Behavioral/Cognitive

Attention Priority Map of Face Images in Human Early Visual Cortex

Ce Mo,[1,2,3,4*] **Dongjun He,**[5*] **and Fang Fang**[1,2,3,4]
[1]School of Psychological and Cognitive Sciences and Beijing Key Laboratory of Behavior and Mental Health, [2]Peking-Tsinghua Center for Life Sciences, [3]Key Laboratory of Machine Perception (Ministry of Education), [4]PKU-IDG/McGovern Institute for Brain Research, Peking University, Beijing 100871, China, and [5]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

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 fined using a standard phase-encoded method (Engel et al., 1997) in which subjects viewed a rotating wedge and an expanding ring that cretransient magnetic saturation effects. For each subject, a general linear 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:

$$\boldsymbol{R}_i(x, y) = \sum_j \boldsymbol{w}_{(i,j)}\, \boldsymbol{G}_j(x, y \mid x_0, y_0, \sigma),\ i \in \{Upright, Inverted\}$$

where $R_i(x, y)$ refers to the stimulus-specific representation intensity at the retinotopic location (x, y) and $G_j(x, y \mid x_0, y_0, \sigma)$ refers to the estimated voxelwise pRF model jointly parameterized by the pRF center (x_0, y_0) and size σ 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:

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

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:

$$Recall = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

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 ± SEM upright face: 0.9967 ± 0.002, inverted face: 0.9492 ± 0.009, scrambled face: 0.9808 ± 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. 3*A*, 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

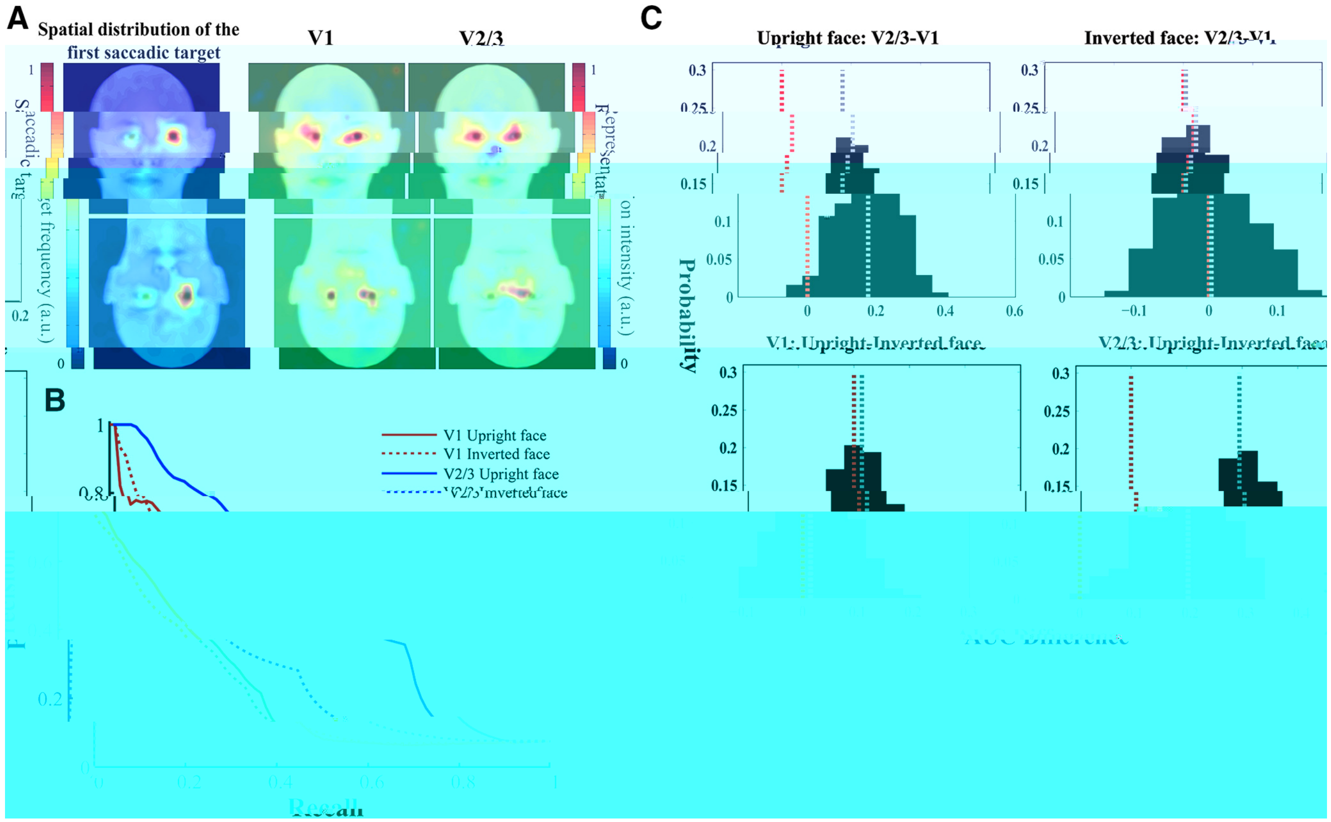

Figure 3. Reconstructed topographic representations of face stimuli and their behavioral relevance. ***A***, Visualization of the spatial patterns of the first saccadic target (left) and the reconstructed representations in V1 (middle) and V2/3 (right). ***B***, Precision-recall curves corresponding 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. 2*B*), 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. 3*A*). 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. 3*B*), 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).

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. 3*C*). 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. 3*C*). 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 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 consistent 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 salience 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, Susstrunk 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