

# Attention Priority Map of Face Images in Human Early Visual Cortex

Ce Mo,<sup>1,2,3,4\*</sup>  Dongjun He,<sup>5\*</sup> and Fang Fang<sup>1,2,3,4</sup>

<sup>1</sup>School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, <sup>2</sup>Peking-Tsinghua Center for Life Sciences, <sup>3</sup>Key Laboratory of Machine Perception (Ministry of Education), <sup>4</sup>PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China, and <sup>5</sup>Sichuan Research Center of Applied Psychology, Chengdu Medical College, Chengdu 610500, China

Attention priority maps are topographic representations that are used for attention selection and guidance of task-related behavior during visual processing. Previous studies have identified attention priority maps of simple artificial stimuli in multiple cortical and subcortical areas, but investigating neural correlates of priority maps of natural stimuli is complicated by the complexity of their spatial structure and the difficulty of behaviorally characterizing their priority map. To overcome these challenges, we reconstructed the topographic representations of upright/inverted face images from fMRI BOLD signals in human early visual areas primary visual cortex (V1) and the extrastriate cortex (V2 and V3) based on a voxelwise population receptive field model. We characterized the priority map behaviorally as the first saccadic eye movement pattern when subjects performed a face-matching task relative to the condition in which subjects performed a phase-scrambled face-matching task. We found that the differential first saccadic eye movement pattern between upright/inverted and scrambled faces could be predicted from the reconstructed topographic representations in V1–V3 in humans of either sex. The coupling between the reconstructed representation and the eye movement pattern increased from V1 to V2/3 for the upright faces, whereas no such effect was found for the inverted faces. Moreover, face inversion modulated the coupling in V2/3, but not in V1. Our findings provide new evidence for priority maps of natural stimuli in early visual areas and extend traditional attention priority map theories by revealing another critical factor that affects priority maps in extrastriate cortex in addition to physical salience and task goal relevance: image configuration.

**Key words:** attention priority map; early visual cortex; face images; first saccadic eye movements

## Significance Statement

Prominent theories of attention posit that attention sampling of visual information is mediated by a series of interacting topographic representations of visual space known as attention priority maps. Until now, neural evidence of attention priority maps has been limited to studies involving simple artificial stimuli and much remains unknown about the neural correlates of priority maps of natural stimuli. Here, we show that attention priority maps of face stimuli could be found in primary visual cortex (V1) and the extrastriate cortex (V2 and V3). Moreover, representations in extrastriate visual areas are strongly modulated by image configuration. These findings extend our understanding of attention priority maps significantly by showing that they are modulated, not only by physical salience and task–goal relevance, but also by the configuration of stimuli images.

## Introduction

In everyday life, our visual system is faced with the critical challenge of selecting the most relevant fraction of visual inputs at the

expense of less relevant information. According to prominent attention priority map theories (Fecteau and Munoz, 2006; Serences and Yantis, 2006; Baluch and Itti, 2011), attention selection is implemented via attention priority maps that signal which part of the visual input should be granted prioritized access and guide the ensuing task-related behavior. Previous studies have identified priority maps in multiple brain regions throughout the visual processing hierarchy, including the frontal eye field (Serences and Yantis, 2007), precentral sulcus (Jerde et al., 2012), lateral

Received May 3, 2017; revised Oct. 28, 2017; accepted Nov. 6, 2017.

Author contributions: C.M., D.H., and F.F. designed research; C.M. and D.H. performed research; C.M., D.H., and F.F. analyzed data; C.M. and F.F. wrote the paper.

This work was supported by the National Natural Science Foundation of China (NSFC 31230029, NSFC 31421003, NSFC 61621136008, NSFC 61527804, NSFC 31671168) and the Ministry of Science and Technology of China (MOST 2015CB351800).

The authors declare no competing financial interests.

\*C.M. and D.H. contributed equally to the work.

Correspondence should be addressed to either Ce Mo or Fang Fang, School of Psychological and Cognitive Science, Peking University, Beijing, P. R. China 100871. E-mail: moce@pku.edu.cn or ffang@pku.edu.cn.

DOI:10.1523/JNEUROSCI.1206-17.2017

Copyright © 2018 the authors 0270-6474/18/380149-09\$15.00/0

intraparietal cortex (Gottlieb et al., 1998; Bisley and Goldberg, 2003, 2010), and V4 (Mazer and Gallant, 2003). More recently, seminal findings by Sprague and Serences (2013) showed that priority maps could be found in early retinotopic areas outside of the frontoparietal regions, including primary visual cortex (V1). However, little is known about the attention priority representation of natural stimuli because previous studies usually used artificial stimuli composed of simple features. Although several pioneering studies have shown that visual search in real-world scenes is achieved by matching incoming visual input to a top-down category-based attentional “template,” an internal object representation with target-diagnostic features (Peelen et al., 2009; Peelen and Kastner, 2011, 2014; Seidl et al., 2012), so far, there is no neural evidence of a topographic profile of attention priority distribution over natural stimuli.

The fundamental theme of identifying neural correlates of attention priority map is to examine the link between the topographic neural representation of visual stimuli and task-related behavior that reflects the spatial pattern of attention priority (i.e., behavioral relevance). However, this is complicated in the case of natural stimuli. First, natural stimuli are highly complex and investigating their topographic representation in the visual cortex is therefore challenging, especially with human brain imaging techniques. Second, it is difficult to characterize the priority map of natural images behaviorally using psychophysical measurements (e.g., contrast sensitivity). Further complicating the matters is that visual processing of natural stimuli is often influenced by image configuration. A well known example is the face inversion effect: face recognition performance is severely impaired by the inversion of the image (Yin, 1969; Rhodes and Tremewan, 1994). As a result, identifying the attention priority representation of natural stimuli remains a critical challenge because no studies have examined the behavioral relevance of topographic representations of natural stimuli while simultaneously taking the influence of image configuration into consideration.

Here, we combined the use of eye tracking and fMRI to address these issues. Face images were chosen as experimental stimuli because the spatial configuration of face components (i.e., eyes, mouth, nose, etc.) is highly consistent across individual faces and the impact of inverted image configuration is more pronounced in faces than other objects (Yin, 1969), which allows effective reconstruction of their topographic neural representation and easy manipulation of their image configuration. We characterized the priority map of faces behaviorally as the differential spatial distribution of the first saccadic targets between intact and phase-scrambled face images during a one-back image-matching task. First saccade after stimulus onset is thought to be a relatively pure signature of attentional guidance when processing complex stimuli (Awh et al., 2006; Einhäuser et al., 2008; Jiang et al., 2014). To reconstruct the topographic representation of face images, we used the voxelwise population receptive field (pRF)-mapping technique (Dumoulin and Wandell, 2008). This technique allows us to identify the corresponding retinotopic location of each voxel in a given visual cortical area and thus enable the reconstruction of the topographic stimulus representation from population activities in the reference frame of subjects' visual field of view (Kok and de Lange, 2014). To examine the behavioral relevance of the reconstructed representation to the priority map, we measured their correspondence using precision-recall curves (Davis and Goadrich, 2006).

## Materials and Methods

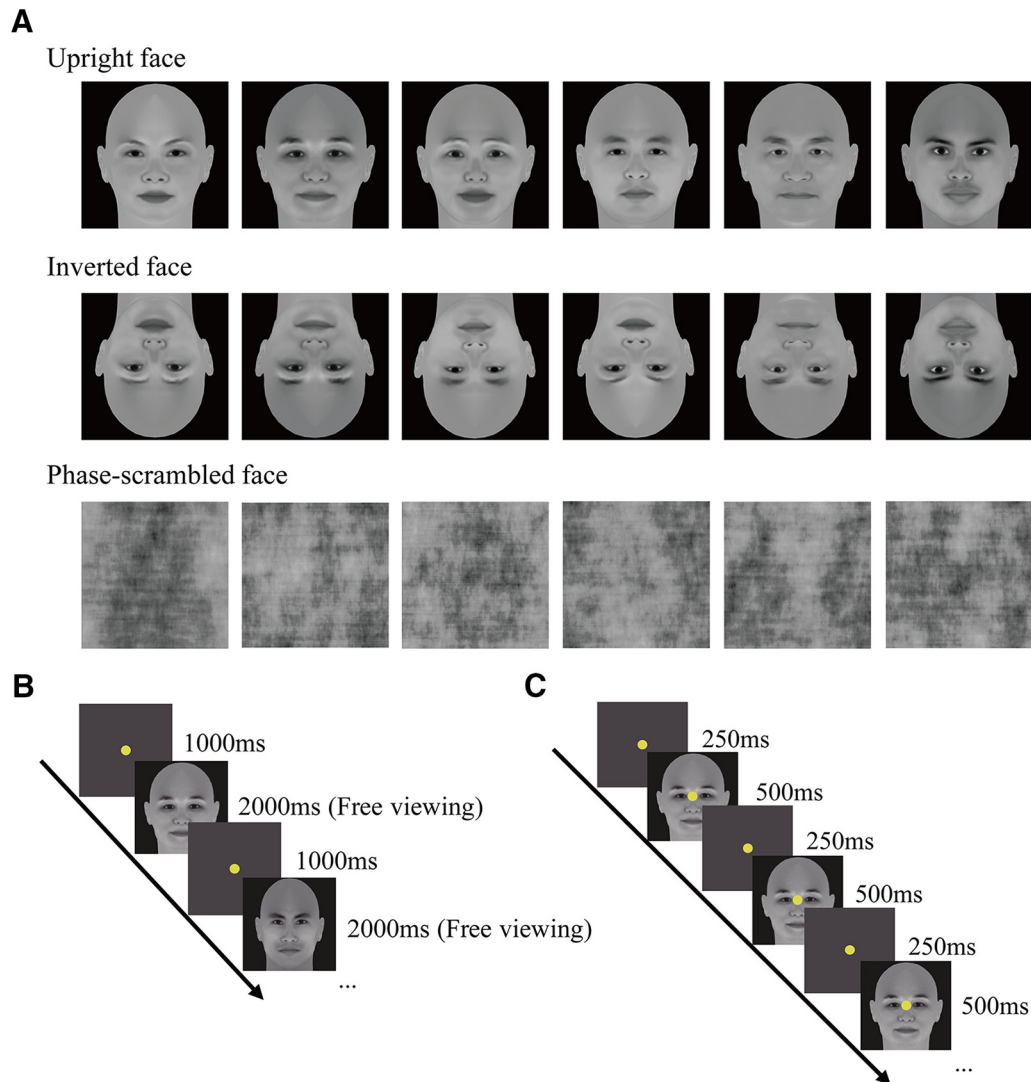
**Participants.** A total of 10 human subjects (5 male, 18–28 years old) were paid to take part in the study. All of them participated in both the eye-tracking and fMRI experiments. All subjects were naive to the purpose of the study. They were right-handed, reported normal or corrected-to-normal vision, and had no known neurological or visual disorders. Written informed consent was collected before the experiments. Experimental procedures were approved by the Human Subject Review Committee at Peking University.

**Stimuli.** Three types of visual stimuli were used in this study, including upright faces, their inverted versions, and phase-scrambled versions (Fig. 1A). The upright face images (45 in total) were generated using FaceGen Modeler (Singular Inversions), subtended 16.6° in visual angle both vertically and horizontally, and were matched for mean luminance (15 cd/m<sup>2</sup>) and root mean square contrast. It is noteworthy that the locations of critical face components (e.g., eyes, mouth, and nose) were almost identical across the faces. The phase spectra of the upright face images were scrambled randomly to generate their phase-scrambled versions. The phase-scrambled images were included as a baseline to exclude potential visual field and/or eccentricity biases during eye-tracking and fMRI measurements (see below).

**Eye-tracking experiment.** The eye-tracking experiment was performed in a behavioral test room. The visual stimuli were presented on an IYAMA HM204DT 22-inch monitor with a spatial resolution of 1024 × 768 and a refresh rate of 85 Hz. Subjects viewed the stimuli from a distance of 70 cm with their head stabilized on a chin rest. Eye movement data were collected using an EyeLink 1000 Plus eye-tracking system. In each trial, subjects fixated at the central fixation point for 1 s and then viewed a stimulus image freely for 2 s. They performed a one-back image-matching task in which they were required to press a button in response to the stimulus that matched the stimulus in the last trial (Fig. 1B). The experiment consisted of six blocks. Each block had 75 trials: 25 trials for each of the three stimulus types. The trial order was randomized, but with a constraint that 15 stimulus images (five for each of the three stimulus types) were presented in two successive trials. For each subject, the spatial distribution of the first saccadic target for the phase-scrambled faces served as a baseline and was subtracted from those for the upright and the inverted faces. The subtraction step minimized any potential effects of subjects' intrinsic visual field and/or eccentricity bias that might be due to the initial central fixation point. Because such biases are not related to image content, they should occur commonly when viewing either the upright/inverted face images or the scrambled face images. Therefore, the first saccadic target pattern after the subtraction is presumably not influenced by these biases. For the sake of clarity, the pattern is referred to as “the differential first saccadic target pattern” because it is the difference between the two patterns from the upright/inverted faces and the scrambled faces. The results after subtraction were convolved with a 2D Gaussian function ( $\sigma = 0.17^\circ$ ) to remove high-frequency noise and to generate smooth patterns. These individual patterns were then averaged to obtain a group-level pattern.

**fMRI experiment.** The MRI experiments were performed on a 3 T Siemens Prisma MRI scanner at the Center for MRI Research at Peking University. MRI data were acquired with a 20-channel phase-array head coil. In the scanner, 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 mounted on the head coil. The viewing distance was 70 cm. BOLD signals were measured using an echoplanar imaging sequence (TE: 30 ms, TR: 2000 ms, flip angle: 90°, acquisition matrix size: 76 × 76, FOV: 152 × 152 mm<sup>2</sup>, slice thickness: 2 mm, gap: 0 mm, number of slices: 33, slice orientation: axial). fMRI slices covered the occipital lobe and a small part of the cerebellum. Before functional runs, a T1-weighted high-resolution 3D structural dataset was acquired for each participant in the same session using a 3D-MPRAGE sequence.

Subjects underwent two sessions, one for retinotopic mapping, ROI localization, and estimation of pRF model parameters and the other for measuring stimulus-evoked activities. In the first session, early retinotopic visual areas (V1 and the extrastriate cortex, V2 and V3) were de-



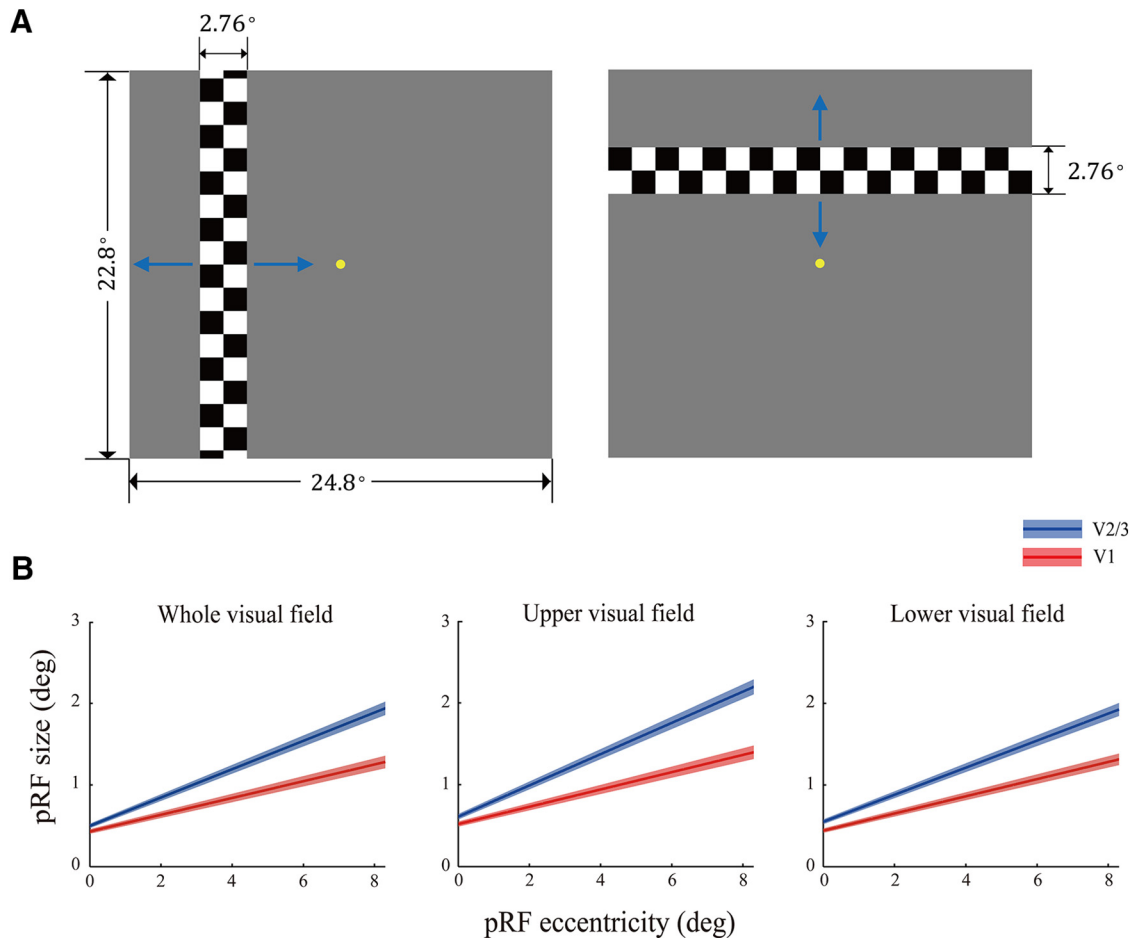
**Figure 1.** Stimuli and experimental protocol. **A**, Exemplar stimuli of three stimulus types. **B**, Protocol of the eye-tracking experiment. Subjects initiated a trial by fixating at the central fixation point. The end point of the first fixation change after stimulus onset was recorded as the first saccadic target. **C**, Protocol of the fMRI experiment. Subjects performed a one-back image-matching task with all three types of stimuli in different blocks of a run.

finned using a standard phase-encoded method (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. An independent block-design run was performed to identify ROIs in the retinotopic areas responding to the stimulus region when subjects fixated at the central fixation point. The run contained eight stimulus blocks of 12 s interleaved with eight blank blocks of 12 s. The stimulus was a full-contrast flickering checkerboard of the same size as the face images. Voxelwise pRF model parameters were estimated using the method described in Dumoulin and Wandell (2008). Specifically, the hemodynamic response function (HRF) was measured for each subject in a separate run containing 12 trials. In each trial, a full-contrast flickering checkerboard with a radius of  $10.94^\circ$  was presented for 2 s, followed by a 30 s blank interval. The HRF was estimated by fitting the convolution of a 6-parameter double-gamma function with a 2 s boxcar function to the BOLD response elicited by the disk. Three pRF mapping runs were performed in which a flickering full-contrast checkerboard swept through the entire visual field. The bar moved through two orientations (vertical and horizontal) in two opposite directions within a given run, giving a total of four different stimulus configurations. The order of the stimulus configurations was randomized. The mapped visual area subtended  $24.8^\circ$  horizontally and  $22.8^\circ$  vertically. The bar was  $2.76^\circ$  in width and its length was either  $24.8^\circ$  or  $22.8^\circ$  (Fig. 2A). Each bar swept through the visual area

in 16 steps within 51 s. The step size was  $1.38^\circ$ . Each pRF mapping run lasted for 204 s. Throughout the session, subjects performed a color discrimination task at fixation point to maintain fixation and control attention.

The second scanning session consisted of four block design runs. In each run, there were 12 stimulus blocks of 12 s (four blocks for each stimulus type) interleaved with 12 blank blocks of 12 s. In a stimulus block, 16 images appeared. Each image was presented for 500 ms, followed by a 250 ms blank interval. Subjects performed the same one-back-matching task as that in the eye-tracking experiment. Throughout the scanning session, subjects were required to fixate at the central fixation point and refrain from any possible eye movements.

fMRI data were processed using BrainVoyager QX (Brain Innovations) and custom scripts written in MATLAB (The MathWorks). The anatomical volume in the first session was transformed into the AC–PC space and then inflated using BrainVoyager QX. Functional volumes in both sessions were preprocessed, including 3D motion correction, linear trend removal, and high-pass filtering (cutoff frequency 0.015 Hz) using BrainVoyager QX. Subjects with excessive head movement ( $>1.2$  mm in translation or  $>0.5^\circ$  in rotation) within any fMRI session were excluded (2 of 10 subjects). These functional volumes were then aligned to the anatomical volume in the first scanning session and transformed into the AC–PC space. The first 6 s of BOLD signals were discarded to minimize



**Figure 2.** pRF mapping. **A**, Stimuli used for pRF mapping. Blue arrows indicate the moving directions of the checkerboard bar in pRF mapping runs. Each bar moved in one direction and switched to the opposite direction when it reached the boundary of the mapped area. **B**, Relationships between pRF size and pRF eccentricity in the whole (left), the upper (middle), and the lower (right) visual field, respectively. The solid lines denote the linear fits relating the pRF eccentricity and the pRF size across subjects. SEs were estimated using a bootstrapped method (Kay et al., 2013), as indicated by the shaded band around each line.

transient magnetic saturation effects. For each subject, a general linear model (GLM) procedure was used to define the functional ROIs and to measure the stimulus-evoked signal intensity for each stimulus type. The ROIs in V1–V3 were defined as the cortical areas that responded more strongly to the checkerboard than to the blank screen ( $p < 10^{-3}$ , uncorrected). Stimulus-specific BOLD signal intensities in each ROI (i.e.,  $\beta$  value) were estimated for individual voxels, subtracted by the mean  $\beta$  value across all the voxels in the ROI, and divided by the maximal absolute. After this normalization step, the  $\beta$  values of all the voxels in the ROI had a zero mean and a maximal absolute value of one. To facilitate the comparison between the primary visual cortex (V1) and the extrastriate cortex (V2 and V3), voxels in V2 and V3 were pooled together to equate their number with the voxel number in V1.

**pRF-based reconstruction.** Reconstruction of the neural representations of the face images in early visual cortex involved two stages. In the first stage, we estimated pRF model parameters for each voxel in all the ROIs using the coarse-to-fine search method described in Dumoulin and Wandell (2008). The predicted BOLD signal was calculated from the known visual stimulus parameters, the HRF, and a model of the joint receptive field of the underlying neuronal population. This model consisted of a 2D Gaussian function with parameters  $x_0$ ,  $y_0$ , and  $\sigma$ , where  $x_0$  and  $y_0$  are the coordinates of the center of the receptive field and  $\sigma$  is its spread (SD) or size. All parameters were stimulus referred and their units were degrees of visual angle. Model parameters were adjusted to obtain the best possible fit of the actual BOLD signal. Only the voxels for which the pRF model could explain at least 10% of the variance of the raw data were included for further analyses (Kok and de Lange, 2014). The pro-

portion of voxels retained after applying this threshold was high and was comparable between the V1 and the V2/3 ROIs (mean proportion  $\pm$  SEM V1:  $0.952 \pm 0.008$ , V2/3:  $0.942 \pm 0.015$ ).

In the second stage, parallel to the eye-tracking experiment, we used a linear regression method to estimate the contribution of the baseline effect (from the phase-scrambled images) to the BOLD signals evoked by the upright and the inverted faces based on the responses of all voxels within an ROI and then removed the contribution accordingly for individual voxels (Kok and de Lange, 2014). Because this regression method uses the data from all voxels in an ROI, compared with the subtraction method, it provides a more robust estimate against outlier voxels and thus improves the signal-to-noise ratio in the reconstructed representations. Specifically, the  $\beta$  values for the upright or the inverted faces were submitted as the dependent variable, whereas the  $\beta$  values for the phase-scrambled images were submitted as the independent variable of the model as follows:

$$\beta_{(i,j)} = r_i \beta_{(\text{Scrambled},j)} + C_i + w_{(i,j)}, i \in \{\text{Upright}, \text{Inverted}\}$$

where subscripts  $i$  and  $j$  refer to the stimulus type (upright or inverted face) and voxel, respectively;  $C$  and  $r$  are the constant term (intercept) and the regression coefficient (slope) of the model, respectively; and  $w$  is the reconstruction weight, which represented the neural activity associated with the upright or inverted faces after removing potential visual field and eccentricity biases. All matrices in the regression equation are  $n$ -by-1, where  $n$  is the number of voxels included in the reconstruction procedure in an ROI. Finally, the voxelwise pRF profiles were multiplied

by the reconstruction weights and summated. The reconstructed representation was therefore a linear sum of the 2D-Gaussian pRF profiles of all voxels weighted by their stimulus-specific BOLD response as follows:

$$R_i(x, y) = \sum_j w_{(i,j)} G_j(x, y | x_0, y_0, \sigma), i \in \{Upright, Inverted\}$$

where  $R_i(x, y)$  refers to the stimulus-specific representation intensity at the retinotopic location  $(x, y)$  and  $G_j(x, y | x_0, y_0, \sigma)$  refers to the estimated voxelwise pRF model jointly parameterized by the pRF center  $(x_0, y_0)$  and size  $\sigma$ . This reconstruction procedure was performed with each subject. Individual representations were subtracted by the mean pixel intensity and divided by the maximal absolute pixel value. They were then averaged to obtain a group-level representation.

**Statistical analysis of behavioral relevance of reconstructed representations.** If a reconstructed cortical representation is behaviorally relevant, then it should perform well at predicting task-related behavior: the regions of high intensity in the representation are more likely to become the target of the first saccade (i.e., the high-priority area). Therefore, measuring the behavioral relevance of a reconstructed representation is equivalent to measuring its ability of predicting the behaviorally measured high-priority areas. Because attention selects only a small portion of visual inputs for further processing at the expense of other less relevant (low-priority) information, high-priority areas are much smaller compared with low-priority areas by nature. We therefore used precision-recall curve to measure the prediction performance of the reconstructed representations in distinguishing the high-priority areas from the low-priority areas (Achanta et al., 2009; Perazzi et al., 2012). Precision-recall curve provides a more informative characterization of classification performance than traditional receiver operating characteristics (ROC) curve that presents an overly optimistic estimation in the context of skewed class distribution (Davis and Goadrich, 2006). In the first step, we defined the top 7.5% pixels in the differential first saccadic target pattern as the high-priority areas and the remaining 92.5% pixels as the low-priority areas. In the second step, the reconstructed representation was continuously thresholded from the lowest to the highest intensity level. For each thresholded representation, we calculated the number of true-positives (TPs), false-positives (FPs), and false-negatives (FNs) to measure its performance in correctly assigning image pixels to the high- and low-priority areas (Chen et al., 2013), as follows:

$$\begin{aligned} TP &= \text{Suprathreshold area} \cap \text{Low priority area} \\ FP &= \text{Suprathreshold area} \cap \text{Low priority area} \\ FN &= \text{Subthreshold area} \cap \text{High priority area} \end{aligned}$$

Where TP refers to the number of pixels of the high priority area that were correctly labeled and FP refers to the number of pixels of the low priority area that were labeled incorrectly. The recall and the precision rates were calculated as follows:

$$\begin{aligned} \text{Recall} &= \frac{TP}{TP + FN} \\ \text{Precision} &= \frac{TP}{TP + FP} \end{aligned}$$

In the final step, we plotted the precision-recall curves from all the recall-precision tuples. Similar to the characteristic of ROC curve, a larger area under the precision-recall curve (AUC) indicate a higher accuracy in distinguishing high- and low-priority areas and thus suggests higher consistency between the reconstructed representation and the differential first saccadic target pattern, which was used to quantify the behavioral relevance of the reconstructed representation.

Statistical significance of the behavioral relevance of the reconstructed representations was examined using permutation tests in which we tested whether the behavioral relevance (i.e., AUC) was significantly above chance level. In this analysis, the estimated pRF positions were shuffled randomly and we performed the same weighted linear summation as described above to obtain the chance-level representation and calculate its corresponding AUC value. This procedure was repeated 1000 times to

derive the null distribution of AUC values for each stimulus type.  $p$ -values were obtained for each reconstructed representation from the corresponding null distribution.

A nonparametric bootstrapping method was used to compare the behavioral relevance of the reconstructed representations under different conditions (Efron and Tibshirani, 1994). In this analysis, subjects were iteratively (1000 times) resampled with replacement in the bootstrapping procedure. Specifically, in each iteration, eight fMRI datasets (corresponding to the eight subjects) were drawn with replacement, so the probability of each subject's data being sampled is equal (Jiang et al., 2006). After resampling, we reconstructed the cortical representations based on the resampled data using the same procedure as described above. The distribution of the AUC difference between the upright and inverted faces was derived by performing the same precision-recall analysis for the two representations, respectively, and measuring their AUC difference in each iteration (Koehn, 2004). Statistical significance of the AUC difference was assessed by calculating the cumulative probability of the positive values from the corresponding distribution. Statistical comparisons between the representations in different cortical areas were conducted in a similar manner. For each condition, the distribution of the AUC difference between the V2/3 representation and the V1 representation was first derived using the same bootstrapping method. We assessed the statistical significance of the AUC difference by calculating the cumulative probability of the positive values from the corresponding distribution.

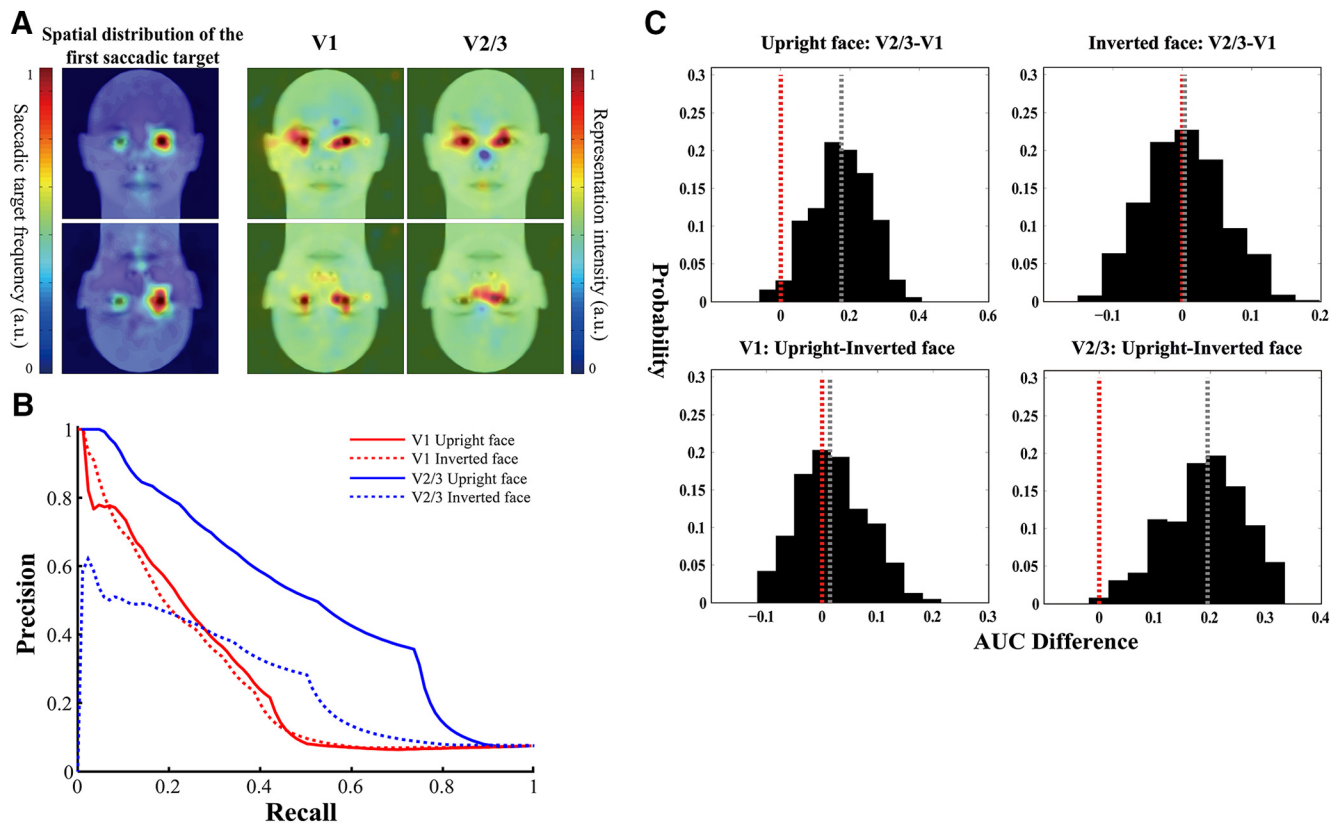
## Results

### Behavioral results

We recorded the location of the first saccadic target after stimulus presentation on a trial-by-trial basis when subjects were required to perform the image-matching task. On average, subjects achieved highly satisfactory performance for all the three stimulus types (mean accuracy  $\pm$  SEM upright face:  $0.9967 \pm 0.002$ , inverted face:  $0.9492 \pm 0.009$ , scrambled face:  $0.9808 \pm 0.005$ ), suggesting that their attention was effectively directed to the stimuli. Moreover, both parametric and nonparametric tests showed that subjects' performance differed among the three stimulus types (parametric one-way ANOVA:  $F_{(2,21)} = 18.93, p < 0.001$ ; nonparametric Kruskal–Wallis test:  $\chi^2_{(2)} = 17.07, p = 0.0002$ ), with significantly better performance achieved for the upright faces compared with the inverted faces ( $t_{(7)} = 5.65, p < 0.05$ , Wilcoxon signed-rank test:  $p < 0.05$ , Bonferroni corrected for multiple comparisons). This is consistent with the classical face inversion effect that recognition of upright faces is better than recognition of inverted faces. We then quantified the priority maps of the upright and the inverted faces by subtracting the distribution of the first saccadic target corresponding to the phase-scrambled face images from those corresponding to the upright and the inverted face images, respectively. At the group level, for both the upright and inverted faces, the differential first saccadic target pattern exhibited high intensity at the eye regions and the mouth region that are more informative in terms of face identity, with a preference for the left eye region of the face images (Fig. 3A, left column). This eyes–mouth triangular pattern is consistent with previous findings that selective sampling of visual information from the eye region is particularly important for recognition of face identity (Yarbus, 1967; Sekuler et al., 2004; Peterson and Eckstein, 2012).

### Attention priority representation of upright and inverted faces in early visual cortex

We reconstructed the topographic representation of the upright and inverted face stimuli in V1 and V2/3 by mapping the stimulus-specific activity patterns to visual space directly via the voxelwise pRF model. This model assumes that the joint receptive field of the neuronal population within a single voxel can be characterized as a 2D isotropic Gaussian function. By fitting the predicted signal based



**Figure 3.** Reconstructed topographic representations of face stimuli and their behavioral relevance. **A**, Visualization of the spatial patterns of the first saccadic target (left) and the reconstructed representations in V1 (middle) and V2/3 (right). **B**, Precision-recall curves corresponding to the reconstructed representations from which the behavioral relevance of these representations was measured as the AUC. **C**, Bootstrapped distributions of behavioral relevance (AUC) difference between the upright and inverted face representations and between the V1 and V2/3 representations. The red dotted line indicates the AUC difference of zero. The gray dotted line indicates the median of the bootstrapped distribution.

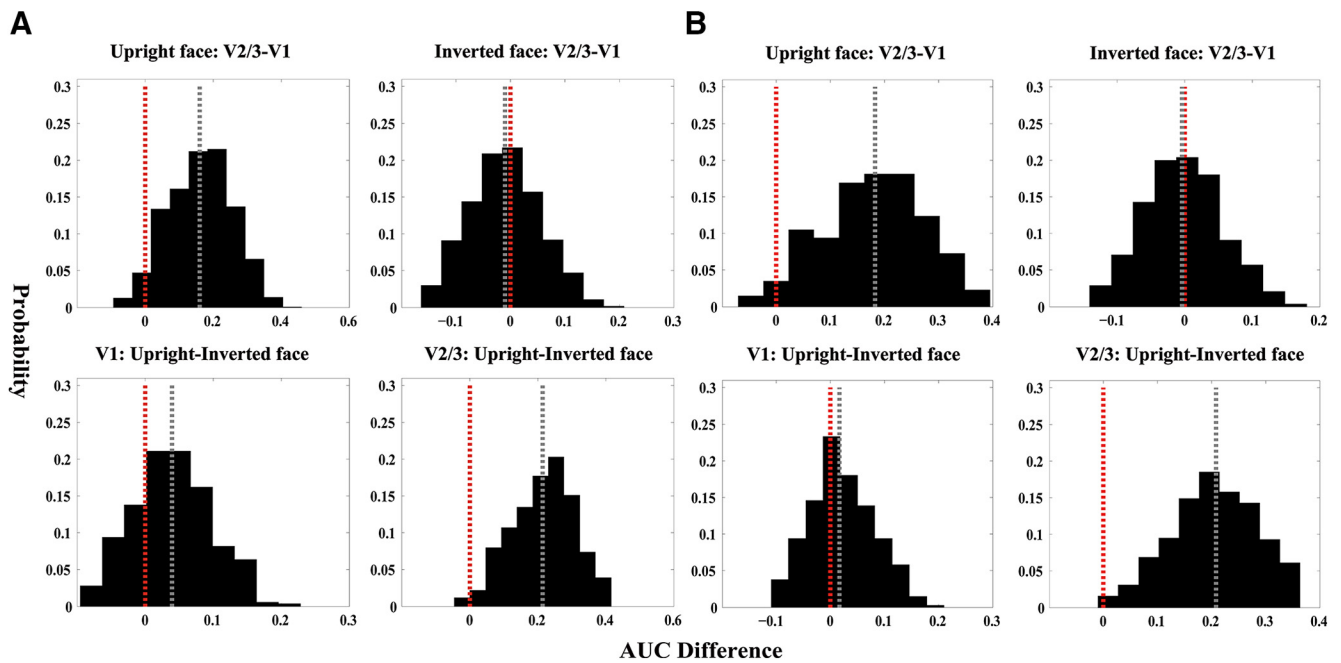
on this model to the measured BOLD signal, the pRF position and size parameters can be estimated for individual voxels, thus providing a full characterization of the receptive field properties of neuronal populations across the visual cortex.

Figure 2 shows the pRF estimation results. We fitted a line relating pRF eccentricity with pRF size in V1 and V2/3 for the whole, upper, and lower visual fields, respectively. Consistent with previous findings (Dumoulin and Wandell, 2008), the pRF size increased with the pRF eccentricity and the size increased faster in V2/3 (slope  $k = 0.174$ , intercept  $b = 0.499$ ) than in V1 ( $k = 0.105$ ,  $b = 0.430$ ). In addition, the relationship between pRF size and eccentricity was very similar across the upper (V1:  $k = 0.106$ ,  $b = 0.520$ ; V2/3:  $k = 0.191$ ,  $b = 0.609$ ) and lower visual fields (V1:  $k = 0.103$ ,  $b = 0.441$ ; V2/3:  $k = 0.166$ ,  $b = 0.550$ ) with no significant difference (Wilcoxon signed-rank test: V1 slope:  $p = 0.31$ ; V1 intercept:  $p = 0.94$ ; V2/3 slope:  $p = 0.20$ ; V2/3 intercept:  $p = 0.55$ ) (Fig. 2B), which would help to rule out potential visual field representation difference explanations for our attention priority map results.

For both the upright and the inverted faces, their cortical representations were reconstructed as the sum of the Gaussians weighted by the stimulus-specific activation level during the image-matching task. It is clear that areas of high representation intensity were mostly located in the image areas that convey important identity information. Behaviorally, these areas were also the regions to which most first saccades were made (Fig. 3A). Importantly, in both primary and extrastriate visual cortex, the reconstructed representations were generally consistent with the differential first saccadic target pattern for the upright and the

inverted faces. These observations suggest that the neural activity patterns in retinotopic visual areas might contribute to the patterns of attention-guided first saccadic eye movement.

We then examined quantitatively the behavioral relevance of the reconstructed representations by measuring how well the reconstructed representations could predict the differential first saccadic target pattern using precision-recall curves. We defined the high-priority areas based on the differential first saccadic target pattern and quantified the behavioral relevance as the area under the precision-recall curves (Fig. 3B), where a larger AUC indicates higher behavioral relevance. Results showed that, for both the upright and inverted faces, AUCs corresponding to the reconstructed representations in primary and extrastriate visual cortex was significantly above chance level (V1 upright face: AUC = 0.273,  $p = 0.001$ ; V1 inverted face: AUC = 0.263,  $p = 0.001$ ; V2/3 upright face: AUC = 0.507,  $p < 0.001$ ; V2/3 inverted face: AUC = 0.267,  $p = 0.002$ ). We performed the same analysis procedure using other criteria for defining the high-priority areas (top 6% and top 4.5%; see Materials and Methods) and obtained similar results [V1 upright face: AUC = 0.282,  $p = 0.001$  (top 6%), AUC = 0.306,  $p < 0.001$  (top 4.5%); V1 inverted face: AUC = 0.269,  $p = 0.001$  (top 6%), AUC = 0.27,  $p < 0.001$  (top 4.5%); V2/3 upright face: AUC = 0.528,  $p < 0.001$  (top 6%), AUC = 0.535,  $p < 0.001$  (top 4.5%); V2/3 inverted face: AUC = 0.258,  $p = 0.001$  (top 6%), AUC = 0.247,  $p = 0.002$  (top 4.5%)]. These results demonstrate the consistency between the reconstructed cortical representations and the differential first saccadic target patterns regardless of face configuration (i.e., orientation).



**Figure 4.** Differences in behavioral relevance obtained using two additional thresholds for defining the high-priority areas. **A**, Bootstrapped distributions of AUC difference using the threshold of 4.5%. **B**, Bootstrapped distributions of AUC difference using the threshold of 6%.

### Behavioral relevance of upright and inverted face representations

In addition to their consistency with the differential first saccadic target patterns, the reconstructed representations exhibited two differences in behavioral relevance as a function of cortical region and stimulus type. First, for the upright faces, the representation in V2/3 was more topographically consistent with the first saccadic target pattern than that in V1, whereas no such difference was observed between V1 and V2/3 for the inverted faces. Second, in V2/3, the representation of the upright faces was more topographically consistent with the differential first saccadic target pattern than that of inverted faces, whereas in V1, the difference between the upright and the inverted faces was less pronounced. We therefore tested whether behavioral relevance differed between the V1 and V2/3 representations for both stimulus types using the nonparametric bootstrapping method. We found that, consistent with our observations, the V2/3 representation predicted the differential first saccadic target pattern better than the V1 representation for the upright faces ( $p < 0.025$ ). In contrast, no significant difference between V1 and V2/3 was found for the inverted faces ( $p = 0.51$ ; Fig. 3C). These findings were robust against difference in criterion for defining the high-priority areas [V2/3 upright face AUC < V1 upright face AUC:  $p < 0.05$  (top 6%),  $p < 0.05$  (top 4.5%); V2/3 inverted face AUC < V1 inverted face AUC:  $p = 0.47$  (top 6%),  $p = 0.44$  (top 4.5%); Figure 4]. The upright face representation predicted the differential first saccadic target pattern better than the inverted face representation in V2/3 ( $p = 0.005$ ), whereas no difference was found in V1 ( $p = 0.4$ ; Fig. 3C). Similar results were obtained using the other two criteria for defining the high-priority areas [V1 upright face AUC < V1 inverted face AUC:  $p = 0.37$  (top 6%),  $p = 0.25$  (top 4.5%); V2/3 upright face AUC < V2/3 upright face AUC:  $p = 0.005$  (top 6%),  $p = 0.012$  (top 4.5%); Fig. 4]. Together, these findings suggested that the V1 representations were mainly driven by the physical and local information of the face images. Compared with V1, the representations in V2/3 were more behav-

ioral relevant and were more profoundly modulated by face configuration, which might be related to feedback signals from higher-order cortical regions that are specialized for processing faces.

### Discussion

We identified attention priority maps of faces in early retinotopic areas by testing whether population activity pattern in these areas were topographically related to the differential first saccadic target pattern. We found that the differential first saccadic target pattern could be predicted from neural activities in the V1 and V2/3 visual cortices, suggesting that these activities are behaviorally relevant. Moreover, in V2/3, cortical representation of upright faces exhibited enhanced behavioral relevance compared with that of inverted faces, suggesting that the attention priority map could be strongly modulated by image configuration in areas beyond V1. Different from previous studies of attention priority maps, which mainly involved artificial stimuli, our study is the first to investigate the attention priority map of complex natural stimuli by examining directly the topographic correspondence of stimuli representation with respect to the pattern of attention-guided behavior.

In the present study, we reconstructed the topographic representations of the face stimuli in V1 and V2/3 and established the link between these representations and the differential first saccadic target patterns. Moreover, we showed that the coupling between the topographic representation and the differential first saccadic target pattern was strengthened as stimulus information was relayed up in the visual hierarchy. It appears that the upright face representation is further “sculpted” in extrastriate visual cortex such that the attention deployment is more precise in terms of guiding task-related behavior. These reconstructed representations thus exhibit several properties compatible with the attention priority map theories that casts attention signals as a series of interacting, hierarchical priority maps in the visual system (Serences and Yantis, 2006; Liu and Hou, 2013), including: (1) the correspondence between perceptual behavior and neural activity

patterns during attention process, (2) a closer link between perceptual behavior and neural activity patterns in higher visual cortex, and (3) the interaction between higher- and lower-level representations in the form of intercortical enhancement of behavioral relevance. Therefore, one promising interpretation of our findings is that attention priority maps of natural stimuli exist in both primary and extrastriate visual cortices. Our findings of enhanced behavioral relevance of the reconstructed representations in extrastriate visual cortex echo the earlier findings by Sprague and Serences (2013). In their study, they reconstructed the topographic representation of a circular checkerboard patch presented at different spatial positions in multiple visual areas using multivariate forward encoding model. They found that the amplitude of these topographic representations systematically increased from low- to high-level visual areas as a result of attention modulation. Together with Sprague and Serences's (2013) study, our findings extend significantly the classical view that attention priority representations are mainly hosted in higher-order brain regions, including parietal regions that are important for integrating top-down and bottom-up signals (Toth and Assad, 2002; Bisley and Goldberg, 2010); the frontal eye field, which is believed to be a critical neural site for controlling eye movement (Fecteau and Munoz, 2006); and the lateral occipital area, which is strongly modulated by top-down signals relevant to object detection (Peelen et al., 2009; Peelen and Kastner, 2011, 2014; Seidl et al., 2012) and target location in scenes (Preston et al., 2013). Interestingly, however, our data showed that the behavioral relevance of the inverted face representation does not increase from primary to extrastriate visual cortex. This suggests a novel property of attention priority maps that, at least for face images, the increase in the strength of functional coupling between neural activities and perceptual behavior along the visual pathway is contingent on stimulus configuration.

Our data showed that, in extrastriate visual cortex, the attention priority representation of upright faces is better than that of inverted faces in terms of predicting the differential first saccadic target pattern. This finding provides a remarkable extension to the conventional view of attention priority map theories that physical salience and task goal relevance are the only major factors constraining attention priority (Fecteau and Munoz, 2006; Serences and Yantis, 2006) because we have demonstrated another critical factor that strongly influences attention priority maps: stimulus configuration. The distinct attention priority patterns of upright and inverted faces suggest that the impaired ability of recognizing inverted faces might be related to the inefficient attention deployment that impedes the early extraction of critical face features (Sekuler et al., 2004), which extends the previous finding that the fusiform face area is the primary neural source of the behavioral face inversion effect (Yovel and Kanwisher, 2005). Conversely, we found no significant difference in behavioral relevance between the upright and the inverted face representations in V1. Because physical salience of the upright and inverted face images should be identical, the absence of a difference in behavioral relevance suggests that V1 neurons are largely driven by the physical salience of visual inputs. This is consistent with previous findings that V1 creates the saliency map of visual inputs that are not perceived consciously by subjects (Li, 1999, 2002; Zhang et al., 2012; Chen et al., 2013). Neurally, it has been suggested that lateral connections between V1 neurons suppress the neuronal response to image parts with homogenous visual features and thus renders the region containing inhomogeneous visual features (i.e., the salient regions) more strongly represented (Gilbert and Wiesel, 1983; Rockland and Lund, 1983). This is also consis-

tent with the "barcode" hypothesis, which postulates that a significant portion of physical information of face images conforms to a horizontal structure consisting of vertically coaligned clusters (Dakin and Watt, 2009). Because the inversion of face images does not alter the horizontal structure, the barcode hypothesis would predict similar behavioral relevance for the upright and inverted face representations mediated by neurons encoding physical salience, which is consistent with our findings in V1. In extrastriate visual cortex, attention priority representations might arise from the competitive circuits in which visual items compete for neural resources. Top-down signals bias the competition in favor of the behaviorally relevant item by increasing its efficacy in driving visual neurons with receptive fields that contain that item (Desimone and Duncan, 1995; Reynolds et al., 1999; Reynolds and Desimone, 2003; Reynolds and Chelazzi, 2004). Together, our findings suggested a critical dissociation between primary and extrastriate visual cortex in terms of the underlying neural mechanism linking the topographic stimulus representations with the task-related behavior. Lateral connection might play a critical role in representing physical salience that constitutes the bulk of the information encoded in attention priority maps in V1, whereas attention priority maps might be mediated by competitive circuits in extrastriate visual cortex that are more susceptible to influences of top-down signals.

A possible explanation of the enhanced behavioral relevance of the upright face representation in V2/3 could be due to increased sensitivity to face features in a typical retinotopic location during normal gaze behavior. In a recent study (de Haas et al., 2016), subjects were presented with isolated face features (e.g., eyes, mouth) either at the typical or the inverted retinotopic location and performed a recognition task. They found that observers performed better at recognizing face features when presented at the typical visual field location than those presented at the inverted location. These findings suggest that the brain representation of face features is not homogeneous across the entire visual field, but rather depends on their retinotopic location regardless of face context. However, if the difference in behavioral relevance between the upright and inverted face representations in V2/3 is indeed caused by the retinotopic position advantage of individual face features, then one might predict such a difference also in V1. This was not observed in our study. Therefore, our findings might be better explained by top-down signals rather than the location-dependent advantage of face features.

The preference for the left eye region in the differential first saccadic target patterns found here is different from the bias of first fixations toward the right eye region in previous findings (Peterson and Eckstein, 2012, 2013, 2015). Three factors might account for this difference. First, in these previous studies, the starting fixation point was placed outside the face, whereas in our study, the starting fixation point was placed at the center of the face image. The starting fixation point difference might have a strong influence on the target of the first saccade. Second, investigators in the previous studies asked subjects to recognize face identity, emotion, or gender, whereas we asked subjects to perform a simpler one-back-matching task. Third, the previous studies presented a face for only 200 ms; the first saccadic target served to optimize information integration for better behavioral performance during this very brief presentation. In our study, a face image was presented for 2 s, in which case the differential first saccadic target pattern might reflect subjects' priority map, as we claim.

In summary, our study demonstrates that attention priority maps of complex natural stimuli such as faces could be found in



both primary and extrastriate visual cortices. We show that attention selection occurs, not only among multiple objects in a scene, but also within a complex object by prioritizing diagnostic object features. Moreover, we show that attention allocation is influenced, not only by physical saliency and task goal relevance, but also by image configuration. Our findings contribute to filling the long-existing blank of attention priority maps of natural stimuli and make headway toward unraveling the mechanisms underlying visual attention selection.

## References

- Achanta R, Hemami S, Estrada F, Sussstrunk S (2009) Frequency-tuned salient region detection. In: *Computer Vision and Pattern Recognition (CVPR)*, 2009 IEEE Conference on, pp 1597–1604: IEEE.
- Awh E, Armstrong KM, Moore T (2006) Visual and oculomotor selection: links, causes and implications for spatial attention. *Trends Cogn Sci* 10: 124–130. [CrossRef Medline](#)
- Baluch F, Itti L (2011) Mechanisms of top-down attention. *Trends Neurosci* 34:210–224. [CrossRef Medline](#)
- Bisley JW, Goldberg ME (2003) Neuronal activity in the lateral intraparietal area and spatial attention. *Science* 299:81–86. [CrossRef Medline](#)
- Bisley JW, Goldberg ME (2010) Attention, intention, and priority in the parietal lobe. *Annu Rev Neurosci* 33:1–21. [CrossRef Medline](#)
- Chen C, Zhang X, Zhou T, Wang Y, Fang F (2013) Neural representation of the bottom-up saliency map of natural scenes in human primary visual cortex. *J Vis* 13:233. [CrossRef](#)
- Dakin SC, Watt RJ (2009) Biological “bar codes” in human faces. *J Vis* 9:2.1–10. [CrossRef Medline](#)
- Davis J, Goadrich M (2006) The relationship between Precision-Recall and ROC curves. In: *Proceedings of the 23rd International Conference on Machine Learning*, pp 233–240: ACM.
- de Haas B, Schwarzkopf DS, Alvarez I, Lawson RP, Henriksson L, Kriegeskorte N, Rees G (2016) Perception and processing of faces in the human brain is tuned to typical feature locations. *J Neurosci* 36:9289–9302. [CrossRef Medline](#)
- Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222. [CrossRef Medline](#)
- Dumoulin SO, Wandell BA (2008) Population receptive field estimates in human visual cortex. *Neuroimage* 39:647–660. [CrossRef Medline](#)
- Efron B, Tibshirani RJ (1994) *An introduction to the bootstrap*. San Diego: CRC.
- Einhäuser W, Spain M, Perona P (2008) Objects predict fixations better than early saliency. *J Vis* 8:18.1–26. [CrossRef Medline](#)
- Engel SA, Glover GH, Wandell BA (1997) Retinotopic organization in human visual cortex and the spatial precision of functional MRI. *Cereb Cortex* 7:181–192. [CrossRef Medline](#)
- Fecteau JH, Munoz DP (2006) Saliency, relevance, and firing: a priority map for target selection. *Trends Cogn Sci* 10:382–390. [CrossRef Medline](#)
- Gilbert CD, Wiesel TN (1983) Clustered intrinsic connections in cat visual cortex. *J Neurosci* 3:1116–1133. [Medline](#)
- Gottlieb JP, Kusunoki M, Goldberg ME (1998) The representation of visual saliency in monkey parietal cortex. *Nature* 391:481–484. [CrossRef Medline](#)
- Jerde TA, Merriam EP, Riggall AC, Hedges JH, Curtis CE (2012) Prioritized maps of space in human frontoparietal cortex. *J Neurosci* 32:17382–17390. [CrossRef Medline](#)
- Jiang Y, Costello P, Fang F, Huang M, He S (2006) A gender-and sexual orientation-dependent spatial attentional effect of invisible images. *Proc Natl Acad Sci U S A* 103:17048–17052. [CrossRef Medline](#)
- Jiang YV, Won BY, Swallow KM (2014) First saccadic eye movement reveals persistent attentional guidance by implicit learning. *J Exp Psychol Hum Percept Perform* 40:1161–1173. [CrossRef Medline](#)
- Kay KN, Winawer J, Mezer A, Wandell BA (2013) Compressive spatial summation in human visual cortex. *J Neurophysiol* 110:481–494. [CrossRef Medline](#)
- Koehn P (2004) Statistical significance tests for machine translation evaluation. In: *Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp 388–395: ACL.
- Kok P, de Lange FP (2014) Shape perception simultaneously up-and down-regulates neural activity in the primary visual cortex. *Curr Biol* 24:1531–1535. [CrossRef Medline](#)
- Li Z (1999) Contextual influences in V1 as a basis for pop out and asymmetry in visual search. *Proc Natl Acad Sci U S A* 96:10530–10535. [CrossRef Medline](#)
- Li Z (2002) A saliency map in primary visual cortex. *Trends Cogn Sci* 6:9–16. [CrossRef Medline](#)
- Liu T, Hou Y (2013) A hierarchy of attentional priority signals in human frontoparietal cortex. *J Neurosci* 33:16606–16616. [CrossRef Medline](#)
- Mazer JA, Gallant JL (2003) Goal-related activity in V4 during free viewing visual search: evidence for a ventral stream visual saliency map. *Neuron* 40:1241–1250. [CrossRef Medline](#)
- Peelen MV, Kastner S (2011) A neural basis for real-world visual search in human occipitotemporal cortex. *Proc Natl Acad Sci U S A* 108:12125–12130. [CrossRef Medline](#)
- Peelen MV, Kastner S (2014) Attention in the real world: toward understanding its neural basis. *Trends Cogn Sci* 18:242–250. [CrossRef Medline](#)
- Peelen MV, Fei-Fei L, Kastner S (2009) Neural mechanisms of rapid natural scene categorization in human visual cortex. *Nature* 460:94–97. [CrossRef Medline](#)
- Perazzi F, Krähenbühl P, Pritch Y, Hornung A (2012) Saliency filters: Contrast based filtering for salient region detection. In: *Computer Vision and Pattern Recognition (CVPR)*, 2012 IEEE Conference on, pp 733–740: IEEE.
- Peterson MF, Eckstein MP (2012) Looking just below the eyes is optimal across face recognition tasks. *Proc Natl Acad Sci U S A* 109:E3314–E3323. [CrossRef Medline](#)
- Peterson MF, Eckstein MP (2013) Individual differences in eye movements during face identification reflect observer-specific optimal points of fixation. *Psychol Sci* 24:1216–1225. [CrossRef Medline](#)
- Peterson MF, Eckstein MP (2015) Initial eye movements during face identification are optimal and similar across cultures. *J Vis* 15:12. [CrossRef Medline](#)
- Preston TJ, Guo F, Das K, Giesbrecht B, Eckstein MP (2013) Neural representations of contextual guidance in visual search of real-world scenes. *J Neurosci* 33:7846–7855. [CrossRef Medline](#)
- Reynolds JH, Chelazzi L (2004) Attentional modulation of visual processing. *Annu Rev Neurosci* 27:611–647. [CrossRef Medline](#)
- Reynolds JH, Desimone R (2003) Interacting roles of attention and visual saliency 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. [Medline](#)
- Rhodes G, Tremewan T (1994) Understanding face recognition: Caricature effects, inversion, and the homogeneity problem. *Visual Cognition* 1:275–311. [CrossRef](#)
- Rockland KS, Lund JS (1983) Intrinsic laminar lattice connections in primate visual cortex. *J Comp Neurol* 216:303–318. [CrossRef Medline](#)
- Seidl KN, Peelen MV, Kastner S (2012) Neural evidence for distracter suppression during visual search in real-world scenes. *J Neurosci* 32:11812–11819. [CrossRef Medline](#)
- Sekuler AB, Gaspar CM, Gold JM, Bennett PJ (2004) Inversion leads to quantitative, not qualitative, changes in face processing. *Curr Biol* 14: 391–396. [CrossRef Medline](#)
- Serences JT, Yantis S (2006) Selective visual attention and perceptual coherence. *Trends Cogn Sci* 10:38–45. [CrossRef Medline](#)
- Serences JT, Yantis S (2007) Spatially selective representations of voluntary and stimulus-driven attentional priority in human occipital, parietal, and frontal cortex. *Cereb Cortex* 17:284–293. [CrossRef Medline](#)
- Sprague TC, Serences JT (2013) Attention modulates spatial priority maps in the human occipital, parietal and frontal cortices. *Nat Neurosci* 16: 1879–1887. [CrossRef Medline](#)
- Toth LJ, Assad JA (2002) Dynamic coding of behaviourally relevant stimuli in parietal cortex. *Nature* 415:165–168. [CrossRef Medline](#)
- Yarbus AL (1967) *Eye movements during perception of complex objects*. New York: Springer.
- Yin RK (1969) Looking at upside-down faces. *J Exp Psychol* 81:141. [CrossRef](#)
- Yovel G, Kanwisher N (2005) The neural basis of the behavioral face-inversion effect. *Curr Biol* 15:2256–2262. [CrossRef Medline](#)
- Zhang X, Zhaoping L, Zhou T, Fang F (2012) Neural activities in V1 create a bottom-up saliency map. *Neuron* 73:183–192. [CrossRef Medline](#)