C M ,[1,2,3,4]

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).

M M

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 (

fined 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 checkered disk with a radius of 10.94° 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 checkered bar 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° horizontally and 22.8° vertically. The bar was 2.76° in width and its length was either 24.8° or 22.8° (Fig. 2*A*). Each bar swept through the visual area in 16 steps within 51 s. The step size was 1.38°. 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 Innova-

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:

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

where R

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: $p =$ [illegible]

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

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.

R

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

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) Salience, 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 salience 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