

Contextual-Dependent Attention Effect on Crowded Orientation Signals in Human Visual Cortex

 Nihong Chen,¹ Pinglei Bao,² and Bosco S. Tjan^{1,3}

¹Department of Psychology, University of Southern California, Los Angeles, California 90089, ²Division of Biology and Biological Engineering, Computation and Neural Systems, Caltech, Pasadena, California 91125, and ³Neuroscience Graduate Program, University of Southern California, Los Angeles, California 90089

A target becomes hard to identify with nearby visual stimuli. This phenomenon, known as crowding, places a fundamental limit on conscious perception and object recognition. To understand the neural representation of crowded stimuli, we used fMRI and a forward encoding model to reconstruct the target-specific feature from multivoxel activation patterns evoked by orientation patches. Orientation-selective response profiles were constructed in V1–V4 for a target embedded in different contexts. Subjects of both sexes either directed their attention over all the orientation patches or selectively to the target. In the context with a weak crowding effect, attending to the target enhanced the orientation selectivity of the response profile; such effect increased along the visual pathway. In the context with a strong crowding effect, attending to the target enhanced the orientation selectivity of the response profile in the earlier visual area, but not in V4. The increase and decrease of orientation selectivity along the visual hierarchy demonstrate a contextual-dependent attention effect on crowded orientation signals: in the context with a weak crowding effect, selective attention gradually resolves the target from nearby distractors along the hierarchy; in the context with a strong crowding effect, while selective attention maintains the target feature in the earlier visual area, its effect decreases in the downstream area. Our findings reveal how the human visual system represents the target-specific feature at multiple stages under the limit of attention selection in a cluttered scene.

Key words: attention; consciousness; crowding; fMRI; object recognition; visual cortex

Significance Statement

Using fMRI and a forward encoding model, we reconstructed orientation-selective response profiles for a target embedded in crowded contexts. In the context with a weak crowding effect, attention gradually resolves the target from nearby distractors along the visual hierarchy. In the context with a strong crowding effect, while the feature of the target is preserved in the early visual cortex, it degrades in the later visual processing stage. The increase and decrease of orientation selectivity along the visual hierarchy reveal how the human visual system strikes to present the target-specific feature under the limit of attention selection in a cluttered scene.

Introduction

Objects rarely appear in isolation in a natural scene. A single glance at the world contains a rich amount of information. Given the limited capacity of the visual system, the processing of an item can be biased among multiple stimuli by two kinds of signal. One is the top-down signal, driven by selective attention (e.g., infor-

mation processing can be facilitated at the attended spatial location) (Moran and Desimone, 1985; Luck et al., 1997; Treue and Martínez Trujillo, 1999; Boynton, 2011). The other is the bottom-up signal, driven by the visual input (e.g., a salient stimulus can be easily detected among the distractors) (Reynolds and Desimone, 2003; Beck and Kastner, 2005).

Even with both top-down and bottom-up processes, some information may not be accessible. It can be hard to identify a single tree when it is surrounded by forest. Visual crowding, a breakdown of target identification when surrounding objects are within a critical distance of the target, places a fundamental limit on conscious perception and object recognition (Whitney and Levi, 2011). Although the neural mechanisms of crowding have been extensively discussed in psychophysical and computational studies (Parkes et al., 2001; Freeman and Simoncelli, 2011; Nandy and Tjan, 2012; Manassi and Whitney, 2018), little is known about the cortical representation of crowded objects: when the

Received March 27, 2018; revised Aug. 3, 2018; accepted Aug. 7, 2018.

Author contributions: N.C. wrote the first draft of the paper; N.C. and P.B. edited the paper; N.C. and B.S.T. designed research; N.C. and B.S.T. performed research; N.C., P.B., and B.S.T. analyzed data; N.C. wrote the paper.

This work was supported by National Institutes of Health R01-EY017707. This paper is dedicated to Prof. Bosco Tjan, who passed away due to a tragic incident on December 2, 2016. We thank Fang Fang, Christopher Baker, Zhong-Lin Lu, and Xiu Yang for helpful discussions and feedback on early versions of the manuscript.

The authors declare no competing financial interests.

Correspondence should be addressed to Dr. Nihong Chen, Department of Psychology, University of Southern California, Los Angeles, CA 90089. E-mail: nihongch@usc.edu.

DOI:10.1523/JNEUROSCI.0805-18.2018

Copyright © 2018 the authors 0270-6474/18/388433-08\$15.00/0

feature of a target cannot be consciously perceived, is it still represented in the visual cortex?

Unlike previous studies, which identified crowding-related changes in the overall BOLD signals (Fang and He, 2008; Bi et al., 2009; Joo et al., 2012; Chen et al., 2014; Millin et al., 2014), we used a forward encoding model in fMRI multivoxel pattern analysis (Brouwer and Heeger, 2011; Saproo and Serences, 2014; Sprague et al., 2015). The reconstructed response profiles at a population level demonstrated feature-selective information of a specific item in a crowded display. Thus, a previously unexplored role of selective attention in resolving the target from nearby distractors can be tested in the human visual cortex.

Our visual stimuli were orientation patches consisting of a target at 2.5° horizontal eccentricity and eight surrounding flankers (see Fig. 1A). During the scan, subjects were asked to attend all the patches, or to attend only the target. We asked how attention tuned the cortical representation to reflect the orientation of the crowded target, with flankers of identical orientation, or flankers of different orientations.

Materials and Methods

Subjects. A total of 12 subjects (5 female, 23–38 years old) were enrolled in the experiment. Eight participated in the first experiment, and eight participated in the second experiment. All subjects had normal or corrected-to-normal vision. They gave written, informed consent in accordance with the procedures and protocols approved by the review committee of University of Southern California.

Stimuli and apparatus. In an array, a Gabor at an eccentricity of 2.5° to the right of fixation was surrounded by eight flanking Gabors (0.75° from the central Gabor). The diameter of each Gabor was 0.625°. All Gabors (spatial frequency = 8 c/deg, Michelson contrast = 100%) were presented on a gray background with their mean luminance (70.15 cd/m²). In the first fMRI experiment, the orientation of all flankers was horizontal. In the second fMRI experiment, the orientation of each flanker was randomly assigned from 0°, 45°, 90°, and 135°. In the psychophysical measurement, the central Gabor was also presented without the flankers.

In the psychophysical measurement, the stimuli were presented on a Trinitron Multiscan G400 22-in monitor (Sony; refresh rate: 85 Hz; spatial resolution: 1024 × 768). Subjects viewed the stimuli from a distance of 80 cm. Their head position was stabilized using a head and chin rest. In the fMRI measurement, the stimuli were back-projected via a video projector (refresh rate: 60 Hz; spatial resolution: 1024 × 768) onto a translucent screen placed inside the scanner bore. Subjects viewed the stimuli through a mirror located above their eyes. The viewing distance was 83 cm.

Behavioral task. Before the fMRI scan, we measured subjects' identification performance on the central Gabor. In a trial, the Gabor array was presented for 120 ms. Subjects were asked to maintain the central fixation and to make a four-alternative-forced-choice judgment of the target orientation (0°, 45°, 90°, or 135°) with a button press. The next trial began after 1000 ms after response. Identification accuracies were measured with and without flankers. For each condition, the performance was averaged over 40 trials. Crowding effect was defined as the difference in identification accuracy between the target-plus-flanker condition and the target-only condition.

Procedures. In each experiment, we measured BOLD signals responding to the Gabor array in block-design fMRI scans over two or three sessions. In a stimulus block, all Gabors in the array were counterphase flickering at 2 Hz with random pauses. Meanwhile, the contrast of the central Gabor decreases at random time points. Subjects maintained fixation while attended to the Gabor stimuli throughout the run. In the Attend_Target condition, subjects detected a reduction in the contrast of the central Gabor in the array. In the Attend_All condition, subjects detected a pause of flickering of the whole Gabor array. The task order was counterbalanced across subjects. For each attention condition, 8 runs were measured. In the first experiment, each run consisted of 12 blocks, with 3 blocks for each target orientation (0°, 45°, 90°, and 135°).

In the second experiment, each run consisted of 16 blocks, with 4 blocks for each target orientation. Each stimulus block lasted for 12 s and was interleaved with 12 s fixation blocks.

MRI data acquisition and preprocessing. MRI data were collected using a 3T Prisma scanner with a 32-channel phase-array coil (Siemens). BOLD signals were measured at a resolution of 3 × 3 × 3 mm³ with a multiband EPI sequence (TE: 35 ms; TR: 1 s; FOV: 192 × 192 mm²; matrix: 64 × 64; flip angle: 63 degrees; slice thickness: 3 mm; gap: 0 mm; number of slices: 42; slice orientation: axial). A high-resolution 3D structural dataset (MPRAGE; 0.8 × 0.8 × 0.8 mm³ resolution) was collected. MRI data analyses were performed using Freesurfer (version 5.3, <http://surfer.nmr.mgh.harvard.edu/>) and FSL (version 4.1, FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl/). The anatomical volume was processed using Freesurfer to reconstruct the inflated cortical surface for each subject. The functional volumes were preprocessed using FSL, including motion-correction and high-pass temporal filtering. All functional datasets were individually registered into 3D space using the subjects' individual high-resolution anatomical images.

Mapping ROIs. Retinotopic mapping of visual areas V1–V4 was performed using standard phase-encoded methods (Serenio et al., 1995; Engel et al., 1997), in which subjects viewed a rotating wedge and an expanding ring that created traveling waves of neural activity in visual cortex. Two independent localizer runs, which were identical to the runs in the main experiment, were used to identify voxels in each area that showed a stronger response to stimulus conditions than fixation ($p < 0.01$).

Multivoxel pattern analysis. A forward encoding model (Brouwer and Heeger, 2011; Saproo and Serences, 2014; Sprague et al., 2015) was used to project the multivoxel response patterns in each ROI to a set of orientation-selective channel responses. The model assumes that BOLD responses from voxels reflect an approximately linear mixture of responses from many subpopulations of neurons with different degrees of selectivity to orientation (Boynton et al., 1996; Logothetis and Wandell, 2004).

For each voxel, β values evoked by the array with four different orientations of the central Gabor were estimated in individual blocks with a GLM procedure. The β values of each voxel were interpreted in the forward encoding model as a linear sum of weighted responses of 12 orientation-selective channels, with the orientation selectivity of the channels linearly spaced between 0° and 180° (0°, 15°, 30°, . . . , 165°). The basis tuning function of each hypothetical channel was modeled using a half-sinusoidal function raised to the fifth power. Raising to the fifth power made the tuning curves narrower: the half-bandwidth at half-maximum height of the basis tuning in the present study is 30°, which is comparable with the physiological findings. Given the considerable amount of variability in the orientation tuning bandwidth from single-unit studies (e.g., Schiller et al., 1976; McAdams and Maunsell, 1999; Ringach et al., 2002), the exponent of the sinusoids used in the forward encoding model has been chosen at 5 (Brouwer and Heeger, 2011; Anderson et al., 2012) or 6 (Scolari et al., 2012). With the half-sinusoidal function raised to the fifth power, at least seven basis functions were needed to cover the orientation space. This minimum number has been generally exceeded in the previous studies to compute orientation-selective channel response (e.g., 9–10 hypothetical channels) (Scolari et al., 2012; Garcia et al., 2013). We used 12 channels to fully cover the orientation space and to include the orientations at 0°, 45°, 90°, and 135°.

The computation of forward encoding model consisted of two stages. The first stage was to estimate the weights on the 12 hypothetical channels separately for each voxel. With these weights, the second stage was to compute the channel outputs associated with the spatial pattern of BOLD signal evoked by the array with different center orientation. Runs from each subject were divided into a training set and a test set using a leave-two-run-out cross-validation scheme. For example, the 24 spatial patterns of voxel response for each target orientation were divided into a training set (18 patterns) and a test set (6 patterns). The training set was used to estimate the channel weights in the first stage, and the test set was used to compute the channel responses in the second stage.

In the first stage, the weight of each channel was estimated using a standard GLM (Eq. 1). Let k be the number of channels, m be the number



Figure 1. Stimuli and behavioral results. **A**, Visual stimuli. A target Gabor (2.5° horizontal eccentricity) is surrounded by eight flanking Gabors. Left, Homogeneous context consisting of flankers with identical orientation at 0°. Right, Heterogeneous context consisting of flankers with mixed orientations at 0°, 45°, 90°, and 135°. **B**, Crowding effect in homogeneous and heterogeneous contexts. Subjects were asked to maintain central fixation and to make a four-alternative-forced-choice judgment of the target orientation (0°, 45°, 90°, or 135°). Crowding effect was defined as the reduction in orientation identification accuracy in percentage induced by the presence of flankers. Error bar indicates 1 SEM across subjects.

of voxels, and n be the number of repeated measurements (i.e., 4 orientation \times 18 patterns in the training set). The matrix of voxel responses in the training set (B_1 , $m \times n$) was related to the matrix of hypothetical channel response (C_1 , $k \times n$) by a weight matrix (W , $m \times k$) as follows:

$$B_1 = WC_1 \quad (1)$$

The ordinary least-squares estimate of W is computed via a pseudoinverse of C_1 (+ indicates pseudoinverse) as follows:

$$\hat{W} = B_1 C_1^+ \quad (2)$$

In the second stage, the pattern of voxel response in the test set B_2 was used to compute the estimated channel responses C_2 using the previously computed weights W .

$$C_2 = (\hat{W}^T \hat{W})^{-1} \hat{W}^T B_2 \quad (3)$$

The training/testing procedure was repeated for all combinations of the patterns. Finally, the channel response profile computed each time was circularly shifted such that the orientation of the target that evoked the response profile was set to 0° offset in the abscissa of Figures 2A and 4A.

It should be noted that a pseudoinverse was used in Equation 2 because C_1 is not a full-rank matrix (the number of the measured orientation conditions was smaller than the number of the channels). In a control analysis, we computed the channel response with 4 hypothetical channels. The orientation selectivity of the channels was linearly spaced between 0° and 180° (0°, 45°, 90°, 135°). In this case, Equation 2 can be written as $\hat{W} = B_1 C_1^T (C_1 C_1^T)^{-1}$. For response profiles computed based on 4 hypothetical channels, see Fig. 2-2 (available at <https://doi.org/10.1523/JNEUROSCI.0805-18.2018.f2-2>) and Fig. 4-2 (available at <https://doi.org/10.1523/JNEUROSCI.0805-18.2018.f4-2>).

Weighted pooling model. We assumed that the channel responses derived above reflect a linear combination of responses from individual orientations. To evaluate the information of the target and the flankers separately, we estimated the weights for the target and the flankers in a linear regression model.

$$T = w_t T_t + w_f \sum T_f \quad (4)$$

The model receives the following inputs: the channel response (T), which was the result of the forward encoding model; the hypothetical channel response (with the same basis function as defined in the forward encoding model) evoked by the physical stimuli. The latter part consisted of the channel response to the target orientation (T_t) and the channel response to the flanking orientation (T_f). Because the target was surrounded by eight flankers, $\sum T_f$ was the sum of eight channel response profiles, each evoked by a flanking orientation. Given these inputs, we estimated the weight of the target (w_t) and the weight of the flankers (w_f). The model was fitted across four target orientations, under each attention condition and each stimulus context.

Experimental design and statistical analysis. To test the change in the channel responses across the visual areas, the centralized channel response profiles were submitted to repeated-measures ANOVA with ROI and channel as two within-subject factors. Because the basis functions were not independent, the assumption of independence when comparing the channel responses across conditions was violated. To evaluate the significance of the reported effects, we first obtained F values from the ANOVA tests. Next, we randomly permuted orientation labels to construct the response profiles. Then we conducted the same statistical analysis on this relabeled dataset. We repeated this procedure for 2000 times to generate a distribution of F values and looked up the probability of obtaining the original F value given this distribution. The p values reported reflect this probability. Differences between experimental conditions were also assessed using one-sample t test and paired-sample t test. In the pooling model, a simple linear regression was calculated. The fitting of the regression equation was assessed by the p value and the R^2 . The weight differences between conditions were assessed using a bootstrap procedure (Fox, 2016). The model statistics, including unstandardized coefficients, standardized coefficients, and p values, were reported.

Results

Behavioral results

Before scanning, subjects were tested on orientation identification at the target with and without surrounding flankers. In the homogeneous context, the crowding effect was weak; the presence of the flankers had little effect on the orientation identification accuracy (3% reduction, $t_{(7)} = 2.31$, $p = 0.06$). In the heterogeneous context, the crowding effect was strong; the presence of the flankers significantly reduced the orientation identification accuracy (20% reduction, $t_{(7)} = 4.25$, $p = 0.004$) (Fig. 1B).

fMRI results

Increased orientation selectivity in a homogeneous context

Using a forward encoding model, we computed channel responses from multivoxel activation patterns evoked by the orientation patches in V1-V4. The sharpness of the tuning reflects the orientation selectivity of the target orientation (Fig. 2A). In the homogeneous context, we found that the response profiles became sharpened along the visual pathway with selective attention; a repeated-measures ANOVA revealed a significant interaction between ROI and channel ($F_{(33,231)} = 2.10$, $p = 0.04$), as well as a significant main effect of channel ($F_{(11,77)} = 8.00$, $p < 0.001$). Without selective attention, the main effect of channel was significant ($F_{(11,77)} = 7.32$, $p = 0.003$), but the interaction was not ($F_{(33,231)} = 1.13$, $p = 0.34$). Similar patterns were observed in the

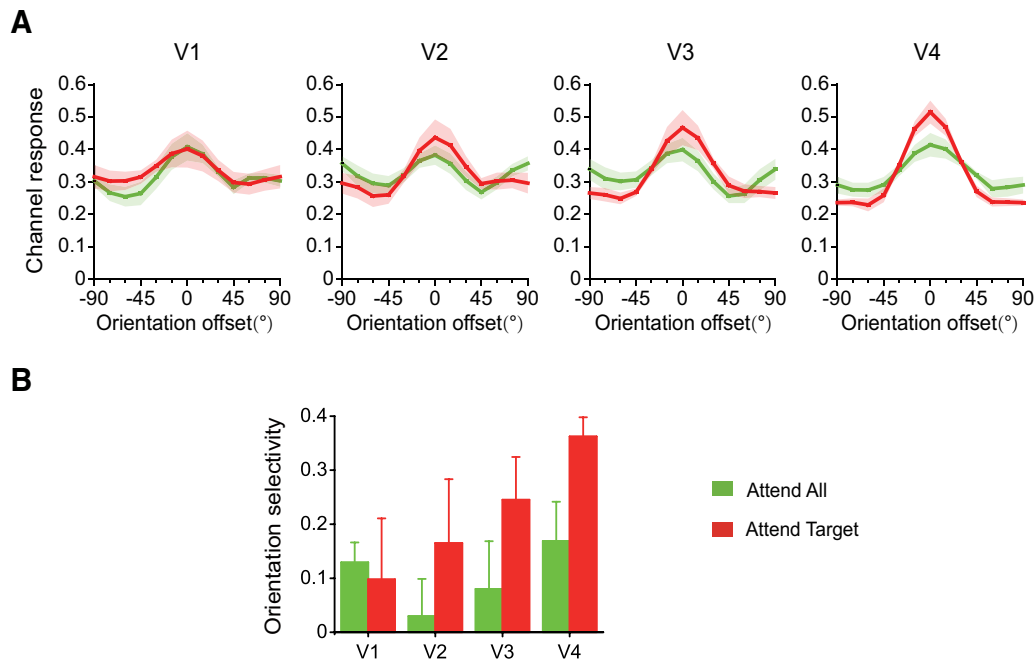


Figure 2. Orientation-selective response profiles in the homogeneous context. **A**, Responses in orientation-selective channels with respect to the target orientation. The response profiles were computed for each target orientation and were circularly shifted such that the orientation of the target was set to 0° offset. For amplitudes and bandwidths of the response profiles, see Figure 2-1 (available at <https://doi.org/10.1523/JNEUROSCI.0805-18.2018.f2-1>). **B**, OSI derived from the response profile, by comparing the responses at 0° and 90° offsets. $OSI = (R_{0^\circ} - R_{90^\circ}) / (R_{0^\circ} + R_{90^\circ})$. Shaded area/error bar represents 1 SEM across subjects. For response profiles computed with four hypothetical channels, see Figure 2-2 (available at <https://doi.org/10.1523/JNEUROSCI.0805-18.2018.f2-2>).

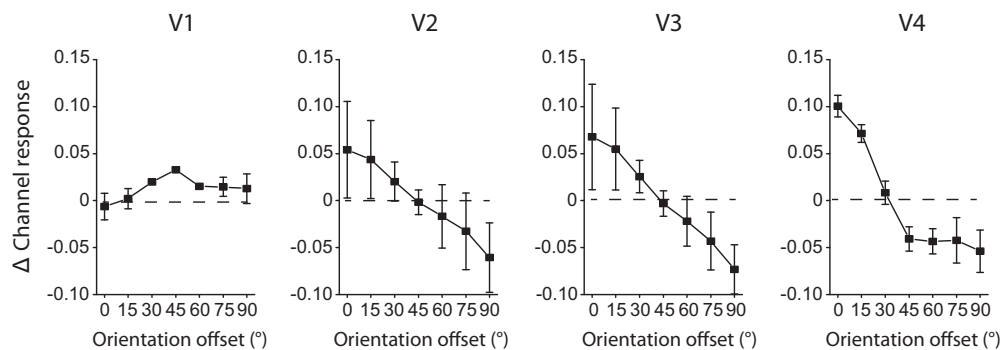


Figure 3. Attention effect on the channel response in the homogeneous context. The attention-related changes in the response profiles in Figure 2A were quantified by subtracting the channel responses in the Attend_All condition from those in the Attend_Target condition. Data were collapsed across channels with the same magnitude of orientation offset. Error bar indicates 1 SEM across subjects.

response profiles constructed with 4 hypothetical channels (Fig. 2-2, available at <https://doi.org/10.1523/JNEUROSCI.0805-18.2018.f2-2>): with selective attention, there was a significant interaction between ROI and channel ($F_{(9,63)} = 2.19, p = 0.03$); without selective attention, the interaction was not significant ($F_{(9,63)} = 0.57, p = 0.81$).

The selectivity of the response profile was quantified using an orientation selectivity index (OSI) (Fig. 2B). With selective attention, the OSI increased from V1 to V4. Along with the sharpened orientation-selective response profile, we found that selective attention enhanced the orientation selectivity in V4 ($t_{(7)} = 4.36, p = 0.003$). The enhancement in the orientation selectivity was also reflected as an increase in the amplitude and a decrease in the bandwidth of the response profiles (Fig. 2-1, available at <https://doi.org/10.1523/JNEUROSCI.0805-18.2018.f2-1>). These results suggest that selective attention gradually tuned the neural representation toward the target orientation along the visual pathway.

To clearly demonstrate the attention effect across orientation channels, we calculated the difference in the channel responses between the two attention conditions in each visual area (Fig. 3). Data were collapsed across channels with the same magnitude of orientation offset based on the symmetric nature of the response profiles. A positive channel difference reflects a larger response with selective attention, and a negative channel difference reflects a smaller response with selective attention. We found that selective attention enhanced the responses in the channels at or near the target orientation, and suppressed the responses in the channels farther from the target orientation in the extrastriate cortex. The magnitude of the enhancement/suppression became largest in V4. These changes underlie the increased orientation selectivity along the visual hierarchy.

Decreased orientation selectivity in a heterogeneous context

A different pattern of response profiles was observed in the heterogeneous context (Fig. 4A). The tuning profiles, which were

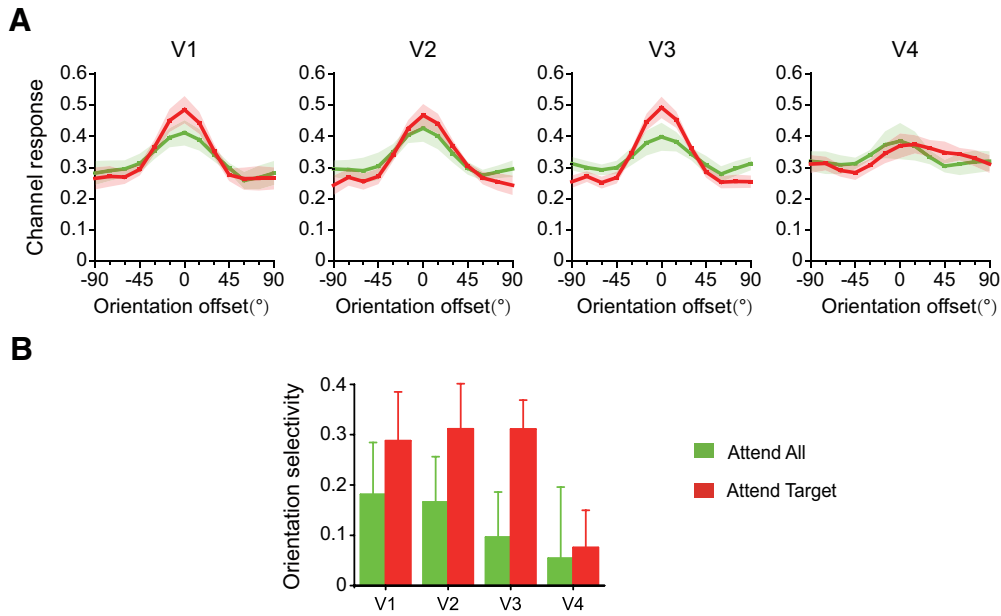


Figure 4. Orientation-selective response profiles in the heterogeneous context. **A**, Responses in orientation-selective channels with respect to the target orientation. For amplitudes and bandwidths of the response profiles, see Figure 4-1 (available at <https://doi.org/10.1523/JNEUROSCI.0805-18.2018.f4-1>). **B**, OSI derived from the response profile, by comparing the responses at 0° and 90° offsets. $OSI = (R_{0^\circ} - R_{90^\circ}) / (R_{0^\circ} + R_{90^\circ})$. Shaded area/error bar represents 1 SEM across subjects. For response profiles computed based on four hypothetical channels, see Figure 4-2 (available at <https://doi.org/10.1523/JNEUROSCI.0805-18.2018.f4-2>).

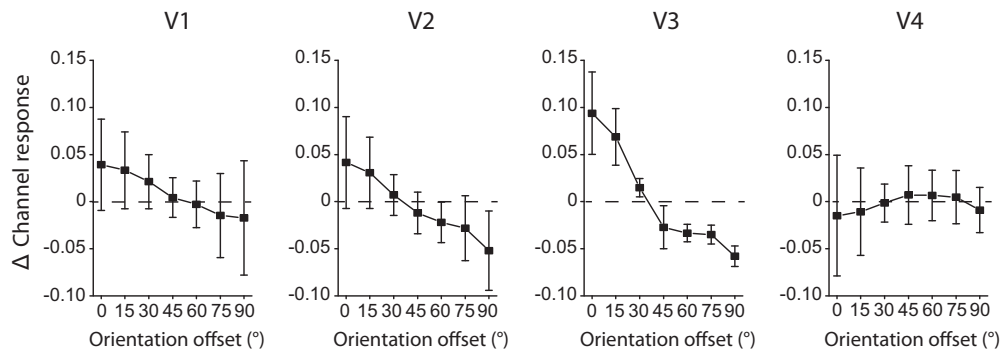


Figure 5. Attention effect on the channel response in the heterogeneous context. The attention-related changes in the response profiles in Figure 4A were quantified by subtracting the channel responses in the Attend_All condition from those in the Attend_Target condition. Data were collapsed across channels with the same magnitude of orientation offset. Error bar indicates 1 SEM across subjects.

sharpened by selective attention in the early visual areas, became less tuned in V4. With selective attention, a repeated-measures ANOVA showed a significant interaction between ROI and channel ($F_{(33,231)} = 3.28, p = 0.006$), as well as a significant main effect of channel ($F_{(11,77)} = 9.54, p < 0.001$). Without selective attention, the main effect of channel was significant ($F_{(11,77)} = 3.43, p = 0.01$), but the interaction was not ($F_{(33,231)} = 0.28, p = 0.98$). Similar patterns were observed in the response profiles constructed with 4 hypothetical channels (Fig. 4-2, available at <https://doi.org/10.1523/JNEUROSCI.0805-18.2018.f4-2>): with selective attention, there was a significant interaction between ROI and channel ($F_{(9,63)} = 2.11, p = 0.04$); without selective attention, the interaction was not significant ($F_{(9,63)} = 0.27, p = 0.98$).

The selectivity of the response profile was quantified using the OSI (Fig. 4B). We found that attention enhanced the orientation selectivity in V3 ($t_{(7)} = 3.78, p = 0.007$), but not in V4 ($t_{(7)} = 0.17, p = 0.87$). The changes of orientation selectivity were also reflected in the amplitudes and the bandwidths of response profiles (Fig. 4-1, available at [18.2018.f4-1\). These results suggest that, in a heterogeneous context, crowding limits the role of selective attention in resolving the target from the flankers, resulting in an impoverished neural representation in V4.](https://doi.org/10.1523/JNEUROSCI.0805-</p>
</div>
<div data-bbox=)

To demonstrate the attention effect across orientation channels, we calculated the difference in the channel responses between the two attention conditions (Fig. 5). With selective attention, there was a trend of enhancement in the channel responses at or near the target orientation and a trend of suppression in the channel responses farther from the target orientation. Such pattern was observed only in the earlier visual areas, but not in V4. These changes underlie the decreased orientation selectivity in the downstream area.

The channel response differences between the two attention conditions in both the homogeneous and the heterogeneous contexts were submitted to an ANOVA with stimulus context and channel as two factors. A significant interaction was observed in V4 ($F_{(6,42)} = 2.51, p = 0.036$). There was no significant interaction in V1-V3 (all $F_{(6,42)} < 0.36, p > 0.90$). Similar results were

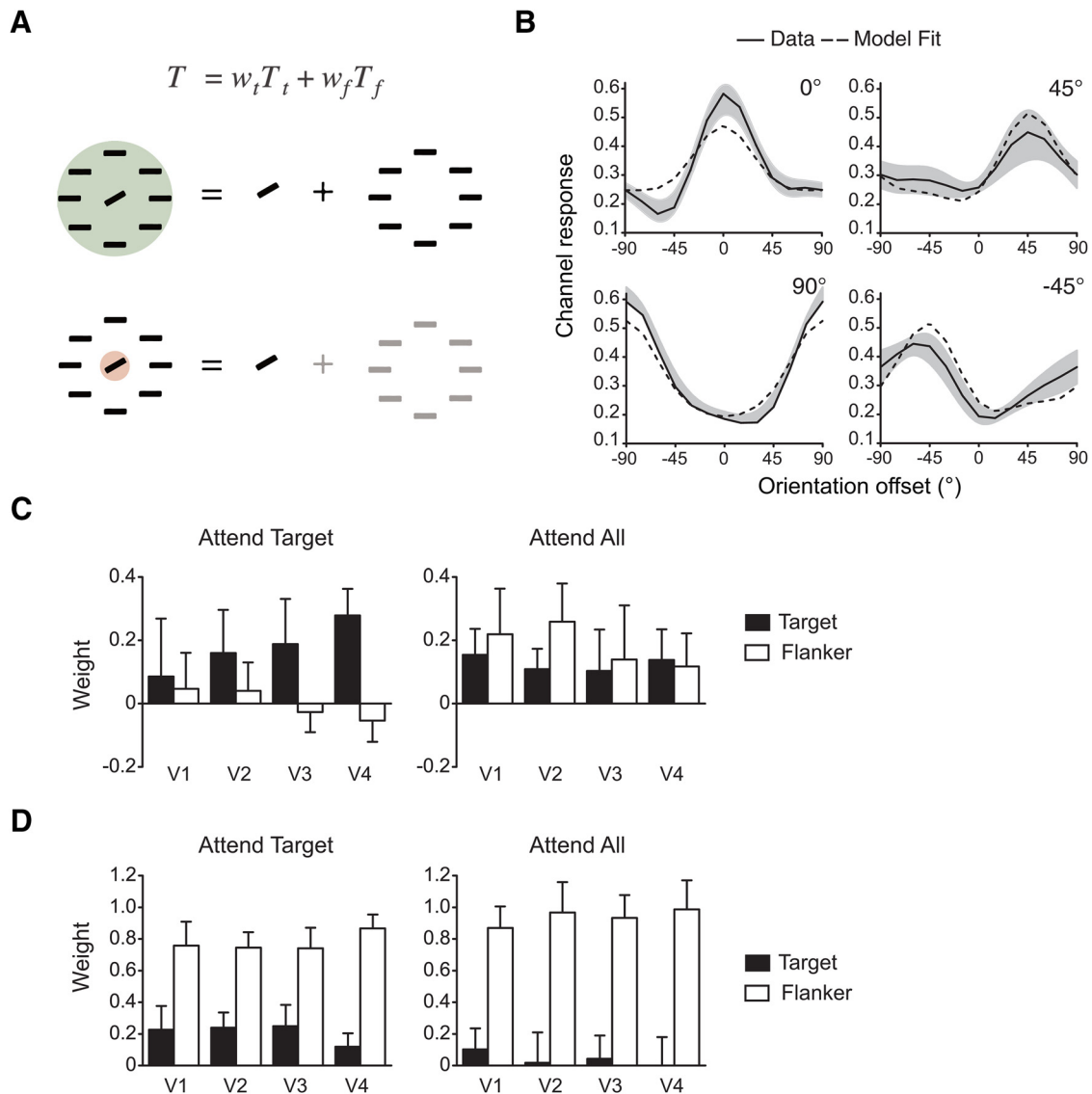


Figure 6. Pooling orientation signals in homogeneous and heterogeneous contexts. **A**, Weighted pooling model. Response profiles evoked by each orientation were summed, given a weight of the target (w_t) and a weight of the flankers (w_f). The model predicts that selective attention increases the weight of the target and decreases the weight of the flankers. **B**, An example of model fit in the Attend_Target condition in the homogeneous context. Solid line and shaded area indicate mean \pm SEM across subjects. Dashed line indicates model fit from 1000 bootstrap samples. **C, D**, Weight estimates in the homogeneous (**C**) and the heterogeneous (**D**) context. For a summary of model statistics, see Figure 6-1 (available at <https://doi.org/10.1523/JNEUROSCI.0805-18.2018.f6-1>) and Figure 6-2 (available at <https://doi.org/10.1523/JNEUROSCI.0805-18.2018.f6-2>). Error bar indicates 95% CI from 1000 bootstrap samples.

found in the analysis with 4 hypothetical channels (V1-V3: all $F_{(2,14)} < 0.44$, $p > 0.65$; V4: $F_{(2,14)} = 7.49$, $p = 0.006$). These results suggest that attention modulates the channel responses in V4 differently between the stimulus contexts.

A pooling model

The above channel responses reflect the neural representation of the orientation ensemble. To understand how crowding arises from individual orientations, we assumed that the visual system generates estimates of the target feature based on a weighted sum of the signals from the target and the flankers. In a pooling model (Fig. 6A), we used hypothetical channel responses evoked by the target and the flankers to predict the channel responses in Figures 2A and 4A. The weight estimates reflect the amount of information of the target and the flankers in each visual area.

In the homogeneous context, the model well explained the response profiles across V1-V4 under both attention conditions (both $R^2 > 0.45$, $p < 0.001$; for ROI-wise results, see Fig. 6-1,

available at <https://doi.org/10.1523/JNEUROSCI.0805-18.2018.f6-1>). Without selective attention, the weights of the target and the flankers were significant in all the ROIs (W_t : all $p < 0.001$; W_f : all $p < 0.001$), suggesting a voluntary pooling of information from both the target and the flankers. With selective attention, we observed increasing weight of the target and decreasing weight of the flankers along the visual pathway (Fig. 6C). The weights of the target were significant in all the ROIs (all $p < 0.001$). The weights of the flankers were not significant in V1-V3 (all $p > 0.05$) and significantly negative in V4 ($p = 0.01$). These results indicate that selective attention enables a selective pooling of the target information. A comparison between two attention conditions further demonstrates that selective attention enhanced the weight of the target ($p = 0.01$, *bootstrap*) and reduced the weight of the flankers ($p = 0.002$, *bootstrap*) in V4.

In the heterogeneous context, the model well explained the response profiles across V1-V4 with selective attention ($R^2 =$

0.41, $p < 0.001$; for ROI-wise results, see Fig. 6-2, available at <https://doi.org/10.1523/JNEUROSCI.0805-18.2018.f6-2>). Along with the decrease in the orientation selectivity, the weight of the target became smaller in V4 than those in V2 ($p = 0.02$, *bootstrap*) and V3 ($p = 0.01$, *bootstrap*) (Fig. 6D). However, without selective attention, the model generated poor fit in the extrastriate areas (V1: $R^2 = 0.21$, $p = 0.005$; V2-V4, all $R^2 < 0.06$, $p > 0.23$), suggesting a lack of automatic linear pooling process under the crowding context.

Discussion

In the context with a weak crowding effect, selective attention increased the orientation selectivity for the target from V1 to V4. It is known that spatial attention biases the visual processing toward the behaviorally relevant location, allowing for enhanced perception of the stimuli at the attended location (Posner, 1980). Many electrophysiological studies have shown that a neuron's response is predominantly determined by the attended stimulus when multiple stimuli fall within its receptive field in macaque extrastriate cortex (Moran and Desimone, 1985; Luck et al., 1997; Reynolds et al., 1999). These results form the basis of the biased-competition theory of attention. By comparing the BOLD responses to sequentially presented stimuli and simultaneously presented stimuli, Kastner et al. (1998) identified the suppressive interaction among nearby stimuli and further demonstrated that spatially directed attention reduces such interaction in the human visual cortex. In line with this finding, we observed an accumulative effect of spatial selective attention along the visual processing hierarchy. The effect of attention in biasing the neural representation toward the target is manifested in multivoxel activations at a population level, as previously reported in the domain of feature-based attention (Reddy et al., 2009; Serences et al., 2009). Our findings bridge the gap between single-neuron recording studies and fMRI multivoxel studies (Boynton, 2011), which constitute important cortical evidence of the role of selective spatial attention in resolving the target from nearby distractors.

In the context with a strong crowding effect, the effect of selective attention became limited. For an item that fails to be recognized in a clutter, there has been a long-standing debate about where the information begins to degrade in the cortex. One hypothesis proposes that crowding arises from lateral/horizontal connections starting in V1 (Pelli and Tillman, 2008; Nandy and Tjan, 2012). Alternatively, it has been suggested that the information is maintained in the early visual cortex and could degrade in later visual processing stages due to a lack of attention resolution (He et al., 1996; Intriligator and Cavanagh, 2001). While crowding-related activation changes have been identified across areas V1–V4 (Fang and He, 2008; Anderson et al., 2012; Chen et al., 2014; Kwon et al., 2014; Millin et al., 2014), these changes were derived by comparing the overall BOLD responses between the target/flanker-only condition and the target-plus-flanker condition. Thus, the neural representation of feature-specific information remains unavailable. Anderson et al. (2012) used fMRI adaptation to reveal the neural correlates of altered orientation perception induced by crowding. A strongest crowding effect was found in V4. This result was consistent with the degradation of orientation selectivity in V4 observed in the current study. These findings suggest that crowding exerts a larger influence in the later visual processing stage, allowing the low-level visual information to persist in the early visual cortex.

Although the orientation of an individual item may not be consciously perceived, it could be pooled by the visual system to generate estimates for the ensemble. A pooling mechanism pro-

poses that the information from the target and flankers in a cluttered display are summarized rather than being lost (Parkes et al., 2001; Freeman and Simoncelli, 2011; Harrison and Bex, 2015). In a linear weighted model, we tested this idea and evaluated the information of the target and surround separately. In the context with a weak crowding effect, the increasing weights of the target and the decreasing weights of the flankers along the visual pathway confirmed the role of selective attention in target facilitation and distractor suppression. In the context with a strong crowding effect, although this model well accounted for the degradation in the target-specific representation with selective attention, it generated poor fit to the neural representation without selective attention. These results indicate that a voluntary linear pooling may not account for crowding under all conditions. Instead, crowding could result from other processes (e.g., feature substitution) in which participants occasionally mistake the target with a distractor (Ester et al., 2014, 2015).

The increase and decrease in the orientation selectivity along the visual pathway reveal an interaction between top-down attention-driven and bottom-up stimulus-driven processes. A recent model proposes that attention reads out the sensory representation in the intrinsic visual circuits via a sparse pooling process (Chaney et al., 2014). This model explains the contextual effect of crowding in the light of grouping (Kooi et al., 1994; Livne and Sagi, 2010). In the present study, the identical flankers can be grouped together, so that a sparse sampling is sufficient to resolve the target from the flankers. However, the heterogeneous flankers cannot be grouped, so that a sparse sampling results in an impoverished neural representation. The sparse selection process could take place at a relatively late stage in the visual processing hierarchy, which is consistent with our findings that the information is carried along before it degraded in V4. The preserved information in the early visual processing stages explains crowding at multiple levels (crowding occurs not only among low-level features; e.g., orientation adaption effect persists with crowding) (He et al., 1996), but also among high-level objects (e.g., holistic object information gets through crowding and influences ensemble perception (Fischer and Whitney, 2011)). Together, these findings illustrate how the human visual system represents information at multiple stages under the limit of attention selection in a cluttered scene.

References

- Anderson EJ, Dakin SC, Schwarzkopf DS, Rees G, Greenwood JA (2012) The neural correlates of crowding-induced changes in appearance. *Curr Biol* 22:1199–1206. [CrossRef Medline](#)
- Beck DM, Kastner S (2005) Stimulus context modulates competition in human extrastriate cortex. *Nat Neurosci* 8:1110–1116. [CrossRef Medline](#)
- Bi T, Cai P, Zhou T, Fang F (2009) The effect of crowding on orientation-selective adaptation in human early visual cortex. *J Vis* 9:13. 1–10. [CrossRef Medline](#)
- Boynton GM (2011) Spikes, BOLD, attention, and awareness: a comparison of electrophysiological and fMRI signals in V1. *J Vis* 11:12. [CrossRef Medline](#)
- Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. *J Neurosci* 16:4207–4221. [CrossRef Medline](#)
- Brouwer GJ, Heeger DJ (2011) Cross-orientation suppression in human visual cortex. *J Neurophysiol* 106:2108–2119. [CrossRef Medline](#)
- Chaney W, Fischer J, Whitney D (2014) The hierarchical sparse selection model of visual crowding. *Front Integr Neurosci* 8:73. [CrossRef Medline](#)
- Chen J, He Y, Zhu Z, Zhou T, Peng Y, Zhang X, Fang F (2014) Attention-dependent early cortical suppression contributes to crowding. *J Neurosci* 34:10465–10474. [CrossRef Medline](#)
- Engel SA, Glover GH, Wandell BA (1997) Retinotopic organization in hu-

- man visual cortex and the spatial precision of functional MRI. *Cereb Cortex* 7:181–192. [CrossRef Medline](#)
- Ester EF, Klee D, Awh E (2014) Visual crowding cannot be wholly explained by feature pooling. *J Exp Psychol Hum Percept Perform* 40:1022–1033. [CrossRef Medline](#)
- Ester EF, Zilber E, Serences JT (2015) Substitution and pooling in visual crowding induced by similar and dissimilar distractors. *J Vis* 15:4. [CrossRef Medline](#)
- Fang F, He S (2008) Crowding alters the spatial distribution of attention modulation in human primary visual cortex. *J Vis* 8:6. 1–9. [CrossRef Medline](#)
- Fischer J, Whitney D (2011) Object-level visual information gets through the bottleneck of crowding. *J Neurophysiol* 106:1389–1398. [CrossRef Medline](#)
- Fox J (2016) *Applied regression analysis and generalized linear models*, pp 647–688. Thousand Oaks, CA: Sage.
- Freeman J, Simoncelli EP (2011) Metamers of the ventral stream. *Nat Neurosci* 14:1195–1201. [CrossRef Medline](#)
- Garcia JO, Srinivasan R, Serences JT (2013) Near-real-time feature-selective modulations in human cortex. *Curr Biol* 23:515–522. [CrossRef Medline](#)
- Harrison WJ, Bex PJ (2015) A unifying model of orientation crowding in peripheral vision. *Curr Biol* 25:3213–3219. [CrossRef Medline](#)
- He S, Cavanagh P, Intriligator J (1996) Attentional resolution and the locus of visual awareness. *Nature* 383:334–337. [CrossRef Medline](#)
- Intriligator J, Cavanagh P (2001) The spatial resolution of visual attention. *Cogn Psychol* 43:171–216. [CrossRef Medline](#)
- Joo SJ, Boynton GM, Murray SO (2012) Long-range, pattern-dependent contextual effects in early human visual cortex. *Curr Biol* 22:781–786. [CrossRef Medline](#)
- Kastner S, De Weerd P, Desimone R, Ungerleider LG (1998) Mechanisms of directed attention in the human extrastriate cortex as revealed by functional MRI. *Science* 282:108–111. [CrossRef Medline](#)
- Kooi FL, Toet A, Tripathy SP, Levi DM (1994) The effect of similarity and duration on spatial interaction in peripheral vision. *Spat Vis* 8:255–279. [CrossRef Medline](#)
- Kwon M, Bao P, Millin R, Tjan BS (2014) Radial-tangential anisotropy of crowding in the early visual areas. *J Neurophysiol* 112:2413–2422. [CrossRef Medline](#)
- Livne T, Sagi D (2010) How do flankers' relations affect crowding? *J Vis* 10:1. 1–14. [CrossRef Medline](#)
- Logothetis NK, Wandell BA (2004) Interpreting the BOLD signal. *Annu Rev Physiol* 66:735–769. [CrossRef Medline](#)
- Luck SJ, Chelazzi L, Hillyard SA, Desimone R (1997) Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J Neurophysiol* 77:24–42. [CrossRef Medline](#)
- Manassi M, Whitney D (2018) Multi-level crowding and the paradox of object recognition in clutter. *Curr Biol* 28:R127–R133. [CrossRef Medline](#)
- McAdams CJ, Maunsell JH (1999) Effects of attention on orientation-tuning functions of single neurons in macaque cortical area V4. *J Neurosci* 19:431–441. [CrossRef Medline](#)
- Millin R, Arman AC, Chung ST, Tjan BS (2014) Visual crowding in V1. *Cereb Cortex* 24:3107–3115. [CrossRef Medline](#)
- Moran J, Desimone R (1985) Selective attention gates visual processing in the extrastriate cortex. *Science* 229:782–784. [CrossRef Medline](#)
- Nandy AS, Tjan BS (2012) Saccade-confounded image statistics explain visual crowding. *Nat Neurosci* 15:463–469, S1–S2. [CrossRef Medline](#)
- Parkes L, Lund J, Angelucci A, Solomon JA, Morgan M (2001) Compulsory averaging of crowded orientation signals in human vision. *Nat Neurosci* 4:739–744. [CrossRef Medline](#)
- Pelli DG, Tillman KA (2008) The uncrowded window of object recognition. *Nat Neurosci* 11:1129–1135. [CrossRef Medline](#)
- Posner MI (1980) Orienting of attention. *Q J Exp Psychol A* 32:3–25. [CrossRef Medline](#)
- Reddy L, Kanwisher NG, VanRullen R (2009) Attention and biased competition in multi-voxel object representations. *Proc Natl Acad Sci U S A* 106:21447–21452. [CrossRef Medline](#)
- Reynolds JH, Desimone R (2003) Interacting roles of attention and visual salience in V4. *Neuron* 37:853–863. [CrossRef Medline](#)
- Reynolds JH, Chelazzi L, Desimone R (1999) Competitive mechanisms subserve attention in macaque areas V2 and V4. *J Neurosci* 19:1736–1753. [CrossRef Medline](#)
- Ringach DL, Shapley RM, Hawken MJ (2002) Orientation selectivity in macaque V1: diversity and laminar independence. *J Neurosci* 22:5639–5651. [CrossRef Medline](#)
- Saproo S, Serences JT (2014) Attention improves transfer of motion information between V1 and MT. *J Neurosci* 34:3586–3596. [CrossRef Medline](#)
- Schiller PH, Finlay BL, Volman SF (1976) Quantitative studies of single-cell properties in monkey striate cortex: II. Orientation specificity and ocular dominance. *J Neurophysiol* 39:1320–1333. [CrossRef Medline](#)
- Scolari M, Byers A, Serences JT (2012) Optimal deployment of attentional gain during fine discriminations. *J Neurosci* 32:7723–7733. [CrossRef Medline](#)
- Serences JT, Saproo S, Scolari M, Ho T, Muftuler LT (2009) Estimating the influence of attention on population codes in human visual cortex using voxel-based tuning functions. *Neuroimage* 44:223–231. [CrossRef Medline](#)
- Sereno MI, Dale AM, Reppas JB, Kwong KK, Belliveau JW, Brady TJ, Rosen BR, Tootell RB (1995) Borders of multiple visual areas in humans revealed by functional magnetic resonance imaging. *Science* 268:889–893. [CrossRef Medline](#)
- Sprague TC, Saproo S, Serences JT (2015) Visual attention mitigates information loss in small- and large-scale neural codes. *Trends Cogn Sci* 19:215–226. [CrossRef Medline](#)
- Treue S, Martínez Trujillo JC (1999) Feature-based attention influences motion processing gain in macaque visual cortex. *Nature* 399:575–579. [CrossRef Medline](#)
- Whitney D, Levi DM (2011) Visual crowding: a fundamental limit on conscious perception and object recognition. *Trends Cogn Sci* 15:160–168. [CrossRef Medline](#)