

Prefrontal Cortex Codes Representations of Target Identity and Feature Uncertainty

 Phillip P. Witkowski^{1,2} and  Joy J. Geng^{1,2}

¹Center for Mind and Brain, University of California, Davis, Davis, California 95618 and ²Department of Psychology, University of California, Davis, Davis, California 95618

Many objects in the real world have features that vary over time, creating uncertainty in how they will look in the future. This uncertainty makes statistical knowledge about the likelihood of features critical to attention demanding processes such as visual search. However, little is known about how the uncertainty of visual features is integrated into predictions about search targets in the brain. In the current study, we test the idea that regions prefrontal cortex code statistical knowledge about search targets before the onset of search. Across 20 human participants (13 female; 7 male), we observe target identity in the multivariate pattern and uncertainty in the overall activation of dorsolateral prefrontal cortex (DLPFC) and inferior frontal junction (IFJ) in advance of the search display. This indicates that the target identity (mean) and uncertainty (variance) of the target distribution are coded independently within the same regions. Furthermore, once the search display appears the univariate IFJ signal scaled with the distance of the actual target from the expected mean, but more so when expected variability was low. These results inform neural theories of attention by showing how the prefrontal cortex represents both the identity and expected variability of features in service of top-down attentional control.

Key words: attention; target template; uncertainty; visual search

Significance Statement

Theories of attention and working memory posit that when we engage in complex cognitive tasks our performance is determined by how precisely we remember task-relevant information. However, in the real world the properties of objects change over time, creating uncertainty about many aspects of the task. There is currently a gap in our understanding of how neural systems represent this uncertainty and combine it with target identity information in anticipation of attention demanding cognitive tasks. In this study, we show that the prefrontal cortex represents identity and uncertainty as unique codes before task onset. These results advance theories of attention by showing that the prefrontal cortex codes both target identity and uncertainty to implement top-down attentional control.

Introduction

Our sensory environment is rich with dynamic objects that change over time, creating uncertainty in their appearance. For example, imagine trying to find your friend in the park without knowing the exact features of her clothes. Will she be the person wearing red, yellow, or some color in between? To find her, you need to account for this uncertainty or else risk dramatic delays in search times, or even failure to locate your friend (Bravo and Farid, 2016; Geng and Witkowski, 2019; Witkowski and Geng, 2019). Situations like this are ubiquitous in the real world, but

previous studies of attention have generally focused on understanding the neural representations of specific search targets, and very little on how the brain integrates uncertainty into predictions about how the target might look. The current experiments seek to identify the neural mechanisms underlying how uncertainty is combined with the expected identity of a target to guide attention.

Theories of attention use the concept of the target template to describe the representation of search targets held in memory (Duncan and Humphreys, 1989; Wolfe, 2021). These representations are encoded by regions of prefrontal cortex that provide top-down signals to modulate sensory processing in visual cortex (Desimone and Duncan, 1995; Hopfinger et al., 2000; Corbetta et al., 2008; Reynolds and Heeger, 2009). For example, the inferior frontal junction (IFJ) maintains target representations that serve as source signals to direct spatial attention, eye-movements, and category selective processing in visual cortex (Baldauf and Desimone, 2014; Bichot et al., 2015; Zhang et al., 2018; Meyyappan et al., 2021). The causal role of IFJ was confirmed by inactivation in this region,

Received June 16, 2023; revised Sep. 4, 2023; accepted Oct. 7, 2023.

Author contributions: P.P.W. and J.J.G. designed research; P.P.W. and J.J.G. performed research; P.P.W. analyzed data; P.P.W. wrote the first draft of the paper; P.P.W. and J.J.G. edited the paper; P.P.W. and J.J.G. wrote the paper.

This work was supported by the National Institute of Mental Health (NIMH) Grant R01MH11385-01.

The authors declare no competing financial interests.

Correspondence should be addressed to Phillip P. Witkowski at pwitkowski@ucdavis.edu.

<https://doi.org/10.1523/JNEUROSCI.1117-23.2023>

Copyright © 2023 the authors

which led to impaired feature-based visual search (Bichot et al., 2019). Similarly, working memory codes in dorsolateral prefrontal cortex (DLPFC) have been found to serve as sources for controlling shifts of attention and object selection (Feredoes et al., 2011; Ester et al., 2015; Panichello and Buschman, 2021). While differences in the roles of IFJ and DLPFC for supporting target templates remain unclear, there are strong functional connections between the two, suggesting they may share information (Gazzaley and Nobre, 2012; Sundermann and Pfeleiderer, 2012; Bedini et al., 2023). Thus, we predict that IFJ and DLPFC will be key regions for supporting representations of the target identity. We additionally hypothesize that these regions will encode the uncertainty of target features and use this information to modulate top-down attentional control.

The idea that IFJ and DLPFC encode target uncertainty is consistent with existing work in the decision literature that finds greater univariate activation in these regions when choice outcomes (Badre et al., 2012; Tomov et al., 2020), or perceptual decisions (Summerfield et al., 2011; Hansen et al., 2012) are uncertain. For example, Summerfield et al. (2011) asked participants to make category judgments on oriented Gabor patches in a task where the mean and variance of the category rule changed over time. The authors reported greater DLPFC activation when the correct category was ambiguous, suggesting that the increase in univariate activation reflected decision uncertainty. Similarly, lateral prefrontal cortex (LPFC) shows uncertainty-weighted error responses following an unexpected stimulus (Iglesias et al., 2013). These findings are consistent with the idea that uncertainty drives activation in LPFC, making it a plausible source for predictions about the expected identity and uncertainty of target features.

The current study tests the idea that IFJ and DLPFC represent target features in conjunction with the feature uncertainty during visual search. To preview the results, we found that IFJ, DLPFC, and visual cortex formed an interconnected network coding target identity in distributed multivariate patterns before search. However, signals related to the uncertainty of the predicted target were only found in the univariate activity within prefrontal regions. These results point to potentially separable neural mechanisms underlying target identity and feature uncertainty. These results illuminate the neural mechanisms that underlie attentional processing by showing how uncertainty is integrated with target identity in advance of visual search.

Materials and Methods

Participants

We recruited a sample of 20 participants (self-reported females = 13, males = 7; mean age = 20.01) from the University of California, Davis SONA system. This sample size was calculated to have 90% power to detect the desired reaction time (RT) effect based on simulations from pilot data, and to match the typical sample size of similar fMRI studies (Summerfield et al., 2011). All participants had normal or corrected-to-normal vision and no impairment in color vision. Participants received a combination of payment and course credit for participation in the experiment. All procedures were approved by the University of California, Davis Internal Review Board (IRB). One block from one participant was removed because of excessive motion during scanning (head movement > 3 mm) but removing this block did not qualitatively change any results.

Experimental design

Participants completed a simplified two-object search task that was adapted for function imagine from previous studies in our lab using multiple object search (Witkowski and Geng, 2019). Each trial began with an auditory tone that cued the target color distribution. A high tone

(750 Hz) indicated that the target color would come from a range of colors centered around an orange-pink value Red-green-blue (RGB): 0.92, 0.58, 0.53); a low tone (375 Hz) indicated the target would come from a range of colors centered around a blue-green value (RGB: 0.18, 0.74, 0.82). The central color of each range (or the mean of the target distribution) was always the most likely to be the target color. The other colors within the range were sampled using a Gaussian-like profile that changed as a function of condition (see below). Tones were presented bilaterally for 250 ms and followed by an 8000-ms interstimulus interval (ISI) of silence. The ISI was followed by a visual search display during which participants were instructed to rapidly “report the location of the object that best matches the central color of cued target distribution.” We call this the “target,” and the difference between the mean of the distribution and the target is the target distance. Participants reported the location of the target by pressing “1” if it was on the left and “2” if it was on the right. The search display was visible for 250 ms, after which the stimuli disappeared, but participants were allowed 1000 ms to respond. Each trial ended with feedback about whether the participant correctly identified the target in time (green circle or a red “X” otherwise). This was followed by a randomly jittered 4000- to 12,000-ms intertrial interval, which followed a γ distribution with a mean of 6600 ms. Each participant completed 160 trials in total during the experiment.

The target color on each trial was drawn from a predetermined distribution of seven possible colors, ranging between -75° and 75° from the mean target color in 25° steps. Participants were told beforehand that the mean of each distribution was the most likely target color, and that the distribution was symmetric. The variance of the target distributions differed across two variability conditions. In the low-variability condition, the target was the mean color value on 80% of trials, $\pm 25^\circ$ away from the mean target color on 5% of trials, $\pm 50^\circ$ away on 10% of trials, and $\pm 75^\circ$ away on 5% of trials. The number of 50° targets were upsampled to increase the number of observations near the tails of the distribution while still maintaining a Gaussian-like profile. In the “high-variability” condition only 25% of target colors were from the mean of the distribution, and each other possible target color occurred on 12.5% of trials. The outcome of this manipulation was greater certainty about the upcoming target’s color in the low-variability condition compared with the high-variability condition. Participants were instructed that “In the low variability condition, the target was likely to be an exact match of the central color and a small chance that it would be different. In the high variability condition, the target was much more likely to be dissimilar to the central color.” Participants were given 20 practice trials of each kind to experience the variability of each distribution.

Finally, distractors on each trial were randomly selected from a uniform distribution between -179° and 180° from mean target color on that trial. The only constraint on selecting the distractor color for a single trial was that it must be at least 20° farther than the actual target from the mean value of the target distribution. This meant that distractors could come from the uncued distribution (i.e., the pink-orange distribution if the blue-green distribution was cued), and accuracy on these trials would be lower if participants could not remember the correct target color.

Participants were explicitly reminded about the mapping between the tone and the target distribution, as well as the variability of each distribution before the beginning of each block. All stimuli had colors taken from a CIE color space color wheel (Bae et al., 2015). Both the target and distractor had a radius of 2° of visual angle and were displayed at 7.5° of eccentricity.

Statistical analysis of behavior

To estimate the effect of variability on RT, we used a linear regression model to test how variability condition moderated participant RT over-and-above that of the target distance. Target distance was calculated as the circular difference between the mean of the target distribution and the target color. We hypothesized that the effect of the target distance would be modulated by expectations about the distribution of target colors. For example, the 50° target should produce a larger RT cost in the low-variability condition than in the high-variability condition. This is

because participants should have high certainty that the target will be the central color value and a 50° target would violate this expectation; in contrast, RT for the same target color should be shorter in the high-variability condition because participants will be more likely to expect any of the possible target colors to appear and a 50° target will be consistent with these expectations. This outcome would produce an interaction in RT between the target distance and variability condition and demonstrate that feature-based attentional selection is modulated by expectations of target variance.

We tested this hypothesis by fitting RT data from trials with correct responses to a γ -distributed hierarchical regression model (Lo and Andrews, 2015), using the lme4 package in R. Trials with incorrect responses or RTs more than three times the interquartile range (IQR) from the median RT were excluded from the analysis (mean = 13 trials total per participant).

The regression model had one regressor for target distance which reflects the slope of the RT increase as the target becomes more dissimilar from the mean target color. We also included a binary regressor for variability condition capturing the overall difference in RT because of uncertainty about the target color, and an interaction term. One final regressor was included to capture the target-distractor distance calculated by the circular difference between target color and the distractor color during the search display, which is known to influence RT (Duncan and Humphreys, 1989; Wolfe and Horowitz, 2017). The random-effects structure included random intercepts for each participant and random by-participant slopes for each of the fixed effects.

All significance testing was done using likelihood-ratio tests between the full model and models with the relevant fixed effect removed, as is appropriate for hierarchical models (Luke, 2017). For example, the significance of the interaction between target distance and variability condition was tested by fitting a model with and without the fixed effect of the interaction.

MRI data acquisition

Data were acquired using a Siemens Skyra 3 Tesla scanner. We used a gradient-echo-planar imaging (EPI) pulse sequence, with a multiband acceleration factor of 2, and aligned the slice angle to the anterior-posterior commissure line. We acquired 48 axial slices, 3 mm thick with the following parameters: repetition time (TR) = 1500 ms, echo time (TE) = 24.6 ms, flip angle = 75°, field of view (FoV) = 210 mm, voxel size = 3 × 3 × 3 mm. Slices were acquired in interleaved order. We also acquired a field map to correct for potential deformations with dual echo-time images covering the whole brain, with the following parameters: TR = 500 ms, TE1 = 4.92 ms, TE2 = 7.38 ms, flip angle = 40°, FoV = 192 mm, voxel size = 3 × 3 × 3 mm. For accurate registration of the EPIs to the standard space, we acquired a T1-weighted structural image using a magnetization-prepared rapid gradient echo sequence (MPRAGE) with the following parameters: TR = 2400 ms, TE = 2.98 ms, flip angle = 7°, FoV = 256 mm, voxel size = 1 × 1 × 1 mm.

Preprocessing

Preprocessing of the data were done in SPM12 (Wellcome Trust Center for Human Neuroimaging) in MATLAB (2019a MathWorks). Data were preprocessed using the default options in SPM. Images were slice-time corrected and realigned to the first volume of each sequence. Realignment was done to correct for motion using a six-parameter rigid body transformation. Inhomogeneities in the field were corrected using the phase of non-EPI gradient echo images at two echo times, which were co-registered with structural maps. Images were then spatially normalized by warping participant specific images to the reference brain in the Montreal Neurologic Institute (MNI) coordinate system with 2-mm isotropic voxels. Finally, for the univariate analyses, images were spatially smoothed using a Gaussian kernel with a full width at half maximum of 8 mm.

ROI selection

Based on our a priori prediction about the role of IFJ, DLPFC, and visual cortex in coding of the attentional template, we defined prefrontal and visual cortical regions of interest to be used in all analyses of the BOLD

data. The IFJ region of interest (ROI) was created using IFJa and IFJb (Index 79 and 80, respectively) from the Human Connectome Project dataset (Glasser et al., 2016). The DLPFC ROI was taken from a probabilistic map of the prefrontal cortex (Sallet et al., 2013). We constructed visual cortex ROIs by using all voxels from visual cortical areas from a probabilistic map of retinotopic regions in the brain (Wang et al., 2015). All voxels generated from the probabilistic maps were included without thresholding. Independent presupplementary motor area (pre-SMA) and anterior insula (AI) ROIs were generated by drawing a 10-mm radius sphere (default size in SPM/Marsbar) around peak activation voxels for frontal regions which supported coding of perceptual uncertainty reported by Geurts et al. (2022).

Univariate fMRI analyses

To model BOLD activity in each voxel we used a GLM with three different regressors; the “cue period” (a boxcar, stimulus presentation time of 250 ms plus 750 ms after the tone ended), the delay period (a boxcar over 7 s of the delay) and the search period (a 1-s boxcar beginning at the onset of the search stimuli). We also included two parametric modulators for the search period: The first was the target distance and the second was the target-distractor distance. Six motion regressors were included as regressors of no interest in the model to account for translation and rotation in head position during the experiment. All temporal regressors were convolved with a standard HRF before computing model coefficients.

From the first-level analysis, we calculated contrast images of the parameter estimates from all regressors (except the motion regressors) and parametric modulators. We then submitted the contrast for each participant into one sample *t* tests in the group-level analyses.

Multivariate fMRI analyses

Multivariate pattern analyses were used to test whether we could decode information about the expected target on each trial. This method of analysis is often used to decode identity information in local neuronal networks (Haynes and Rees, 2005; Kamitani and Tong, 2005; Kriegeskorte et al., 2008). To do this, we estimated the trial-by-trial BOLD activity pattern during the delay period using unsmoothed preprocessed images and a GLM with regressors for each trial's delay period. The delay period was modeled as a boxcar that had a constant duration lasting 7 s from the end of the cue period. No parametric modulators were added. Pattern-based classifiers [linear support vector machines (SVMs)] were trained using LIBSVM (Chang and Lin, 2011) to distinguish between trials where the predicted target color was pink-orange and trials where the predicted target color was blue-green.

We used a searchlight procedure to identify voxels in each ROI that differentiated the two target identities. Each searchlight consisted of a 5 × 5 × 5 voxel cube and was required to contain at least 10 nonempty voxels to be considered for further processing. The searchlight was then standardized by z-scoring the β values across voxels. We tested voxel accuracy using an iterative leave-one-out procedure. Trials from all but one block were used to train the SVM, then tested on the left-out block. Accuracy was averaged across iterations and compared with chance performance (50%). The resulting maps were then spatially smoothed using a Gaussian kernel with full width half maximum of 8 mm. Group-level analyses were performed using a one-sample *t* test on accuracy maps across participant.

To calculate trial-by-trial decoding strength of representations, we calculated the distance of each trial to the hyperplane separating the two categories using the equation specified on the LIBSVM webpage (<https://www.csie.ntu.edu.tw/~cjlin/libsvm/faq.html>). Patterns that are more distant from the hyperplane can be thought of as having more information about a category, and those that are closer to the hyperplane as having less information (Schuck and Niv, 2019; Witkowski et al., 2022). We then signed the distance of each point according to whether the predicted category label was correct (+ for correct, – for incorrect).

Group-level statistical inference

Group-level testing was done using a one-sample *t* test (*df* = 19) on the cumulative functional maps generated by the first level analysis. All first level maps were smoothed before being combined and tested at the

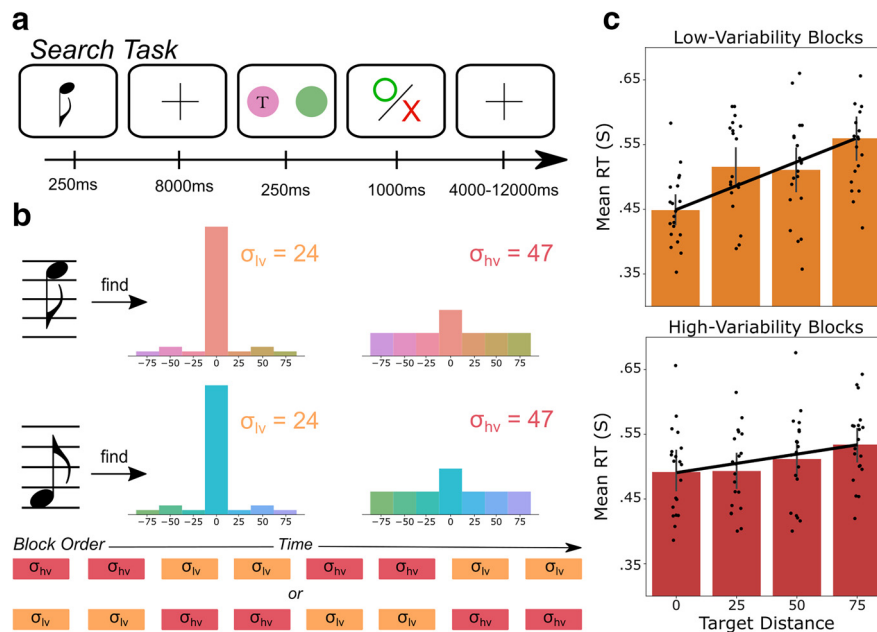


Figure 1. Search task schematic and behavioral results. *a*, Each trial began with the presentation of a cue-tone which informed participants of the relevant target color. High tones (750 Hz) indicated a pink-orange central color (RGB 0.92, 0.58, 0.53) while low tones (350 Hz) indicated a blue-green color (RGB 0.18, 0.74, 0.82). After the interstimulus interval (8000 ms) two colored circles, one target and one distractor, were displayed on the screen for 250 ms. The circle with “T” illustrates the target on this trial. The “T” was not visible to participants during the task. Responses indicating the target location (left/right) were allowed for up to 1 s. Participants then received visual feedback about the correctness of their response: an “O” if correct and an “X” if incorrect. *b*, Blocks were divided into two conditions based on the variability of the target distribution. In the low-variability condition, the target distributions had a low standard deviation (24°) centered around the mean target color. In high-variability blocks, the target distribution had a high standard deviation (47°). *c*, Response time (RT) during the search period followed the expected distribution of target stimuli.

group level. To correct for multiple comparisons, we used threshold-free cluster enhancement (TFCE) which uses permutation testing and accounts for both the height and extent of the cluster (Smith and Nichols, 2009). All parameters were set to the default ($H = 2$, $E = 0.5$) and we used 5000 permutations. In all ROI-based analyses and whole brain analyses we report effects below a $p_{TFCE} < 0.05$ threshold. We first performed group-level inference on independent anatomic ROIs, then performed exploratory whole brain analyses. For ROI analyses, we first extracted voxels from each ROI in each participant’s first-level activation map, averaged the maps together, then applied small volume TFCE correction. All other analyses were corrected for multiple comparisons at the whole brain level.

Code availability

Raw data are available on the NIMH DataArchive (collection #2922; experiment ID 2367), and all original code is publicly available on Open Science Framework (https://osf.io/cndzx/?view_only=f704c0b266324202b983472bf2dd1039).

Results

Search performance is determined by the expected variability of search targets

We first assessed the extent to which the variability of each target distribution affected participant search performance. Specifically, we hypothesized that RT would increase as the target distance increased, but that the slope of this effect would be less steep in the high-variability condition compared with the low-variability condition. This hypothesis was confirmed by the significant interaction between target distance and variability condition ($\beta = 0.79 \text{ ms}/^\circ/\text{condition}$, model comparison $\chi^2(1) = 12.94$, $p < 0.001$; Fig. 1). This means that the same target (e.g., the 50° target distance) elicited longer RTs in the low-variability condition than the high-variability condition because participants were less likely to expect it in the low-variability condition. The

same analysis in accuracy resulted in no interaction between target distance and variability condition ($\chi^2(1) = 0.12$, $p = 0.73$), nor overall effects of target distance ($\chi^2(1) = 0.329$, $p = 0.57$) and variability condition ($\chi^2(1) = 3.56$, $p = 0.059$). Mean accuracy was high in both the low-variability (mean = 0.929, SD = 0.052) and high-variability (mean = 0.900, SD = 0.064) conditions. Together, these results show that participants could accurately identify all target colors in the target distribution, but knowledge about the variability of this distribution modulated the time it took to locate and identify the target.

In contrast to the target distance, while there was a significant main effect of target-to-distractor difference on RT ($\beta = -0.26 \text{ ms}/^\circ$ difference, $\chi^2(1) = 13.55$, $p < 0.001$), there was no interaction with variability condition ($\chi^2(1) = 0.001$, $p = 0.98$). This pattern was also present in the accuracy data: the main effect was significant ($\chi^2(1) = 13.86$, $p < 0.001$), but there was no interaction with variability condition ($\chi^2(1) = 2.50$, $p = 0.11$). These results show that similarity between the target and distractor affected performance but equally so across the expected variability of the targets.

Multivariate representations of target identity are coded in a prefrontal and visual network during the delay period independent of expected uncertainty

The behavioral data showed that there was a relationship between expected variability and target distance. However, the main goal of this study was to test whether the neural representations of the target identity and its expected variability were encoded together or separately, and whether the two types of information were represented in the prefrontal cortex in anticipation of visual search. To better understand how the target identity is coded we first performed a multivariate decoding analysis of activation from the delay period using a search-light

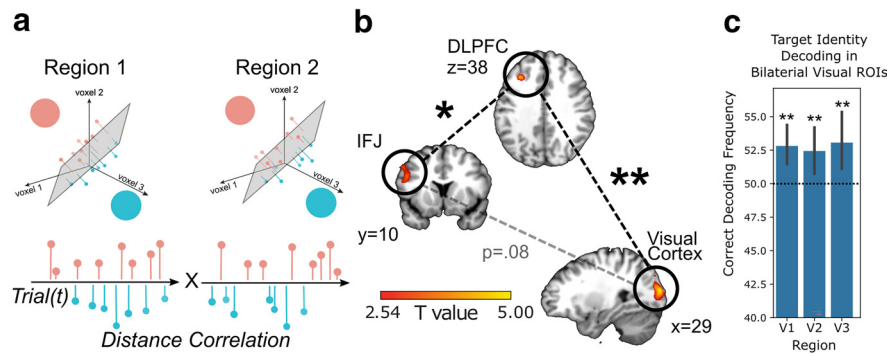


Figure 2. Multivariate coding of target-identity in network spanning prefrontal and visual cortices. **a**, Analysis scheme for data shown in **b**. We calculated the trial-by-trial strength of information about target identities in each of the regions of interest. The strength of information about a target identity was taken as the distance between a voxel pattern in multidimensional space and the hyperplane separating categories. We then correlated these distance measures together as a measure of information connectivity. **b**, Network of regions involved in coding multivariate representations of the target identity across conditions. Lines connecting each area represent the results of information connectivity between regions (** $p < 0.01$, * $p < 0.05$). For illustration, we display the voxels in each region that survive at a threshold of $t_{(19)} = 2.54$, $p < 0.01$, uncorrected. **c**, Results of analysis using bilateral visual ROIs to decode target identity. Dotted horizontal line indicated chance decoding. Error bars represent 95% CI.

procedure over each ROI. This method of analysis has been used in previous work to decode identity information in local neuronal networks (Witkowski et al., 2022). A linear pattern classifier was trained to distinguish trials in which the target came from the pink-orange distribution or blue-green distribution independently for the low-variability and high-variability conditions.

This analysis identified a significant cluster of voxels in IFJ ($[x,y,z] = [-50,12,30]$, $t_{(19)} = 3.48$, $p_{\text{TFCE}} = 0.041$ ROI corrected; Fig. 2*b*), and in left DLPFC ($[x,y,z] = [-32,28,38]$, $t_{(19)} = 4.20$, $p_{\text{TFCE}} = 0.047$ ROI corrected). We also found a significant effect correcting over all voxels in visual cortex ($[x,y,z] = [36,-94,20]$, $t_{(19)} = 5.70$, $p_{\text{TFCE}} = 0.003$ ROI corrected). In an additional analysis, we tested specific subregions of visual cortex by using all voxels in each bilateral region consistent with other approaches in vision science (Haynes and Rees, 2005; Kamitani and Tong, 2005; Kriegeskorte et al., 2008). The results showed significant decoding in V1, V2, and V3 (V1 $t_{(19)} = 3.49$, $p = 0.001$; V2 $t_{(19)} = 2.67$, $p = 0.008$; and V3 $t_{(19)} = 2.80$, $p = 0.006$), in line with work showing that these regions maintain task-relevant information in working memory (Ester et al., 2016; Li et al., 2021). These results replicate previous findings showing that DLPFC, IFJ, and visual cortex encode target features in working memory in anticipation of an upcoming visual search. However, none of these regions showed significant differences between variability conditions (all $p_{\text{TFCEs}} > 0.19$), suggesting that identity information was consistently coded in the multivariate pattern across conditions. Additional exploratory whole-brain corrected analyses using the same multivariate approach also did not find any regions sensitive to target identity, variability condition, or a difference in the strength target identity decoding between conditions (minimum $p_{\text{TFCE}} = 0.25$). Similarly, we observed no significant decoding of the expected target identity after the onset of search (minimum $p_{\text{TFCE}} = 0.38$). This indicates that the cue-induced representations of the target identity held during the delay period were not affected by the predictive uncertainty of exact target values manipulated across variability conditions.

To test whether significant decoding of target identity in IFJ, DLPFC, and visual cortex was because of shared information between regions during the delay, we used information connectivity, a multivariate version of functional connectivity (Coutanche and Thompson-Schill, 2013). We applied a threshold to the searchlights in each region, such that only searchlights that

coded target identity were included (threshold $t_{(19)} = 2.54$, $p < 0.01$, uncorrected). We then extracted the distance of the multivariate pattern on each trial to the hyperplane which separates trials with different targets (see Materials and Methods). These distances can be taken as a measure of the “decoding strength” of target identities on each trial (Schuck and Niv, 2019). Trial-by-trial decoding strength was significantly correlated between IFJ and DLPFC ($t_{(19)} = 2.07$, $p = 0.026$), and between DLPFC and in visual cortex ($t_{(19)} = 3.19$, $p = 0.002$). IFJ and visual cortex showed a marginally significant correlation of decoding strength ($t_{(19)} = 1.67$, $p = 0.055$). However, none of these connections were found to be modulated by the variability condition (all $t_{(19)s} < 1.47$, all $ps > 0.15$). This suggests that the strength of information connectivity within the network of regions that code target identity is not affected by uncertainty. IFJ, DLPFC, and visual cortex comprised a network of regions that code the target identity during the delay period but the multivariate information in these regions was independent of the uncertainty in the prediction of target color.

Coding of predictive uncertainty in response to target variability is supported by univariate signals in prefrontal cortex

Although the behavioral data indicated that both target identity and expected variability impacted visual search performance, the multivariate analyses did not find any regions in the brain that jointly encoded target identity and expected variability. However, previous work indicated that uncertainty may be coded via an increase in univariate activity in DLPFC and IFJ (Summerfield et al., 2011; Badre et al., 2012; Tomov et al., 2020). As such, we tested the idea that target variability is encoded by overall activations (univariate signal) in these regions. We began by constructing a GLM (see Materials and Methods) and then compared differences in the univariate activation during the delay period between the high-variability and low-variability conditions. We hypothesized that we would observe higher univariate activation before high-variability search trials compared with low-variability search trials in the same prefrontal areas in which we decoded target identity, namely IFJ and DLPFC. Our results confirmed this prediction, showing higher activity in the left IFJ ($[x,y,z] = [-44,2,22]$, $t_{(19)} = 3.38$, $p_{\text{TFCE}} = 0.014$ ROI corrected) and left DLPFC ($[x,y,z] = [-48,30,32]$, $t_{(19)} = 4.76$, $p_{\text{TFCE}} = 0.006$ ROI corrected; Fig. 3*a*) during the delay period

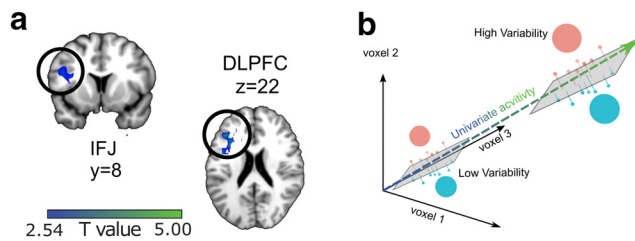


Figure 3. Univariate uncertainty codes overlay multivariate target-identity codes in prefrontal cortex. **a**, Results from univariate contrast between conditions showing higher activity in high-variability blocks compared to low-variability blocks. **b**, Hypothetical effect of how a univariate representation of uncertainty might be overlaid on multivariate sensitivity to target identity within the same region. Note that uncertainty does not change efficiency of decoding but scales only with voxel activity.

in the high-variability condition compared with low-variability condition. This pattern held true even after limiting the voxels included in the analysis to only those for which the multivariate pattern of target identity was also significant (both p TFCEs < 0.05 , ROI corrected). However, we found no effect of variability, positive or negative, in the visual cortex ROI (all p TFCEs > 0.3). We also conducted an exploratory analysis to see whether any ROI continued to code uncertainty into the search period, and found a significant effect in IFJ ($[x,y,z] = [-52,14,28]$, $t_{(19)} = 3.43$, p TFCE = 0.023 ROI corrected). Together, these results suggest that prefrontal areas specifically encoded the predictive uncertainty of the target features before visual search.

IFJ and DLPFC code target distance at the time of search

Because IFJ and DLPFC held information about the target's expected identity and feature variability during the delay period, we also expected processing in these regions to reflect how well the current search target matches those prior expectations. Specifically, we expected activation in these regions to mirror the behavioral data and scale with the variability-weighted target distance. If found, this result would suggest that these frontal regions resolve the uncertainty of target features during active attentional guidance and target decisions.

We first tested for regions with greater modulation of activity by target distance in the low-variability condition compared with the high-variability condition. We found that only IFJ encoded this interaction pattern ($[x,y,z] = [42,12,26]$, $t_{(19)} = 3.13$, p TFCEs < 0.047 ROI corrected; Fig. 4). Neither DLPFC nor visual cortex voxels reached significance (both p TFCE > 0.1 ROI corrected) and no other regions were significant at the whole brain level (all p TFCE > 0.2). IFJ coded the variability-weighted target distance at the time of search, reflecting the pattern of behavioral RT. This suggests that IFJ may play a unique role in deciding whether a target stimulus matches the target template in memory.

Prefrontal cortex codes target-to-distractor relationship at the time of search

In a complimentary analysis to our target distance analysis, we looked for regions that coded the target-to-distractor distance at the time of search. Regions sensitive to this are involved in the resolution of competition between targets and distractors, and therefore should show greater activity when the target-to-distractor difference is smaller (i.e., harder to resolve). We observed the predicted effect in IFJ ($[x,y,z] = [-40,8,28]$, $t_{(19)} = 4.30$, p TFCE < 0.05 ROI corrected) and DLPFC ($[x,y,z] = [-30,8,34]$, $t_{(19)} = 4.20$, p TFCE < 0.05 ROI corrected; Fig. 5b). As a *post hoc*

analysis, we defined ROIs for pre-SMA and AI (see Materials and Methods, ROI selection) based on previous work showing that these regions code perceptual decision uncertainty when viewing stimuli (Michael et al., 2015; Geurts et al., 2022). The ROIs were used to define a small volume for a searchlight analysis. We found significant coding of target-to-distractor distance in both regions (pre-SMA ($[x,y,z] = [0,20,52]$, $t_{(19)} = 3.89$; left AI $[x,y,z] = [-34,20,-2]$, $t_{(19)} = 4.44$; right AI $[x,y,z] = [36,22,-2]$, $t_{(19)} = 4.18$, all p TFCE < 0.05 ROI corrected). We found no other associations between brain activity and target-to-distractor distance at the whole brain level in the positive (all p TFCEs > 0.15) or negative directions (all p TFCEs > 0.35). The fact that IFJ codes both target-distractor distance as well as uncertainty-weighted target distance further suggests a special role for IFJ in integrating the target in memory with the perceptual display in service of discriminating the target.

Discussion

Understanding how uncertainty is integrated into target templates is essential to understanding how attention operates in the real world. As we search for objects in dynamic environments, attention systems must consider the natural variability of objects and make predictions based on this uncertainty. The current experiments add to a literature on how attentional systems use statistical knowledge by looking at the neural mechanisms that underlie the representation of uncertainty and its integration with the target template. We find that frontal regions coded both multivariate representations of predicted targets and a univariate representation tracking the uncertainty of the target's actual features. These results add to theories of attention by showing that IFJ and DLPFC are critical for integrating the known variability of dynamic search targets with predictions of the target identity.

Our multivariate analysis showed that a network of regions spanning IFJ, DLPFC, and visual cortex coded the predicted target identity during the delay period before search. This result is consistent with previous reports of these regions being involved in generating sensory predictions (Iglesias et al., 2013; Kok et al., 2013, 2017) and top-down signals during attentionally demanding tasks (Giesbrecht et al., 2003; Baldauf and Desimone, 2014; Jackson et al., 2021). Similarly, the decoding of target identity in visual cortex is consistent with previous work showing that visual cortex supports the reactivation of remembered visual information from long-term or working memory or when imagining previously seen stimuli (Harrison and Tong, 2009; Naselaris et al., 2015; Vo et al., 2022). The reinstatement in our study involved information connectivity with frontal regions, implicating a network in the recall of visual information in preparation for visual search. The current experiments expand on these findings by showing that the strength of target representations fluctuated in tandem within this network. This suggests that these regions are not only coactive during attentionally demanding tasks but share information about the identity of the target before the onset of search.

Our key aim in the study was to test which regions of the brain coded both the identity of the predicted target and the uncertainty of those predictions. We hypothesized that DLPFC and IFJ would represent the uncertainty of visual features analogous to the way in which these regions encode decision uncertainty (Summerfield et al., 2011; Badre et al., 2012; Michael et al., 2015; Tomov et al., 2020). Our results confirmed these predictions by showing that uncertainty was found as a univariate signal that increased when the distribution of possible targets

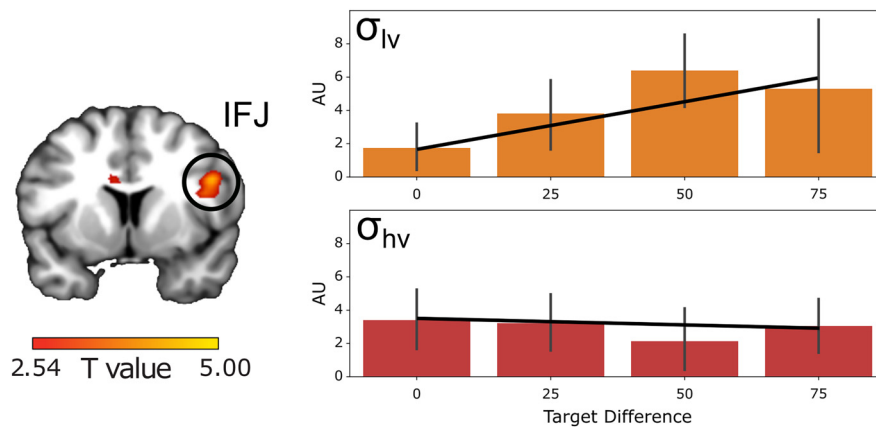


Figure 4. Univariate BOLD Response (arbitrary units; AU) to variability-weighted target distance. Coronal slice of IFJ showing voxels with magnitude differences in response to target distance each trial. Right side shows the BOLD response in this region binned by target distance. Error bars show 95% confidence interval. BOLD responses for target distance in the low-variability condition shown in orange (top panel), and in the high-variability condition shown in red (bottom panel). For illustration, we display the voxels in each region that survive at a threshold of $t_{(19)} = 2.54$, $p < 0.01$, uncorrected.

became more variable. Interestingly, we did not observe any significant modulation of the multivariate identity code by uncertainty, standing in contrast to recent reports of joint coding of working memory contents and uncertainty in this region (Van Bergen and Jehee, 2019; Li et al., 2021). One important difference between their studies and ours is that our participants were cued to the target with a tone whereas their participants directly viewed the target stimulus. Thus, uncertainty about the target in our study concerned predictions about an upcoming target while theirs involved uncertainty about the category of a remembered stimulus. In our study, this signal may reflect uncertainty itself or the anticipated difficulty of resolving the target identity in the upcoming search because of uncertainty. Because the expected identity of the target remained the same in our study regardless of the variability condition, coding expected uncertainty separately from target identity might offer a flexible way to represent the demands of the upcoming search task. Future work will need to specify when and how univariate and multivariate coding schemes are used to represent task variables, and their role in overall task performance. Despite these open questions, our results add to theories of attention by showing how abstract, real-world attributes such as predictive uncertainty are integrated into target representations.

One interesting caveat to the symmetry of our finding for IFJ and DLPFC concerns the role of these two regions during active search. While both DLPFC and IFJ coded the target identity and uncertainty before search, only IFJ continued to code uncertainty through the search period. Additionally, IFJ coded the variance-weighted target distance once the search display appeared. Together, this suggests that IFJ maintained information about how variable the target was likely to be, and increased activation when those expectations were violated in the low-variability condition once the target appeared. DLPFC, however, did not show significant coding of variance-weighted target distance. This divergence suggests that IFJ may be more closely tied to global feature-based attentional guidance and target identification decisions during search (Zhang et al., 2018; Meyyappan et al., 2021). This is supported by the observation that the pattern in IFJ reflected the same pattern of results as the behavioral responses to the search display. DLPFC's role in search may be more specific to maintaining information over long delays before search.

Our behavioral results showed that both variability weighted cue-to-target difference and target-to-distractor difference moderated

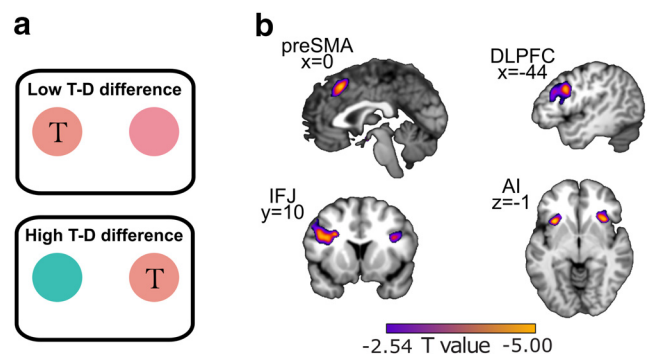


Figure 5. Univariate BOLD activity associated with target-to-distractor difference. **a**, Example trials showing low target-to-distractor difference (top) and high target-to-distractor differences (bottom). The circle with “T” is the hypothetical target. **b**, regions showing significant increases in neural activation in response to low target-distractor difference. For illustration, we display the voxels in each region that survive at a threshold of $t_{(19)} = 2.54$, $p < 0.01$, uncorrected.

search times. While the main purpose of this paper was to determine the neural mechanism of predictive uncertainty, captured by the variability condition, the target-to-distractor difference captures decision uncertainty. That is, uncertainty about which one of the two current options is the true target because of perceptual similarity. Our analysis of the neural activation moderated by target-to-distractor difference illuminated a network of regions including DLPFC, pre-SMA, and AI, which replicates previous work showing that these regions are sensitive to perceptual uncertainty during ongoing decision-making (Michael et al., 2015; Geurts et al., 2022). However, we also identified the IFJ as the only region that coded both the uncertainty of the target distribution and the target-to-distractor difference at search, suggesting a link between IFJ and search behavior (Bedini et al., 2023). Together, these results highlight the IFJ as a region that integrates remembered and perceived sources of uncertainty to support feature-based attention.

Overall, this study adds knowledge to theories of attention by showing how uncertainty is integrated with target information in anticipation of, and during, visual search. We report that IFJ and DLPFC encode predictions of target information and expected variability during the delay period through multivariate and univariate representations, respectively. Furthermore, the activity of IFJ during search was also moderated by target distance.

Together, this work highlights the distinct roles of IFJ and DLPFC in supporting predictive representations of dynamic targets based on their identity and uncertainty.

References

- Badre D, Doll BB, Long NM, Frank MJ (2012) Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration. *Neuron* 73:595–607.
- Bae GY, Olkkonen M, Allred SR, Flombaum JI (2015) Why some colors appear more memorable than others: a model combining categories and particulars in color working memory. *J Exp Psychol Gen* 144:744–763.
- Baldauf D, Desimone R (2014) Neural mechanisms of object-based attention. *Science* 344:424–427.
- 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 Struct Funct* 228:997–1017.
- Bichot NP, Heard MT, DeGennaro EM, Desimone R (2015) A source for feature-based attention in the prefrontal cortex. *Neuron* 88:832–844.
- Bichot NP, Xu R, Ghadooshahy A, Williams ML, Desimone R (2019) The role of prefrontal cortex in the control of feature attention in area V4. *Nat Commun* 10:5727.
- Bravo MJ, Farid H (2016) Observers change their target template based on expected context. *Atten Percept Psychophys* 78:829–837.
- Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2:1–27.
- Corbetta M, Patel G, Shulman GL (2008) The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58:306–324.
- Coutanche MN, Thompson-Schill SL (2013) Informational connectivity: identifying synchronized discriminability of multi-voxel patterns across the brain. *Front Hum Neurosci* 7:15.
- Desimone R, Duncan J (1995) Neural mechanisms of selective visual attention. *Annu Rev Neurosci* 18:193–222.
- Duncan J, Humphreys GW (1989) Visual search and stimulus similarity. *Psychol Rev* 96:433–458.
- Ester EF, Sprague TC, Serences JT (2015) Parietal and frontal cortex encode stimulus-specific mnemonic representations during visual working memory. *Neuron* 87:893–905.
- Ester EF, Sutterer DW, Serences JT, Awh E (2016) Feature-selective attentional modulations in human frontoparietal cortex. *J Neurosci* 36:8188–8199.
- Feredoes E, Heinen K, Weiskopf N, Ruff C, Driver J (2011) Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. *Proc Natl Acad Sci U S A* 108:17510–17515.
- Gazzaley A, Nobre AC (2012) Top-down modulation: bridging selective attention and working memory. *Trends Cogn Sci* 16:129–135.
- Geng JJ, Witkowski P (2019) Template-to-distractor distinctiveness regulates visual search efficiency. *Curr Opin Psychol* 29:119–125.
- Geurts LS, Cooke JRH, van Bergen RS, Jehee JFM (2022) Subjective confidence reflects representation of Bayesian probability in cortex. *Nat Hum Behav* 6:294–305.
- Giesbrecht B, Woldorff MG, Song AW, Mangun GR (2003) Neural mechanisms of top-down control during spatial and feature attention. *Neuroimage* 19:496–512.
- Glasser MF, Smith SM, Marcus DS, Andersson JLR, Auerbach EJ, Behrens TEJ, Coalson TS, Harms MP, Jenkinson M, Moeller S, Robinson EC, Sotiropoulos SN, Xu J, Yacoub E, Ugurbil K, Van Essen DC (2016) The Human Connectome Project's neuroimaging approach. *Nat Neurosci* 19:1175–1187.
- Hansen KA, Hillenbrand SF, Ungerleider LG (2012) Effects of prior knowledge on decisions made under perceptual vs. categorical uncertainty. *Front Neurosci* 6:163.
- Harrison SA, Tong F (2009) Decoding reveals the contents of visual working memory in early visual areas. *Nature* 458:632–635.
- Haynes JD, Rees G (2005) Predicting the orientation of invisible stimuli from activity in human primary visual cortex. *Nat Neurosci* 8:686–691.
- Hopfinger JB, Buonocore MH, Mangun GR (2000) The neural mechanisms of top-down attentional control. *Nat Neurosci* 3:284–291.
- Iglesias S, Mathys C, Brodersen KH, Kasper L, Piccirelli M, den Ouden HEM, Stephan KE (2013) Hierarchical prediction errors in midbrain and basal forebrain during sensory learning. *Neuron* 80:519–530.
- Jackson JB, Feredoes E, Rich AN, Lindner M, Woolgar A (2021) Concurrent neuroimaging and neurostimulation reveals a causal role for dlPFC in coding of task-relevant information. *Commun Biol* 4:588.
- Kamitani Y, Tong F (2005) Decoding the visual and subjective contents of the human brain. *Nat Neurosci* 8:679–685.
- Kok P, Brouwer GJ, van Gerven MAJ, de Lange FP (2013) Prior expectations bias sensory representations in visual cortex. *J Neurosci* 33:16275–16284.
- Kok P, Mostert P, de Lange FP (2017) Prior expectations induce prestimulus sensory templates. *Proc Natl Acad Sci U S A* 114:10473–10478.
- Kriegeskorte N, Mur M, Bandettini P (2008) Representational similarity analysis – connecting the branches of systems neuroscience. *Front Syst Neurosci* 2:4.
- Li HH, Sprague TC, Yoo AH, Ma WJ, Curtis CE (2021) Joint representation of working memory and uncertainty in human cortex. *Neuron* 109:3699–3712.e6.
- Lo S, Andrews S (2015) To transform or not to transform: using generalized linear mixed models to analyse reaction time data. *Front Psychol* 6:1171–1187.
- Luke SG (2017) Evaluating significance in linear mixed-effects models in R. *Behav Res Methods* 49:1494–1502.
- Meyyappan S, Rajan A, Mangun GR, Ding M (2021) Role of inferior frontal junction (IFJ) in the control of feature versus spatial attention. *J Neurosci* 41:8065–8074.
- Michael E, De Gardelle V, Nevado-Holgado A, Summerfield C (2015) Unreliable evidence: 2 sources of uncertainty during perceptual choice. *Cereb Cortex* 25:937–947.
- Naselaris T, Olman CA, Stansbury DE, Ugurbil K, Gallant JL (2015) A voxel-wise encoding model for early visual areas decodes mental images of remembered scenes. *Neuroimage* 105:215–228.
- Panichello MF, Buschman TJ (2021) Shared mechanisms underlie the control of working memory and attention. *Nature* 592:601–605.
- Reynolds JH, Heeger DJ (2009) The normalization model of attention. *Neuron* 61:168–185.
- Sallet J, Mars RB, Noonan MP, Neubert F-X, Jbabdi S, O'Reilly JX, Filippini N, Thomas AG, Rushworth MF (2013) The organization of dorsal frontal cortex in humans and macaques. *J Neurosci* 33:12255–12274.
- Schuck NW, Niv Y (2019) Sequential replay of nonspatial task states in the human hippocampus. *Science* 364:eaaw5181.
- Smith S, Nichols TE (2009) Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. *Neuroimage* 44:83–98.
- Summerfield C, Behrens TE, Koehlin E (2011) Perceptual classification in a rapidly-changing environment. *Neuron* 71:725–736.
- Sundermann B, Pfeleiderer B (2012) Functional connectivity profile of the human inferior frontal junction: involvement in a cognitive control network. *BMC Neurosci* 13:119.
- Tomov MS, Truong VQ, Hundia RA, Gershman SJ (2020) Dissociable neural correlates of uncertainty underlie different exploration strategies. *Nat Commun* 11:2371.
- Van Bergen RS, Jehee JFM (2019) Probabilistic representation in human visual cortex reflects uncertainty in serial decisions. *J Neurosci* 39:8164–8176.
- Vo VA, Sutterer DW, Foster JJ, Sprague TC, Awh E, Serences JT (2022) Shared representational formats for information maintained in working memory and information retrieved from long-term memory. *Cereb Cortex* 32:1077–1092.
- Wang L, Mruczek REB, Arcaro MJ, Kastner S (2015) Probabilistic maps of visual topography in human cortex. *Cereb Cortex* 25:3911–3931.
- Witkowski PP, Geng JJ (2019) Learned feature variance is encoded in the target template and drives visual search. *Vis Cogn* 27:487–501.
- Witkowski PP, Park SA, Boorman ED (2022) Neural mechanisms of credit assignment for inferred relationships in a structured world. *Neuron* 110:2680–2690.e9.
- Wolfe JM (2021) Guided search 6.0: an updated model of visual search. *Psychon Bull Rev* 28:1060–1092.
- Wolfe JM, Horowitz TS (2017) Five factors that guide attention in visual search. *Nat Hum Behav* 1:0058.
- Zhang X, Mlynaryk N, Ahmed S, Japee S, Ungerleider LG (2018) The role of inferior frontal junction in controlling the spatially global effect of feature-based attention in human visual areas. *PLoS Biol* 16:e2005399.