

# Idiosyncratic Patterns of Representational Similarity in Prefrontal Cortex Predict Attentional Performance

Jeongmi Lee and  Joy J. Geng

Center for Mind and Brain, University of California, Davis, Davis, California 95618

The efficiency of finding an object in a crowded environment depends largely on the similarity of nontargets to the search target. Models of attention theorize that the similarity is determined by representations stored within an “attentional template” held in working memory. However, the degree to which the contents of the attentional template are individually unique and where those idiosyncratic representations are encoded in the brain are unknown. We investigated this problem using representational similarity analysis of human fMRI data to measure the common and idiosyncratic representations of famous face morphs during an identity categorization task; data from the categorization task were then used to predict performance on a separate identity search task. We hypothesized that the idiosyncratic categorical representations of the continuous face morphs would predict their distractability when searching for each target identity. The results identified that patterns of activation in the lateral prefrontal cortex (LPFC) as well as in face-selective areas in the ventral temporal cortex were highly correlated with the patterns of behavioral categorization of face morphs and search performance that were common across subjects. However, the individually unique components of the categorization behavior were reliably decoded only in right LPFC. Moreover, the neural pattern in right LPFC successfully predicted idiosyncratic variability in search performance, such that reaction times were longer when distractors had a higher probability of being categorized as the target identity. These results suggest that the prefrontal cortex encodes individually unique components of categorical representations that are also present in attentional templates for target search.

**Key words:** attentional template; fMRI; individual difference; prefrontal cortex; RSA; visual search

## Significance Statement

Everyone’s perception of the world is uniquely shaped by personal experiences and preferences. Using functional MRI, we show that individual differences in the categorization of face morphs between two identities could be decoded from the prefrontal cortex and the ventral temporal cortex. Moreover, the individually unique representations in prefrontal cortex predicted idiosyncratic variability in attentional performance when looking for each identity in the “crowd” of another morphed face in a separate search task. Our results reveal that the representation of task-related information in prefrontal cortex is individually unique and preserved across categorization and search performance. This demonstrates the possibility of predicting individual behaviors across tasks with patterns of brain activity.

## Introduction

Efficiently allocating attention to currently relevant information is important for the survival of the observer. A consensus among theories of attention is that the contents of the “attentional template,” the collection of task-related features held in working memory

(Bundesen, 1990; Carlisle et al., 2011), determine the efficiency of attentional allocation by adjusting sensory gain to distinguish target features from nontarget features (Wolfe, 1994; Desimone and Duncan, 1995; Moore and Egeth, 1998; Martinez-Trujillo and Treue, 2004; Malcolm and Henderson, 2009; Reynolds and Heeger, 2009; Hout and Goldinger, 2015). Models of attention frequently assume that target information in the attentional template is accurate and uniform across individuals and have not addressed how individual differences in the precision of the attentional template affect selection. However, the presence of considerable individual differences in the quality of the attentional template should not be surprising, since everyone’s representation of the world is uniquely shaped by personal experiences and preferences (Charest et al., 2014).

Understanding individual variation in the precision of the template is critical for models of attention, because the precision

Received April 28, 2016; revised Nov. 8, 2016; accepted Dec. 9, 2016.

Author contributions: J.L. and J.J.G. designed research; J.L. performed research; J.L. and J.J.G. analyzed data; J.L. and J.J.G. wrote the paper.

This work was supported by NSF Grant BCS 1230377-0 to J.J.G. We thank Marlene Behrmann and Arne Ekstrom for their helpful comments.

The authors declare no competing financial interests.

Correspondence should be addressed to Jeongmi Lee, Center for Mind and Brain, University of California, Davis, 267 Cousteau Place, Davis, CA 95618. E-mail: jmslee@ucdavis.edu.

DOI:10.1523/JNEUROSCI.1407-16.2016

Copyright © 2017 the authors 0270-6474/17/371257-12\$15.00/0

defines the effective “similarity” of stimuli to the target. Decades of behavioral research have shown that search performance is critically dependent on the similarity between objects. For example, objects that are physically more similar to the target capture attention and interfere more with search (Treisman and Gelade, 1980; Duncan and Humphreys, 1989; Nagy and Sanchez, 1990; Folk et al., 1992). Likewise, objects that share the same category label with the target are perceived to be more similar and get prioritized (Nosofsky, 1986; Goldstone et al., 2001; Yang and Zelinsky, 2009; Althaus and Westermann, 2016). The importance of considering individually unique representations of similarity is illustrated by previous results showing that psychological representations of similarity are not entirely determined by physical properties, but are also dependent on idiosyncratic experiences and preferences (Alexander and Zelinsky, 2011; Kriegeskorte and Mur, 2012; Mur et al., 2013; Hout et al., 2016). For example, a previous study by Charest et al. (2014) found individually unique representational geometries for complex objects in human ventral temporal cortex (VTC) that accounted for idiosyncratic differences in the similarity judgments. Interestingly, they found that individual differences were greater for personally meaningful objects than for unfamiliar objects.

Despite clear evidence for individual differences in representations of similarity, the neural substrates of encoding the individually unique aspects of the attentional template remain unclear. Previous functional magnetic resonance imaging (fMRI) studies have found representations for object categories (Martin et al., 1996; Kanwisher et al., 1997; Haxby et al., 2001; Peelen et al., 2009) and individual exemplars within the same category (Kriegeskorte et al., 2007; Kay et al., 2008) in the ventral temporal cortex. However, more abstract and ad hoc categorical task representations that are used for behavioral decisions have been observed in the posterior parietal cortex and prefrontal cortex (Freedman and Assad, 2006; Feredoes et al., 2007; Gazzaley and Nobre, 2012; Lee et al., 2013; Sarma et al., 2016). The different representations presumably reflect different stages of storage and use of information for task goals, but it remains unknown which of these brain areas encode the idiosyncratic representations that can predict the degree of competition between objects during search.

Here, we used the representational similarity analysis (RSA) of fMRI data to test two questions regarding the relationship between category representations in the brain and attentional performance: (1) Where in the brain are the common and individually unique category representations encoded? (2) Can idiosyncrasies in the representational geometries of stimuli predict individually unique variance in an independent search task? We hypothesized that measurements of the unique and common aspects of stimulus representations could be used to predict the precision of the attentional template and the efficiency of search performance.

## Materials and Methods

### Participants

Fourteen subjects (nine females) ranging in age from 19 to 31 (mean age, 22.57) participated in a 2 h session and received monetary compensation. Data from one participant were excluded from analyses due to poor behavioral performance [lower than 80% accuracy rate, slower than 1500 ms mean reaction time (RT)], which resulted in a final group of 13 subjects (nine females). All had normal or corrected-to-normal vision and no neurological or psychiatric history. Informed consent was obtained according to procedures approved by the Institutional Review Board of the University of California, Davis.

### Apparatus

Stimuli were presented on a Dell 2408WFP monitor using Presentation software (version 16.5; <http://neurobs.com>). For the identity search task outside the scanner, participants viewed the monitor from a distance of 60 cm in a dimly lit room. To make sure that participants fixated on the center throughout the experiment, eye position data were collected using an EyeLink 1000 version 4.56 (SR Research; sampling rate, 500 Hz), initiating each trial only if 100 ms of continuous central fixation was detected.

### Identity categorization task

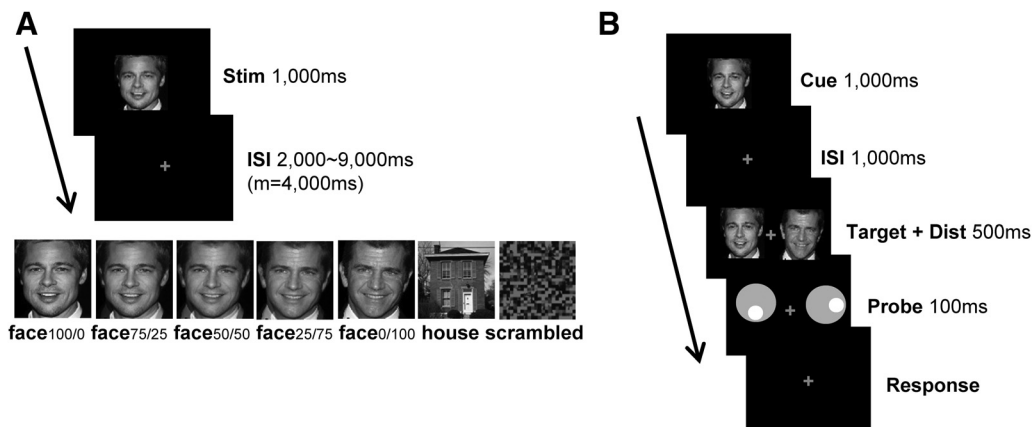
An identity categorization task (Fig. 1A) was performed inside the scanner to obtain each individual's representational geometries for the task stimuli. The initial stimuli consisted of four portraits of famous celebrities (two males, two females). The pictures were gray scaled, cropped, resized into  $220 \times 250$  pixels, and divided into two gender-matched pairs. The faces in each pair were then morphed into five images ( $\text{face}_{100/0}$ ,  $\text{face}_{75/25}$ ,  $\text{face}_{50/50}$ ,  $\text{face}_{25/75}$ ,  $\text{face}_{0/100}$ ) that represented gradual transitions from one original face to the other, in steps of 25% (0 to 100%), using Fantamorph 5.4.5 (Abrosoft). In addition to the face morphs, two nonface stimuli of the same size were also included in the stimulus set: an achromatic picture of a house and a scrambled image, which was created by scrambling  $10 \times 10$  pixel patches of the intact  $\text{face}_{50/50}$  such that no facial features were discernible. Thus, there were two stimulus sets (the male set and the female set), with each set containing five face morphs and two nonface stimuli. It was randomly chosen which stimulus set would be presented to each subject in the identity categorization task.

On each trial, one image from the stimulus set was shown in the center of the screen for 1000 ms, followed by a fixation display for a jittered duration between 2000 and 9000 ms (mean, 4000 ms). Participants classified the stimulus into three possible categories: Face morphs were categorized as one identity or the other by pressing button 1 or 2, respectively. A house or a scrambled image was categorized as a nonface by pressing button 3. Participants held the button box in their right hand and pressed the three buttons with their index, middle, and ring fingers, respectively. Each scan contained 28 trials (four repetitions of each of the seven stimuli) presented in a random order, and a total of 12 scans were obtained.  $\beta$ -Coefficients were extracted by using a general linear model and a design matrix that modeled the response to the stimulus onset event.

We chose a small set of face morphs, instead of a larger set of different classes of stimuli, for the following reasons: First, our study involved pinpointing the sources of idiosyncratic variations. Since the RSA calculates correlation coefficients based on all possible pairs of stimuli, blind to the particular pairs of stimuli that determine the observed correlation coefficients, having a large set of stimuli would have made it difficult to identify the source of individually unique representational variation. Second, we used the same stimuli in an independent search task as target-and-distractor pairs. Thus, a large set of stimuli in the categorization task would have exponentially increased the pairs of stimuli required in the search task.

### Identity search task

An identity search task (Fig. 1B) was performed outside the scanner to observe how the representational similarity structure obtained in the identity categorization task influenced the efficiency of attentional selection. The five face morphs each participant saw in the identity categorization task were used as the target and distractor stimuli. On each trial, a cue face (one of the two original 100% faces) was presented on the center of the screen for 1000 ms, followed by a fixation display of 1000 ms. During this time, the cue face was presumably stored as an attentional template within working memory. Then, the target and a distractor face were presented bilaterally for 500 ms, requiring participants to discriminate the two faces and move spatial attention to the location of the target. The target face was the same face as the cue face, and the distractor face was one of the other face morphs. The physical similarity of the distractor face to the target was manipulated by the percentage of the target face within the morph (75, 50, 20, and 0%). Next, two probes were presented bilaterally for 100 ms, followed by a fixation display until response. There



**Figure 1.** Experiment design. **A**, Sequence of events and time course of a trial in the identity categorization task. Participants categorized each stimulus as identity 1, identity 2, or nonface. In each stimulus set, there were five face morphs gradually changing from one identity to another in steps of 25% and two nonface stimuli (a house and a scrambled image). The original photo information is as follows: identity 1 is from <https://www.flickr.com/photos/smithsoccasional/1459959152/in/photolist-3e1EZw-31r2eh-fPRHKv-eboneNeN-5sme2p-flrc1j-3a7Dch-eWW4G4-dx3n1K-5ArYqi-6sXPNU-ebi7SH-4eU7HD-eoeKlJ-akQ3Gz-dE3aNz-avwXBU-dVJQXS-cde57w-8kyVuH-59Wna-9rT5Q5-nPUdVa-fPRGBD-dBoD5d-rmPFSB-9AN59Z-afGzLV-EKfYA-bvH3L5-2j5zJE-bjqWAK-5nQ6NB-an64PA-zLXuH-5RE5cg-AJieAU-65VcvS-633FZc-bsBeUu-ETb4c-yW7Htq-6jFfEj-jEEjs-dVtkCQ-BtpV5c-7GZaSi-pBoGyj-9qpQj5-gLpqq> (creator, L. Smith; license, <https://creativecommons.org/licenses/by-nc-sa/4.0/>); identity 2 is from an online album from [driffordchoi](http://driffordchoi.com) (no longer available online); the house is from <https://www.flickr.com/photos/broc7/108307858/in/photolist-az79S-6r3VGD-roUZGa-6QqrGr-3ZyA9Z-bx5XyL-9sQguK-3QCa4u-5V6wGp-zcDiE-985Pt-wpxDu3-3esBRD-5zP8Fs-526AeC-esSkc6-aDLALm-7gSC7c-zcCeq-cpY1u-ssdB8-J8Mz6-cbDxE-bDKDbK-7jVNww-qSLfw6-az6ti-5PEfWv-5Ka4cT-2dVVD5F-2h3HD-cLn3Q-egipk-4RSTvn-zcCC-2jvxd-C1Jij-4G4Wen-8vK2id-4Dev4z-4w6DEo-e2Gg1-6WF6PP-jvHAY-ck1ou-6xaYD-6UPaWx-r2vJV5-3Nzkt-3kftpX> (creator, broc7; license, <https://creativecommons.org/licenses/by-nc-sa/4.0/>). All images were gray scaled, cropped, resized to  $220 \times 250$  pixels, and then morphed (Fantamorph 5.4.5, Abrosoft). **B**, Sequence of events and time course of a trial in the identity search task. The face morphs each participant saw in the identity categorization task were used as the target and distractor stimuli. The physical similarity of the distractor face to the target was manipulated by the percentage of the target face within the morph (75, 50, 20, or 0%). The task was to pay attention to the target face (cued face, identity 1 or 2) and report the location of the hole inside the probe that appeared on the target side. ISI, interstimulus interval.

was a hole inside each probe, with four possible states: the hole in the probe at the cued (target) location was either up or down; the hole in the distractor probe was either right or left. The task was to report the location of the hole in the probe that appeared at the cued location by pressing “k” or “m” on a keyboard to indicate up or down, respectively. This enabled us to make sure that participants’ responses were based on the characteristics of the probe at the cued location, rather than the probe at the opposite location, while equating for low-level perceptual effects. Since the probe was presented very briefly (100 ms), it was difficult to do the task correctly unless participants covertly attended to the cued location in advance. Thus, participants were specifically instructed to move their attention to the location of the target face while fixating on the center fixation cross to do the task successfully.

#### Event-related fMRI

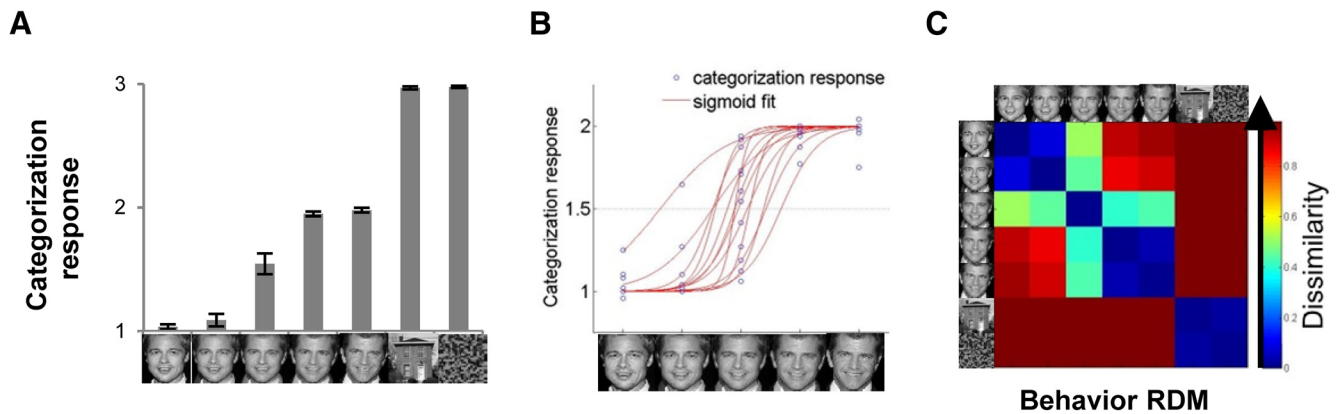
MRI scanning was performed on a 3-Tesla Siemens Skyra scanner with a 32 channel phased-array head coil at the imaging research center at the University of California, Davis. A T2-weighted echoplanar imaging (EPI) sequence was used to acquire whole-brain volumes of 60 axial slices of 2.2 mm thickness (TR, 1805 ms; TE, 28 ms). Each scan acquired 86 volumes (155.2 s) and consisted of 28 experimental trials (jittered interstimulus interval with a mean of 4000 ms). A total of 12 scans were acquired to decrease scan specific noise in the analyses. An MPRAGE T1-weighted structural image (TR, 1800 ms; TE, 2.97 ms;  $1 \times 1 \times 1$  mm<sup>3</sup> resolution; 208 slices) was acquired for visualizing the associated anatomy. The structural image was coregistered to the mean of the EPI images. Image data were analyzed using SPM12 (Wellcome Trust Centre for Neuroimaging). The EPI volumes were spatially realigned and unwarped and then normalized to a standard MNI reference brain.

#### Representational similarity analysis

We used RSA (Kriegeskorte et al., 2008; Kriegeskorte and Kievit, 2013) to obtain behavioral and neural representational geometries of each individual. Behavioral representational geometries were measured by constructing a representational dissimilarity matrix (RDM) from identity categorization responses. The value in each cell in the behavioral RDM was calculated by the absolute difference in mean categorization responses between each pair of stimuli, with the constraint that the difference between face and nonface stimuli was normalized to 1. To localize

brain regions whose intrinsic representation structure resembles that of a behavioral RDM, we did a whole-brain searchlight analysis using the RSA toolbox (Nili et al., 2014). A spherical searchlight with a radius of 8 mm (257 voxels) was moved throughout the brain. At each searchlight location, the neural activity pattern for each of the seven stimuli were Pearson product-moment correlated with that for each of the seven stimuli, creating a  $7 \times 7$  brain RDM. This brain RDM was compared with the individual’s  $7 \times 7$  behavioral RDM using spearman correlation. The resulting correlation coefficient was converted to a z-value using Fisher transformation to conform to the statistical assumptions (normality) required for second-level parametric statistical tests. The results formed a continuous statistical whole-brain map reflecting how well the behavioral model fits in each of the local brain regions. Each subject’s Fisher-transformed whole-brain map was submitted to a second-level one-sample *t* test to identify voxels in which the correlation value was greater than zero. The resulting statistical map was thresholded at the cluster level of  $p < 0.05$ , FWE corrected for whole-brain for multiple comparisons.

The resulting clusters from the group analysis served as candidate brain regions for encoding task-relevant representations. To select regions of interest (ROIs) that best reflect each individual’s idiosyncratic patterns of activity, we used each of the clusters as a mask and did a second searchlight analysis for each subject, moving a spherical searchlight with a radius of 8 mm looking for a location inside the mask that had the highest correlation with the individual’s behavioral RDM with only five face stimuli ( $5 \times 5$  behavioral RDM). We used the behavioral RDM with only face stimuli, since we were interested in identifying each individual’s anatomical locations that best reflected the individual’s dissociations between two identities, rather than dissociations between face versus nonface stimuli. As a result, each individual had an ROI with varying center coordinates within a group cluster, from which brain RDMs were constructed. It is possible that selection of the ROIs was driven mostly by variation between faces and nonfaces, rather than variation between identities, since the face stimuli were more similar to each other than to nonfaces. The more specific hypothesis that the individual ROIs would reflect identity categorization was tested separately by correlating behavioral and brain RDMs with only five face stimuli within and between subjects (see “Statistical analyses” below).



**Figure 2.** Representational geometry in the identity categorization task. **A**, Mean identity categorization response for each stimulus averaged across participants. Error bars represent the SEM. **B**, Sigmoidal function fitted to each participant's identity categorization responses for the face morphs. **C**, Behavioral RDM averaged across participants.

### Statistical analyses

**Testing the significance of within- and between-subject correlations.** To quantify the unique aspects and the shared variance of the RDMs, within- and between-subject correlations between RDMs were calculated (Charest et al., 2014). When comparing an RDM ( $5 \times 5$ ) with the identity search efficiency matrix (ISEM;  $2 \times 4$ ), cells in the RDM that corresponded to the identical stimulus pairs in the ISEM were extracted and used for calculating correlations. First, a subject similarity matrix was constructed, by comparing each subject's one RDM of interest (e.g., behavioral RDM) with another RDM of interest (e.g., brain RDM) using the Pearson correlation. Thus, the diagonal entries of the subject similarity matrix indicate the within-subject correlations, and the off-diagonal entries indicate the between-subject correlations. Then, we computed the average within-subject and between-subject correlations between the two RDMs of interest. To control for the stimulus-set dependencies, the average of between-subject correlations was restricted to pairs of subjects who had viewed the identical stimulus set (female or male set). The significance of the within-subject and between-subject correlations was tested by a permutation test with randomized stimulus labels. Specifically, under the null hypothesis that all stimuli elicited the same response patterns, we rearranged one of the two RDMs of interest in a randomly permuted stimulus order for each subject, constructed a randomized subject similarity matrix correlating each subject's randomized RDM with another RDM of interest, and computed the average within-subject and between-subject correlations from the randomized subject similarity matrix. This step was repeated 10,000 times, creating a distribution of permuted within-subject and between-subject correlations under the null hypothesis that the two RDMs of interest were not related. We estimated the  $p$  values of the actual within- and between-subject correlations as the rank of the actual correlation in the permuted distribution. For instance, if the actual within-subject correlation was greater than any of the 10,000 permuted within-subject correlations, then the  $p$  value was estimated as  $p < 0.0001$ . We rejected the null hypothesis of unrelated RDMs if the actual correlations were higher than any of the top 500 permuted correlations ( $p < 0.05$ ). For correlations between RDMs involving brain ROIs,  $p$  values were Bonferroni corrected for four tests for the four ROIs.

**Testing the significance of the individuation index.** To test for the individually unique relationships between two RDMs of interest, an individuation index ( $i$ -index) was calculated by the difference between the average within-subject correlation and average between-subject correlation (Charest et al., 2014). The significance of the actual  $i$ -index was tested by a permutation test with randomized subject labels. Under the null hypothesis that the representational geometry was the same for all subjects, the subject labels of the vertical dimension of the subject similarity matrix were randomized, such that the label match between vertical and horizontal dimensions of the subject similarity matrix was destroyed. To control for the stimulus-set dependencies, the subject-label randomization was restricted to groups of subjects who had viewed the identical

stimulus set (female or male set). From the randomized subject similarity matrix, the average within- and between-subject correlations were computed, and the difference between the two correlations was stored as a randomized  $i$ -index. This step was repeated 10,000 times, creating a distribution of randomized  $i$ -indices under the null hypothesis that the relationship between the two RDMs of interest is the same regardless of subject labels. The  $p$  value of the actual  $i$ -index was estimated as the rank of the actual  $i$ -index in the randomized  $i$ -index distribution. For instance, if the actual  $i$ -index was greater than any of the 10,000 randomized  $i$ -indices, then the  $p$  value was estimated as  $p < 0.0001$ . We rejected the null hypothesis of no individually unique relationships if the actual  $i$ -index was higher than any of the top 500 randomized  $i$ -indices ( $p < 0.05$ ). For  $i$ -indices involving brain ROIs,  $p$  values were Bonferroni corrected for four tests for the four ROIs.

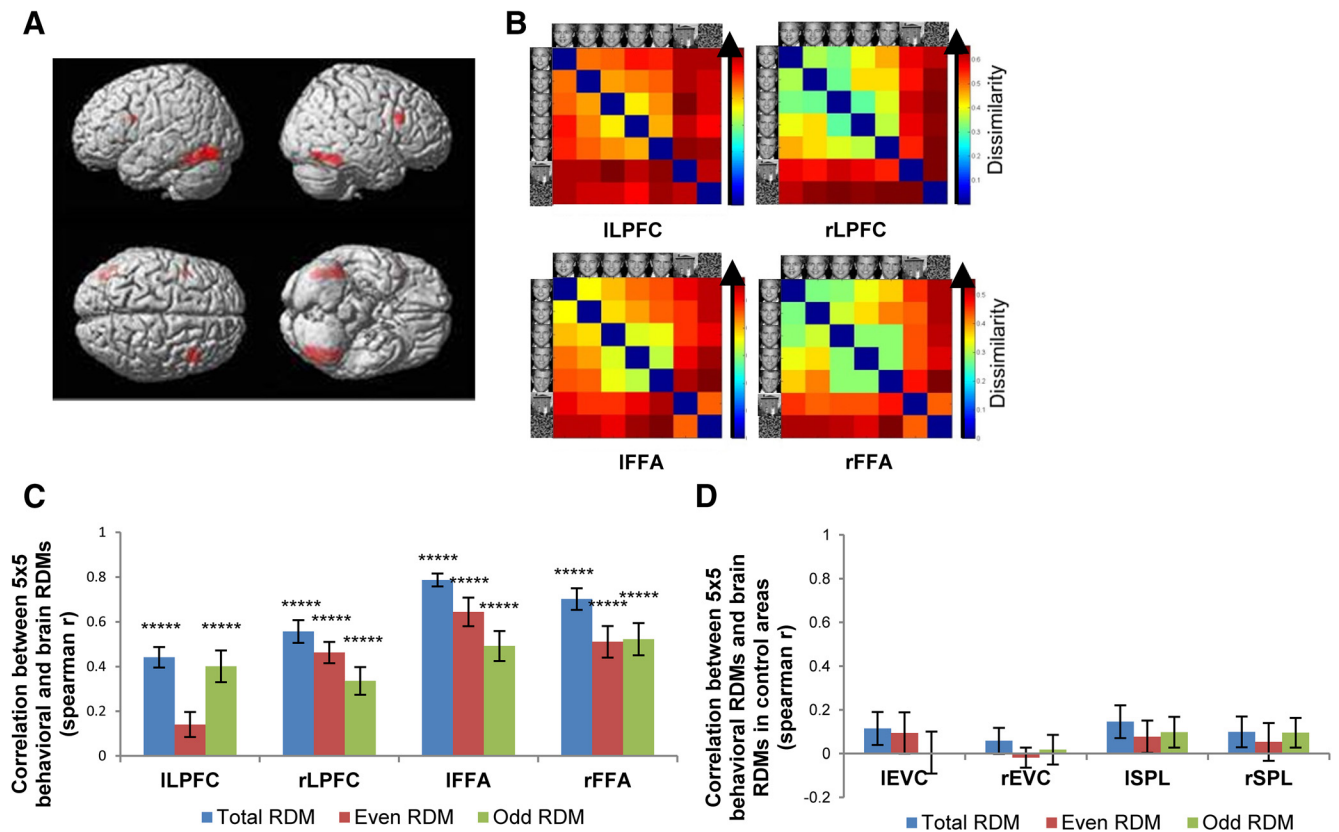
## Results

### Representational geometry in the identity categorization task

To obtain each individual's representational geometries for task stimuli, we measured brain activity using fMRI while participants performed an identity categorization task (Fig. 1A). The stimulus set included five face morphs that represented gradual transitions from one famous identity to another in steps of 25% ( $face_{100/0}$ ,  $face_{75/25}$ ,  $face_{50/50}$ ,  $face_{25/75}$ ,  $face_{0/100}$ ) and two nonface stimuli (a house, a scrambled image). There were two stimulus sets (the male set, the female set), and it was randomly chosen which stimulus set would be presented to each participant. Participants categorized each stimulus as one identity or the other, or as a nonface.

### Behavioral representational geometry

Preliminary analysis revealed no effect involving the stimulus set (male and female sets), and we therefore collapsed the data across this factor. The mean identity categorization responses for each of the seven stimuli are shown in Figure 2A. A repeated-measures ANOVA with stimulus as a within-subject factor and categorization response as the dependent measure revealed a significant main effect of stimulus ( $F_{(1.86, 22.36)} = 474.99$ ,  $p < 0.0001$ , Greenhouse-Geisser corrected). Pairwise comparisons showed significant differences between all pairs ( $p$  values  $< 0.0001$ ), except for between  $face_{100/0}$  and  $face_{75/25}$ ,  $face_{25/75}$  and  $face_{0/100}$ , and the house and scrambled image. This result indicates that the identity categorization of face stimuli did not change linearly, even though the five face morphs represented linear physical transitions from one original face to another. Confirming this, response profiles for the five face morphs fit by polynomial regression revealed a significant cubic component ( $F_{(1,12)} = 94.52$ ,  $p < 0.0001$ ), which captured a sigmoidal function, in ad-



**Figure 3.** Brain representational geometry in the identity categorization task. **A**, Brain maps showing the clusters identified in the  $7 \times 7$  whole-brain searchlight analysis, located in the left and right LPFC as well as in the left and right fusiform gyrus. **B**, Brain RDMs averaged across subjects in the left LPFC, right LPFC, left FFA, and right FFA. **C**, Mean within-subject correlation between the  $5 \times 5$  behavioral RDM and the  $5 \times 5$  brain RDM constructed from all the data, or on either half of the data in each of the four group clusters. The significance of the correlations was assessed by randomization of the stimulus labels.  $****p < 0.0005$  (Bonferroni corrected for four tests for the four ROIs). Error bars represent the SEM, estimated by bootstrap resampling of subjects. **D**, Mean within-subject correlation between the  $5 \times 5$  behavioral RDM and the  $5 \times 5$  brain RDM constructed from all the data, or on either half of the data in each of the four control areas (IEVC, rEVC, ISPL, and rSPL).

dition to a linear component ( $F_{(1,12)} = 682.50, p < 0.0001$ ). The observed nonlinear pattern in the response suggests that facial identity has a categorical representation, which is consistent with previous studies of the neural representation of faces (Rotshtein et al., 2005).

To closely examine each participant's response pattern, we fitted the following sigmoidal function to each subject's responses to the five face morphs:  $y = 1 + 1/(1 + e^{-[(x - \alpha)/\beta]})$ . In the sigmoidal function,  $\alpha$  indicates the point at which the stimulus is categorized as identity 1 or 2 with 50% probability, and  $\beta$  indicates the steepness of the slope of the function. The fitting results (Fig. 2B) indicated that each subject's mean categorization response (blue dots) for the most ambiguous face morph ( $\text{face}_{50/50}$ ) was widely dispersed between 1 and 2, causing big individual differences in the shape of the fitted sigmoidal function (red lines) for each participant's responses. Upon further examination, we found that the between-subjects variance in categorization for  $\text{face}_{50/50}$  was much greater (SD, 0.31) than those for other stimuli (average SD, 0.08). Indeed, the Mauchly's test of sphericity indicated that the assumption of sphericity was violated ( $\chi^2_{(20)} = 96.01, p < 0.0001$ ), indicating that the variances were not equal. Unlike the other stimuli, there was disagreement about whether  $\text{face}_{50/50}$  should be categorized as identity 1 or 2: about half of individuals categorized  $\text{face}_{50/50}$  as identity 1 more often, whereas the other half categorized it as identity 2 more often.

Next, we quantified the representational distance between

each pair of stimuli by constructing a representational dissimilarity matrix for each subject. The RDM is a symmetric matrix with a diagonal of zeroes and off-diagonal values that represent the dissimilarity in categorization responses for each pair of stimuli (Kriegeskorte et al., 2008; Fig. 2C). For instance, the representational dissimilarity between  $\text{face}_{100/0}$  and  $\text{face}_{50/50}$  was calculated as the absolute difference between the individual's mean categorization responses for  $\text{face}_{100/0}$  and  $\text{face}_{50/50}$ . The group RDM (Fig. 2C) averaged across individuals showed distinctive clusters for identity 1 ( $\text{face}_{100/0}$  and  $\text{face}_{75/25}$ ), identity 2 ( $\text{face}_{25/75}$  and  $\text{face}_{0/100}$ ), and nonface stimuli (house, scrambled), but  $\text{face}_{50/50}$  did not consistently cluster with either identity.

### Brain representational geometry

Since there was no a priori hypothesis regarding the brain regions that encode the categorical representations of task stimuli, we first conducted a whole-brain searchlight analysis (Nili et al., 2014) for each subject using the  $7 \times 7$  behavioral RDM of each subject as a model. The searchlight identified brain regions in which the neural activity pattern similarity for each pair of stimuli was highly correlated with the behavioral RDM. The group analysis identified four significant brain regions located in the left and right fusiform gyrus [including the fusiform face area (FFA)], as well as in the left (ILPFC) and right lateral prefrontal cortex (rLPFC; encompassing parts of dorsolateral and ventrolateral prefrontal cortices; Fig. 3A). Statistics and the MNI coordinates of the peak voxels in the four clusters are reported in Table 1.

**Table 1. Brain regions in which the neural activity pattern similarity was significantly correlated with the behavioral RDM in the whole-brain searchlight analysis**

Anatomical regions	x	y	z	# of voxels	z-value
Left fusiform gyrus	−38	−70	−14	317	6.93
Right fusiform gyrus	42	−58	−20	181	6.05
Left lateral prefrontal cortex	−44	8	26	16	5.63
Right lateral prefrontal cortex	46	20	24	76	5.62

Coordinates (x, y, and z) are reported in MNI space.

To select ROIs that best reflect each individual's idiosyncratic patterns of activity, we used each of the four clusters as a mask and did a second searchlight analysis for each subject. A spherical searchlight with a radius of 8 mm was moved inside the mask, looking for the location that had the highest correlation with the individual's behavioral RDM with only five face stimuli. We used the behavioral RDM with only face stimuli, since we were interested in identifying the brain locations that best reflected the individual's dissociations between two identities, rather than between face versus nonface stimuli. The resulting locations served as each individual's ROIs, from which brain RDMs were constructed (Fig. 3B). Similar to the behavioral RDM, the brain RDMs also showed roughly two clusters for the two face identities and another cluster for nonface stimuli.

To confirm that the selected individual ROIs contained meaningful information regarding variation between two identities, we performed a split-half cross-validation procedure. We split the fMRI data into halves (even and odd runs) and did a second searchlight analysis for each subject looking for a location inside the group mask that had the highest correlation with the individual's 5 × 5 behavioral RDM, separately for each half of the data. In this way, we obtained two ROIs per cluster for each subject: an ROI that showed the best correlations with the data in even runs and another that showed the best correlations with the data in odd runs. Next, we constructed the brain RDM within each ROI with the activation patterns in the other half of the data. If the selected ROIs that best correlated with the one half of the data contained only ranked noise, then the brain RDMs constructed with the other half of the data would not have meaningful relationships with the behavioral RDMs. We tested this by calculating within-subject correlations between the behavioral and brain RDMs, separately for RDMs constructed based on even, odd, or total data. The results revealed that there were significant positive correlations between each individual's behavioral and brain RDMs in all of the four areas (Fig. 3C). The only exception was a statistically marginal ( $p < 0.1$ ) correlation observed in the left lateral prefrontal cortex RDM constructed based on the data in even runs, which indicated the existence of variability between runs. However, the consistent pattern of correlation results observed across all four brain regions argues against the possibility that the individually selected ROIs contained only sorted noise.

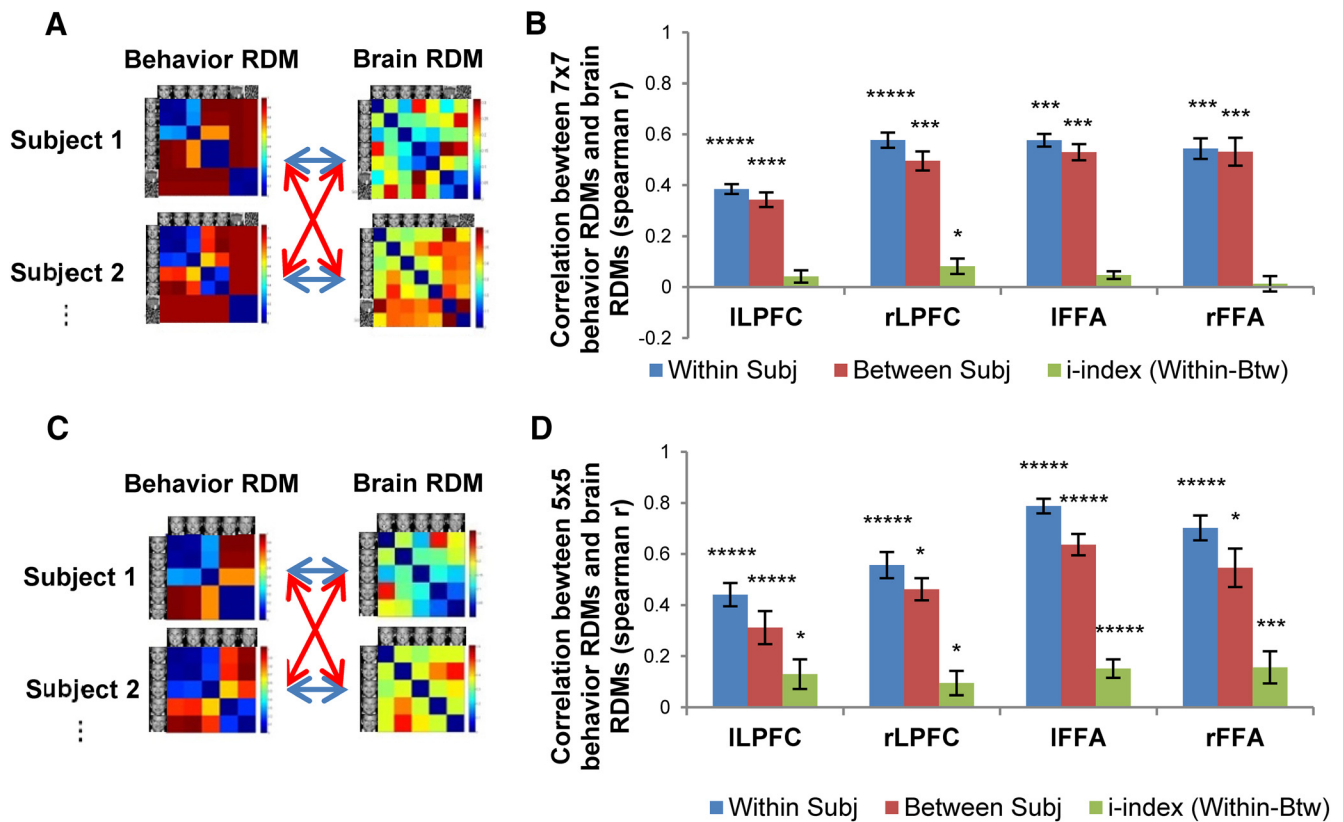
As further test of the reliability of the previous analysis, we selected four "control" areas in the left early visual cortex (LEVC), right early visual cortex (REVC), left superior parietal lobe (LSPL), and right superior parietal lobe (RSPL). These control areas were created to be similar in size to the original clusters obtained in the 7 × 7 group searchlight analysis. These control areas were then treated identically to the data within the main group clusters (i.e., individual ROIs that best correlated with the individual's behavioral RDM were selected, behavioral and brain RDMs were correlated using all the data or with either the odd or even half of the data). The results indicated that in all of the four control areas, the

correlations between the behavioral RDMs and brain RDMs were not significant, regardless of whether the RDM was based on all the data or on either half of the data (Fig. 3D). This is in stark contrast to the significant correlations observed in the ROIs selected from the group clusters [lLPFC, rLPFC, left FFA (lFFA), and right FFA (rFFA)] and argues against the possibility that the ROIs selected from the group clusters contained only sorted noise. Having consistently observed the validity of the individual ROIs selected from the group clusters, we used the RDMs based on all the data (total RDM) in subsequent analyses, rather than the RDMs based on either half of the data, for statistical power and simplicity.

To examine the specificity of the relationship between behavioral and brain representational geometries within and across subjects, we constructed a subject similarity matrix, which compares each subject's 7 × 7 behavioral RDM with each subject's 7 × 7 brain RDM using the Pearson correlation (Fig. 4A). The diagonal entries of the subject similarity matrix indicated the within-subject correlations between behavioral RDMs and brain RDMs, and the off-diagonal entries indicated the between-subject correlations. The significance of within- and between-subject correlations was tested using a stimulus-label randomization test, by randomly permuting stimulus order for each subject 10,000 times, creating a distribution of permuted within-subject and between-subject correlations under the null hypothesis that the two RDMs of interest were not related (see Materials and Methods). This was done separately for each ROI. The results indicated that behavioral RDMs and brain RDMs were significantly correlated in all four brain ROIs (Fig. 4B) both within subjects (lLPFC,  $r = 0.38$ ,  $p < 0.0005$ ; rLPFC,  $r = 0.58$ ,  $p < 0.0005$ ; lFFA,  $r = 0.58$ ,  $p < 0.005$ ; rFFA,  $r = 0.54$ ,  $p < 0.005$ ) and between subjects (lLPFC,  $r = 0.34$ ,  $p < 0.001$ ; rLPFC,  $r = 0.50$ ,  $p < 0.005$ ; lFFA,  $r = 0.53$ ,  $p < 0.005$ ; rFFA,  $r = 0.53$ ,  $p < 0.005$ ). This indicates that the shared variance in the behavioral categorization of the seven task stimuli was also reflected in the brain representational geometries, confirming that these four brain regions encode the generic categorical representations of the task stimuli.

To quantify the individually unique relationships between behavioral and brain representational geometries, we compared the within-subject correlations with the between-subject correlations (Fig. 4B). If there is an individually unique component to the relationship between behavioral and brain RDMs, then within-subject correlations should be higher than between-subjects correlations. Thus, an individuation index was calculated by the difference between within-subject correlation and between-subject correlation (average within-subject  $r$  value minus average between-subject  $r$  value; Charest et al., 2014). The significance of the i-index was tested using a subject-label randomization test by randomly permuting the subject labels of the vertical dimension of the subject similarity matrix 10,000 times, creating a distribution of randomized i-indices under the null hypothesis that the relationship between the two RDMs of interest is the same regardless of subject labels (see Materials and Methods). Among the four ROIs, the i-index was significant only in rLPFC (i-index = 0.08,  $p < 0.05$ ), indicating that the relationship between behavioral and brain representational geometries was individually unique in rLPFC. Considering that the biggest source of individual differences in behavioral representational geometry was categorization of face<sub>50/50</sub>, this suggested that rLPFC contained a unique representation of the task stimuli that reflected an individual's decision bias to categorize the most ambiguous face as either identity.

It is possible that the selected ROIs could have been driven mostly by variation between face and nonfaces rather than vari-



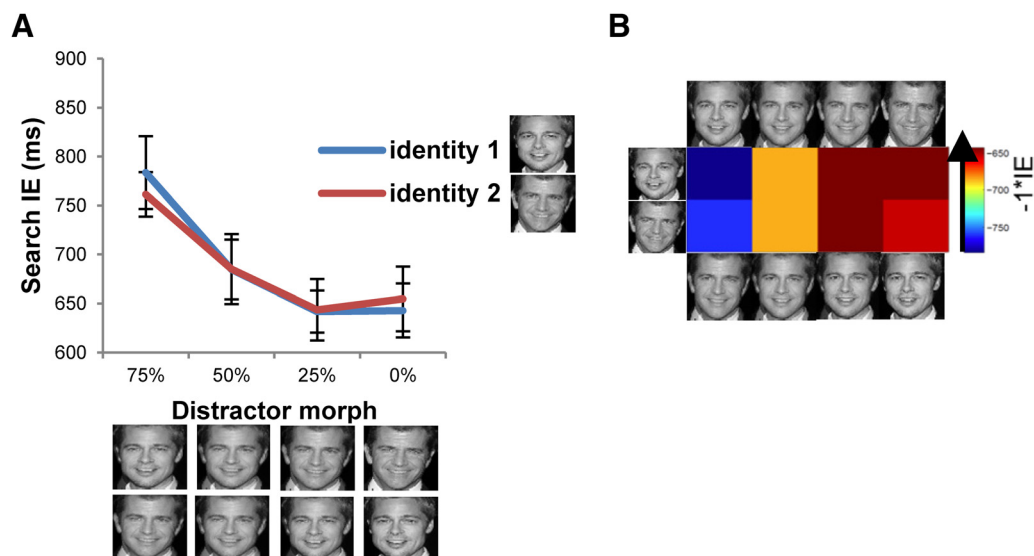
**Figure 4.** The relationship between the behavioral RDM and the brain RDM in each of the four ROIs. **A**, Each subject's  $7 \times 7$  behavioral RDM was correlated with each subject's  $7 \times 7$  brain RDM. Blue arrows indicate within-subject correlations, and red arrows indicate between-subject correlations. **B**, Mean within-subject correlation, between-subject correlation, and i-index, between the  $7 \times 7$  behavioral RDM and the  $7 \times 7$  brain RDM in each of the four ROIs. The significance of the within- and between-subject correlations was assessed by randomization of the stimulus labels. The significance of the i-index was assessed by randomization of the subject labels. \* $p < 0.05$ ; \*\*\* $p < 0.005$ ; \*\*\*\* $p < 0.001$ ; \*\*\*\*\* $p < 0.0005$  (Bonferroni corrected for four tests for the four ROIs). Error bars represent the SEM, estimated by bootstrap resampling of subjects. **C**, Each subject's  $5 \times 5$  behavioral RDM (only the five face morphs) was correlated with each subject's  $5 \times 5$  brain RDM. Blue arrows indicate within-subject correlations, and red arrows indicate between-subject correlations. **D**, Mean within-subject correlation, between-subject correlation, and i-index between the  $5 \times 5$  behavioral RDM and the  $5 \times 5$  brain RDM in each of the four ROIs.

ation between identity 1 and identity 2, since face morphs were physically more similar to each other than to nonfaces. If the ROIs did not contain information about the variation of identities, then it should be impossible to find significant correlations between behavioral categorization pattern and neural activation patterns for the five face morphs. We directly tested this hypothesis by correlating the behavioral RDMs ( $5 \times 5$ ) and the brain RDMs ( $5 \times 5$ ) with only five face stimuli (Fig. 4C). The results indicated that behavioral RDMs and brain RDMs were significantly correlated in all four brain ROIs (Fig. 4D) both within subjects (ILPFC,  $r = 0.44$ ,  $p < 0.0005$ ; rLPFC,  $r = 0.56$ ,  $p < 0.0005$ ; IFFA,  $r = 0.79$ ,  $p < 0.0005$ ; rFFA,  $r = 0.70$ ,  $p < 0.0005$ ) and between subjects (ILPFC,  $r = 0.31$ ,  $p < 0.0005$ ; rLPFC,  $r = 0.46$ ,  $p < 0.05$ ; IFFA,  $r = 0.64$ ,  $p < 0.0005$ ; rFFA,  $r = 0.55$ ,  $p < 0.05$ ). This strongly indicates that the selected brain ROIs encoded not only the variation between faces and nonfaces, but also the variation across face morphs. To quantify the individually unique relationships between the  $5 \times 5$  behavioral and brain representational geometries, the i-index in each of the four ROIs was calculated using the same procedure described before. The i-index was significant in all four ROIs (ILPFC, i-index = 0.13,  $p < 0.05$ ; rLPFC, i-index = 0.09,  $p < 0.05$ ; IFFA, i-index = 0.15,  $p < 0.0005$ ; rFFA, i-index = 0.16,  $p < 0.005$ ), indicating that the relationship between behavioral and brain representational geometries with the five face stimuli was individually unique in all four ROIs (Fig. 4D). Particularly, the idiosyncratic components

of the categorization behavior were reliably decoded in rLPFC across different analyses ( $7 \times 7$  and  $5 \times 5$  RDMs), suggesting that rLPFC plays a key role in encoding unique categorical representations. In summary, the shared and unique variances in the behavioral representational geometry were encoded in stimulus-specific perceptual areas as well as in the prefrontal cortex. We next examined whether the idiosyncratic difference in representational geometry would predict the contents of the attentional template during an identity search task; if the categorization RDM is replicated in the attentional template, we would expect face morphs to compete with the target for attention commensurate with their probability of being categorized as the target identity.

#### Identity search task

An identity search task (Fig. 1B) was performed outside the scanner to observe how the representational geometry obtained in the identity categorization task influenced the efficiency of attentional selection. The five face morphs each participant saw in the identity categorization task were used as the target and distractor stimuli. The target (one of the two original 100% faces) was cued on each trial, and therefore the contents of the attentional template changed between identity 1 or 2 on a trial-by-trial basis. The distractor face was one of the other face morphs. Although the physical similarity of the distractor face to the target was manipulated (75, 50, 20, and



**Figure 5.** Results in the identity search task. **A**, Search inverse efficiency for each target (identity 1, identity 2) as a function of the distractor similarity (75, 50, 25, or 0%) to the target. Error bars represent the SEM. **B**, ISEM in which the value in each cell indicates search IE multiplied by  $-1$  for each pair of target and distractor faces.

0%), we hypothesized that the psychological similarity would be predicted by the representational geometry from the categorization task. The task was to report a property of a probe stimulus that appeared at the same location of the target face (see Materials and Methods).

To combine the effects in RT and accuracy rate, the inverse efficiency (IE; correct RT/accuracy rate) was calculated for each condition and subject (Townsend and Ashby, 1983). The identity search data were analyzed in two stages: First, we examined the overall behavioral pattern across individuals using ANOVA with target face and distractor similarity as fixed effects. Second, we constructed an identity search efficiency matrix for each individual that characterized the distractibility of each face morph for each target, which could be compared with the categorization RDMs of each individual.

The data were entered into a repeated-measures ANOVA with target face (identity 1, identity 2) and distractor similarity to the target (75, 50, 25, and 0%) as within-subject factors and IE as the dependent measure (Fig. 5A). The analysis revealed a significant main effect of distractor similarity to the target ( $F_{(3, 36)} = 15.30, p < 0.0001$ ), with the mean IE decreasing as the distractor similarity to the target decreased. Specifically, pairwise comparisons revealed significant differences between all pairs of distractors ( $p$  values  $< 0.05$ ), except for between the distractors with 25 and 0% similarity to the target. This result is consistent with the previous results where the similarity between targets and nontargets plays an important role in determining search efficiency (Duncan and Humphreys, 1989; Nagy and Sanchez, 1990; Becker, 2011). Notably, the 25 and 0% distractor faces were uniformly categorized as a different identity from the target face in the categorization task. Thus, the lack of difference in search IEs between 25 and 0% distractor faces suggests the possibility that distractors interfere with search uniformly regardless of physical similarity, once they cross a categorical boundary. The main effect of target face and the interaction between target face and distractor similarity were not significant ( $F$  values  $< 1$ ).

To compare the identity search data with the categorization RDMs, the IE data from each individual were transformed into an ISEM (Fig. 5B). Each cell of the ISEM contained the IE data for a single target–distractor face pair multiplied by  $-1$ , such that

values closer to 0 represent more efficient search. Thus, the value in each cell reflected the efficiency of search with the given pair of target and distractor faces. We hypothesized that the greater the dissimilarity value in the categorization RDM for a given pair of stimuli, the greater the search efficiency value would be in the ISEM for that pair of target and distractor faces.

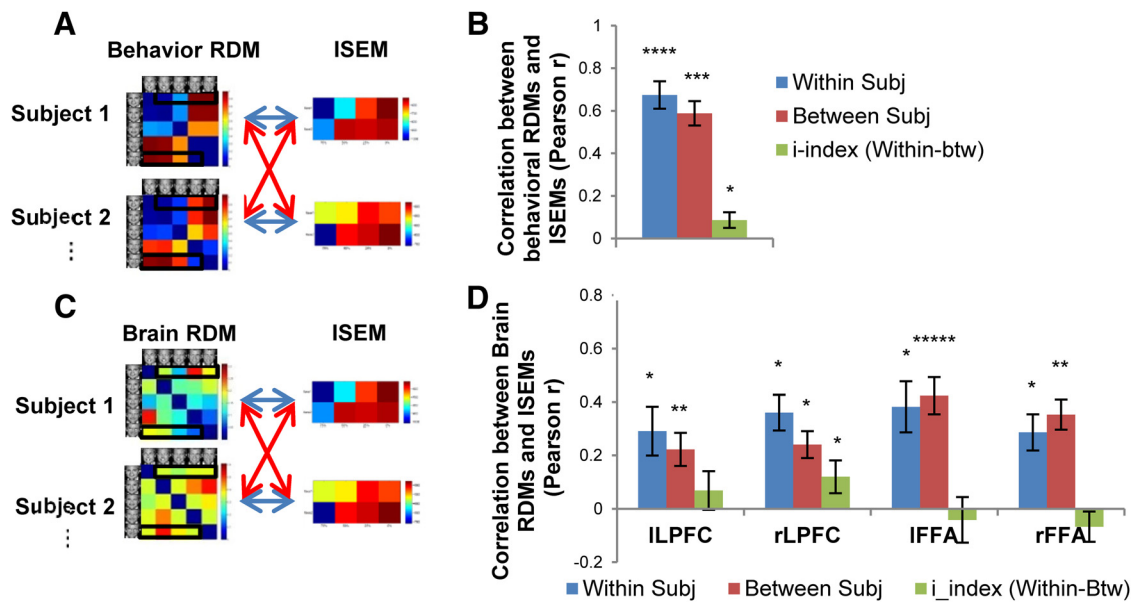
### Representational geometry predicts search performance

To investigate whether the categorical representational geometry can predict search performance, we compared behavioral and brain RDMs from the identity categorization task with the ISEM (Fig. 6A, C). To match the ISEM data to the categorization RDM, cells in the RDM that corresponded to the identical stimulus pairs in the ISEM were extracted and used for calculating correlations. For both data from brain and behavior, a subject similarity matrix was constructed, which compared each subject's categorization RDM with each subject's ISEM using the Pearson correlation. The significance of within- and between-subject correlations was tested using a stimulus-label randomization test (see Materials and Methods).

The results indicated that behavioral RDMs and ISEMs were highly correlated both within subjects (average  $r = 0.67, p < 0.001$ ) and between subjects (average  $r = 0.59, p < 0.005$ ; Fig. 6B). Similarly, the brain RDMs and ISEMs were significantly correlated in all four ROIs (Fig. 6D) both within subjects (ILPFC, average  $r = 0.29, p < 0.05$ ; rLPFC, average  $r = 0.36, p < 0.05$ ; IFFA, average  $r = 0.38, p < 0.05$ ; rFFA, average  $r = 0.29, p < 0.05$ ) and between subjects (ILPFC, average  $r = 0.22, p < 0.01$ ; rLPFC, average  $r = 0.24, p < 0.05$ ; IFFA, average  $r = 0.42, p < 0.0005$ ; rFFA, average  $r = 0.35, p < 0.01$ ). The representational geometry from behavior and all four brain ROIs could predict the shared variances in the identity search performance, indicating that greater representational dissimilarity between two stimuli in the categorization RDM predicted more efficient search when one was the target and the other the distractor.

To test whether the within-subject correlation between representational geometry and identity search efficiency was significantly greater than the between-subjects correlations, we calculated an  $i$ -index that reflects the degree to which the relationship between





**Figure 6.** The relationship between representational geometries and visual search efficiency. **A**, Each subject's behavioral RDM was correlated with each subject's ISEM. Blue arrows indicate within-subject correlations, and red arrows indicate between-subject correlations. **B**, Mean within-subject correlation, between-subject correlation, and i-index between the behavioral RDM and the ISEM. The significance of the within- and between-subject correlations was assessed by randomization of the stimulus labels. The significance of the i-index was assessed by randomization of the subject labels. **C**, Each subject's brain RDM was correlated with each subject's ISEM. **D**, Mean within-subject correlation, between-subject correlation, and i-index between the ISEM and the brain RDM in each of the four ROIs. \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.005$ ; \*\*\*\* $p < 0.001$ ; \*\*\*\*\* $p < 0.0005$  (Bonferroni corrected for four tests for the four ROIs). Error bars represent the SEM, estimated by bootstrap resampling of subjects.

representational geometry and identity search efficiency is unique to individuals, and tested the significance of the i-index using the same methods described before. The i-index was significant for the behavioral RDM and ISEM (i-index = 0.09;  $p < 0.05$ ), indicating that the individually unique aspects of behavioral categorization successfully predicted idiosyncratic elements of identity search performance (Fig. 6B). In the brain, on the other hand, the i-index was significant only in rLPFC (i-index = 0.12;  $p < 0.05$ ) among the four ROIs (Fig. 6D). This indicates that the representational geometry in rLPFC reflects the individually unique aspects of categorization, and it can also predict idiosyncratic variance of identity search performance.

The most likely source of the individually unique relationship between the representational geometry and identity search efficiency was the categorization of face<sub>50/50</sub>, the most ambiguous morph. Recall that we previously found categorization of face<sub>50/50</sub> to be the most variable among subjects, suggesting that individuals had different biases to categorize face<sub>50/50</sub> as either identity 1 or identity 2. Therefore, the representational geometry of face<sub>50/50</sub> provided a window of opportunity to discover how individual difference in brain representational similarity is related to the contents of individual's target template. To test whether the variance in categorization predicted distractability during target search, we correlated each individual's categorization bias with face<sub>50/50</sub> distractability (i.e., how much face<sub>50/50</sub> interfered with search for each target identity). We hypothesized that individuals with a greater bias to categorize face<sub>50/50</sub> as identity 1 should also be more distracted by face<sub>50/50</sub> when searching for identity 1 compared to identity 2.

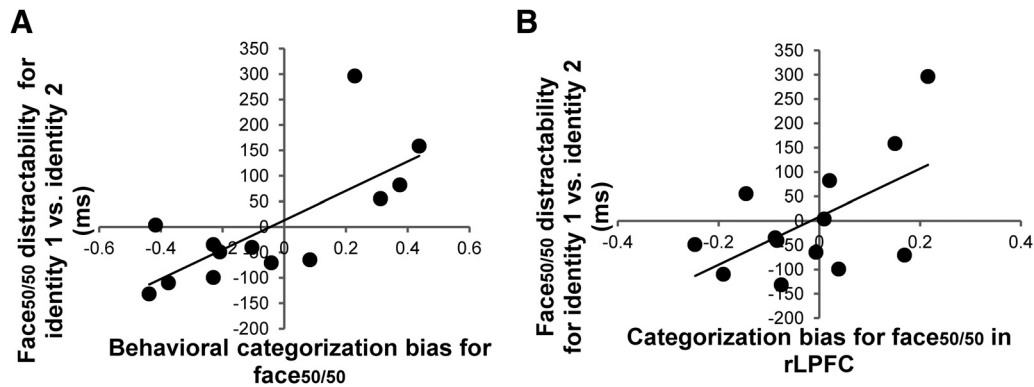
The categorization bias for face<sub>50/50</sub> was calculated by subtracting each individual's mean categorization response for face<sub>50/50</sub> from 1.5 (the point of no bias), with positive values indicating bias toward identity 1 and negative values toward identity 2. Face<sub>50/50</sub> distractability was calculated by subtracting each individual's mean IE on trials with identity 2 targets from the mean IE for identity 1 targets, when the distractor was face<sub>50/50</sub>.

Hence, positive values indicated more distraction by face<sub>50/50</sub> when searching for identity 1 compared to identity 2, and negative values indicated the reverse. The correlation between each individual's behavioral categorization bias and face<sub>50/50</sub> distractability (Fig. 7A) was significant and positive ( $r = 0.73$ ,  $p < 0.005$ ), indicating that the degree of behavioral categorization bias for face<sub>50/50</sub> predicted the extent of interference caused by face<sub>50/50</sub> for different target identities: as hypothesized, individuals with a greater bias toward identity 1 were more distracted by face<sub>50/50</sub> when searching for identity 1 versus identity 2 and vice versa.

We next used the same approach using the activity patterns in rLPFC to calculate the brain categorization bias for face<sub>50/50</sub>. Each individual's brain categorization bias for face<sub>50/50</sub> was calculated by subtracting the representational dissimilarity between face<sub>50/50</sub> and face<sub>100/0</sub> from that of face<sub>50/50</sub> and face<sub>0/100</sub>. Thus, positive values indicated bias toward identity 1, and negative values toward identity 2. The direct relationship was examined by correlating each individual's brain categorization bias and the face<sub>50/50</sub> distractability in the identity search task (Fig. 7B). There was a significant positive correlation ( $r = 0.57$ ,  $p < 0.05$ ), indicating that the extent to which face<sub>50/50</sub> interfered with search was predicted by each individual's unique brain representational geometry for face<sub>50/50</sub> in rLPFC: individuals who had more similar brain representations for face<sub>50/50</sub> and identity 1 were also more distracted by face<sub>50/50</sub> when searching for identity 1 versus identity 2 and vice versa. This replicates the results observed with the behavioral categorization bias, further confirming the close relationship between behavior and neural representational geometry in rLPFC.

## Discussion

Previous studies have shown that the representation of task-related information determines the speed of attentional selection and distractor suppression (Duncan and Humphreys, 1989;



**Figure 7.** *A*, A scatter plot (with best-fitting regression line) illustrating the relationship between the magnitude of behavioral categorization bias for face<sub>50/50</sub> and the difference in distraction by face<sub>50/50</sub> for target identity 1 versus identity 2. *B*, A scatter plot (with best-fitting regression line) illustrating the relationship between the magnitude of brain categorization bias for face<sub>50/50</sub> in rLPFC and the difference in distraction by face<sub>50/50</sub> for target identity 1 versus identity 2.

Wolfe, 1994; Becker, 2011; Hout and Goldinger, 2015). Although the task-relevant representations were observed in multiple regions of the brain, including stimulus-specific perceptual areas and higher cognitive areas (Kanwisher et al., 1997; Epstein and Kanwisher, 1998; Freedman et al., 2001; Haxby et al., 2001; Sarma et al., 2016), most of the prior studies were interested in the common representational structure across individuals, leaving the neural substrates of individually unique representations unknown. Here, we tested where in the brain the idiosyncratic and the common portions of the task-relevant representations are encoded and how individual differences in the representational geometry affect efficiency in attentional allocation. The results demonstrated that the activity patterns in the rLPFC reflected the unique aspects as well as the common portion of the representational geometry and, moreover, predicted the individual variances in target search. This result suggests that the rLPFC plays a crucial role in encoding individually unique representations and that idiosyncrasies in the perception of objects influence the efficiency of attentional allocation.

The current study provides two important novel findings that elucidate the underlying neural mechanisms of how task-relevant representations are encoded and influence the individual's efficiency in attentional selection. First, the individually unique aspects of the representational geometries of task stimuli were reflected in behavioral categorization, and the unique encoding consistent with the behavioral responses was observed in the lateral prefrontal cortex as well as in the stimulus-specific perceptual area (FFA). In the categorization task, all participants commonly demonstrated categorical representations for the two identities, but there were individual differences in the shape of the sigmoidal function, based on whether the most ambiguous face morph was classified as identity 1 or 2 more often. This difference in the inclusiveness of face<sub>50/50</sub> within the representation of identity 1 or 2 can be thought of as the width of a representational tuning curve for each identity. Patterns of activation in the LPFC as well as in face-selective areas in the ventral temporal cortex were highly consistent with the categorical representation of face identity, in line with the previous studies showing the common representational structure across individuals in stimulus-specific perceptual areas and higher cognitive areas (Martin et al., 1996; Gauthier and Tarr, 1997; Kanwisher et al., 1997; Freedman and Assad, 2006; Crowe et al., 2013). Most importantly, the RDMs in LPFC and FFA had individually unique relationships with the behavioral RDMs, suggesting that these areas also encode the unique aspects of representational geometries. In particular, the individually

unique components of the categorization behavior were reliably decoded in rLPFC across different analyses ( $7 \times 7$  and  $5 \times 5$  RDMs).

Second, idiosyncrasies in the representational geometries predicted the variance in the efficiency of search for a face identity. The behavioral and brain representational geometries were highly correlated with the search performance, indicating that the common patterns of categorization responses could predict the shared variance in the search performance. This replicates previous research that search becomes inefficient as the similarity between target and nontargets increases (Neider and Zelinsky, 2006; Zelinsky, 2008; Hwang et al., 2009; Becker, 2011). In addition to the shared variance in the identity search performance, representational geometries in rLPFC also had individually unique relationships with the search performance: the extent to which a distractor interfered with search was predicted by each individual's unique brain representational geometry for the most ambiguous face morph in rLPFC. This is consistent with the results that the individually unique aspects of the behavioral response were reliably reflected only in rLPFC and provides a novel insight that the extent to which a distractor interferes with search is not only dependent on the physical similarity to the target, but also dependent on each individual's unique representational geometry. Our data suggest that rLPFC plays a key role in encoding the unique aspects of the task-relevant representations and also in modulating the efficiency of attentional allocation based on its similarity structure.

Our data indicated that although both LPFC and the stimulus-specific perceptual areas in the ventral temporal cortex encoded individually unique aspects of identity categorization, only LPFC predicted the idiosyncratic variance in performance during the attentionally demanding identity search task. This result appears on the surface to conflict with a previous study by Charest et al. (2014) that demonstrated activity patterns in ventral temporal cortex correlated with the individually unique object similarity judgments. However, the results are likely due to the many differences in the characteristics of the stimuli and the task demands between studies. First, Charest et al. (2014) had a priori hypotheses regarding the brain locations that encode idiosyncratic object representations and restricted their analyses to the ventral temporal cortex and early visual cortex. On the other hand, we performed a whole-brain searchlight analysis first to look for all the brain regions that significantly correlated with the behavioral categorization and identified both LPFC and a subregion of VTC. Second, Charest et al. (2014) used a large VTC ROI that encompassed brain regions specialized for encoding different classes of

stimuli (e.g., FFA, PPA, EBA, etc.), which was appropriate for their large set of heterogeneous objects (faces, places, bodies, etc.), but we focused on smaller ROIs that were more appropriate for our interest in individual differences in response to a limited set of homogeneous face stimuli. Finally, unlike their similarity judgment task, our task required identity categorization, which added a decision component to stimulus processing. This could explain the greater sensitivity of rLPFC to our task.

This raises the interesting possibility that representations in LPFC reflect task-general categorical representations necessary for action and identity selection (Gottlieb, 2007; Avidan and Behrmann, 2009; Thomas et al., 2009; Bichot et al., 2015; Sarma et al., 2016) that are absent in the long-term semantic representations of objects in ventral temporal cortex. Therefore, we conclude that rLPFC is at least an important part of the neural mechanisms that maintain unique categorical representations, and moreover, the same representation is instantiated as the attentional template in working memory during target search. This suggests that the “tuning” of the target template is composed of individually unique category representations. This conclusion is consistent with the biased competition model of attention (Desimone and Duncan, 1995) in which the contents of the target template held in working memory is the source of the attentional bias instantiated within sensory cortex (Peelen et al., 2009; Schafer and Moore, 2011). Here, we extend those results by demonstrating that individually unique category representations decoded from rLPFC can predict performance in an independent search task, suggesting that the contents of the attentional template reflect the categorical structure of information that is specific to an individual.

Together, our results provide clear evidence that task-relevant representations encoded in the prefrontal cortex reflect idiosyncrasies in the person’s perceptual experiences and also predict the unique portion of the variance in attentional selection. This work demonstrates that subtle individual differences in the representational structure can be captured using fMRI and RSA, and contributes to understanding the neural mechanisms underlying the relationship between categorical representations and attentional allocation at the level of individuals.

## References

- Alexander RG, Zelinsky GJ (2011) Visual similarity effects in categorical search. *J Vis* 11(8) pii: 9. Medline
- Althaus N, Westermann G (2016) Labels constructively shape object categories in 10-month-old infants. *J Exp Child Psychol* 151:5–17. Medline
- Avidan G, Behrmann M (2009) Functional MRI reveals compromised neural integrity of the face processing network in congenital prosopagnosia. *Curr Biol* 19:1146–1150. CrossRef Medline
- Becker SI (2011) Determinants of dwell time in visual search: similarity or perceptual difficulty? *PLoS One* 6:e17740. CrossRef Medline
- Bichot NP, Heard MT, DeGennaro EM, Desimone R (2015) A source for feature-based attention in the prefrontal cortex. *Neuron* 88:832–844. CrossRef Medline
- Bundesden C (1990) A theory of visual attention. *Psychol Rev* 97:523–547. CrossRef Medline
- Carlisle NB, Arita JT, Pardo D, Woodman GF (2011) Attentional templates in visual working memory. *J Neurosci* 31:9315–9322. CrossRef Medline
- Charest I, Kievit RA, Schmitz TW, Deca D, Kriegeskorte N (2014) Unique semantic space in the brain of each beholder predicts perceived similarity. *Proc Natl Acad Sci U S A* 111:14565–14570. CrossRef Medline
- Crowe DA, Goodwin SJ, Blackman RK, Sakellaridi S, Sponheim SR, MacDonald AW 3rd, Chafee MV (2013) Prefrontal neurons transmit signals to parietal neurons that reflect executive control of cognition. *Nat Neurosci* 16:1484–1491. CrossRef Medline
- Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222. CrossRef Medline
- Duncan J, Humphreys GW (1989) Visual search and stimulus similarity. *Psychol Rev* 96:433–458. CrossRef Medline
- Epstein R, Kanwisher N (1998) A cortical representation of the local visual environment. *Nature* 392:598–601. CrossRef Medline
- Feredoes E, Tononi G, Postle BR (2007) The neural bases of the short-term storage of verbal information are anatomically variable across individuals. *J Neurosci* 27:11003–11008. CrossRef Medline
- Folk CL, Remington RW, Johnston JC (1992) Involuntary covert orienting is contingent on attentional control settings. *J Exp Psychol Hum Percept Perform* 18:1030–1044. CrossRef Medline
- Freedman DJ, Riesenhuber M, Poggio T, Miller EK (2001) Categorical representation of visual stimuli in the primate prefrontal cortex. *Science* 291:312–316. CrossRef Medline
- Freedman DJ, Assad JA (2006) Experience-dependent representation of visual categories in parietal cortex. *Nature* 443:85–88. CrossRef Medline
- Gauthier I, Tarr MJ (1997) Becoming a “Greeble” expert: exploring mechanisms for face recognition. *Vision Res* 37:1673–1682. CrossRef Medline
- Gazzaley A, Nobre AC (2012) Top-down modulation: bridging selective attention and working memory. *Trends Cogn Sci* 16:129–135. CrossRef Medline
- Goldstone RL, Lippa Y, Shiffrin RM (2001) Altering object representations through category learning. *Cognition* 78:27–43. CrossRef Medline
- Gottlieb J (2007) From thought to action: the parietal cortex as a bridge between perception, action, and cognition. *Neuron* 53:9–16. CrossRef Medline
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P (2001) Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* 293:2425–2430. CrossRef Medline
- Hout MC, Goldinger SD (2015) Target templates: the precision of mental representations affects attentional guidance and decision-making in visual search. *Atten Percept Psychophys* 77:128–149. CrossRef Medline
- Hout MC, Godwin HJ, Fitzsimmons G, Robbins A, Menneer T, Goldinger SD (2016) Using multidimensional scaling to quantify similarity in visual search and beyond. *Atten Percept Psychophys* 78:3–20. CrossRef Medline
- Hwang AD, Higgins EC, Pomplun M (2009) A model of top-down attentional control during visual search in complex scenes. *J Vis* 9(5):25 1–18. Medline
- Kanwisher N, McDermott J, Chun MM (1997) The fusiform face area: a module in human extrastriate cortex specialized for face perception. *J Neurosci* 17:4302–4311. Medline
- Kay KN, Naselaris T, Prenger RJ, Gallant JL (2008) Identifying natural images from human brain activity. *Nature* 452:352–355. CrossRef Medline
- Kriegeskorte N, Kievit RA (2013) Representational geometry: integrating cognition, computation, and the brain. *Trends Cogn Sci* 17:401–412. CrossRef Medline
- Kriegeskorte N, Mur M (2012) Inverse MDS: inferring dissimilarity structure from multiple item arrangements. *Front Psychol* 3:245. Medline
- Kriegeskorte N, Formisano E, Sorger B, Goebel R (2007) Individual faces elicit distinct response patterns in human anterior temporal cortex. *Proc Natl Acad Sci U S A* 104:20600–20605. CrossRef Medline
- Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis – connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4. CrossRef Medline
- Lee SH, Kravitz DJ, Baker CI (2013) Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nat Neurosci* 16:997–999. CrossRef Medline
- Malcolm GL, Henderson JM (2009) The effects of target template specificity on visual search in real-world scenes: evidence from eye movements. *J Vis* 9(11):8 1–13. Medline
- Martin A, Wiggs CL, Ungerleider LG, Haxby JV (1996) Neural correlates of category-specific knowledge. *Nature* 379:649–652. CrossRef Medline
- Martinez-Trujillo JC, Treue S (2004) Feature-based attention increases the selectivity of population responses in primate visual cortex. *Curr Biol* 14:744–751. CrossRef Medline
- Moore CM, Egeth H (1998) How does feature-based attention affect visual processing? *J Exp Psychol Hum Percept Perform* 24:1296–1310. CrossRef Medline
- Mur M, Meys M, Bodurka J, Goebel R, Bandettini PA, Kriegeskorte N (2013) Human object-similarity judgments reflect and transcend the primate-IT object representation. *Front Psychol* 4:128. Medline

- Nagy AL, Sanchez RR (1990) Critical color differences determined with a visual search task. *J Opt Soc Am A* 7:1209–1217. [CrossRef Medline](#)
- Neider MB, Zelinsky GJ (2006) Searching for camouflaged targets: Effects of target-background similarity on visual search. *Vision Res* 46:2217–2235. [CrossRef Medline](#)
- Nili H, Wingfield C, Walther A, Su L, Marslen-Wilson W, Kriegeskorte N (2014) A toolbox for representational similarity analysis. *PLoS Comput Biol* 10:e1003553. [CrossRef Medline](#)
- Nosofsky RM (1986) Attention, similarity, and the identification-categorization relationship. *J Exp Psychol Gen* 115:39–61. [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](#)
- Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61:168–185. [CrossRef Medline](#)
- Rotshtein P, Henson RN, Treves A, Driver J, Dolan RJ (2005) Morphing Marilyn into Maggie dissociates physical and identity face representations in the brain. *Nat Neurosci* 8:107–113. [CrossRef Medline](#)
- Sarma A, Masse NY, Wang XJ, Freedman DJ (2016) Task-specific versus generalized mnemonic representations in parietal and prefrontal cortices. *Nat Neurosci* 19:143–149. [Medline](#)
- Schafer RJ, Moore T (2011) Selective attention from voluntary control of neurons in prefrontal cortex. *Science* 332:1568–1571. [CrossRef Medline](#)
- Thomas C, Avidan G, Humphreys K, Jung KJ, Gao F, Behrmann M (2009) Reduced structural connectivity in ventral visual cortex in congenital prosopagnosia. *Nat Neurosci* 12:29–31. [CrossRef Medline](#)
- Townsend JT, Ashby FG (1983) Stochastic modelling of elementary psychological processes. New York: Cambridge UP.
- Treisman AM, Gelade G (1980) A feature-integration theory of attention. *Cognit Psychol* 12:97–136. [CrossRef Medline](#)
- Wolfe JM (1994) Guided Search 2.0: A revised model of visual search. *Psychon Bull Rev* 1:202–238. [CrossRef Medline](#)
- Yang H, Zelinsky GJ (2009) Visual search is guided to categorically defined targets. *Vision Res* 49:2095–2103. [CrossRef Medline](#)
- Zelinsky GJ (2008) A theory of eye movements during target acquisition. *Psychol Rev* 115:787–835. [CrossRef Medline](#)