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 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

 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;

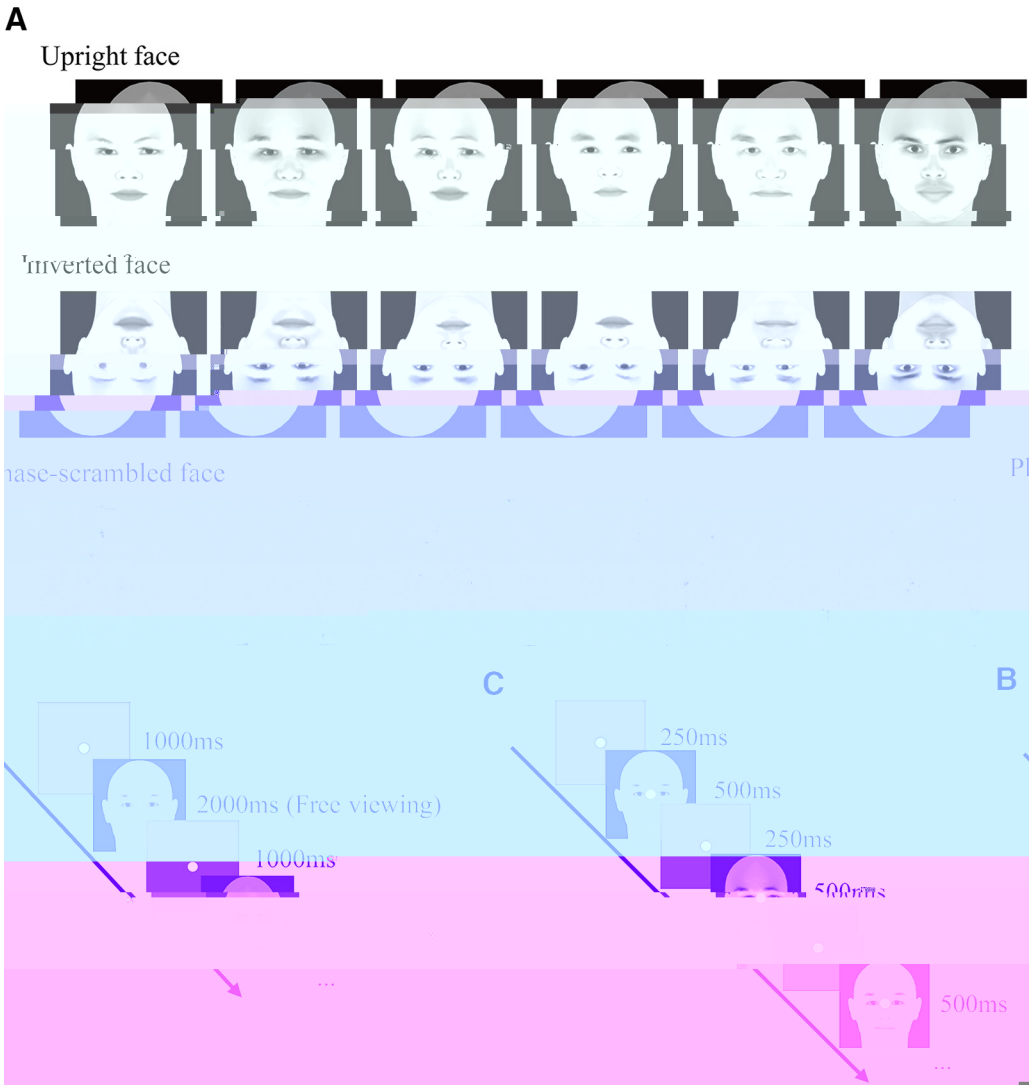

Figure 1. Stimuli and experimental protocol. ***A***, Exemplar stimuli of three stimulus types. ***B***, Protocol of the eye-tracking experiment. Subjects initiated a trial by fixating at the central fixation point. The end point of the first fixation change after stimulus onset was recorded as the first 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.

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 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° 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

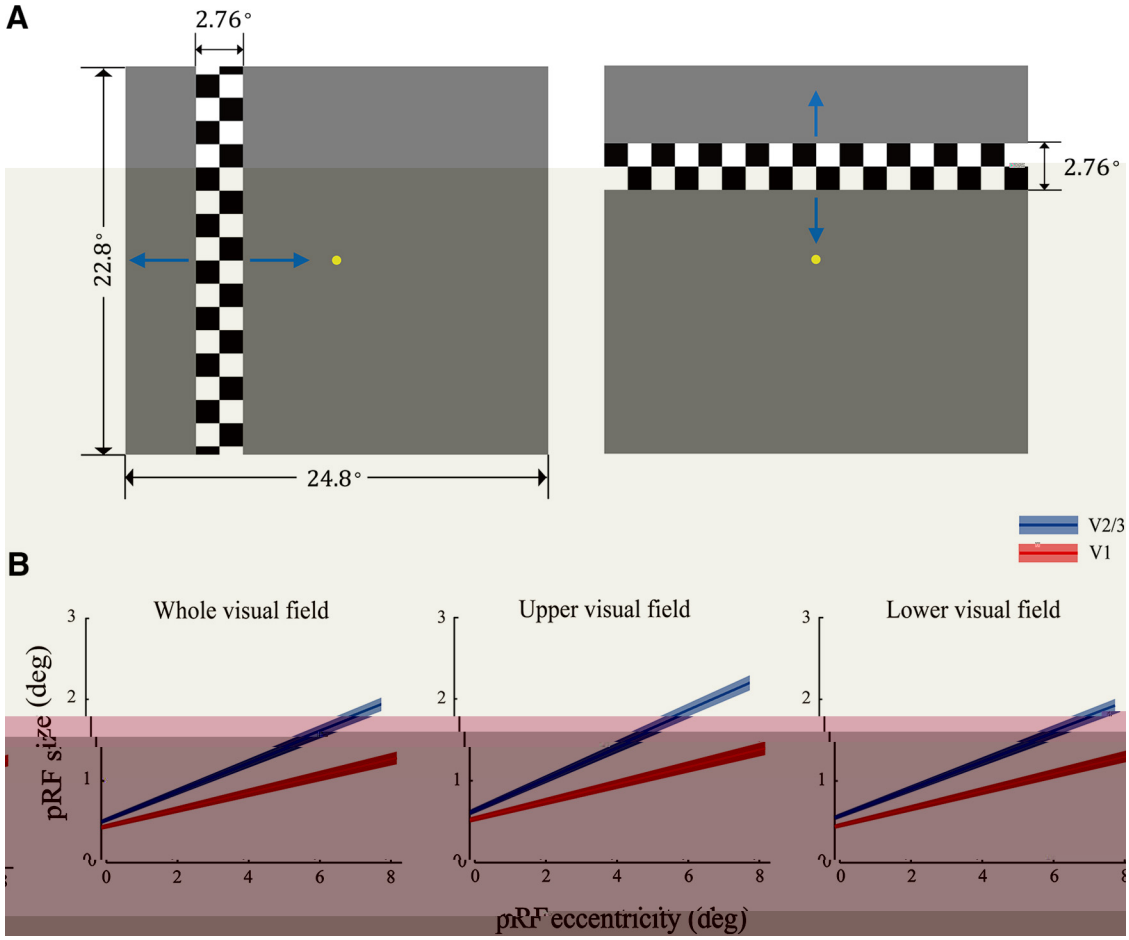

Figure 2. pRF mapping. ***A***, Stimuli used for pRF mapping. Blue arrows indicate the moving directions of the checkered bar in pRF mapping runs. Each bar moved in one direction and switched to the opposite direction when it reached the boundary of the mapped area. ***B***, Relationships between pRF 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., β value) were estimated for individual voxels, subtracted by the mean β value across all the voxels in the ROI, and divided by the maximal absolute. After this normalization step, the β 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 σ, where x_0 and y_0 are the coordinates of the center of the receptive field and σ 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 proportion of voxels retained after applying this threshold was high and was comparable between the V1 and the V2/3 ROIs (mean proportion ± SEM V1: 0.952 ± 0.008, V2/3: 0.942 ± 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 β values for the upright or the inverted faces were submitted as the dependent variable, whereas the β values for the phase-scrambled images were submitted as the independent variable of the model as follows:

$$\boldsymbol{\beta}_{(i,j)} = r_i \boldsymbol{\beta}_{(Scrambled,j)} + C_i + \boldsymbol{w}_{(i,j)}, i \in \{Upright, Inverted\}$$

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

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

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 (

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 circu3opo

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

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

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 2014 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 salience 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