Preparatory attentional templates in prefrontal and sensory cortex encode target-associated information

Zhiheng Zhou[1,2] and Joy J. Geng[2,3]

[1]College of Psychology, Sichuan Normal University, Chengdu, Sichuan 610066, China
[2]Center for Mind and Brain, University of California Davis, Davis, CA 95618, USA
[3]Department of Psychology, University of California Davis, Davis, CA 95616, USA

Correspondence concerning this article should be addressed to Zhiheng Zhou, College of Psychology, Sichuan Normal University, 5 Jing'an Road, Chengdu, Sichuan 610066, China, E-mail: zhou@sicnu.edu.cn, or Joy J. Geng, Center for Mind and Brain, University of California Davis, 267 Cousteau Place, Davis, CA 95618, USA, E-mail: jgeng@ucdavis.edu.

Acknowledgements:
This research received funding supported by the National Institutes of Health under Grant R01-MH113855 to Joy J. Geng.

CRediT authorship contribution statement:
Zhiheng Zhou: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing, Data curation, Formal analysis, Visualization.
Joy J. Geng: Conceptualization, Methodology, Investigation, Writing – original draft, Writing – review & editing, Supervision.

The deidentified data and supplemental materials are available on Open Science Framework (https://osf.io/xw8hm/).

Abstract

Visual search relies on the ability to use information about the target in working memory to guide attention and make target-match decisions. This memory representation is referred to as the “attentional” or “target” template and is typically thought to contain veridical target information. However, more recent studies have shown that in complex visual environments, target-associated information is often used to guide attention (Battistoni et al., 2017; de Lange et al., 2018; Peelen et al., 2024; Vo et al., 2019; Yu et al., 2023). This is particularly true when the target is difficult to discriminate (Zhou & Geng, 2024). Here, we use fMRI and multivariate pattern analysis, to test if attentional guidance by target-associated information is explicitly represented in the preparatory period before search begins, either in conjunction with the target or even in place of it. Participants were first trained on four face-scene category pairings. After learning, they engaged in a cued visual search task that began with a face cue, followed by a delay, and then a search display with two lateralized faces, each superimposed on a scene image. The target face appeared on its previously associated scene on 75% of “scene-valid” trials. Our results show that in the cue period, face information was decoded in the fusiform face area (FFA), superior parietal lobule (SPL), and dorsolateral prefrontal cortex (dLPFC). However, during the delay period, face information was no longer decoded from FFA; instead, scene information was now decoded in the parahippocampal place area (PPA) and inferior frontal junction (IFJ). These findings demonstrate the dynamic nature of template information during visual search and suggest that target-associated information can be used as a guiding template, even replacing actual target information in the active attentional template.

Introduction

We engage in visual search repeatedly throughout the day whenever we look for something, such as a friend in a café, a phone on the table, or ingredients at a store. When visual search is efficient, behavior feels frictionless, but failures result in frustration or even the derailment of behavioral goals. The canonical mechanism supporting visual search is thought to involve a network of frontal regions that maintain target information in a working memory "template" that is used to adjust sensory gain in neurons that encode matching features (Chelazzi et al., 1993; Robert Desimone & John Duncan, 1995; Liu et al., 2007; O'Connor et al., 2002; O'Craven et al., 1999; Reynolds & Heeger, 2009; Serences & Boynton, 2007; Stokes et al., 2009; Sylvester et al., 2007; Treue & Trujillo, 1999). Changes in sensory gain lead to increases in baseline activity that subsequently enhances the stimulus-evoked response generated by target-matching stimuli (Boettcher et al., 2020; Chelazzi et al., 1993; R. Desimone & J. Duncan, 1995; Gayet & Peelen, 2022; O'Connor et al., 2002; Peelen & Kastner, 2011; Serences & Boynton, 2007; Stokes et al., 2009; Sylvester et al., 2007; van Loon et al., 2018; Witkowski & Geng, 2023). In this way, information in the target template is thought to dictate sensory processing priorities that impact the ability to perceive and act on behaviorally relevant objects. Despite its importance for behavior, the bulk of work on target templates in the brain has focused on target veridical features (e.g., the color "red" or small round shapes when looking for an apple). Fewer studies have examined how non-target information might also be held in the template and used as a proxy for the target. The current study tests if objects that are statistically associated with the target are ever prioritized in the target template in conjunction with or even instead of the actual target in anticipation of visual search (Boettcher et al., 2018; Helbing et al., 2022; Mack & Eckstein, 2011; Yu et al., 2023; Zhou & Geng, 2024).

Information in the target template is considered a source of goal-directed sensory-motor control (Nee, 2021). Recently, the posterior end of the inferior frontal sulcus and its junction with the precentral sulcus has been identified as holding object and feature-based information consistent with the definition of target template (Baldauf & Desimone, 2014; Bedini et al., 2023; Bichot et al., 2015; Gong et al., 2022; O'Reilly, 2010; Soyuhos & Baldauf, 2023; Witkowski & Geng, 2023; Zanto et al., 2010). These regions include a posterior-to-anterior cortical organization labeled as premotor eye field (PEF), posterior inferior frontal junction (IFJp), and anterior inferior frontal junction (IFJa) in the multi-modal parcellation atlas (Bedini et al., 2023;

Glasser et al., 2016). For example, Baldauf and Desimone (2014) found that attention to superimposed faces versus houses resulted in induced gamma synchrony between the IFJ and fusiform face area (FFA) or parahippocampal place area (PPA), respectively; this was coupled with selective sensory enhancements in stimulus processing. A likely homologue of this region has also been identified in non-human primates in the ventral prearcuate region (VPA) of prefrontal cortex (Bichot et al., 2015; Bichot et al., 2019). Consistent with its role as a source for feature-based attention, inactivation of VPA resulted in poorer visual search performance and reduced V4 visual responses to the target. These results provide evidence for a causal link between the target representation in IFJ and sensory biases in V4 that facilitate visual search behavior. Together the data suggest that regions encompassed by IFJ maintain target information and provide feedback signals to visual neurons encoding those same features. The role IFJ plays for feature-based attention appears to be analogous to that of the frontal eye fields (FEF) for spatial attention (Bichot et al., 2015; Corbetta et al., 2008; Moore et al., 2003; Ruff et al., 2006; Thompson et al., 2005).

In addition to IFJ, which appears to encode the source of target information, visual search involves a number of other cognitive control functions that recruit other frontal and parietal regions. In particular, the dorsolateral and ventrolateral prefrontal cortices (dLPFC and vLPFC) are involved in proactive and reactive cognitive control including the maintenance and manipulation of goal-relevant information across stimulus types, and flexibly updating those goal-based representations when unexpected errors are encountered (Badre & Nee, 2018; Bettencourt & Xu, 2016; Braver et al., 2009; Christophel et al., 2018; Christophel et al., 2017; Emrich et al., 2013; Ester et al., 2015; Finn et al., 2019; Lee et al., 2013; Liu et al., 2003; Long & Kuhl, 2018; Nee & D'Esposito, 2017; Soon et al., 2013; Stokes et al., 2013). Working memory representations in lateral prefrontal and parietal regions are engaged in cognitive control computations that task non-specific but essential to their functioning (Bettencourt & Xu, 2016; Kwak & Curtis, 2022; Long & Kuhl, 2018; Panichello & Buschman, 2021). The ability to flexibly update goals is also an important aspect of visual search because finding a target requires an iterative cycle of choosing an object based on the current template to attend to or look at, and then deciding if it is the correct target or not (Alexander & Zelinsky, 2011; Hout & Goldinger, 2014; Malcolm & Henderson, 2010; Yu et al., 2022; Yu et al., 2023). When it is the target, search is terminated, but when it is not, attention is guided to a new potential target and

the cycle continues. Thus, targets must be maintained in memory over time, but control mechanisms flexibly adjust goals based on sensory outcomes (Badre & Wagner, 2007; Egner, 2023; Kurtin et al., 2023; Poskanzer & Aly, 2023; Rossi et al., 2009; Tunnermann et al., 2021).

In contrast to the relatively straightforward idea that veridical target features are held in working memory and used to bias sensory processing, there is substantial evidence that during naturalistic scene viewing, attention is often guided by objects that are associated with the target but not the target itself (Boettcher et al., 2018; Helbing et al., 2022; Mack & Eckstein, 2011; Zhou & Geng, 2024). For example, people are faster to find targets (e.g., sand toys) that are near associated "anchor" objects (e.g., sandbox), because the anchor serves as a large and more easily perceived spatial predictor of the target's location (Boettcher et al., 2018; Helbing et al., 2022; Turini & Võ, 2022). Using such proxy information is beneficial because it reduces the space over which the target must be searched for (Boettcher et al., 2018; Castelhano & Krzys, 2020; Hall & Geng, 2024; Josephs et al., 2016). A recent study found that target-associated anchor objects implied by a scene could be decoded in parietal and visual cortex before the target appeared (Lerebourg et al., 2024), supporting the notion that objects associated with the target are used to bias sensory processing and guide attention (Gayet & Peelen, 2022; Peelen et al., 2024; Yu et al., 2023).

In this study, we go one step further to ask if associated information in memory can replace target information in IFJ, frontoparietal attention network, and category-selective sensory cortex during the delay period in anticipation of visual search. Such an outcome would suggest that information predictive of the target is prioritized in preparation for attentional guidance, even at the expense of the target itself. We base our paradigm on our previous behavioral work showing that participants can learn new face-scene category associations and use this information to guide target search, but only when target-distractor discrimination is difficult (Zhou & Geng, 2024). In this previous study, we trained participants on four face-scene associations in advance of a cued-visual search task. During visual search, a single face was cued to indicate the target on each trial. Next, a search display appeared with two lateralized faces superimposed on scenes. The target appeared on its associated scene on 75% of "scene-valid" trials. Different sub-experiments used a variety of scene-invalid trials, including ones in which the associated scene was never present, was co-located with the distractor only, or appeared with both the target and the distractor. Here, we only use one of the "scene-invalid" conditions

because the main interest is to test the hypothesis that the associated scene is represented within the target template during the preparatory delay period prior to target search. Specifically, we hypothesize that scene information will be decoded from frontoparietal control and sensory regions, specifically IFJ, dLPFC, and FFA/PPA, in advance of the visual search display and that this anticipatory activation serves to rapidly guide attention to locations that are statistically likely to contain the target.

Results

Behavioral tasks

All participants (n = 26) were first trained outside the scanner on four face-scene category associations (Figure 1A). First, they saw the displays introducing a face with a scene, for example, the text "Emma is a lifeguard and can be found on the beach" presented with an image of a face, "Emma", superimposed on a beach scene (see Methods). Second, participants were given a match-nonmatch decision task in which they indicated if a specific face-scene pair was a previously learned pair (Figure 1B). All participants reached a decision accuracy greater than 90%, with mean accuracy = 97.4% ± 2.7% (Figure 1C).

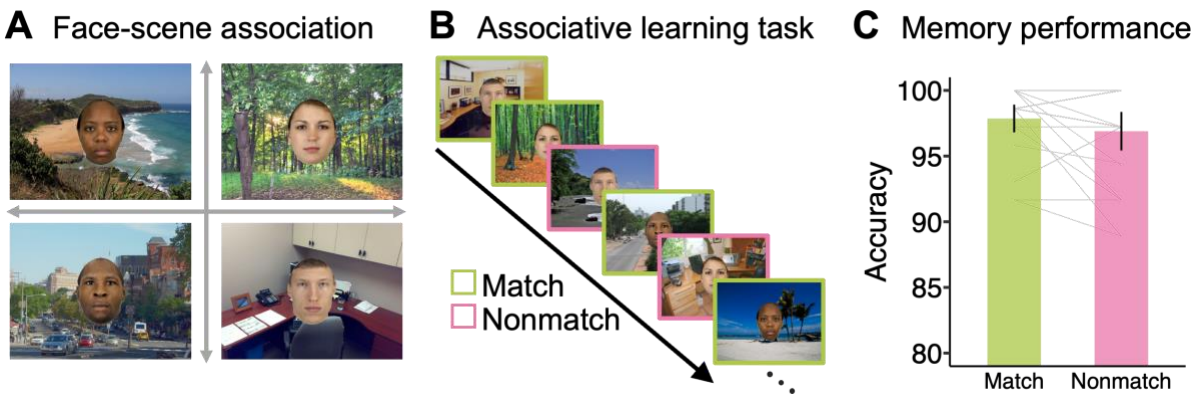

Figure 1. Associative learning task. (A) The four target face stimuli and their associated scenes. (B) Associative learning task to test memory for face-scene pairs. Participants viewed a series of face-scene pairs and made a judgment about whether the face and scene were matched or not. (C) Memory performance for both match and nonmatch conditions was high suggesting a strong association was formed before the face search task.

During the scan session, participants were given a cued visual search task (Figure 2A). Each trial began with a *search cue* of the target face for that trial, followed by an 8-second blank *search delay* period, and then a search display for 250ms. Search trials were composed of two face stimuli superimposed on scene images. The target appeared on the associated scene on 75% of trials (scene-valid) and on one of the three other scenes on 25% of trials (scene-invalid). Thus, the scene was probabilistically predictive of the target's location but not deterministic. The distractor face was a race-gender match, which we previously found made target discrimination difficult enough to make the scene useful for guiding attention (Zhou & Geng, 2024). After the main search task, participants completed a face/scene 1-back task (Figure 2B). Data from this task was used as a benchmark test set for neural activation patterns trained in response to faces and scenes from the main visual search task.

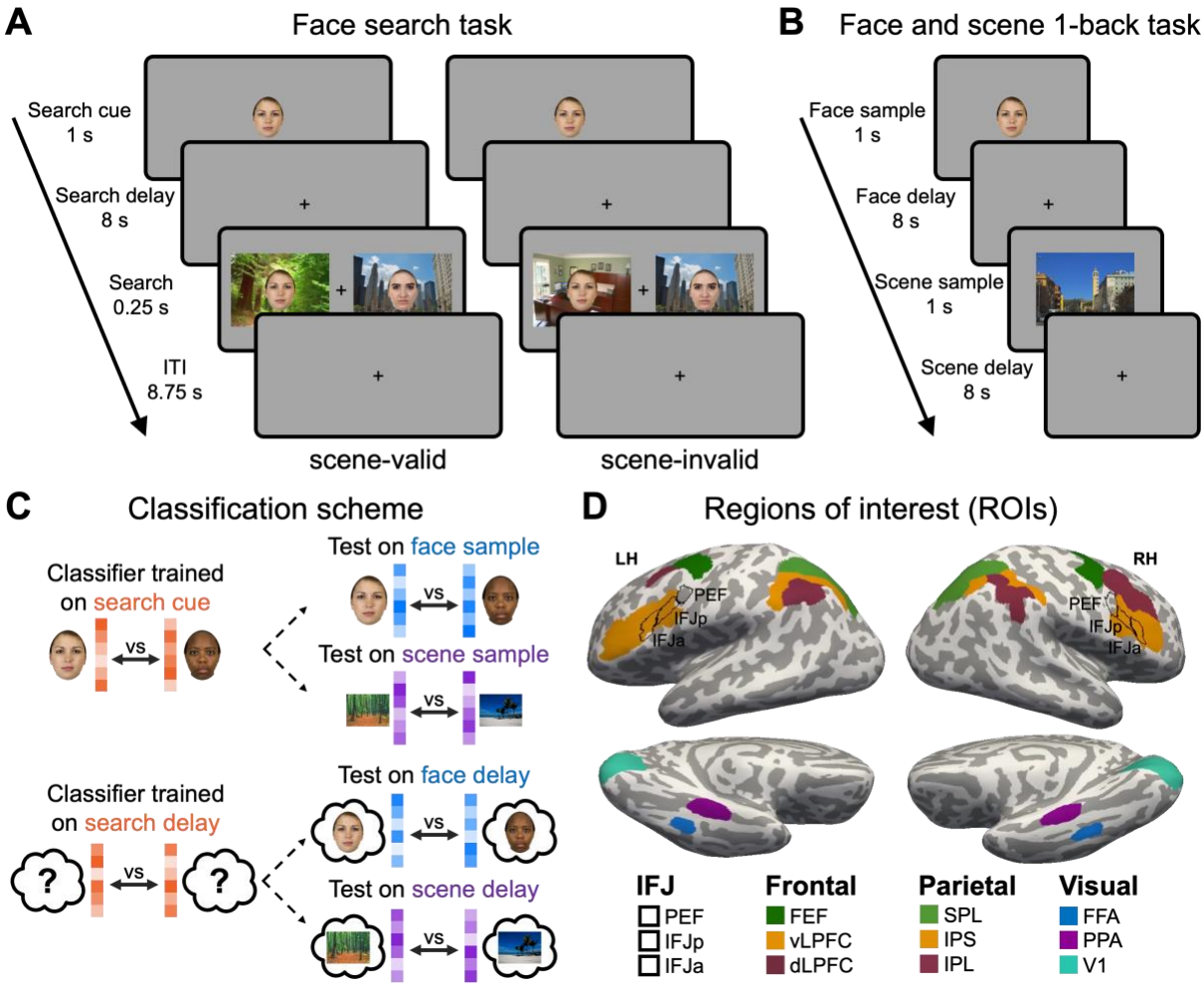

Figure 2. (A) Face search task. Each trial started with a 1-second *search cue* indicating the target face for that trial. This was followed by an 8-second blank *search delay* period, and then the search display for 250 ms. Participants pressed a button to indicate the location of the target on the left or right. (B) Illustration of the separate 1-back task used for cross-task classifier testing. Consistent with the main visual search task, the image was presented for 1 second followed by an 8-second *delay* period. (C) Classification scheme. Classifiers were trained on the neural response patterns from the *search cue* face stimulus and *delay* periods from the visual search task; the classifier was tested on face or scene *sample* stimulus and *delay* period from the 1-back task. (D) Visualization of the twelve functional ROIs on the cortical surface of a representative participant. All ROIs were defined in an individual's native space.

Cue-evoked face information is translated into a preparatory scene template prior to visual search

To test if face and/or scene information was present in the IFJ, attentional control networks, and category-selective visual cortex during the cue and delay periods prior to visual search, we adopted a cross-classification multivariate pattern analysis (MVPA) approach (Figure 2C). Separate classifiers were created for the cue and delay periods, but both classifiers were trained on data from the main search task and tested on data from the separate 1-back task using trials with the same face and scene stimuli. Analyses were focused on twelve regions of interest (ROIs) within the frontoparietal network and the sensory cortex that are known to be involved in representing task structure, target templates for feature-based attention, and category-selective visual processing (Figure 2D). Frontoparietal ROIs were defined by the HCP-MMP1 (Glasser et al., 2016) and the 17-network atlases (Schaefer et al., 2018). Category-selective visual regions for faces (FFA) and scenes (PPA) in the ventral visual pathway were identified for each participant using the independent functional localizer task (Stigliani et al., 2015).

The cross-classification procedure for the cue stimulus involved training the classifier on activation patterns in response to the *search cue* in the face search task, and testing on activation patterns in response to *face* and *scene* stimulus *samples* seen during an independent 1-back task (Figure 2C). We expected the *search cue* classifier to successfully decode face stimuli but not necessarily scene stimuli since this time period should primarily reflect the stimulus-evoked sensory processing of the face cue. The *delay* period cross-classification procedure involved

training the classifier on patterns of activation during the 8-second *search delay* period after the face *search cue* had disappeared and testing the classifier on activation patterns from the 8-second *delay* following the *face* or *scene* sample stimulus from the 1-back task (Figure 2C). Here, we hypothesized that we would see decoding of scene information in the IFJ-PPA network either in conjunction with face information, or even in place of it.

Application of this approach to the *search cue* period (Figure 3A) revealed significant decoding of face information in bilateral FFA (left $p = 0.0074$, uncorrected; right $p < 0.005$). As expected, each of the four face cues could be distinguished from each other in FFA. Faces were also significantly decoded in bilateral SPL (left $p = 0.018$; right $p < 0.005$), right dLPFC ($p < 0.005$), and left IFJp ($p = 0.012$, uncorrected) consistent with the storage of face stimuli in working memory for task-based control. In contrast, scene information was only decoded in bilateral vLPFC (left $p = 0.0066$, uncorrected; right $p = 0.023$); there was no significant decoding of scene information in the scene-selective regions PPA (left $p = 0.12$; right $p = 0.51$). In addition to the ROI results, we also conducted a whole-brain searchlight using a cluster-corrected threshold of pTFCE < .005 to identify additional areas that may have encoded face or scene information during the *search cue* period. This revealed additional significant clusters where the face cue could be decoded in the left dLPFC (Figure 3B and Table S1). These results indicate that the *search cue* period is dominated by face information in a widespread network of sensory and frontoparietal regions, with limited scene information in vLPFC, perhaps related to the cognitive control of memory (Badre & Wagner, 2007).

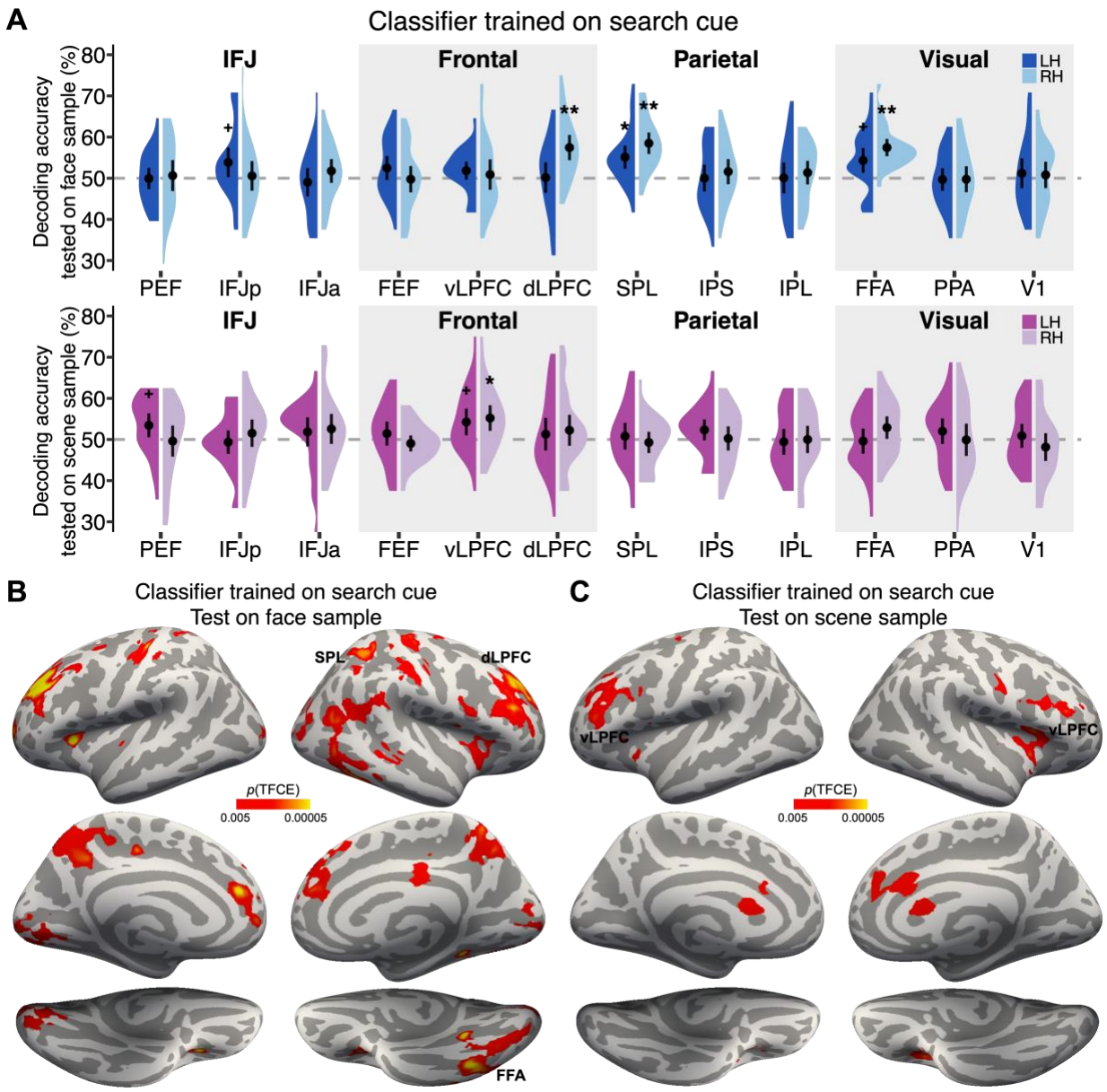

Figure 3. Decoding of face and scene information during the *search cue* period. (A) Evidence of face information in *a priori* defined ROIs. Greater than chance-level classification accuracies were found in the dLPFC, SPL, and FFA. Evidence for scene information was found in vLPFC. (B and C) Significant brain regions revealed by a whole-brain searchlight procedure with information about the face cue (B) or associated scene (C).

In contrast to the *search cue* period, face information during the *search delay* period (Figure 4A) could only be decoded in the left IPS (p = 0.019, uncorrected), perhaps reflecting working memory storage of the target face (Bettencourt & Xu, 2016; Jeong & Xu, 2016).

However, the associated scene was now reliably decoded in the bilateral PPA (left $p = 0.007$; right $p < 0.005$), bilateral PEF (left $p = 0.03$; right $p < 0.005$), and right IFJp ($p = 0.027$, uncorrected). Additional searchlight analyses revealed clusters (pTFCE < .05) in bilateral retrosplenial cortex (Figure 4B and Table S2). The results show robust decoding of scene information in the IFJ-PPA network only during the delay period, suggesting that the face cue was translated into a "guiding template" of the associated scene in anticipation of visual search. Strikingly, the overall pattern shows a double dissociation between face information in both the prefrontal and category-selective visual cortex during the *search cue* period and scene information during the *search delay* period. This is strong evidence for the hypothesis that target-associated information is activated in preparation for guiding attention during target search. Moreover, the associated scene was activated during the delay period and was more prominent in category-selective visual cortex than even the target face.

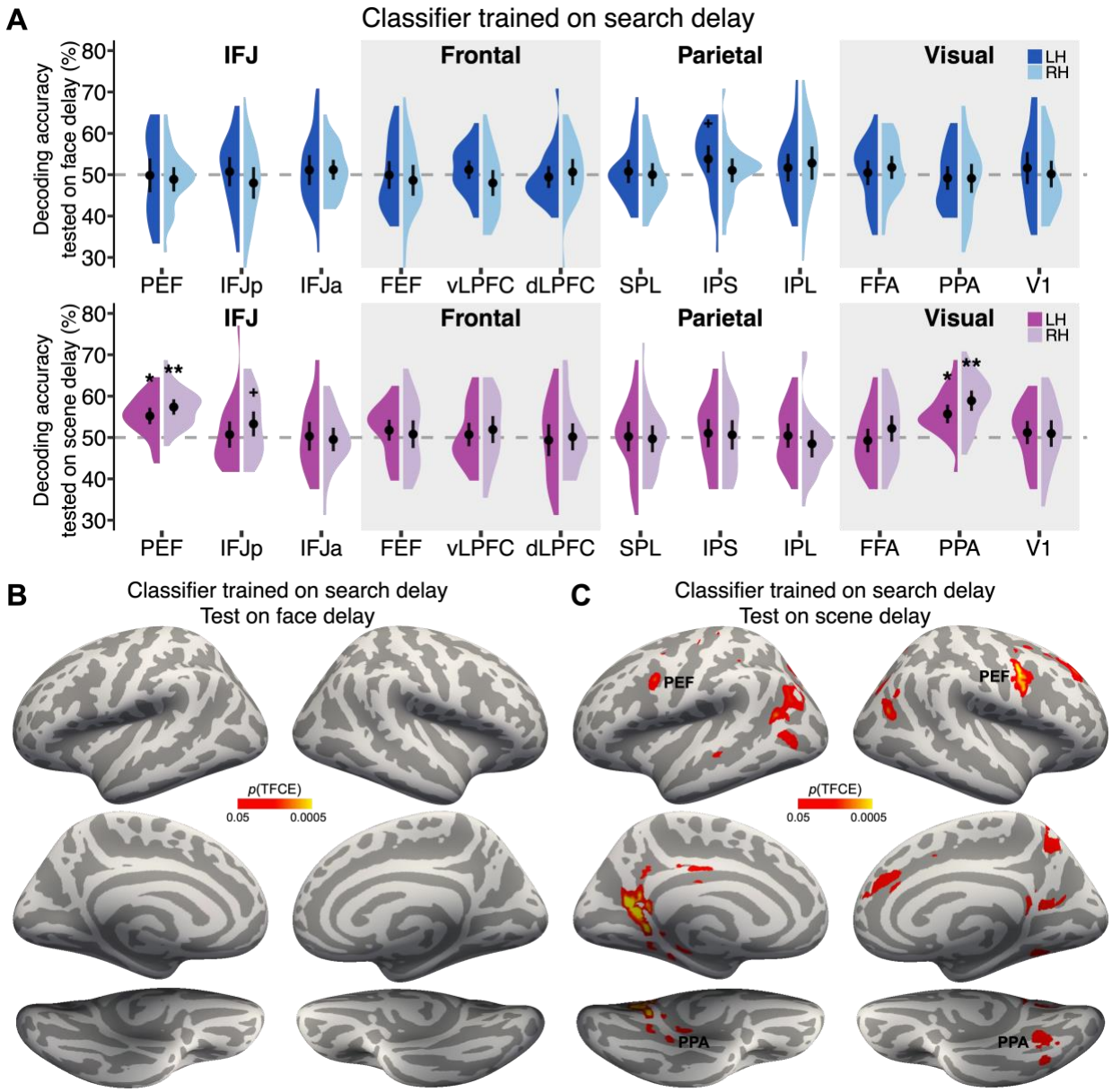

Figure 4. Decoding of face and scene information during the *search delay* period. (A) Evidence of face information in *a priori* defined ROIs was only found in the left IPS. However, evidence of scene information was found in both IFJ and PPA. (B and C) Whole-brain searchlight analyses showed no additional brain regions carried significant information about the face (B), but additional scene information was found in the retrosplenial cortex (C).

Scene-invalid trials require more cognitive control and delay target localization

The decoding results from the *search delay* period suggest that scene information is prioritized over that of the target face in preparation for visual search. If true, then scene-invalid

trials should produce a prediction error and require reactive cognitive control to switch the search template back to the target face. The behavioral results support this interpretation: participants were more accurate ($t(25) = 4.96$, $p < 0.001$, $d = 0.97$) and had shorter RTs ($t(25) = -3.46$, $p = 0.002$, $d = 0.68$) on scene-valid compared to scene-invalid trials (Figure 5A).

To test for prediction error and reactive cognitive control, we conducted a univariate analysis contrasting scene-invalid and scene-valid trials. This "validity effect" contrast showed greater activation on scene-invalid trials in bilateral inferior frontal junction (IFJ) extending to the middle frontal gyrus, bilateral IPS, insula, and the anterior cingulate cortex (ACC; Figure 5B, Figure S1, and Table S3). These regions are similar to those reported in previous studies when cognitive control is needed for task switching, attentional shifting, and resolving cognitive conflict (Corbetta et al., 2008; Dombert et al., 2016; Guerin et al., 2012; Liu et al., 2003). In particular, the combined activation of the lateral frontal cortex and ACC is indicative of conflict detection in ACC and the updating of goals in lateral PFC necessary for successful goal adaptation (Egner, 2023).

Moreover, looking specifically within our *a priori* ROIs, we find significant activation in bilateral IFJp, IFJa, vLPFC, and the left dLPFC and IPS during scene-invalid compared to scene-valid trials (Figure 5C). Notably, scene-valid trials did not produce significantly greater activation in any of the ROIs, suggesting that scene-invalid trials required cognitive control to reset attentional priorities when expectations that the scene could be used to find the target were violated. Together these results are consistent with the idea that scene-invalid trials required the active reinstatement of the target face into working memory and the target template to make a response. These results are consistent with scene-invalid trials producing a prediction error that required flexible, reactive, cognitive control mechanisms to switch search templates (Cole et al., 2013), a process that slows responses and reduces accuracy. Finally, It is worth noting that both the univariate contrast maps and ROI analyses revealed a potential distinction between the posterior IFJ subregion, PEF, and the relatively anterior IFJp and IFJa in feature-based attention and cognitive control (Soyuhos & Baldauf, 2023).

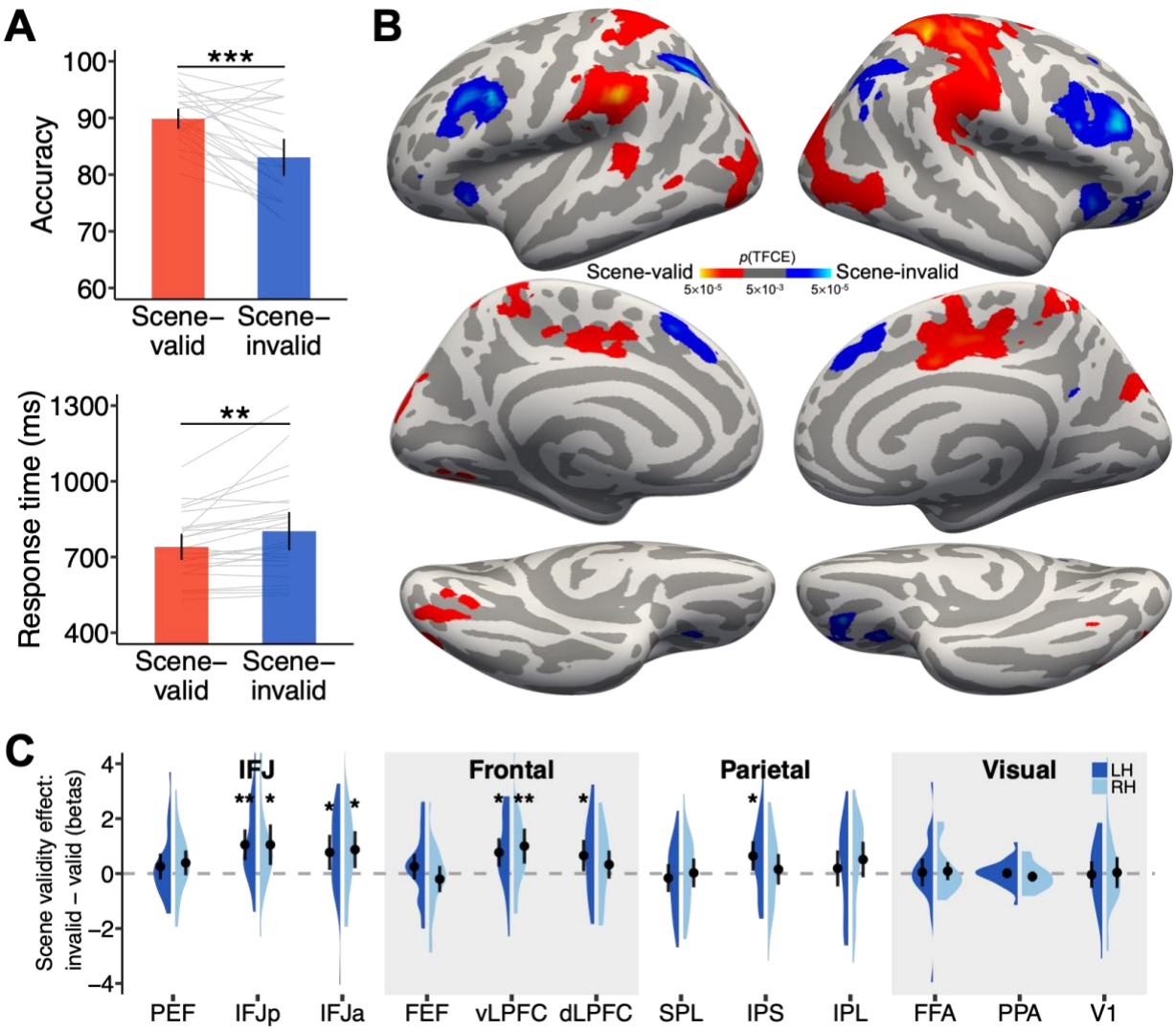

Figure 5. Behavioral and brain results from the face *search* period. (A) Behavioral accuracy and RT both showed a scene-validity effect, suggesting scene information was used to guide attention during search. $^{**}p < .01$, $^{***}p < .001$. Error bars refer to 95% confidence intervals. (B) Whole-brain group-level univariate contrast results showing significantly greater activations for the scene-invalid than scene-valid conditions are illustrated in blue (cold colors), and the reverse contrast in red (hot colors). (C) Contrast betas from the scene-invalid minus scene-valid conditions within each of the *a priori* ROIs, $^{*}p < .05$, $^{**}p < .01$.

Discussion

The canonical mechanism for visual attention involves preparatory changes in sensory gain that increases stimulus-driven responsivity when a matching stimulus appears (Chelazzi et

al., 1993; Corbetta et al., 2008; Kastner et al., 1999; Luck et al., 1997; Reynolds & Heeger, 2009; Serences & Boynton, 2007; Vossel et al., 2014). This sensory modulation is thought to have its source in a network of frontoparietal regions that encode and hold target information as a "template" for sensory modulation (Chelazzi et al., 1993; Ester et al., 2016; Esterman & Yantis, 2009; Gong et al., 2022; Liu et al., 2003; Peelen & Kastner, 2011; Stokes et al., 2009; van Loon et al., 2018). These studies = provide a foundation for understanding how working memory representations distributed across frontal and parietal regions are used to bias visual processing towards relevant information for achieving behavioral goals. However, most of the work on this topic has focused on target-veridical features. Much less is known about how these mechanisms support adaptive updates to information used to find the target during active visual search.

This stands in contrast to the considerable evidence showing that preparatory attention is flexible, highly adaptive, and sensitive to changing contexts. Many behavioral studies have shown that scene structure, statistically co-occurring object pairs, and large, stable, predictive "anchor objects" are all used to guide attention and locate smaller objects (Battistoni et al., 2017; Boettcher et al., 2018; Castelhano & Krzys, 2020; Castelhano et al., 2009; Collegio et al., 2019; de Lange et al., 2018; Gayet & Peelen, 2022; Hall & Geng, 2024; Helbing et al., 2022; Josephs et al., 2016; Mack & Eckstein, 2011; Malcolm & Shomstein, 2015; Nah & Geng, 2022; Peelen et al., 2024; Vo et al., 2019; Yu et al., 2023; Zhou & Geng, 2024). For example, in a previous behavioral study, we showed that when the target is hard to find, scene information is used as a proxy in the target template to guide attention toward the likely target location more efficiently (Zhou & Geng, 2024).

Despite the vast behavioral evidence for the use of non-target information in target-guided visual search, there has been little evidence so far for the presence of target-associated information in preparatory "target templates" encoded in frontal-parietal-sensory attentional control networks. One exception is a recent study that found preparatory activation in the lateral occipital complex (LOC) for a target-associated anchor object in advance of target detection (Lerebourg et al., 2024). In this study, participants were asked to find either a book or a bowl located on two visually unique tables within two living room scenes. The scene context predicted which table would hold each object. When participants were given a target cue and shown a scene with the table regions masked, the associated table, but not the target, could be decoded from LOC. This study provided compelling evidence that target-associated information is

preferentially represented in advance of object detection. However, because the tables were implied by the scene preview, it is somewhat unclear if decoding of the tables reflected the preparatory reinstatement of a target-associated memory or priming of task-relevant information by the scene. There was also no information about template source regions in frontal cortex, such as IFJ (Baldauf & Desimone, 2014; Bedini et al., 2023; Soyuhos & Baldauf, 2023).

Our study builds on this prior literature by teaching participants novel face-scene associations and then using the faces as cues and targets in a visual search task. Scenes were technically task-irrelevant in the search task and were not shown during the cue period. Thus, the presence of any pre-search preparatory scene information in the brain has to be reinstated from memory based on a previously learned association without direct visual prompting. First, however, we confirmed that the identity of the face shown as the cue was decoded in bilateral FFA, right dLPFC, bilateral SPL, and left IFJp using a priori ROIs. Decoding in left dLFPC was found in subsequent searchlight analyses, which revealed symmetric clusters of voxels located bilaterally in the middle frontal gyrus extending to the superior frontal gyrus (Figure 3B, Table S1). Overall, these results were as expected, and showed that the onset of the face cue was represented in sensory and task-based control regions (Chiu et al., 2011; Contreras et al., 2013; Guntupalli et al., 2017; Jeong & Xu, 2016; Long & Kuhl, 2018; Nestor et al., 2011).

In contrast, the only brain region containing scene category information during the cue period was in vLPFC. In prior studies, scene decoding has been found in vLPFC during perceptual scene categorization tasks (Jung et al., 2018; Jung & Walther, 2021; Qin et al., 2011; Walther et al., 2009), but more generally, vLPFC is engaged when cognitive control is required to adjudicate task-based memories (Badre & Wagner, 2007; Levy & Wagner, 2011; Neubert et al., 2014; Tamber-Rosenau et al., 2018). This suggests that the face cue evoked the retrieval of scene category information from memory. The most important question is therefore whether this cue related scene information in vLPFC would be translated into a preparatory search template.

We predicted that, if the associated scene was used as the target template, we should observe successful decoding of scene categories in scene-selective sensory cortex during the delay period reflecting enhancements in sensory gain, and in parietal and lateral prefrontal cortex reflecting scene information in working memory and cognitive control. It is well-documented that template information can be decoded in distributed cortical regions, but that each region likely plays a different computational role (Ester et al., 2015; Ester et al., 2016; Gong et al.,

2022; Lee & Geng, 2017; Long & Kuhl, 2018; Peelen & Kastner, 2011; Soon et al., 2013; Stokes et al., 2009; van Loon et al., 2018; Witkowski & Geng, 2023). Consistent with this, we the associated scene category was decoded in IFJ regions and PPA during the delay period. This suggests that scene information decoded in vLPFC during the cue period triggered retrieval of the associated scene, which was turned into a scene template in IFJ and PPA in preparation for target search (Badre & Wagner, 2007; Xu et al., 2022). Additional searchlight analyses revealed clusters in bilateral retrosplenial cortex that contained scene information (Figure 4C, Table S2). Interestingly, no face information was decodable from FFA or IFJ during the search delay period, suggesting it was not actively maintained in the target template. However, there was evidence of decoding in IPS, indicating that face information was still held in working memory (Bettencourt & Xu, 2016; Christophel et al., 2018; Ester et al., 2015; Favila et al., 2018; Gong et al., 2022; Jeong & Xu, 2016; Kwak & Curtis, 2022; Olmos-Solis et al., 2021; Yu & Postle, 2021).

The consequences of prioritizing associated scenes over the target face were seen in longer RT and more errors on scene-invalid trials. The univariate results suggested that scene-invalid trials triggered a prediction error in ACC that led to an updating of what information should be used to find the target in vLPFC (including IFJp/IFJa) and IPS. This pattern of data supports the notion that preparatory scene activity in the delay period supplanted the target in the "target" template in service of facilitating behavior; when it was inaccurate, however, reactive cognitive control was necessary to update search adaptively (Botvinick et al., 2004; Braver, 2012; Egner, 2023; Kolling et al., 2016; O'Reilly et al., 2013).

In summary, our study builds on a growing body of literature on preparatory attention, focusing specifically on what information is used to generate the most predictive target template. We provide evidence that target-associated information can even supplant the actual target in the "target" template when the associated information is expected to more easily facilitate behavioral goals. In our case, the cued target face was translated into a template for a probabilistically predictive associated scene in IFJ and sensory cortex in preparation for search. This suggests that when there is uncertainty in target detection and decisions, additional information is dynamically recalled to facilitate performance (Hansen et al., 2012; Summerfield et al., 2011; Witkowski & Geng, 2023). The results are consistent with decades of behavioral work showing the adaptive

nature of attention in supporting goal-directed behavior. It goes farther in demonstrating how this mechanism unfolds in frontal-parietal-sensory networks in anticipation of the search display.

Materials and methods

Participants

The sample size was chosen based on a power analysis of a previous behavioral study (Zhou & Geng, 2024) and similar fMRI MVPA studies (Albers et al., 2013; van Loon et al., 2018). Using the more conservative estimate based on the behavioral validity effect size (*d*z = .678), the estimated minimum sample size to achieve significant effects (p = .05, two-tailed) with a power .9 was N = 25. A final number of twenty-six participants (mean age = 21.1 years, range = 18 – 29 years, females = 14, males = 12) were recruited from the University of California, Davis, and were given a combination of course credit and monetary compensation ($50) for a 2-hour MRI session. All participants were right-handed and native English speakers or had spoken English before the age of 5. None of them reported a history of neurological or psychiatric disorders. In addition, they all had normal or corrected-to-normal vision and passed the color blindness test (https://colormax.org/color-blind-test/). Two additional participants were removed due to excessive motion during scanning (head movement > 3mm), and a third participant was removed due to falling asleep in the scanner. The study adhered to the ethical principles of the Declaration of Helsinki, and all consent forms and procedures were approved by the Institutional Review Board of the University of California, Davis.

Stimuli

Stimuli for all experiments consisted of 16 faces (Figure 1A and Figure S2) selected from the Chicago Face Database (CFD, Ma et al., 2015) and 64 scenes (Figure 1A and Figure S3) selected from a scene categorization study (Jung et al., 2018). The ratings for gender and race for all faces were restricted to be larger than 0.9 (based on normalized rating scores ranging between 0 to 1 provided by the CFD; a higher score indicates being more representative of that category). As in our previous study (Zhou & Geng, 2024), we used four target faces derived from crossings between two genders (woman and man) and races (black and white). An additional three faces were selected from each category to serve as distractors for the study. An oval mask was used to

crop all selected faces to reduce the visibility of extra-facial features, e.g., hair. The scene categories were crossed between content (nature and urban) and layout (open and closed). There were 16 exemplar images from each scene category used in the experiment. The resolution of the face images was 150 × 200 pixels (2° × 2.7° visual angles) and the scenes were 800 × 600 pixels (10.7° × 8° visual angles). All stimuli were presented against a gray background (RGB: 128, 128, 128). The low-level luminance and contrast of all face and scene images were controlled with the SHINE_color toolbox (Willenbockel et al., 2010).

Experimental design

Face-scene associative learning task. Before the MRI scan session, participants first learned four face-scene category associations (Figure 1A). During the learning phase, each of the face-scene pairs was shown along with a narrative, for example, "This is Holly. She is a ranger, so you will find her in the forest." Next, participants were tested for their memory on the newly learned associations in a face-scene pair matching task (Figure 1B). On each test trial, both a face and scene appeared on the screen, and participants were asked to judge whether the face was associated with the scene by pressing the "m" for a match, or "n" for a nonmatch. Feedback was given with the word "correct" or "incorrect" presented for 0.5 s after each response. The inter-trial interval (ITI) was 0.5 s. Each of the four faces was tested on 18 scene-match trials and 9 scene-nonmatch trials. There was a total of 108 trials presented in a pseudorandomized order. For the 18 match trials, participants viewed the target face on 18 different scene exemplars from the same category. For the 9 nonmatch trials, participants viewed the target face on 3 different scene images (randomly selected) from the remaining three scene categories.

Face search task. After the learning task, participants completed the face search task in the first scan session (Figure 2A). At the beginning of each trial, a face cue was presented for 1 s followed by an 8 s fixation cross. Next, the search display appeared for 0.25 s. The search display was followed by an 8.75 s ITI. The search display was composed of the target face and one randomly selected distractor face. Each face was superimposed on a scene image. The face-scene pairs appeared on the left and right of the screen split by a central fixation cross. The faces were centered in each of the scenes and the distance from the inner edge of the scene to the central fixation was 1° visual angle. The target face was always presented while the distractor face was randomly drawn from the set of three faces that matched the target in both gender and

race. Participants indicated the location of the target face by pressing the “1” or “2” key on the MRI-compatible keypad corresponding to the left or right side. The response was limited to within 2 s.

Search trials were divided into scene-valid and scene-invalid conditions. The scene was valid on 75% of trials and invalid on the remaining 25%. On the scene-valid trials, the target face was always superimposed on its associated scene while the distractor face was superimposed on a scene from one of the three unassociated categories. On scene-invalid trials, the scenes that appeared with the target and distractor faces were from two different scene categories, both of which were unassociated with the target face.

All participants completed 8 runs of the face search task. There were 16 trials (12 scene-valid trials and 4 scene-invalid trials) in each run and 4 trials with each of the four target faces. The participants were not informed about the exact probability of the valid to invalid trials in the experiment. The trial order as well as the location of the target face in the search display was counterbalanced and pseudorandomized. Participants completed a full practice run in the scanner at the beginning to familiarize them with the task and the testing environment.

Face and scene 1-back task. After completing the face search task, a 6-min T1-weighted structural scan was collected. Then, each participant completed a 1-back task in a second scan session with 4 runs (Figure 2B). On each trial, one stimulus was presented at the center for 1 s, followed by an 8 s fixation. Participants were instructed to make a button press when the current stimulus was the same as the previous one. There were four trials for each target face and scene category. Different exemplars were used for each repetition of the same scene category. An additional 6 trials were designated as the 1-back trial with a repeated stimulus. The trial sequence was counterbalanced and pseudorandomized. Participants completed a full practice run in the scanner before the formal test.

Functional category-selective localizer. We identified two regions of interest (ROIs), the FFA and the PPA for each individual participant in an independent functional localizer task (Stigliani et al., 2015). Stimuli were 2-D grayscale images (~16° × 16° for all stimuli) consisting of faces, houses, corridors, cars, instruments, and phase-scrambled images. Eight images from the same object category were on for 0.4 s and off for 0.1 s in a 4-s mini-block. Participants performed a 1-back memory task and repeats occurred in half of the blocks in each run. Each run consisted of 54 blocks evenly divided by the 6 stimulus categories, and the block order was

counterbalanced and pseudo-randomized across runs. Two runs were collected for each participant.

Stimulus apparatus

The *Face-scene associative learning task* was conducted on a 14-inch MacBook Pro laptop with a spatial resolution of 1920 × 1080, and all stimuli were presented using the Testable platform (www.testable.org). The remaining experimental tasks were conducted while the participants were in the scanner. All stimuli for the scanner experiments were generated using Psychotoolbox-3 installed on a Dell desktop PC and displayed on a 24-inch BOLDscreen LCD monitor with a spatial resolution of 1920 × 1200 pixels. Participants viewed stimuli through a mirror attached to the head coil which projected the monitor ~ 120 cm away outside of the scanner bore.

MRI acquisition and preprocessing

All scans were performed on a 3-Tesla Siemens Skyra scanner with a 32-channel phased-array head coil at University of California, Davis. Functional scans using T2-weighted echoplanar imaging (EPI) sequence with an acceleration factor of 2 were acquired with whole-brain volumes of 48 axial slices of 3 mm thickness (TR/TE 1500/24.6ms, flip angle 75°, base/phase resolution 70/100, 3 × 3 mm^2, FOV 210 mm). High-resolution structural MPRAGE T1-weighted images (TR/TE 1800/2.97 ms, flip angle 7°, base/phase resolution 256/100, FOV 256 mm, 208 sagittal slices) were acquired and were used for anatomical normalization, co-registration, and cortical surface reconstruction. The whole MRI session was finished within 2 hours.

Functional and structural data were preprocessed using SPM12 (Wellcome Department of Imaging Neuroscience), FreeSurfer (Dale et al., 1999; Fischl et al., 1999), and in-house MATLAB (MathWorks) code. The first six initial functional scans of each run were discarded to allow for equilibrium effects. The preprocessing for functional EPI data included slice-time correction and spatial realignment. Using a two-pass procedure, fMRI data from all three scan sessions were aligned to the mean of the EPI images of the first session run. Participants with head motion > 3 mm were excluded. The structural image was coregistered with the mean image. Cortical hemispheric surface reconstruction was performed using the ‘recon-all’ commend in the

FreeSurfer. All fMRI analyses were performed in the native individual space without spatial smoothing, except for a 4 mm full-width at half-maximum (FWHM) smoothing was applied to the *functional category-selective localizer* runs.

The segmentation utility in SPM12 was applied to estimate gray and white matter boundary parameters for spatial normalization. Then, the analyzed native space functional data (i.e., univariate beta images and searchlight classification accuracy images) was normalized into the standard $2 \times 2 \times 2$ mm^3 MNI reference space and smoothed with an 8 mm FWHM isotropic kernel for group-based whole-brain analyses.

fMRI general linear model (GLM).

For each scan session, a GLM was generated by convolving the canonical hemodynamic response functions with experimental conditions for each voxel. Six motion realignment parameters in each experimental run were included as nuisance regressors to control for head motion confounds. For the *face search task*, the GLM was based on different time segments of the task corresponding to the face *search cue*, *search delay* period, and search display. The GLM was constructed with 11 regressors for each experimental run. The first four regressors had a duration of 1 s and were used to estimate the four different face search cue identities. The next four regressors were 8 s long and estimated the search delay period following each face. Three additional regressors with a duration of 1 s were used to model the three types of search displays: scene-valid trials, scene-invalid trials, and error response trials.

A separate GLM was used to model the *face and scene 1-back task*. The GLM was constructed with 17 regressors for each experimental condition in each run. There were 8 regressors for each of the four face images and its corresponding delay period, and another 8 regressors for each scene category image and its corresponding delay period. The BOLD responses to the stimulus image and its corresponding delay period were modeled separately, just as it was for the *face search task*. The same model parameters were used so that the two tasks would be compatible for our cross-task decoding procedure. The last regressor with 1 s duration was used to define the 1-back event, i.e., response to a repeated image, which was not a trial type of interest.

A third GLM was used to model the *functional category-selective localizer* in order to identify FFA and PPA in each participant. The GLM was constructed with 6 regressors with a 4 s

duration to estimate all object categories (faces, houses, corridors, cars, instruments, and phase-scrambled images) within two runs. Contrast beta images between specific types of stimulus categories were used to identify category-selective regions in the visual cortex at the individual level.

Regions of interest (ROIs) definition

The ROIs of primary interest focused on subregions of the frontal, parietal, and visual cortex that are known to be involved in attentional and cognitive control. As shown in Figure 2D, A total of twelve ROIs were defined in the native space of each participant through a combination of the Human Connectome Project Multi-Modal Parcellation (HCP-MMP1; Glasser et al., 2016), the resting-state network (Schaefer et al., 2018), and our independent category-selective localizer. Each ROI selected from the surface-based HCP-MMP1 atlas and the resting-state network was first transformed from the grouped-based fsaverage surface into an individual surface, and then remapped onto the volumetric native space using the 'mri_surf2vol'.

The HCP-MMP1 atlas was used to define three ROIs corresponding to the IFJ (Bedini et al., 2023), located along the posteriorly-to-anteriorly along the inferior frontal sulcus, namely the premotor eyefield (PEF), posterior IFJ (IFJp), anterior IFJ (IFJa). Using the 17-network resting-state network, we extracted the dorsal attention network, the ventral attention network, and the frontoparietal control network ROIs the lateral frontal and parietal cortex. The cortical parcellations in the three networks in the lateral frontal cortex correspond to the frontal eye fields (FEF), ventrolateral prefrontal cortex (vLPFC), and dorsolateral prefrontal cortex (dLPFC); the cortical parcellations in the lateral parietal cortex correspond to the superior parietal lobule (SPL), intraparietal sulcus (IPS), and inferior parietal lobule (IPL). In addition, the early visual cortex (V1) ROI was defined from the visual network located in areas labeled the striate cortex and striate calcarine. The last two visual ROIs, the fusiform face area (FFA) and parahippocampal place area (PPA), were identified from an independent functional category-selective localizer. The preprocessed data of the two functional localizer runs were fitted with a first-level GLM to estimate BOLD responses to each of the stimulus categories in the individual volumetric brain. The FFA was defined by a contiguous cluster of voxels in the fusiform gyrus from the contrast between faces vs. all remaining categories, at $p < 0.001$, uncorrected; the PPA was defined as the contiguous cluster of voxels in the parahippocampal gyrus from the contrast

between scenes/corridors vs. all remaining categories, at p(FWE) < 0.005. No activation was observed at these thresholds for six participants, thus a more lenient threshold was used to allow sufficient voxels in FFA and PPA to be identified for analyses. The average number of voxels for each ROI is reported in the Supplementary Materials, Table S4. Note that a large portion of IFJp and IFJa from the HCP-MMP1 atlas overlapped with the vLPFC from the 17-network atlas. In addition, PEF from the HCP-MMP1 atlas highly overlapped with the precentral label from the dorsal attention network in the left hemisphere and the ventral attention network in the right using the 17-network atlas (Figure S4).

Univariate whole-brain analysis

Although our primary goal was to identify cortical involvement in holding the search template during the face cue and delay periods, we also examined the scene validity effect during the search display period of the *face search task*. To do this, the normalized and smoothed data for the scene-valid and scene-invalid conditions were entered into a second-level GLM with random effects modeled at the group level. Linear contrasts between the two conditions were used to reveal selective regions with significant BOLD activations to either the scene-valid (valid > invalid) or scene-invalid (invalid > valid) conditions. The resultant t-value maps were corrected for multiple comparisons at the cluster level using the threshold-free cluster enhancement (TFCE; Smith & Nichols, 2009; Spisák et al., 2019) implemented in SPM12. The threshold was set at pTFCE < 0.005 for clusters with greater than 50 voxels.

Multivariate pattern analysis (MVPA)

The estimated beta parameters from the *face search task* and the *face and scene 1-back task* first-level GLMs were used to decode whether the target face and/or its associated scene was held as the search template during the *search cue* or *search delay* periods. The beta parameters extracted from each of the 12 ROIs were normalized to remove univariate differences between conditions before being submitted to a binary linear SVM classifier implemented in LIBSVM (Chang & Lin, 2011), with a default cost parameter of 1. Figure 2C illustrates the single-step cross-classification scheme (Brandman & Peelen, 2017; Gayet & Peelen, 2022) adopted in the current study. For the decoding of faces during the search cue period, the classifiers were trained to discriminate between pairs of faces (with a full combination of 6 pairs

based on 4 types of face cues) on the *face search task* runs. These classifiers were then tested using *face and scene* information from the 1-back task stimulus sample period. Following the same procedure for the delay period, classifiers were trained to discriminate between pairs of *search delay* period information following each face cue (6 pairs of delay periods)*;* the classifiers were then tested on the delay period following face or scene sample stimuli from the *face and scene 1-back task*. The classification accuracy was estimated by averaging performance from the six binary classifiers for each type of cross-classification.

Statistical inference

Group-level statistical significance testing was established based on a nonparametric permutation method in which data labels were randomized. The null hypothesis classification accuracy distribution was generated by 10,000 iterations of each type of cross-classification for each participant. The group-level null distribution was then calculated by averaging these classification accuracies across participants. The permuted p-value was defined by the proportion of counts in the null distribution that were equal to or higher than the observed real group average classification accuracy, (n + 1)/(10,000 + 1). This p-value was used as the assessment for statistical significance, and all resulting p-values controlled for family-wise error by using Bonferroni correction.

Searchlight analysis

We also investigated the whole-brain multivariate decoding results using the searchlight approach (Kriegeskorte et al., 2006). At the individual level, the linear SVM decoding accuracy was assigned to the center voxel within a 9 mm radius sphere in volumetric space. This procedure estimates classification accuracies for all voxels within a whole-brain mask. The pair-wise classification schemes were identical to that implemented for the ROI multivariate decoding analysis. The resulting individual participant whole-brain information maps for different classifications were first smoothed using a 4 mm FWHM Gaussian kernel and then normalized and tested for statistical significance against 50% chance decoding using a one-sample t-test (one-tailed). The statistical threshold of all group-level searchlight maps was corrected for multiple comparisons at the cluster level using TFCE.

References

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared Representations for Working Memory and Mental Imagery in Early Visual Cortex. *Current Biology*, *23*(15), 1427-1431. https://doi.org/10.1016/j.cub.2013.05.065

Alexander, R. G., & Zelinsky, G. J. (2011). Visual similarity effects in categorical search. *J Vis*, *11*(8). https://doi.org/10.1167/11.8.9

Badre, D., & Nee, D. E. (2018). Frontal Cortex and the Hierarchical Control of Behavior. *Trends in Cognitive Sciences*, *22*(2), 170-188. https://doi.org/10.1016/j.tics.2017.11.005

Badre, D., & Wagner, A. D. (2007). Left ventrolateral prefrontal cortex and the cognitive control of memory. *Neuropsychologia*, *45*(13), 2883-2901. https://doi.org/https://doi.org/10.1016/j.neuropsychologia.2007.06.015

Baldauf, D., & Desimone, R. (2014). Neural Mechanisms of Object-Based Attention. *Science*, *344*(6182), 424-427. https://doi.org/10.1126/science.1247003

Battistoni, E., Stein, T., & Peelen, M. V. (2017). Preparatory attention in visual cortex. *Annals of the New York Academy of Sciences*, *1396*(1), 92-107. https://doi.org/10.1111/nyas.13320

Bedini, M., Olivetti, E., Avesani, P., & Baldauf, D. (2023). Accurate localization and coactivation profiles of the frontal eye field and inferior frontal junction: an ALE and MACM fMRI meta-analysis. *Brain Structure & Function*. https://doi.org/10.1007/s00429-023-02641-y

Bettencourt, K. C., & Xu, Y. (2016). Decoding the content of visual short-term memory under distraction in occipital and parietal areas. *Nature Neuroscience*, *19*(1), 150-157. https://doi.org/10.1038/nn.4174

Bichot, N. P., Heard, M. T., DeGennaro, E. M., & Desimone, R. (2015). A Source for Feature-Based Attention in the Prefrontal Cortex. *Neuron*, *88*(4), 832-844. https://doi.org/https://doi.org/10.1016/j.neuron.2015.10.001

Bichot, N. P., Xu, R., Ghadooshahy, A., Williams, M. L., & Desimone, R. (2019). The role of prefrontal cortex in the control of feature attention in area V4. *Nature Communications*, *10*(1), 5727. https://doi.org/10.1038/s41467-019-13761-7

Boettcher, S. E. P., Draschkow, D., Dienhart, E., & Vo, M. L. (2018). Anchoring visual search in scenes: Assessing the role of anchor objects on eye movements during visual search. *Journal of Vision*, *18*(13), 11. https://doi.org/10.1167/18.13.11

Boettcher, S. E. P., Stokes, M. G., Nobre, A. C., & van Ede, F. (2020). One Thing Leads to Another: Anticipating Visual Object Identity Based on Associative-Memory Templates. *Journal of Neuroscience*, *40*(20), 4010-4020. https://doi.org/10.1523/Jneurosci.2751-19.2020

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: an update. *Trends Cogn Sci*, *8*(12), 539-546. https://doi.org/10.1016/j.tics.2004.10.003

Brandman, T., & Peelen, X. V. (2017). Interaction between Scene and Object Processing Revealed by Human fMRI and MEG Decoding. *Journal of Neuroscience*, *37*(32), 7700-7710. https://doi.org/10.1523/Jneurosci.0582-17.2017

Braver, T. S. (2012). The variable nature of cognitive control: a dual mechanisms framework. *Trends Cogn Sci*, *16*(2), 106-113. https://doi.org/10.1016/j.tics.2011.12.010

Braver, T. S., Paxton, J. L., Locke, H. S., & Barch, D. M. (2009). Flexible neural mechanisms of cognitive control within human prefrontal cortex. *Proceedings of the National Academy of Sciences*, *106*(18), 7351-7356. https://doi.org/doi:10.1073/pnas.0808187106

Castelhano, M. S., & Krzys, K. (2020). Rethinking Space: A Review of Perception, Attention, and Memory in Scene Processing. *Annual Review of Vision Science*, *6*, 563-586. https://doi.org/10.1146/annurev-vision-121219-081745

Castelhano, M. S., Mack, M. L., & Henderson, J. M. (2009). Viewing task influences eye movement control during active scene perception. *Journal of Vision*, *9*(3), 6-6. https://doi.org/10.1167/9.3.6

Chang, C.-C., & Lin, C.-J. (2011). LIBSVM: a library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)*, *2*(3), 1-27.

Chelazzi, L., Miller, E. K., Duncan, J., & Desimone, R. (1993). A neural basis for visual search in inferior temporal cortex. *Nature*, *363*(6427), 345-347. https://doi.org/10.1038/363345a0

Chiu, Y. C., Esterman, M., Han, Y., Rosen, H., & Yantis, S. (2011). Decoding task-based attentional modulation during face categorization. *J Cogn Neurosci*, *23*(5), 1198-1204. https://doi.org/10.1162/jocn.2010.21503

Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C., & Haynes, J. D. (2018). Cortical specialization for attended versus unattended working memory. *Nature Neuroscience*, *21*(4), 494-+. https://doi.org/10.1038/s41593-018-0094-4

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J. D. (2017). The Distributed Nature of Working Memory. *Trends in Cognitive Sciences*, *21*(2), 111-124. https://doi.org/10.1016/j.tics.2016.12.007

Cole, M. W., Reynolds, J. R., Power, J. D., Repovs, G., Anticevic, A., & Braver, T. S. (2013). Multi-task connectivity reveals flexible hubs for adaptive task control. *Nature Neuroscience*, *16*(9), 1348-1355. https://doi.org/10.1038/nn.3470

Collegio, A., Nah, J. C., Scotti, P. S., & Shomstein, S. (2019). Attention scales according to inferred real-world object size. *Nature Human Behaviour*, *3*(1), 40-47. https://doi.org/10.1038/s41562-018-0485-2

Contreras, J. M., Banaji, M. R., & Mitchell, J. P. (2013). Multivoxel patterns in fusiform face area differentiate faces by sex and race. *PLoS One*, *8*(7), e69684. https://doi.org/10.1371/journal.pone.0069684

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron*, *58*(3), 306-324. https://doi.org/10.1016/j.neuron.2008.04.017

Dale, A. M., Fischl, B., & Sereno, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage*, *9*(2), 179-194. https://doi.org/10.1006/nimg.1998.0395

de Lange, F. P., Heilbron, M., & Kok, P. (2018). How Do Expectations Shape Perception? *Trends in Cognitive Sciences*, *22*(9), 764-779. https://doi.org/10.1016/j.tics.2018.06.002

Desimone, R., & Duncan, J. (1995). Neural Mechanisms of Selective Visual Attention. *Annual Review of Neuroscience*, *18*(Volume 18, 1995), 193-222. https://doi.org/https://doi.org/10.1146/annurev.ne.18.030195.001205

Desimone, R., & Duncan, J. (1995). Neural mechanisms of selective visual attention. *Annu Rev Neurosci*, *18*, 193-222. https://doi.org/10.1146/annurev.ne.18.030195.001205

Dombert, P. L., Kuhns, A., Mengotti, P., Fink, G. R., & Vossel, S. (2016). Functional mechanisms of probabilistic inference in feature- and space-based attentional systems. *Neuroimage*, *142*, 553-564. https://doi.org/10.1016/j.neuroimage.2016.08.010

Egner, T. (2023). Principles of cognitive control over task focus and task switching. *Nature Reviews Psychology*, *2*(11), 702-714. https://doi.org/10.1038/s44159-023-00234-4

Emrich, S. M., Riggall, A. C., LaRocque, J. J., & Postle, B. R. (2013). Distributed Patterns of Activity in Sensory Cortex Reflect the Precision of Multiple Items Maintained in Visual Short-Term Memory. *Journal of Neuroscience*, *33*(15), 6516-6523. https://doi.org/10.1523/Jneurosci.5732-12.2013

Ester, E. F., Sprague, T. C., & Serences, J. T. (2015). Parietal and Frontal Cortex Encode Stimulus-Specific Mnemonic Representations during Visual Working Memory. *Neuron*, *87*(4), 893-905. https://doi.org/10.1016/j.neuron.2015.07.013

Ester, E. F., Sutterer, D. W., Serences, J. T., & Awh, E. (2016). Feature-Selective Attentional Modulations in Human Frontoparietal Cortex. *Journal of Neuroscience*, *36*(31), 8188-8199. https://doi.org/10.1523/Jneurosci.3935-15.2016

Esterman, M., & Yantis, S. (2009). Perceptual Expectation Evokes Category-Selective Cortical Activity. *Cerebral Cortex*, *20*(5), 1245-1253. https://doi.org/10.1093/cercor/bhp188

Favila, S. E., Samide, R., Sweigart, S. C., & Kuhl, B. A. (2018). Parietal Representations of Stimulus Features Are Amplified during Memory Retrieval and Flexibly Aligned with Top-Down Goals. *J Neurosci*, *38*(36), 7809-7821. https://doi.org/10.1523/JNEUROSCI.0564-18.2018

Finn, E. S., Huber, L., Jangraw, D. C., Molfese, P. J., & Bandettini, P. A. (2019). Layer-dependent activity in human prefrontal cortex during working memory. *Nature Neuroscience*, *22*(10), 1687-1695. https://doi.org/10.1038/s41593-019-0487-z

Fischl, B., Sereno, M. I., & Dale, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage*, *9*(2), 195-207. https://doi.org/10.1006/nimg.1998.0396

Gayet, S., & Peelen, M. V. (2022). Preparatory attention incorporates contextual expectations. *Current Biology*, *32*(3), 687-+. https://doi.org/10.1016/j.cub.2021.11.062

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., Smith, S. M., & Van Essen, D. C. (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171-178. https://doi.org/10.1038/nature18933

Gong, M. Y., Chen, Y. L., & Liu, T. S. (2022). Preparatory attention to visual features primarily relies on non-sensory representation. *Scientific Reports*, *12*(1), 21726. https://doi.org/10.1038/s41598-022-26104-2

Guerin, Scott A., Robbins, Clifford A., Gilmore, Adrian W., & Schacter, Daniel L. (2012). Interactions between Visual Attention and Episodic Retrieval: Dissociable Contributions of Parietal Regions during Gist-Based False Recognition. *Neuron*, *75*(6), 1122-1134. https://doi.org/https://doi.org/10.1016/j.neuron.2012.08.020

Guntupalli, J. S., Wheeler, K. G., & Gobbini, M. I. (2017). Disentangling the Representation of Identity from Head View Along the Human Face Processing Pathway. *Cereb Cortex*, *27*(1), 46-53. https://doi.org/10.1093/cercor/bhw344

Hall, E. H., & Geng, J. J. (2024). Object-based attention during scene perception elicits boundary contraction in memory. *Memory & Cognition*. https://doi.org/10.3758/s13421-024-01540-9

Hansen, K., Hillenbrand, S., & Ungerleider, L. (2012). Effects of Prior Knowledge on Decisions Made Under Perceptual vs. Categorical Uncertainty [Original Research]. *Frontiers in Neuroscience*, *6*. https://doi.org/10.3389/fnins.2012.00163

Helbing, J., Draschkow, D., & Vo, M. L. H. (2022). Auxiliary Scene-Context Information Provided by Anchor Objects Guides Attention and Locomotion in Natural Search Behavior. *Psychological Science*, *33*(9), 1463-1476. https://doi.org/10.1177/09567976221091838

Hout, M. C., & Goldinger, S. D. (2014). Target templates: the precision of mental representations affects attentional guidance and decision-making in visual search. *Attention, Perception, & Psychophysics*, *77*(1), 128-149. https://doi.org/10.3758/s13414-014-0764-6

Jeong, S. K., & Xu, Y. (2016). Behaviorally Relevant Abstract Object Identity Representation in the Human Parietal Cortex. *J Neurosci*, *36*(5), 1607-1619. https://doi.org/10.1523/JNEUROSCI.1016-15.2016

Josephs, E., Drew, T., & Wolfe, J. (2016). Shuffling your way out of change blindness. *Psychonomic Bulletin & Review*, *23*(1), 193-200. https://doi.org/10.3758/s13423-015-0886-4

Jung, Y., Larsen, B., & Walther, D. B. (2018). Modality-Independent Coding of Scene Categories in Prefrontal Cortex. *Journal of Neuroscience*, *38*(26), 5969-5981. https://doi.org/10.1523/JNEUROSCI.0272-18.2018

Jung, Y., & Walther, D. B. (2021). Neural Representations in the Prefrontal Cortex Are Task Dependent for Scene Attributes But Not for Scene Categories. *J Neurosci*, *41*(34), 7234-7245. https://doi.org/10.1523/JNEUROSCI.2816-20.2021

Kastner, S., Pinsk, M. A., De Weerd, P., Desimone, R., & Ungerleider, L. G. (1999). Increased Activity in Human Visual Cortex during Directed Attention in the Absence of Visual Stimulation. *Neuron*, *22*(4), 751-761. https://doi.org/10.1016/S0896-6273(00)80734-5

Kolling, N., Behrens, T., Wittmann, M., & Rushworth, M. (2016). Multiple signals in anterior cingulate cortex. *Curr Opin Neurobiol*(37), 36-43. https://doi.org/doi: 10.1016/j.conb.2015.12.007

Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proc Natl Acad Sci U S A*, *103*(10), 3863-3868. https://doi.org/10.1073/pnas.0600244103

Kurtin, D. L., Araña-Oiarbide, G., Lorenz, R., Violante, I. R., & Hampshire, A. (2023). Planning ahead: Predictable switching recruits task-active and resting-state networks. *Human Brain Mapping*, *44*(15), 5030-5046. https://doi.org/https://doi.org/10.1002/hbm.26430

Kwak, Y., & Curtis, C. E. (2022). Unveiling the abstract format of mnemonic representations. *Neuron*, *110*(11), 1822-1828. https://doi.org/10.1016/j.neuron.2022.03.016

Lee, J., & Geng, J. J. (2017). Idiosyncratic Patterns of Representational Similarity in Prefrontal Cortex Predict Attentional Performance. *J Neurosci*, *37*(5), 1257-1268. https://doi.org/10.1523/JNEUROSCI.1407-16.2016

Lee, S. H., Kravitz, D. J., & Baker, C. I. (2013). Goal-dependent dissociation of visual and prefrontal cortices during working memory. *Nature Neuroscience*, *16*(8), 997-U935. https://doi.org/10.1038/nn.3452

Lerebourg, M., de Lange, F. P., & Peelen, M. V. (2024). Preparatory activity during visual search reflects attention-guiding objects rather than search targets. *bioRxiv*, 2024.2002.2002.578555.

Levy, B. J., & Wagner, A. D. (2011). Cognitive control and right ventrolateral prefrontal cortex: reflexive reorienting, motor inhibition, and action updating. *Ann N Y Acad Sci*, *1224*(1), 40-62. https://doi.org/10.1111/j.1749-6632.2011.05958.x

Liu, T., Larsson, J., & Carrasco, M. (2007). Feature-Based Attention Modulates Orientation-Selective Responses in Human Visual Cortex. *Neuron*, *55*(2), 313-323. https://doi.org/10.1016/j.neuron.2007.06.030

Liu, T., Slotnick, S. D., Serences, J. T., & Yantis, S. (2003). Cortical mechanisms of feature-based attentional control. *Cereb Cortex*, *13*(12), 1334-1343. https://doi.org/10.1093/cercor/bhg080

Long, N. M., & Kuhl, B. A. (2018). Bottom-Up and Top-Down Factors Differentially Influence Stimulus Representations Across Large-Scale Attentional Networks. *J Neurosci*, *38*(10), 2495-2504. https://doi.org/10.1523/JNEUROSCI.2724-17.2018

Luck, S. J., Chelazzi, L., Hillyard, S. A., & Desimone, R. (1997). Neural Mechanisms of Spatial Selective Attention in Areas V1, V2, and V4 of Macaque Visual Cortex. *Journal of Neurophysiology*, *77*(1), 24-42. https://doi.org/10.1152/jn.1997.77.1.24

Ma, D. S., Correll, J., & Wittenbrink, B. (2015). The Chicago face database: A free stimulus set of faces and norming data. *Behavior Research Methods*, *47*(4), 1122-1135. https://doi.org/10.3758/s13428-014-0532-5

Mack, S. C., & Eckstein, M. P. (2011). Object co-occurrence serves as a contextual cue to guide and facilitate visual search in a natural viewing environment. *Journal of Vision*, *11*(9). https://doi.org/10.1167/11.9.9

Malcolm, G. L., & Henderson, J. M. (2010). Combining top-down processes to guide eye movements during real-world scene search. *J Vis*, *10*(2), 4 1-11. https://doi.org/10.1167/10.2.4

Malcolm, G. L., & Shomstein, S. (2015). Object-based attention in real-world scenes. *Journal of Experimental Psychology: General*, *144*(2), 257-263. https://doi.org/10.1037/xge0000060

Moore, T., Armstrong, K. M., & Fallah, M. (2003). Visuomotor Origins of Covert Spatial Attention. *Neuron*, *40*(4), 671-683. https://doi.org/https://doi.org/10.1016/S0896-6273(03)00716-5

Nah, J. C., & Geng, J. J. (2022). Thematic object pairs produce stronger and faster grouping than taxonomic pairs. *Journal of Experimental Psychology: Human Perception and Performance*, *48*(12), 1325-1335. https://doi.org/https://doi.org/10.1037/xhp0001031

Nee, D. E. (2021). Integrative frontal-parietal dynamics supporting cognitive control. *Elife*, *10*, e57244. https://doi.org/10.7554/eLife.57244

Nee, D. E., & D'Esposito, M. (2017). Causal evidence for lateral prefrontal cortex dynamics supporting cognitive control. *Elife*, *6*, e28040. https://doi.org/10.7554/eLife.28040

Nestor, A., Plaut, D. C., & Behrmann, M. (2011). Unraveling the distributed neural code of facial identity through spatiotemporal pattern analysis. *Proc Natl Acad Sci U S A*, *108*(24), 9998-10003. https://doi.org/10.1073/pnas.1102433108

Neubert, F.-X., Mars, Rogier B., Thomas, Adam G., Sallet, J., & Rushworth, Matthew F. S. (2014). Comparison of Human Ventral Frontal Cortex Areas for Cognitive Control and Language with Areas in Monkey Frontal Cortex. *Neuron*, *81*(3), 700-713. https://doi.org/https://doi.org/10.1016/j.neuron.2013.11.012

O'Connor, D. H., Fukui, M. M., Pinsk, M. A., & Kastner, S. (2002). Attention modulates responses in the human lateral geniculate nucleus. *Nature Neuroscience*, *5*(11), 1203-1209. https://doi.org/10.1038/nn957

O'Craven, K. M., Downing, P. E., & Kanwisher, N. (1999). fMRI evidence for objects as the units of attentional selection. *Nature*, *401*(6753), 584-587. https://doi.org/10.1038/44134

O'Reilly, J. X., Schuffelgen, U., Cuell, S. F., Behrens, T. E., Mars, R. B., & Rushworth, M. F. (2013). Dissociable effects of surprise and model update in parietal and anterior cingulate cortex. *Proc Natl Acad Sci U S A*, *110*(38), E3660-3669. https://doi.org/10.1073/pnas.1305373110

O'Reilly, R. C. (2010). The What and How of prefrontal cortical organization. *Trends in Neurosciences*, *33*(8), 355-361. https://doi.org/https://doi.org/10.1016/j.tins.2010.05.002

Olmos-Solis, K., van Loon, A. M., & Olivers, C. N. L. (2021). Content or status: Frontal and posterior cortical representations of object category and upcoming task goals in working memory. *Cortex*, *135*, 61-77. https://doi.org/10.1016/j.cortex.2020.11.011

Panichello, M. F., & Buschman, T. J. (2021). Shared mechanisms underlie the control of working memory and attention. *Nature*, *592*(7855), 601-605. https://doi.org/10.1038/s41586-021-03390-w

Peelen, M. V., Berlot, E., & de Lange, F. P. (2024). Predictive processing of scenes and objects. *Nature Reviews Psychology*, *3*(1), 13-26. https://doi.org/10.1038/s44159-023-00254-0

Peelen, M. V., & Kastner, S. (2011). A neural basis for real-world visual search in human occipitotemporal cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(29), 12125-12130. https://doi.org/10.1073/pnas.1101042108

Poskanzer, C., & Aly, M. (2023). Switching between External and Internal Attention in Hippocampal Networks. *The Journal of Neuroscience*, *43*(38), 6538-6552. https://doi.org/10.1523/jneurosci.0029-23.2023

Qin, S., van Marle, H. J., Hermans, E. J., & Fernandez, G. (2011). Subjective sense of memory strength and the objective amount of information accurately remembered are related to distinct neural correlates at encoding. *J Neurosci*, *31*(24), 8920-8927. https://doi.org/10.1523/JNEUROSCI.2587-10.2011

Reynolds, J. H., & Heeger, D. J. (2009). The Normalization Model of Attention. *Neuron*, *61*(2), 168-185. https://doi.org/10.1016/j.neuron.2009.01.002

Rossi, A. F., Pessoa, L., Desimone, R., & Ungerleider, L. G. (2009). The prefrontal cortex and the executive control of attention. *Experimental Brain Research*, *192*(3), 489-497. https://doi.org/10.1007/s00221-008-1642-z

Ruff, C. C., Blankenburg, F., Bjoertomt, O., Bestmann, S., Freeman, E., Haynes, J.-D., Rees, G., Josephs, O., Deichmann, R., & Driver, J. (2006). Concurrent TMS-fMRI and Psychophysics Reveal Frontal Influences on Human Retinotopic Visual Cortex. *Current Biology*, *16*(15), 1479-1488. https://doi.org/https://doi.org/10.1016/j.cub.2006.06.057

Schaefer, A., Kong, R., Gordon, E. M., Laumann, T. O., Zuo, X.-N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2018). Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cerebral Cortex*, *28*(9), 3095-3114. https://doi.org/10.1093/cercor/bhx179

Serences, J. T., & Boynton, G. M. (2007). Feature-Based Attentional Modulations in the Absence of Direct Visual Stimulation. *Neuron*, *55*(2), 301-312. https://doi.org/10.1016/j.neuron.2007.06.015

Smith, S. M., & Nichols, T. E. (2009). Threshold-free cluster enhancement: Addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage*, *44*(1), 83-98. https://doi.org/https://doi.org/10.1016/j.neuroimage.2008.03.061

Soon, C. S., Namburi, P., & Chee, M. W. L. (2013). Preparatory patterns of neural activity predict visual category search speed. *Neuroimage*, *66*, 215-222. https://doi.org/10.1016/j.neuroimage.2012.10.036

Soyuhos, O., & Baldauf, D. (2023). Functional connectivity fingerprints of the frontal eye field and inferior frontal junction suggest spatial versus nonspatial processing in the prefrontal cortex. *European Journal of Neuroscience*, *57*(7), 1114-1140. https://doi.org/10.1111/ejn.15936

Spisák, T., Spisák, Z., Zunhammer, M., Bingel, U., Smith, S., Nichols, T., & Kincses, T. (2019). Probabilistic TFCE: A generalized combination of cluster size and voxel intensity to increase statistical power. *Neuroimage*, *185*, 12-26. https://doi.org/https://doi.org/10.1016/j.neuroimage.2018.09.078

Stigliani, A., Weiner, K. S., & Grill-Spector, K. (2015). Temporal Processing Capacity in High-Level Visual Cortex Is Domain Specific. *J Neurosci*, *35*(36), 12412-12424. https://doi.org/10.1523/JNEUROSCI.4822-14.2015

Stokes, M., Thompson, R., Nobre, A. C., & Duncan, J. (2009). Shape-specific preparatory activity mediates attention to targets in human visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(46), 19569-19574. https://doi.org/10.1073/pnas.0905306106

Stokes, Mark G., Kusunoki, M., Sigala, N., Nili, H., Gaffan, D., & Duncan, J. (2013). Dynamic Coding for Cognitive Control in Prefrontal Cortex. *Neuron*, *78*(2), 364-375. https://doi.org/10.1016/j.neuron.2013.01.039

Summerfield, C., Behrens, Timothy E., & Koechlin, E. (2011). Perceptual Classification in a Rapidly Changing Environment. *Neuron*, *71*(4), 725-736. https://doi.org/https://doi.org/10.1016/j.neuron.2011.06.022

Sylvester, C. M., Shulman, G. L., Jack, A. I., & Corbetta, M. (2007). Asymmetry of Anticipatory Activity in Visual Cortex Predicts the Locus of Attention and Perception. *The Journal of Neuroscience*, *27*(52), 14424-14433. https://doi.org/10.1523/jneurosci.3759-07.2007

Tamber-Rosenau, B. J., Asplund, C. L., & Marois, R. (2018). Functional dissociation of the inferior frontal junction from the dorsal attention network in top-down attentional control. *Journal of Neurophysiology*, *120*(5), 2498-2512. https://doi.org/10.1152/jn.00506.2018

Thompson, K. G., Biscoe, K. L., & Sato, T. R. (2005). Neuronal Basis of Covert Spatial Attention in the Frontal Eye Field. *The Journal of Neuroscience*, *25*(41), 9479-9487. https://doi.org/10.1523/jneurosci.0741-05.2005

Treue, S., & Trujillo, J. C. M. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*(6736), 575-579. https://doi.org/10.1038/21176

Tunnermann, J., Chelazzi, L., & Schubo, A. (2021). How Feature Context Alters Attentional Template Switching. *Journal of Experimental Psychology-Human Perception and Performance*, *47*(11), 1431-1444. https://doi.org/10.1037/xhp0000951

Turini, J., & Võ, M. L.-H. (2022). Hierarchical organization of objects in scenes is reflected in mental representations of objects. *Scientific Reports*, *12*(1), 20068. https://doi.org/10.1038/s41598-022-24505-x

van Loon, A. M., Olmos-Solis, K., Fahrenfort, J. J., & Olivers, C. N. L. (2018). Current and future goals are represented in opposite patterns in object-selective cortex. *Elife*, *7*. https://doi.org/10.7554/eLife.38677

Vo, M. L. H., Boettcher, S. E. P., & Draschkow, D. (2019). Reading scenes: how scene grammar guides attention and aids perception in real-world environments. *Current Opinion in Psychology*, *29*, 205-210. https://doi.org/10.1016/j.copsyc.2019.03.009

Vossel, S., Geng, J. J., & Fink, G. R. (2014). Dorsal and Ventral Attention Systems:Distinct Neural Circuits but Collaborative Roles. *The Neuroscientist*, *20*(2), 150-159. https://doi.org/10.1177/1073858413494269

Walther, D. B., Caddigan, E., Fei-Fei, L., & Beck, D. M. (2009). Natural Scene Categories Revealed in Distributed Patterns of Activity in the Human Brain. *Journal of Neuroscience*, *29*(34), 10573-10581. https://doi.org/10.1523/Jneurosci.0559-09.2009

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: the SHINE toolbox. *Behavior Research Methods*, *42*(3), 671-684. https://doi.org/10.3758/BRM.42.3.671

Witkowski, P. P., & Geng, J. J. (2023). Prefrontal Cortex Codes Representations of Target Identity and Feature Uncertainty. *The Journal of Neuroscience*, *43*(50), 8769-8776. https://doi.org/10.1523/jneurosci.1117-23.2023

Xu, R., Bichot, N. P., Takahashi, A., & Desimone, R. (2022). The cortical connectome of primate lateral prefrontal cortex. *Neuron*, *110*(2), 312-327 e317. https://doi.org/10.1016/j.neuron.2021.10.018

Yu, Q., & Postle, B. R. (2021). The Neural Codes Underlying Internally Generated Representations in Visual Working Memory. *Journal of Cognitive Neuroscience*, *33*(6), 1142-1157. https://doi.org/10.1162/jocn_a_01702

Yu, X., Hanks, T. D., & Geng, J. J. (2022). Attentional Guidance and Match Decisions Rely on Different Template Information During Visual Search. *Psychological Science*, *33*(1), 105-120. https://doi.org/10.1177/09567976211032225

Yu, X., Zhou, Z., Becker, S. I., Boettcher, S. E. P., & Geng, J. J. (2023). Good-enough attentional guidance. *Trends in Cognitive Sciences*, *27*(4), 391-403. https://doi.org/10.1016/j.tics.2023.01.007

Zanto, T. P., Rubens, M. T., Bollinger, J., & Gazzaley, A. (2010). Top-down modulation of visual feature processing: The role of the inferior frontal junction. *Neuroimage*, *53*(2), 736-745. https://doi.org/10.1016/j.neuroimage.2010.06.012

Zhou, Z., & Geng, J. J. (2024). Learned associations serve as target proxies during difficult but not easy visual search. *Cognition*, *242*, 105648. https://doi.org/https://doi.org/10.1016/j.cognition.2023.105648