickly and with what strength a response is elicited in the target region, given in units of Hz (Friston et al., 2003). The resulting networks reflect the Bayesian model family comparisons, with a full recurrent network in the single condition, a recurrent subnetwork including regions V2 to V4 in the 2-flanker condition, a top-down network excluding V1 in the 4-flanker condition and a recurrent network excluding V1 in the annulus-flanker condition. The particular strength of DCM is that it not only allows model and model family comparisons, but it also provides us with estimates of the effective connectivity parameters, representing the connection strengths between regions. Thus, we are able not only to identify the system as local or global and feedforward or feedback, but also to examine the computational details of the excitatory or inhibitory nature of its connections.

## 4. Discussion

The mechanisms governing crowding have been an issue of contention in vision research for more than half a century. Most studies of crowding are theory-driven, attempting to match models – more or less inspired by the visual cortical hierarchy – to behavioral data. The majority of the models proposed thus far have been feedforward and local,

a model class which is mathematically tractable and relatively homogeneous. However, these models – and even the majority of global ones – fail when confronted with uncrowding (Doerig et al., 2019). Clearly, the rejection of local feedforward models is not sufficient in explaining (un)crowding, as the universe of recurrent models is vast. New, data-driven techniques are needed to determine the computational details of cortical processing in (un)crowding. Here, we took advantage of state-of-the-art methods in neuroimaging analysis to investigate the complex cortical interactions underlying crowding through dynamic causal modeling. We provide unique empirical evidence of how exactly recurrent processing operates in (un)crowding. Our results reveal the modulatory effect of context on recurrent connectivity across the visual cortical hierarchy, from low-level to high-level visual areas.

Previous neuroimaging studies of crowding have taken the approach of comparing percent BOLD signal change in crowded and non-crowded conditions and localizing the regions in the brain where the difference in PSC is greatest (Fang and He, 2008; Bi et al., 2009; Anderson et al., 2012; Millin et al., 2014). While they disagree on the exact locus of crowding (V1 or V2), all aforementioned studies agree that crowding attenuates the BOLD response. As a first analysis, we used this established method and replicated previous results, finding that crowding attenuates the BOLD response as early as V2 and that the attenuation persists through V3, V4 and LOC (see Fig. 2A and B).

While we were able to replicate previous findings regarding the “neural locus” of crowding, the standard approach failed when applied to uncrowding. Following the logic that PSC reflects the magnitude of the crowding effect, one would expect to see an increase of PSC in the uncrowding conditions relative to the crowding condition. Instead, we found that uncrowding further suppressed the BOLD response. This finding suggests an important role for top-down connections. Moreover, it suggests that identifying the “neural correlates” of crowding and uncrowding is more complex than just looking for a reflection of behavioral patterns in BOLD attenuation at the group level.

In a second analysis, we sought to link PSC with the DDM parameter of drift rate, a proxy for stimulus difficulty, at the participant level. Our analysis revealed that the difference between the crowding and no crowding conditions was best reflected by the signal in V3, followed closely by the LOC. For all other pairwise condition comparisons, either V4 or the LOC reflected the stimulus difficulty best. This result recalls a recent electroencephalography (EEG) study of (un)crowding (Chicherov et al., 2014), where in a Vernier discrimination task, the N1 component was suppressed in crowding with respect to uncrowding. Interestingly and in accordance with our current findings, this suppression was found to occur in a high-level visual area such as the LOC, as revealed through EEG source localization. These findings support the notion that higher-level areas determine perceptual grouping and send a suppressing or enhancing signal to lower-level ROIs depending on the extent to which the target groups with the flankers.

Visual processing is dynamic, and so we posit that the standard approach – which relates behavioral performance to a static measure of BOLD activation – cannot provide the full picture. In our final analysis, we used dynamic causal modeling to explore the crowding-induced changes in inter-areal effective connectivity. Our results show that recurrence is critical, not only for crowding and uncrowding, but even in the no crowding condition (see Fig. 4). Moreover, global (between-ROI) processing is vastly more important than local (within-ROI) processing. These findings shed light on the possible reasons behind discrepancies in previous PSC analyses of crowding: given that recurrent processing is crucial, a neural signature of crowding can in theory be detected at any point throughout the visual hierarchy.

Interestingly, we found that whereas the preferred model in the no crowding condition includes modulation of connectivity between all ROIs, the preferred models in the (un)crowding conditions consistently exclude area V1. This result is consistent with our finding of crowding-related BOLD signal suppression starting in V2 and persisting in areas V3, V4 and the LOC. V2 has been suggested to play a critical role in

crowding (Fang and He, 2008; Bi et al., 2009; He et al., 2019), which is in accordance with the two-stage model of crowding (Levi, 2008; Pelli and Tillman, 2008), in which simple feature detection takes place in V1 while feature integration takes place beyond V1 – perhaps in V2. While classical theories posit that this feature integration is a feedforward mechanism, it is completely feasible that feature integration is mediated by top-down mechanisms.

Our finding of recurrent connectivity modulation further strengthens our argument of the importance of perceptual grouping in (un)crowding, which originates from behavioral studies (Manassi et al., 2012; Hermens and Bell, 2014; Herzog et al., 2015b; Manassi et al., 2015) and is supported by our aforementioned results of V4/LOC as the site of between-condition differences in performance. The role of recurrent connections in crowding can be seen through the lens of visual processing as predictive coding (Srinivasan et al., 1982; Mumford, 1992; Rao and Ballard, 1999). We speculate that feedback connections from higher to lower visual areas convey predictions as to the noteworthiness of the target, i.e., either it is part of a group and thus provides only redundant information or it is unique among its surroundings and thus provides important information. This top-down information then enhances or suppresses the feedforward information, in our case, the tilt direction, leading to better or worse behavioral performance.

It is also interesting to note here the omission of the LOC from the winning model in the crowding condition, while the LOC is present in the case of both uncrowding conditions. While V4 and the LOC both respond preferentially to images of objects as opposed to scrambled images (Grill-Spector et al., 1998), V4 is retinotopically organized, while the LOC is largely non-retinotopic (Grill-Spector et al., 2001). In the uncrowding conditions, the stimuli are arranged in such a way that the flankers form two circles (Fig. 1A). In the crowding condition, on the other hand, the global shape is not a good Gestalt. Thus, we speculate that the omission of the LOC from the winning model in crowding and not in uncrowding may be related to the complexity of the global shapes in the corresponding stimulus configurations. However, we would like to note that good Gestalt per se is not an indicator for uncrowding since when a target is part of a good Gestalt, it can be strongly crowded. Likewise, crowding occurs also when the target is embedded in scrambled elements.

Here, it is worth noting several possible criticisms of our study. First, our study is somewhat limited by the small number of participants ( $n = 12$ ). However, our use of Bayesian statistical methods – for the behavioral data and PSC data analyses, as well as throughout the DCM analysis – should alleviate any power concerns. A second limitation pertains to DCM itself. The main issues of DCM are well described elsewhere (Roebroeck et al., 2011a, 2011b; Valdes-Sosa et al., 2011; David, 2011; Friston, 2011; Lohmann et al., 2012), but we will mention a couple of the more relevant limitations here. In contrast to the other leading effective connectivity estimation technique – Granger causality, DCM analysis is model-based rather than data-driven. As such, DCM analysis typically entails testing a few chosen models, which express specific hypotheses. In our study, this restriction is tempered by the use of Bayesian model reduction, which enables the rapid estimation of nested models. Our use of BMR allowed us to test a relatively large model space of 82 models. Another noteworthy limitation of DCM is the fact of model parameters being conditional on the tested model (Daunizeau et al., 2011). This fact needs to be kept in mind when interpreting the connectivity parameter estimates. A final possible criticism of our study is the perhaps predictable outcome of the importance of both inter- and intra-regional connections in predicting the data. However, it is important to keep in mind that Bayesian model selection follows the principle of parsimony (see Methods Section 2.7), making it nontrivial that a model that is not the simplest would be chosen. The fact that, in all experimental conditions, the winning model family includes top-down connections strongly supports our initial hypothesis that top-down connections are crucial in both crowding and uncrowding.

## 5. Conclusion

While classifying a model as global vs. local and feedforward vs. feedback is a first step, it is the computational details of the model – i.e., the specific combinations of excitatory and inhibitory connections between stages of processing – which determine its robustness in predicting behavioral data (Clarke et al., 2014). It is precisely for this reason that methods such as DCM – and the Bayesian generative modeling paradigm in general – are so valuable. The DCM framework allows us not only to compare models whose architectures are defined to match specific hypotheses but also to estimate the computational details of the model in the form of the connection strengths between regions. Our findings can be used as a basis for theoretical models of crowding, both in terms of information representation (global rather than local) and function (feedback rather than feedforward). Moreover, our study paves the way for future DCM investigations of crowding, which would generalize our results to other stimulus classes and visual paradigms.

## Declaration of Competing Interest

The authors declare no competing interests.

## Credit authorship contribution statement

**Maya A. Jastrzębowska:** Methodology, Formal analysis, Investigation, Data curation, Writing – original draft, Visualization, Writing – review & editing. **Vitaly Chicherov:** Conceptualization. **Bogdan Draganski:** Writing – review & editing, Supervision, Project administration. **Michael H. Herzog:** Conceptualization, Resources, Writing – review & editing, Supervision, Project administration, Funding acquisition.

## Acknowledgments

We would like to thank Marc Repnow for his technical support and Antoine Lutti for setting up the MRI sequences. M.A.J. and M.H.H. are supported by the Swiss National Science Foundation grant ‘Basics of visual processing: from elements to figures’ (176153, <http://p3.snf.ch/Project-176153>). B.D. is supported by the Swiss National Science Foundation (NCCR Synapsy, project grant numbers 32003B\_135679, 32003B\_159780, 324730\_192755 and CRSK-3\_190185) and the Leenaards Foundation. LREN is very grateful to the ROGER DE SPOELBERCH and Partridge Foundations for their generous financial support.

## Data and code availability

The data from the current study has not been deposited in a public repository due to participant privacy concerns but is available from the corresponding author when accompanied by a formal data sharing agreement. The code used to pre-process and analyze the data is available at <https://github.com/mayajas/Unraveling-brain-interactions-in-vision-the-example-of-crowding>.

## Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:[10.1016/j.neuroimage.2021.118390](https://doi.org/10.1016/j.neuroimage.2021.118390).

## References

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

Bach, M., 1996. The Freiburg visual acuity test—automatic measurement of visual acuity. *Optom. Vis. Sci. Off. Publ. Am. Acad. Optom.* 73 (1), 49–53. doi:[10.1097/00006324-199601000-00008](https://doi.org/10.1097/00006324-199601000-00008).

Balas, B., Nakano, L., Rosenholtz, R., 2009. A summary-statistic representation in peripheral vision explains visual crowding. *J. Vis.* 9 (12), 1–18. doi:[10.1167/9.12.1](https://doi.org/10.1167/9.12.1).

Banks, W.P., Larson, D.W., Prinzmetal, W., 1979. Asymmetry of visual interference. *Percept. Psychophys.* 25 (6), 447–456. doi:[10.3758/BF03213822](https://doi.org/10.3758/BF03213822).

Banks, W.P., White, H., 1984. Lateral interference and perceptual grouping in visual detection. *Percept. Psychophys.* 36 (3), 285–295. doi:[10.3758/BF03206370](https://doi.org/10.3758/BF03206370).

Bastos, A.M., Usrey, W.M., Adams, R.A., Mangun, G.R., Fries, P., Friston, K.J., 2012. Canonical microcircuits for predictive coding. *Neuron* 76 (4), 695–711. doi:[10.1016/j.neuron.2012.10.038](https://doi.org/10.1016/j.neuron.2012.10.038).

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. doi:[10.1167/9.11.13](https://doi.org/10.1167/9.11.13), –13.

Brainard, D.H., 1997. The psychophysics toolbox. *Spat. Vis.* 10 (4), 433–436. doi:[10.1163/156856897X00357](https://doi.org/10.1163/156856897X00357).

Brett, M., Anton, J.-L., Valabregue, R., Poline, J.-B., 2002. Region of interest analysis using the MarsBar toolbox for SPM 99. *Neuroimage* 16, 497.

Cao, Y., Grossberg, S., 2005. A laminar cortical model of stereopsis and 3D surface perception: closure and da Vinci stereopsis. *Spat. Vis.* 18 (5), 515–578. doi:[10.1163/156856805774406756](https://doi.org/10.1163/156856805774406756).

Castillo, I., Schmidt-Hieber, J., Van der Vaart, A., 2015. Bayesian linear regression with sparse priors. *Ann. Stat.* 43 (5), 1986–2018.

Chakravarthi, R., Pelli, D.G., 2011. The same binding in contour integration and crowding. *J. Vis.* 11 (8), 1–12. doi:[10.1167/11.8.10](https://doi.org/10.1167/11.8.10).

Chaney, W., Fischer, J., Whitney, D., 2014. The hierarchical sparse selection model of visual crowding. *Front. Integr. Neurosci.* 8, 73. doi:[10.3389/fnint.2014.00073](https://doi.org/10.3389/fnint.2014.00073).

Chicherov, V., Plomp, G., Herzog, M.H., 2014. Neural correlates of visual crowding. *Neuroimage* 93 Pt 1, 23–31. doi:[10.1016/j.neuroimage.2014.02.021](https://doi.org/10.1016/j.neuroimage.2014.02.021).

Clarke, A.M., Herzog, M.H., Francis, G., 2014. Visual crowding illustrates the inadequacy of local vs. global and feedforward vs. feedback distinctions in modeling visual perception. *Front. Psychol.* 5, 1193. doi:[10.3389/fpsyg.2014.01193](https://doi.org/10.3389/fpsyg.2014.01193).

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* 9 (2), 179–194. doi:[10.1006/nimg.1998.0395](https://doi.org/10.1006/nimg.1998.0395).

Daunizeau, J., David, O., Stephan, K.E., 2011. Dynamic causal modelling: a critical review of the biophysical and statistical foundations. *Neuroimage* 58 (2), 312–322. doi:[10.1016/j.neuroimage.2009.11.062](https://doi.org/10.1016/j.neuroimage.2009.11.062).

David, O., 2011. fMRI connectivity, meaning and empiricism. Comments on: Roebroeck et al. The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *Neuroimage* 58 (2), 306–309. doi:[10.1016/j.neuroimage.2009.09.073](https://doi.org/10.1016/j.neuroimage.2009.09.073).

Ding, M., Chen, Y., Bressler, S.L., 2006. Granger causality: basic theory and application to neuroscience. In: Schelter, B., Winterhalder, M., Timmer, J. (Eds.), *Handbook of Time Series Analysis: Recent Theoretical Developments and Applications*. Wiley, pp. 437–460. doi:[10.1002/9783527609970.ch17](https://doi.org/10.1002/9783527609970.ch17).

Doerig, A., Bornet, A., Rosenholtz, R., Francis, G., Clarke, A.M., Herzog, M.H., 2019. Beyond Bouma’s window: how to explain global aspects of crowding? *PLoS Comput. Biol.* 15 (5), 1–28. doi:[10.1371/journal.pcbi.1006580](https://doi.org/10.1371/journal.pcbi.1006580).

Doerig, A., Schmittwilken, L., Sayim, B., Manassi, M., Herzog, M.H., 2020. Capsule networks as recurrent models of grouping and segmentation. *PLoS Comput. Biol.* 16 (7), e1008017. doi:[10.1371/journal.pcbi.1008017](https://doi.org/10.1371/journal.pcbi.1008017).

Dougherty, R.F., Koch, V.M., Brewer, A.A., Fischer, B., Modersitzki, J., Wandell, B.A., 2003. Visual field representations and locations of visual areas V1/2/3 in human visual cortex. *J. Vis.* 3 (10), 1. doi:[10.1167/3.10.1](https://doi.org/10.1167/3.10.1).

Duncan, R.O., Boynton, G.M., 2003. Cortical magnification within human primary visual cortex correlates with acuity thresholds. *Neuron* 38, 659–671. doi:[10.1167/2.7.129](https://doi.org/10.1167/2.7.129).

Fang, F., He, S., 2008. Crowding alters the spatial distribution of attention modulation in human primary visual cortex. *J. Vis.* 8 (9), 1–9. doi:[10.1167/8.9.6](https://doi.org/10.1167/8.9.6).

Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* 1 (1), 1–47. doi:[10.1093/cercor/1.1.1](https://doi.org/10.1093/cercor/1.1.1).

Francis, G., Manassi, M., Herzog, M.H., 2017. Neural dynamics of grouping and segmentation explain properties of visual crowding. *Psychol. Rev.* 124 (4), 483–504. doi:[10.1037/rev0000070](https://doi.org/10.1037/rev0000070).

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. *Neuroimage* 19 (4), 1273–1302. doi:[10.1016/s1053-8119\(03\)00202-7](https://doi.org/10.1016/s1053-8119(03)00202-7).

Friston, K.J., 2011. Dynamic causal modeling and Granger causality comments on: the identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *Neuroimage* 58 (2), 303–305. doi:[10.1016/j.neuroimage.2009.09.031](https://doi.org/10.1016/j.neuroimage.2009.09.031).

Friston, K.J., Litvak, V., Oswal, A., Razi, A., Stephan, K.E., van Wijk, B., Ziegler, G., Zeidman, P., 2016. Bayesian model reduction and empirical Bayes for group (DCM) studies. *Neuroimage* 128, 413–431. doi:[10.1016/j.neuroimage.2015.11.015](https://doi.org/10.1016/j.neuroimage.2015.11.015).

Friston, K.J., Zeidman, P., Litvak, V., 2015. Empirical Bayes for DCM: a group inversion scheme. *Front. Syst. Neurosci.* 9, 164. doi:[10.3389/fnsys.2015.00164](https://doi.org/10.3389/fnsys.2015.00164).

Geweke, J., 1982. Measurement of linear dependence and feedback between multiple time series. *J. Am. Stat. Assoc.* 77 (378), 304–313. doi:[10.1080/01621459.1982.10477803](https://doi.org/10.1080/01621459.1982.10477803).

Granger, C.W.J., 1980. Testing for causality. A personal viewpoint. *J. Econ. Dyn. Control* 2, 329–352. doi:[10.1016/0165-1889\(80\)90069-X](https://doi.org/10.1016/0165-1889(80)90069-X).

Greenwood, J.A., Bex, P.J., Dakin, S.C., 2009. Positional averaging explains crowding with letter-like stimuli. *Proc. Natl. Acad. Sci. U.S.A.* 106 (31), 13130–13135. doi:[10.1073/pnas.0901352106](https://doi.org/10.1073/pnas.0901352106).

Grill-Spector, K., Kushnir, T., Hendler, T., Edelman, S., Itzhak, Y., Malach, R., 1998. A sequence of object-processing stages revealed by fMRI in the human occipital lobe. *Hum. Brain Mapp.* 6 (4), 316–328. doi:[10.1002/\(SICI\)1097-0193\(1998\)6:4<316::AID-HBM9>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1097-0193(1998)6:4<316::AID-HBM9>3.0.CO;2-6).

Grill-Spector, K., Kourtzi, Z., Kanwisher, N., 2001. The lateral occipital complex and its role in object recognition. *Vis. Res.* 41 (10–11), 1409–1422. doi:[10.1016/S0042-6989\(01\)00073-6](https://doi.org/10.1016/S0042-6989(01)00073-6).

- Hansen, M.H., Yu, B., 2001. Model selection and the principle of minimum description length. *J. Am. Stat. Assoc.* 96, 746–774.
- He, D., Wang, Y., Fang, F., 2019. The critical role of V2 population receptive fields in visual orientation crowding. *Curr. Biol.* 29 (13), 2229–2236. doi:10.1016/j.cub.2019.05.068, e3.
- Hermens, F., Bell, A., 2014. Speeded classification in simultaneous masking. *J. Vis.* 14 (6), 6. doi:10.1167/14.6.6.
- Herzog, M.H., Manassi, M., 2015a. Uncorking the bottleneck of crowding: a fresh look at object recognition. *Curr. Opin. Behav. Sci.* 1, 86–93. doi:10.1016/j.cobeha.2014.10.006.
- Herzog, M.H., Sayim, B., Chicherov, V., Manassi, M., 2015b. Crowding, grouping, and object recognition: a matter of appearance. *J. Vis.* 15 (6), 5. doi:10.1167/15.6.5.
- Herzog, M.H., Thunell, E., Ögmen, H., 2016a. Putting low-level vision into global context: why vision cannot be reduced to basic circuits. *Vis. Res.* 126, 9–18. doi:10.1016/j.visres.2015.09.009.
- Herzog, M.H., Sayim, B., Manassi, M., Chicherov, V., 2016b. What crowds in crowding? *J. Vis.* 16 (11), 25. doi:10.1167/16.11.25.
- Hoeting, J.A., Madigan, D., Raftery, A.E., Volinsky, C.T., 1999. Bayesian model averaging: a tutorial. *Stat. Sci.* 382–401.
- Jeffreys, H., 1961. *Theory of Probability*, 3rd ed. Oxford University Press, Oxford, UK.
- Joo, S.J., Boynton, G.M., Murray, S.O., 2012. Long-range, pattern-dependent contextual effects in early human visual cortex. *Curr. Biol.* 22 (9), 781–786. doi:10.1016/j.cub.2012.02.067.
- Lee, M.D., Wagenmakers, E.J., 2014. *Bayesian Cognitive Modeling: A Practical Course*. Cambridge University Press, Cambridge.
- Levi, D.M., 2008. Crowding — an essential bottleneck for object recognition: a mini-review. *Vis. Res.* 48 (2008), 635–654. doi:10.1016/j.visres.2007.12.009.
- Levi, D.M., Carney, T., 2009. Crowding in peripheral vision: why bigger is not always better [Abstract]. *J. Vis.* 9 (8), 982. doi:10.1167/9.8.982, 982ahttp://journalofvision.org/9/8/982/.
- Litvak, V., Garrido, M., Zeidman, P., Friston, K.J., 2015. Empirical Bayes for group (DCM) studies: a reproducibility study. *Front. Hum. Neurosci.* 9, 1–12. doi:10.3389/fnhum.2015.00670.
- Livne, T., Sagi, D., 2007. Configuration influence on crowding. *J. Vis.* 7 (2), 1–12. doi:10.1167/7.2.4.
- Livne, T., Sagi, D., 2010. How do flankers' relations affect crowding? *J. Vis.* 10 (3), 1–14. doi:10.1167/10.3.1.
- Lohmann, G., Erfurth, K., Müller, K., Turner, R., 2012. Critical comments on dynamic causal modelling. *Neuroimage* 59 (3), 2322–2329. doi:10.1016/j.neuroimage.2011.09.025.
- 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), 1–11. doi:10.1167/7.2.24.
- Love, J., et al., 2019. JASP: graphical statistical software for common statistical designs. *J. Stat. Softw.* 88, 1–17.
- Malania, M., Herzog, M.H., Westheimer, G., 2007. Grouping of contextual elements that affect vernier thresholds. *J. Vis.* 7 (2), 1–7. doi:10.1167/7.2.1.
- Manassi, M., Sayim, B., Herzog, M.H., 2012. Grouping, pooling, and when bigger is better in visual crowding. *J. Vis.* 12 (10), 1–14. doi:10.1167/12.10.13.Introduction.
- Manassi, M., Sayim, B., Herzog, M.H., 2013. When crowding of crowding leads to uncrowding. *J. Vis.* 13 (10), 1–10. doi:10.1167/13.10.10.doi.
- Manassi, M., Hermens, F., Francis, G., Herzog, M.H., 2015. Release of crowding by pattern completion. *J. Vis.* 15 (8), 1–15. doi:10.1167/15.8.16.doi.
- Manassi, M., Lonchamps, S., Clarke, A., Herzog, M.H., 2016. What crowding can tell us about object representations. *J. Vis.* 16, 1–13. doi:10.1167/16.3.35.doi.
- McIntosh, A.R., Gonzalez-Lima, F., 1994. Structural equation modeling and its application to network analysis in functional brain imaging. *Hum. Brain Mapp.* 2 (1–2), 2–22. doi:10.1002/hbm.460020104.
- Millin, R., Arman, A.C., Chung, S.T., Tjan, B.S., 2014. Visual crowding in V1. *Cereb. Cortex* 24 (12), 3107–3115. doi:10.1093/cercor/bht159.
- Mumford, D., 1992. On the computational architecture of the neocortex. II. The role of cortico-cortical loops. *Biol. Cybern.* 66 (3), 241–251. doi:10.1007/BF00198477.
- Oberfeld, D., Stahn, P., Kuta, M., 2014. Why do forward maskers affect auditory intensity discrimination? Evidence from "molecular psychophysics". *PLoS One* 9 (6), e99745. doi:10.1371/journal.pone.0099745.
- Overvliet, K.E., Sayim, B., 2016. Perceptual grouping determines haptic contextual modulation. *Vis. Res.* 126, 52–58. doi:10.1016/j.visres.2015.04.016.
- Pelli, D.G., Palomares, M., Majaj, N., 2004. Crowding is unlike ordinary masking: distinguishing feature integration from detection. *J. Vis.* 4 (12), 1136–1169. doi:10.1167/14.12.12, 12.
- Pelli, D.G., Tillman, K.A., 2008. The uncrowded window of object recognition. *Nat. Neurosci.* 11 (10), 1129–1135. doi:10.1038/nn.2187.
- Penny, W., Mattout, J., Trujillo-Barreto, N., 2006. *Bayesian Model Selection and Averaging. Statistical Parametric Mapping: The Analysis of Functional Brain Images*. Elsevier, London.
- Pöder, E., 2006. Crowding, feature integration, and two kinds of "attention". *J. Vis.* 6 (2), 163–169. doi:10.1167/6.2.7.
- Rao, R.P., Ballard, D.H., 1999. Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nat. Neurosci.* 2 (1), 79–87. doi:10.1038/4580.
- Roebroeck, A., Formisano, E., Goebel, R., 2011a. The identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *Neuroimage* 58 (2), 296–302. doi:10.1016/j.neuroimage.2009.09.036.
- Roebroeck, A., Formisano, E., Goebel, R., 2011b. After comments on: the identification of interacting networks in the brain using fMRI: model selection, causality and deconvolution. *Neuroimage* 58 (2), 310–311. doi:10.1016/j.neuroimage.2009.10.077.
- Rossion, B., Pourtois, G., 2004. Revisiting Snodgrass and Vanderwart's object pictorial set: the role of surface detail in basic-level object recognition. *Perception* 33 (2), 217–236. doi:10.1068/p5117.
- Ratcliff, R., McKoon, G., 2008. Drift diffusion decision model: theory and data for two-choice decision tasks. *Neural Comput.* 20 (4), 873–922. doi:10.1162/neco.2008.12.06.420.
- Robol, V., Casco, C., Dakin, S.C., 2012. The role of crowding in contextual influences on contour integration. *J. Vis.* 12 (7), 1–18. doi:10.1167/12.7.3.
- Rosenholtz, R., Yu, D., Keshvari, S., 2019. Challenges to pooling models of crowding: implications for visual mechanisms. *J. Vis.* 19 (7), 1–25. doi:10.1167/19.7.15.
- Saarela, T.P., Sayim, B., Westheimer, G., Herzog, M.H., 2009. Global stimulus configuration modulates crowding. *J. Vis.* 9 (2), 1–11. doi:10.1167/9.2.5.
- Schwarzkopf, D.S., de Haas, B., Alvarez, I., 2019. SamSrf 6 - Toolbox for pRF Modelling doi:10.17605/OSF.IO/2RGSM.
- Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Van Der Linde, A., 2002. Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B (Stat. Methodol.)* 64, 583–639. doi:10.1111/1467-9868.00353.
- Srinivasan, M.V., Laughlin, S.B., Dubs, A., 1982. Predictive coding: a fresh view of inhibition in the retina. *Proc. R. Soc. Lond., B Biol. Sci.* 216 (1205), 427–459. doi:10.1098/rspb.1982.0085.
- Strasburger, H., Harvey Jr, L.O., Rentschler, I., 1991. Contrast thresholds for identification of numeric characters in direct and eccentric view. *Percept. Psychophys.* 49 (6), 495–508. doi:10.3758/bf03212183.
- Strasburger, H., 2005. Unfocused spatial attention underlies the crowding effect in indirect form vision. *J. Vis.* 5 (11), 1024–1037. doi:10.1167/5.11.8.
- Valdes-Sosa, P.A., Roebroeck, A., Daunizeau, J., Friston, K.J., 2011. Effective connectivity: influence, causality and biophysical modeling. *Neuroimage* 58 (2), 339–361. doi:10.1016/j.neuroimage.2011.03.058.
- Whitney, D., Levi, D.M., 2011. Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends Cognit. Sci. (Regul. Ed.)* 15 (4), 160–168. doi:10.1016/j.tics.2011.02.005.
- Wiecki, T.V., Sofer, I., Frank, M.J., 2013. HDDM: hierarchical Bayesian estimation of the drift-diffusion model in Python. *Front. Neuroinform.* 7, 1–10. doi:10.3389/fninf.2013.00014.
- Wilson, H.R., Cowan, J.D., 1973. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetik* 13 (2), 55–80. doi:10.1007/BF00288786.
- Wolford, G., Chambers, L., 1983. Lateral masking as a function of spacing. *Percept. Psychophys.* 33 (2), 129–138. doi:10.3758/BF03202830.
- Yeotikar, N.S., Khuu, S.K., Asper, L.J., Suttle, C.M., 2011. Configuration specificity of crowding in peripheral vision. *Vis. Res.* 51 (11), 1239–1248. doi:10.1016/j.visres.2011.03.016.
- Zhaoping, L., 2003. V1 mechanisms and some figure-ground and border effects. *J. Physiol. Paris* 97 (4–6), 503–515. doi:10.1016/j.jphysparis.2004.01.008.