Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory

Brice A. Kuhl [a,*], Jesse Rissman [b,d], Anthony D. Wagner [b,c]

[a] Department of Psychology, Yale University, New Haven, CT 06511, United States
[b] Department of Psychology, Stanford University, Stanford, CA 94305, United States
[c] Neurosciences Program, Stanford University, Stanford, CA 94305, United States
[d] Department of Psychology, University of California, Los Angeles, CA 90095, United States

ABSTRACT

Successful encoding of episodic memories is thought to depend on contributions from prefrontal and temporal lobe structures. Neural processes that contribute to successful encoding have been extensively explored through univariate analyses of neuroimaging data that compare mean activity levels elicited during the encoding of events that are subsequently remembered vs. those subsequently forgotten. Here, we applied pattern classification to fMRI data to assess the degree to which distributed patterns of activity within prefrontal and temporal lobe structures elicited during the encoding of word–image pairs were diagnostic of the visual category (Face or Scene) of the encoded image. We then assessed whether representation of category information was predictive of subsequent memory. Classification analyses indicated that temporal lobe structures contained information robustly diagnostic of visual category. Information in prefrontal cortex was less diagnostic of visual category, but was nonetheless associated with highly reliable classifier-based evidence for category representation. Critically, trials associated with greater classifier-based estimates of category representation in temporal and prefrontal regions were associated with a higher probability of subsequent remembering. Finally, consideration of trial-by-trial variance in classifier-based measures of category representation revealed positive correlations between prefrontal and temporal lobe representations, with the strength of these correlations varying as a function of the category of image being encoded. Together, these results indicate that multi-voxel representations of encoded information can provide unique insights into how visual experiences are transformed into episodic memories.

© 2011 Elsevier Ltd. All rights reserved.

1. Introduction

For more than a decade, functional neuroimaging studies of human memory have considered how neural responses elicited during encoding relate to later memory outcomes. Most frequently, this has been addressed through univariate analysis of functional magnetic resonance imaging (fMRI) data, testing for individual voxels (or clusters of voxels) that show greater mean activity during the encoding of items that will be later remembered relative to items that will be forgotten—a *subsequent memory effect* (Brewer, Zhao, Desmond, Glover, & Gabrieli, 1998; Wagner, Schacter, et al., 1998). Such studies have regularly implicated lateral prefrontal cortex and the medial temporal lobe in successful memory formation (for reviews, see Blumenfeld & Ranganath, 2007; Kim, 2011; Paller & Wagner, 2002; Spaniol et al., 2009). These observations are complemented by neuropsychological investigations that demonstrate

the necessity of prefrontal (Shimamura, 1995; Wheeler, Stuss, & Tulving, 1995) and medial temporal lobe structures for episodic memory (Cohen & Eichenbaum, 1994; Scoville & Milner, 1957; Squire, 1992). Despite the obvious importance of these structures for event memory, there remains considerable ambiguity regarding the cognitive processes and neural mechanisms that are reflected by greater fMRI activation during the encoding of subsequently remembered items. One approach that offers the potential for new insight into these processes is multi-voxel pattern analysis (MVPA). By considering distributed patterns of neural activity, MVPA represents a highly sensitive method for fMRI data analysis and is ideally suited for assessing the similarities or differences between neural states across events (Norman, Polyn, Detre, & Haxby, 2006).

To date, only a handful of studies have applied MVPA to evaluate distributed patterns of neural activity that give rise to episodic encoding success (for review, see Rissman & Wagner, in press). In one recent study, Watanabe et al. (2011) demonstrated that multi-voxel patterns within the medial temporal lobe could be used to classify individual stimuli as subsequently remembered vs. forgotten. Two additional studies used multivariate approaches to

* Corresponding author. Tel.: +1 650 796 6405; fax: +1 203 432 9621.
E-mail address: brice.kuhl@yale.edu (B.A. Kuhl).

consider more subtle questions about how neural pattern similarity across repetitions of a stimulus, or across different stimuli, relate to later memory. In one study, Xue et al. (2010) found that neural pattern similarity across repeated presentations of a stimulus was positively associated with later memory for that stimulus (c.f., Wagner, Maril, & Schacter, 2000). In another study, Jenkins and Ranganath (2010) found that when an encoding event was associated with patterns of neural activity that were relatively dissimilar to neighboring events, that event was more likely to be later associated with successful memory for its temporal context.

An alternative, and to our knowledge unexplored, application of MVPA to the study of episodic encoding success is to consider how the neural representation of stimulus features during encoding relates to later memory. That is, does the strength with which an event is represented positively relate to later memory for that event? If so, are representations in some neural structures more predictive of later memory success than representations in other structures? In the present study, we addressed these questions by using MVPA to (a) measure the neural representation of the visual category of an encoded stimulus, and (b) assess how representational strength within prefrontal and temporal lobe structures relates to subsequent memory outcomes.

We scanned subjects as they formed memories for arbitrary associations between cue words and images of either well-known people (Faces) or well-known locations (Scenes). Using a subset of the encoding data, we trained an MVPA classifier to discriminate fMRI activity patterns associated with Face vs. Scene trials. This classifier was then used to index the relative manifestation of these category-selective activity patterns on each of the remaining encoding trials, and this process was iteratively repeated until all trials had been a part of both the training and testing sets. By performing these pattern classification analyses on the data from each of a set of anatomically defined regions of interest (ROIs) within prefrontal cortex and the temporal lobes, we assessed how classifier-based evidence for target information (Face vs. Scene representations) related to the likelihood that subjects would later recall the relevant Face/Scene image when probed with its associated cue word. We predicted that the degree to which encoding trials were associated with category-specific activity patterns would be an indicator of stimulus representational strength at encoding, and hence a predictor of subsequent memory. We also assessed whether this putative relationship between representational strength and subsequent memory differed across prefrontal and temporal lobe structures.

On the one hand, the representation of visual categories, such as Faces and Scenes, has been most clearly established within temporal lobe structures (Epstein & Kanwisher, 1998; Haxby et al., 2001; Kanwisher, McDermott, & Chun, 1997; Puce, Allison, Asgari, Gore, & McCarthy, 1996; Weiner & Grill-Spector, 2010), and there is some evidence for category-selective subsequent memory effects in temporal lobe areas (e.g., Kirchhoff, Wagner, Maril, & Stern, 2000; Prince, Dennis, & Cabeza, 2009). On the other hand, prefrontal cortex is regularly implicated in successful episodic encoding and, while visual category representation in prefrontal cortex has not been well defined through fMRI studies, recordings from individual neurons in monkeys have provided compelling evidence for category-level representations of visual stimuli in lateral prefrontal cortex (e.g., Freedman, Riesenhuber, Poggio, & Miller, 2001). Thus, while we predicted that category representation would be most robust within temporal lobe structures, we also anticipated that category representation would be observed in prefrontal cortex and closely tied to subsequent memory outcomes. Moreover, consistent with the view that prefrontal cortex operates upon the products of—and potentially influences—posterior representations (e.g., Miller & Cohen, 2001), we asked whether, on a trial-by-trial basis, the strength of representations in temporal lobe structures

was correlated with the strength of representations within prefrontal structures.

2. Methods

2.1. Subjects

Eighteen subjects (10 female) participated in the study. All were right-handed native English speakers between the ages of 18 and 27 yrs. Subjects received $20/h for their participation. Informed consent was obtained according to procedures approved by the Stanford Institutional Review Board.

2.2. Materials and procedure

The experiment consisted of alternating blocks of encoding and retrieval, all conducted during fMRI scanning. During encoding, subjects viewed nouns (cues; e.g., 'flag' 'couch') presented above grayscale photographs of well-known people (Faces; e.g., 'Tom Cruise' or 'Julia Roberts') or well-known locations (Scenes; e.g., 'Eiffel Tower' or 'Grand Canyon'). Nouns were drawn from the Medical Research Council Psycholinguistic Database (http://www.psy.uwa.edu.au/MRCDataBase/uwa_mrc.htm) and ranged in length from four to eight letters ($M = 5.4$). All nouns had a Kucera–Francis written frequency of at least five ($M = 20.7$) and a concreteness rating of at least 500 ($M = 600$). Faces and Scenes were grayscale images, 225×225 pixels, with a resolution of 150 pixels/in. Faces included hair and varied in emotional expression, but were selected and cropped such that background objects or scenes were not visible. Scenes were selected and cropped such that they did not contain any faces or prominent people. Beneath each image was a label providing a specific name for that image (e.g., 'Tom Cruise,' or 'Eiffel Tower'). Half of all Faces were male; half were female. Half of all Scenes were manmade structures (e.g., 'Eiffel Tower'); half were natural landscapes (e.g., 'Grand Canyon').

Across seven blocks, a total of 72 cue–Face pairs and 72 cue–Scene pairs were studied (an additional 8 pairs were generated as fillers). For each image category (Faces and Scenes), 48 of the 72 pairs consisted of novel cues paired with novel images; the remaining 24 pairs consisted of repeated cues paired with novel images. In other words, of the 48 novel pairs, half of the cues were later paired with a second image (always from the opposite image category) to create overlapping pairs. The overlapping pairs were intended to elicit interference during retrieval—a topic that is not the focus of the present study but is described elsewhere (Kuhl, Rissman, Chun, & Wagner, 2011). The novel pairs were distributed across blocks 1–6; the overlapping pairs were distributed across blocks 2–7. Given the aims of the present study, overlapping pairs were excluded from all of the analyses reported here except for classification of image sub-category, as described below.

Each encoding trial lasted 4 s and was followed by an 8 s inter-trial baseline period. The baseline period began with the presentation of a fixation cross for 800 ms, followed by six randomly left/right-oriented arrows (800 ms each). Each arrow was followed by a 400 ms fixation cross and subjects were instructed to indicate the left/right orientation of the arrow via a button box held in their right hand. The arrow task was included in order to disrupt or eliminate continued encoding of pairs during the baseline period—which would otherwise be likely given subjects' knowledge of the forthcoming retrieval phase—and to therefore allow elicited hemodynamic responses to subside before the onset of the next trial. Subjects did not make any response during the encoding trial itself and were not provided with specific instructions on how to form the cue–image associations; however, subjects were made aware of how their memory would be tested before beginning any of the encoding blocks.

During the retrieval blocks, subjects were presented with cues that had appeared in the immediately preceding encoding block and attempted to retrieve the corresponding image. For cues that were to be paired with a second image (the overlapping pairs), this re-pairing did not occur until the ensuing encoding block; thus, each of the 48 novel Face pairs and each of the 48 novel Scene pairs was presented in an encoding round and probed during a retrieval round *before* the overlapping pair was encoded. Critically, all of the retrieval data reported in the present study concern performance for the novel pairs.

Each retrieval trial lasted 5 s and consisted of a single cue presented above a square equal in size to the Face/Scene images. The interior of the square was black, matching the background screen color, thus giving the impression of an empty box. The outline of the square was white for the first 4 s of the trial, changing to red for the last 1 s to indicate that the trial was about to end. Subjects were instructed to covertly recall the target image and to make one of five responses using a button in their right hand to indicate their retrieval success: (1) "face-specific" indicated that they were able to recall the specific image that was paired with the cue, and that it was a Face; (2) "face-general" indicated that they were not able to recall the specific image paired with the cue, but that they had a general memory of the image being a Face; likewise for (3) "scene-specific," and (4) "scene-general," and (5) "Don't Know" indicated that the subject could not remember anything about the target image. Subjects could respond at any point during the 5-s duration of the trial and no emphasis was placed on responding quickly. Retrieval trials were followed by a 7-s baseline period during which a fixation cross was presented. No response was required during the baseline period. A 'passive' baseline was used during retrieval—as opposed to the active baseline at encoding—because subjects

did not have an obvious incentive to continue processing stimuli in between trials, thereby reducing the need to distract subjects during this period.

All cue–image pairings were randomized for each subject. Each encoding and retrieval block contained an equal number of Face and Scene trials, arranged in pseudo-random order to control for average serial position of Faces and Scenes.

Following the last retrieval block, subjects completed a face/scene localizer task during which novel, non-famous faces and scenes were presented, one at a time, and subjects were instructed to make a button response whenever an image repeated on consecutive trials. The scan consisted of 153 volumes (5 min, 6 s) and contained 8 blocks of faces and 8 blocks of scenes (7 images/block).

2.3. fMRI methods

fMRI scanning was conducted at the Lucas Center at Stanford University on a 3.0 T GE Signa MRI system (GE Medical Systems). Functional images were obtained using a T2*-weighted 2D gradient echo spiral-in/out pulse sequence; repetition time (TR) = 2 s; echo time (TE) = 30 ms; flip angle = 75°; 30 slices, 3.4 mm × 3.4 mm × 4 mm; axial oblique sequential acquisition. The seven encoding blocks consisted of 940 total volumes. Image preprocessing and data analyses were performed using SPM5 (Wellcome Department of Cognitive Neurology, London).

Functional data were corrected for slice timing and head motion. Subjects' structural images were coregistered to functional images and segmented into gray matter, white matter, and cerebrospinal fluid. Gray matter images were stripped of remaining skull and normalized to the Montreal Neurological Institute (MNI) gray matter template. Parameters generated during normalization of the gray matter images were applied to the non-segmented structural and functional images. Images were resampled to 3 mm cubic voxels and smoothed with an 8 mm FWHM Gaussian kernel.

Univariate data analyses were conducted under the assumptions of the general linear model (GLM). Individual trials were modeled using a canonical hemodynamic response function and its first order temporal derivative. Encoding data were modeled with scan session (block) treated as a covariate. Linear contrasts were applied for each subject to obtain subject-specific estimates of effects of interest. These estimates were then entered into a second-level, random-effect analysis for which one-sample t tests were applied against a contrast value of zero for each voxel; a five-voxel extent threshold was applied. Contrast maps were overlaid on a normalized canonical brain using MRIcron (http://www.sph.sc.edu/comd/rorden/mricron/). To test for univariate effects of subsequent memory, a GLM was constructed that consisted of four regressors representing two levels of image category (Faces vs. Scenes) and two levels of subsequent memory (subsequently Remembered vs. subsequently Forgotten items). Only novel pairs were included in this model. A fifth regressor represented filler trials and overlapping pairs.

2.4. Multi-voxel pattern analysis methods

Pattern classification analyses were conducted using the Princeton Multi-Voxel Pattern Analysis Toolbox (http://www.pni.princeton.edu/mvpa) and custom code implemented in MATLAB (The MathWorks, Natick, MA). All fMRI data used for classification analyses were pre-processed in the same way that data for the univariate analyses were pre-processed (including normalization and spatial smoothing). Additionally, data used for pattern classification analyses were high-pass filtered (0.01 Hz), detrended, and z-scored (first across all trials within each run, then across runs but only for those trials used for classification). Classifier analyses were based on penalized logistic regression using L2-norm regularization and a penalty parameter of 100. All classification analyses used a cross-validation approach where all but two trials—one from each condition—comprised the training set and the two left-out trials comprised the testing set. Training and testing were repeated iteratively so that all trials were part of the testing set for one iteration.

All of the classifications in the present study were binary. The main classification analysis corresponded to classification of Faces vs. Scenes (image category); secondary analyses were conducted for classification of Male vs. Female Faces (Face sub-category) and for classification of Manmade vs. Natural Scenes (Scene sub-category). For classification of image category, only novel pairs were included. Overlapping pairs were excluded because of the likelihood that subsequent memory effects might differ for novel vs. overlapping pairs. For classification of image sub-categories, novel and overlapping pairs were combined because (a) subsequent memory analyses were not considered with respect to classification of image sub-category, and (b) classification of image sub-category was a very subtle distinction that benefitted from the additional power. Notably, for classification of both image category and sub-category, the cells were balanced; that is, there were an equal number of Face and Scene trials for the image category classification, an equal number of Male and Female trials for the Face sub-category classification, and an equal number of Manmade and Natural trials for the Scene sub-category classification.

For each trial in the testing set, the logistic regression classifier generated a scalar probability estimate that the trial corresponded to category A vs. category B (by construction, these probability estimates summed to unity). On each trial, the classifier's 'guess' corresponded to the category with the higher probability and was coded as 'correct' or 'incorrect' based on whether the guess matched the actual category for that trial. Classification accuracy thus represented the percentage of trials that the classifier correctly categorized. Additionally, we also computed mean *classifier evidence*––that is, the mean of the scalar probability estimate that the classifier assigned to the relevant category for each trial. This continuous measure of classifier performance capitalizes on the fact that the classifier's predictions were probabilistic rather than binary, and potentially provides a more sensitive index of category discriminability than classification accuracy. All classification analyses were performed on a trial-by-trial basis, as opposed to a volume-by-volume basis. Trial-level classifier data were obtained by averaging temporally contiguous volumes that corresponded to the expected peak of the hemodynamic response function (i.e., TRs 3–4, corresponding to 4–8 s post stimulus onset).

Anatomical ROIs were generated using the Anatomical Automatic Labeling (AAL) atlas (http://www.cyceron.fr/web/aal_anatomical_automatic_labeling.html), which provides anatomical masks in MNI space. Three temporal and five prefrontal ROIs were generated by summing the left and right masks corresponding to regions of a priori interest (Fig. 1A). Any voxels that were part of more than one mask, according to the AAL atlas, were excluded so that each mask contained an independent set of voxels. The three temporal ROIs corresponded to the AAL masks for fusiform gyrus (FG; 1686 voxels), parahippocampal gyrus (PHG; 659 voxels), and hippocampus (HIPP; 633 voxels). For the prefrontal ROIs, an ROI representing inferior frontal gyrus (IFG; 3398 voxels) was generated by summing the AAL masks corresponding to *pars orbitalis*, *pars triangularis*, and *pars opercularis* ('Frontal_Inf_Orb' + 'Frontal_Inf_Tri' + 'Frontal_Inf_Oper'). ROIs representing middle frontal gyrus (MFG; 3052 voxels) and superior frontal gyrus (SFG; 2314 voxels) corresponded to the AAL masks 'Frontal_Mid' and 'Frontal_Sup,' respectively. An ROI representing medial prefrontal cortex (mPFC; 2980 voxels) was generated by summing the AAL masks corresponding to anterior cingulate cortex, medial superior frontal gyrus, and medial orbitofrontal gyrus ('Cingulum_Ant' + 'Frontal_Sup_Medial' + 'Frontal_Med_Orb'). An ROI representing orbitofrontal cortex (OFC; 1198 voxels) was generated by summing the AAL masks corresponding to the orbital extent of the middle and superior frontal gyri ('Frontal_Sup_Orb' + 'Frontal_Mid_Orb'). Our use of anatomical masks was intended to characterize information representation within specific anatomical structures; no additional feature selection was applied.

To visualize the contribution of individual voxels to classifier performance, an importance map was generated based on a classifier applied to a meta-ROI that was comprised the sum of the three temporal and five prefrontal ROIs. To generate the importance map, the classifier-assigned weight for each voxel—which could be positive or negative, depending on a voxel's relation to category A vs. B trials—was multiplied by its mean activity level for category A trials. Voxels for which the weight and mean activity were both positive were assigned positive importance values; voxels for which the weight and mean activity were both negative were assigned negative importance values; all other cases were assigned importance values of zero (McDuff, Frankel, & Norman, 2009; Rissman, Greely, & Wagner, 2010). Importance maps were averaged across classifier iterations for each subject and then across subjects to generate a single group-level importance map.

3. Results

3.1. Behavioral results

Subjects were able to recall the target image category (either "specific" or "general" memory for the image) on the majority of retrieval trials ($M = 79.2\%$); hereinafter 'Remembered' items. 'Forgotten' items corresponded to trials on which subjects responded "Don't Know" ($M = 13.2\%$) or responded with the incorrect category ($M = 5.7\%$). Trials for which subjects failed to respond ($M = 2.0\%$) were excluded from subsequent memory analyses. The percentage of items Remembered did not differ for Face vs. Scene trials ($M = 80.1\%$ vs. $M = 78.2\%$, respectively, $t(17) = 1.14$, $p = .27$).

3.2. Category information during encoding

3.2.1. Classification of image category

Across subjects, mean classification accuracy for the category of the encoded image (Face vs. Scene) was above chance (50%) in each of the temporal and prefrontal ROIs (all $t(17)$s > 4.40, all ps < .001 and significant following Bonferroni correction; Fig. 2A). Classification based on temporal ROIs (averaged across HIPP, FG, and PHG) yielded markedly higher accuracy than classification based on prefrontal ROIs (averaged across IFG, MFG, SFG, mPFC, and OFC) ($M = 91.1\%$ vs. $M = 69.6\%$, respectively, $t(17) = 19.77$, $p < .001$). It is of
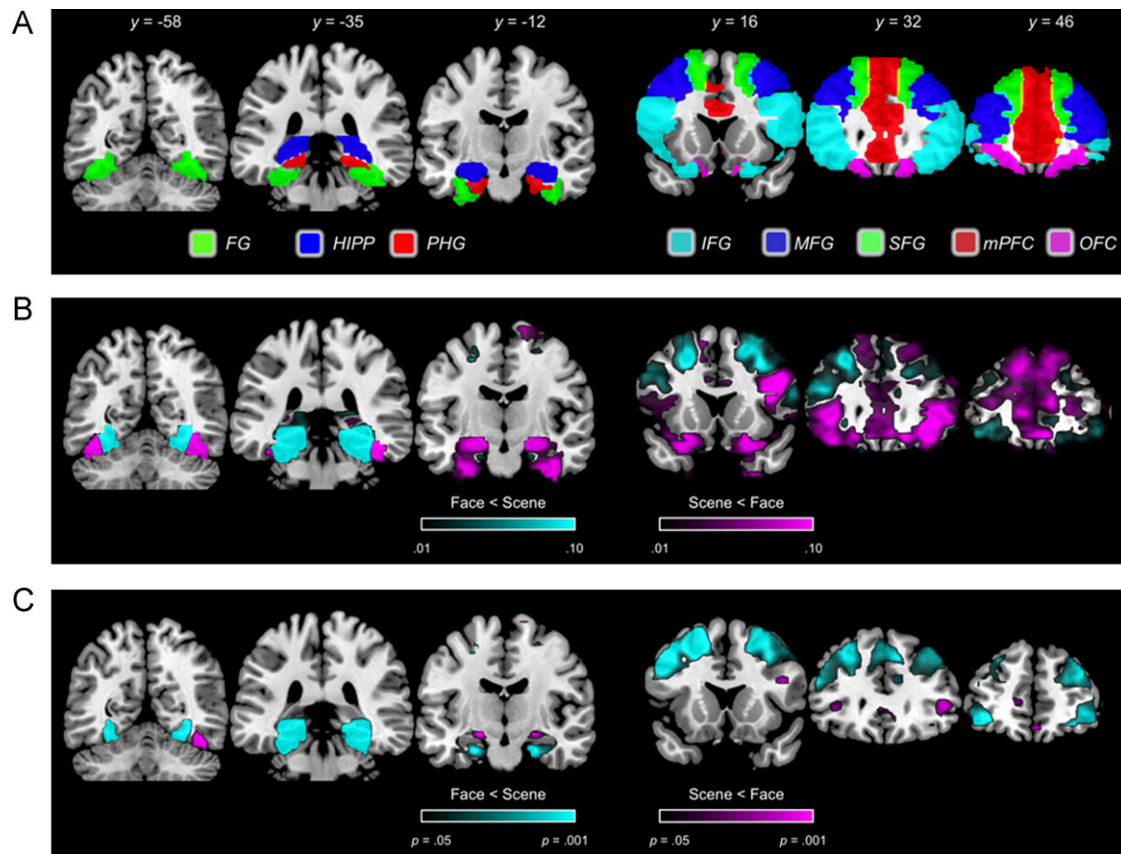
Fig. 1. (A) ROI specification for temporal (left) and prefrontal (right) regions. (B) Importance map showing voxels that positively contributed to Face vs. Scene encoding; arbitrarily thresholded at .01. (C) Univariate contrast of Face vs. Scene encoding trials.

note that this difference was in spite of the fact that the temporal ROIs were generally much smaller (number of voxels) than the prefrontal ROIs. To visualize the distribution of voxels that positively contributed to Face vs. Scene classification we ran a separate

Fig. 2. (A) Classification accuracy for image category (Face vs. Scene) across temporal and prefrontal ROIs. (B) Classification accuracy for image sub-category (Male vs. Female for Face trials; Manmade vs. Natural for Scene trials) across temporal and prefrontal ROIs. Error bars indicate standard error of the mean.

classification analysis using a single meta-ROI that combined each of the eight temporal and prefrontal ROIs (mean accuracy = 94.7%) and generated an importance map from this classification. As can be seen in Fig. 1B, within prefrontal cortex, voxels that positively contributed to Face classification were most prevalent in IFG and mPFC, and, to a lesser extent, in SFG and OFC. Voxels that positively contributed to Scene classification were more prevalent in MFG and, to a lesser extent in SFG and anterior portions of IFG. In the temporal lobes, voxels that positively contributed to Face classification were evident in posterior and anterior FG, as well as anterior HIPP. Voxels that positively contributed to Scene classification were evident in posterior PHG, posterior FG, and posterior HIPP.
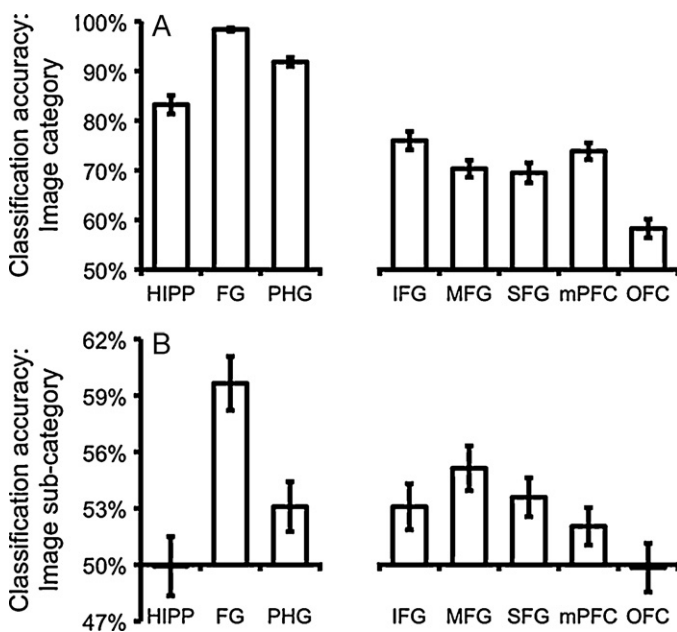
3.2.2. Classification of image sub-categories

We next tested for evidence of sub-category representation in prefrontal and temporal ROIs (Male vs. Female for Faces; Manmade vs. Natural for Scenes). Collapsing across Face and Scene sub-categories, classification accuracy was significantly above chance for temporal ROIs ($t(17) = 3.51$, $p < .005$) and prefrontal ROIs ($t(17) = 3.51$, $p < .005$) (Fig. 2B). However, sub-category classification differed robustly across temporal ROIs ($F(2,34) = 24.73$, $p < .001$), with accuracy markedly higher for FG ($M = 59.6\%$) than PHG ($M = 53.1\%$) or HIPP (49.9%). Accuracy also differed across prefrontal ROIs ($F(4,68) = 4.32$, $p < .005$), with accuracy tending to be higher in lateral prefrontal ROIs (IFG, MFG, SFG) than mPFC or OFC. For the temporal ROIs, an interaction was observed between ROI and image category ($F(2,34) = 6.36$, $p < .01$), reflecting greater classification of Face sub-category in HIPP and PHG but better classification of Scene sub-category in FG (Table 1). For the prefrontal ROIs, there was no interaction between ROI and image category ($F < 1$; Table 1).

Table 1
Classification accuracy for image sub-categories: Faces (Male vs. Female); Scenes (Manmade vs. Natural). SEM, standard error of the mean.

	Temporal			Prefrontal				
	HIPP	FG	PHG	IFG	MFG	SFG	mPFC	OFC
Faces								
Mean	51.7%	56.9%	54.3%	54.7%	55.1%	54.0%	53.5%	50.2%
SEM	2.6%	2.4%	2.0%	1.7%	1.9%	1.5%	1.8%	1.9%
Scenes								
Mean	48.1%	62.3%	51.9%	51.5%	55.2%	53.2%	50.6%	49.5%
SEM	2.1%	1.3%	1.8%	1.3%	1.8%	1.6%	1.5%	1.5%

3.3. Relationship between category information at encoding and subsequent memory

3.3.1. Trial-by-trial variability

To assess the relationship between trial-by-trial variability in the representation of category information during encoding and subsequent memory, we separated all encoding trials according to whether the target image was later Remembered vs. Forgotten. Classification accuracy and classifier evidence were then considered for Remembered vs. Forgotten items. This was done separately for Face and Scene trials, so that image category was not confounded with subsequent memory, but all data reported below were averaged across image category.

To first address the relationship between classification accuracy at encoding and subsequent memory, two ANOVAs were generated: one for the temporal ROIs and one for the prefrontal ROIs. Each ANOVA contained two factors: ROI and subsequent memory. For the ANOVA on the temporal ROIs, the subsequent memory effect was not significant ($F(1,17) = 1.37$, $p = .26$; Fig. 3A), nor did subsequent memory interact with ROI ($F < 1$). Thus, there was no evidence from the temporal ROIs that subsequently Remembered items were better classified as Faces vs. Scenes, relative to subsequently Forgotten items. Rather, both mnemonic classes of items were classified with extremely high accuracy; this was particularly evident for FG, where subsequently Remembered and Forgotten

Fig. 3. (A) Difference in classification accuracy for items subsequently Remembered vs. subsequently Forgotten across temporal and prefrontal ROIs. (B) Difference in continuous measure of classifier evidence for items subsequently Remembered vs. subsequently Forgotten across temporal and prefrontal ROIs. Error bars indicate standard error of the mean.

items were each classified with near-perfect accuracy ($M = 98.5\%$ and $M = 98.9\%$, respectively).

For the ANOVA on the prefrontal ROIs, the main effect of subsequent memory was significant, reflecting greater classification accuracy for Remembered vs. Forgotten items ($F(1,17) = 8.13$, $p < .05$; Fig. 3A). This subsequent memory effect did not interact with ROI ($F < 1$). Thus, in contrast to classification accuracy based on temporal ROIs, Face vs. Scene classification accuracy based on the prefrontal ROIs was significantly higher for items that would later be Remembered. A separate ANOVA indicated that the difference in the subsequent memory effect for prefrontal vs. temporal ROIs was marginally significant ($F(1,17) = 4.21$, $p < .06$).

The preceding analyses relating classification accuracy to subsequent memory outcomes point to a potential dissociation in terms of how diagnostic the distributed encoding activity within prefrontal vs. temporal cortex is of memory outcomes. On the one hand, these data suggest that category information is highly discriminable in ventral temporal regions, with classification accuracy approaching ceiling levels (Fig. 2A), whereas the representation of category information in prefrontal cortex is more variable and, critically, predictive of memory outcomes. On the other hand, it is important to note that our measure of classification accuracy only reflects whether neural evidence on a given trial favored the target image category or not, but does not capture potential gradations in the strength of evidence. It is possible that Forgotten items were associated with weaker temporal lobe representations than Remembered items, but that these weaker representations were nevertheless sufficient to allow for very high classification accuracy.

To address whether more subtle differences in representational strength in temporal regions were related to memory outcomes, we replicated the analyses described above—generating one ANOVA for temporal ROIs and one for prefrontal ROIs—except that, instead of considering the binary measure of classification accuracy, we now considered the continuous measure of classifier evidence. Critically, for the ANOVA on the temporal ROIs, we now observed a significant main effect of subsequent memory, reflecting greater evidence for Remembered vs. Forgotten items ($F(1,17) = 5.84$, $p < .05$; Fig. 3B). This effect did not interact with ROI ($F < 1$). For the ANOVA on the prefrontal ROIs, the effects were consistent with those based on classification accuracy: there was a main effect of subsequent memory ($F(1,17) = 6.76$, $p < .05$; Fig. 3B) and this effect did not interact with ROI ($F < 1$). While the effect was, numerically, larger for prefrontal than temporal ROIs, a separate ANOVA indicated that the difference was not significant ($F(1,17) = 1.73$, $p = .21$). Thus, these data indicate that the continuous measure of classifier evidence captured differences in category information in temporal regions that were not reflected in the categorical measure of classification accuracy (the latter null result may have partially stemmed from a restricted range due to ceiling effects).

3.3.2. Individual differences

In the preceding section, we presented evidence that trial-by-trial differences in classifier-based measures of category information were related to subsequent memory outcomes. We next asked whether cross-subject variability in the strength of category information at encoding was related to individual differences in retrieval success. Specifically, we tested for a correlation between mean classifier evidence based on prefrontal and temporal ROIs for the Face vs. Scene discrimination at encoding and the percentage of Remembered items for each subject. This was done separately for Face trials (i.e., correlating mean classifier evidence across Face trials with mean retrieval success for Face trials) and Scene trials.
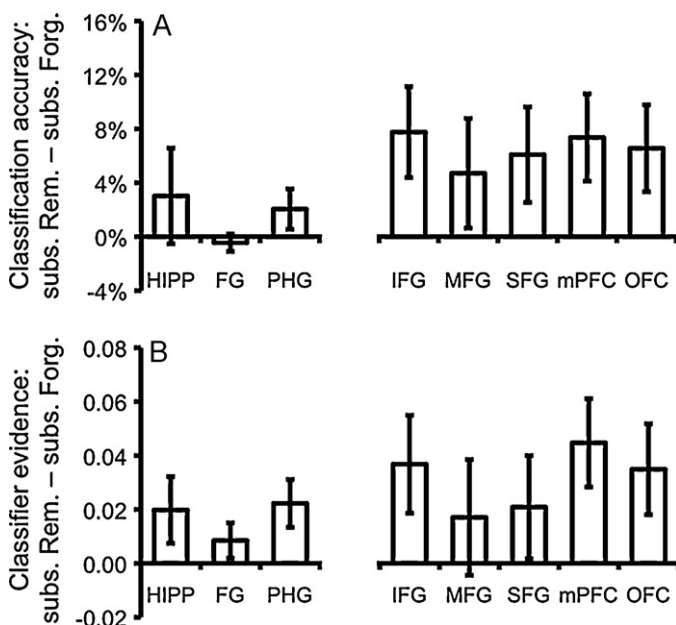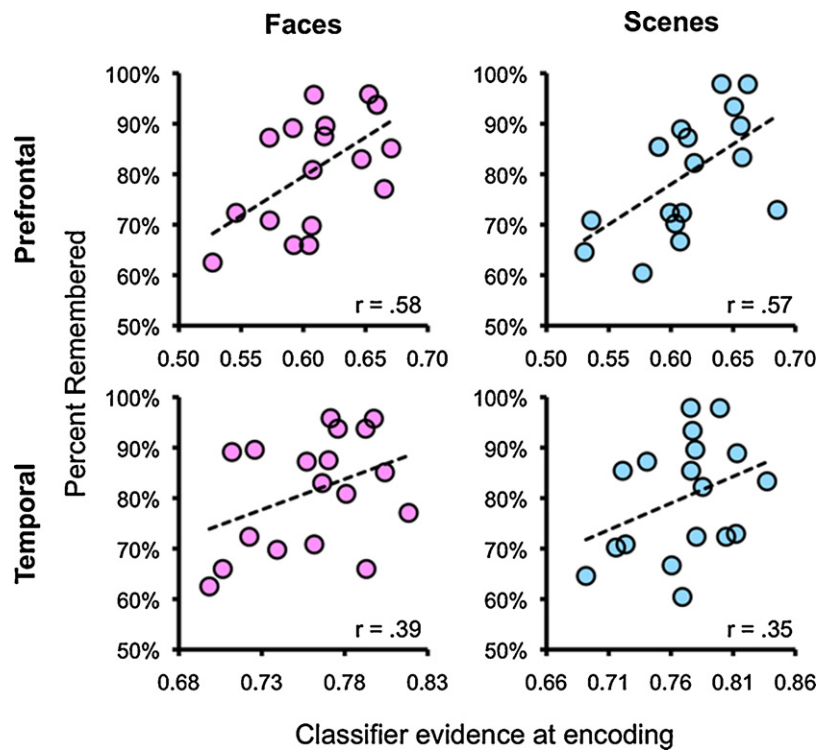
Fig. 4. Across-subject correlations showing relationship between mean classifier evidence at encoding and mean success rate at retrieval. Correlations are separately shown for prefrontal ROIs (top row; classifier evidence averaged across the five prefrontal ROIs) and temporal ROIs (bottom row; classifier evidence averaged across the three temporal ROIs) and for Face trials (left column) and Scene trials (right column).

Using classifier evidence from temporal ROIs, we observed a positive, but non-significant relationship between classifier evidence at encoding and subsequent memory for Faces and Scenes ($ps > .1$; Fig. 4). For classifier evidence from the prefrontal ROIs, the correlations for Face and Scene trials were each significant ($ps < .05$; Fig. 4). These correlations reflected a positive relationship between the discriminability of Faces vs. Scenes at encoding and later retrieval success. While statistical significance was only considered for data averaged across prefrontal ROIs vs. data averaged across temporal ROIs (to avoid excessive hypothesis testing), correlation coefficients for each ROI within the temporal and prefrontal groups are reported in Table 2. It also is worth noting that while the correlations were significant across the prefrontal ROIs, but not the temporal ROIs, the former correlations were not significantly greater than the latter (William's test: $ts < 1.4$; $ps > .20$).

3.4. Correlations between temporal and prefrontal category information

The data presented thus far indicate that, during encoding, category information was robustly represented in distributed patterns of activity in both temporal and prefrontal structures. Although these representations were clearly more robust in temporal regions, they were predictive of memory outcomes in both prefrontal and temporal regions. While it is possible that prefrontal and temporal regions represent independent forms of information, extant evidence suggests that these regions interact, with perceptual representations feeding forward from temporal to prefrontal regions (e.g., Simons & Spiers, 2003) and prefrontal regions influencing temporal representations via top-down control (e.g., Miller & Cohen, 2001; Miller, Vytlacil, Fegen, Pradhan, & D'Esposito, 2010; Tomita, Ohbayashi, Nakahara, Hasegawa, & Miyashita, 1999; Zanto, Rubens, Thangavel, & Gazzaley, 2011). To the extent that such interactions occur, trial-by-trial variability in the strength of category information within temporal regions should be correlated with variability in the strength of category information within prefrontal cortex.

To test the hypothesis that category information within temporal regions is correlated with such information within prefrontal cortex, we used the continuous measure of classifier evidence, generating a within-subject correlation coefficient for each pairing of temporal vs. prefrontal ROIs. Correlation coeffi-

Table 2
Correlation coefficients (r) representing across-subject relationship between classifier evidence for Face vs. Scene discrimination during encoding and the percentage of items later Remembered. Results are reported separately for Face trials (i.e., the correlation between classifier evidence from Face trials and subsequent memory for Face trials) and Scene trials. For descriptive purposes, correlation coefficients are reported for each temporal and prefrontal ROI, but p values are only reported for the correlations that used data averaged across temporal ROIs or averaged across prefrontal ROIs.

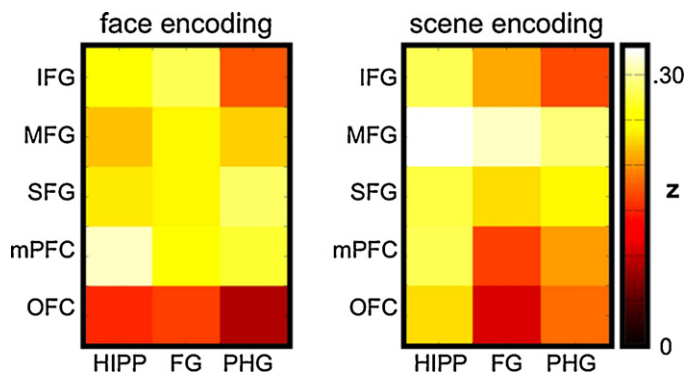|	Temporal				Prefrontal					
	HIPP	FG	PHG	Avg.	IFG	MFG	SFG	mPFC	OFC	Avg.
Face										
r	.36	.27	.43	.39	.33	.54	.62	.47	.35	.58
p	–	–	–	.11	–	–	–	–	–	.01
Scene										
r	.30	.33	.37	.35	.44	.57	.59	.41	.33	.57
p	–	–	–	.15	–	–	–	–	–	.01

Fig. 5. Correlation matrices for Face and Scene encoding trials showing the mean strength of correlations between trial-level classifier evidence in individual prefrontal and temporal ROIs. Correlation coefficients were transformed to Fisher's z prior to averaging.

cients were separately generated for Face and Scene trials and transformed to z-scores (Fischer's z). The z-scores were then considered across subjects and across prefrontal–temporal pairings for group-level statistical analyses. An ANOVA was generated with three levels: image category (Face vs. Scene trials), prefrontal ROI (IFG, MFG, SFG, mPFC, OFC), and temporal ROI (HIPP, FG, PHG).

Collapsing across image category, individual t tests confirmed that each of the prefrontal–temporal correlations (15 pairings total) was significantly greater than 0 ($ts > 5.11$, $ps < .001$, significant following Bonferroni correction), reflecting a general positive relationship between evidence within temporal and prefrontal ROIs (Fig. 5). The main effect of image category was not significant ($F < 1$), indicating that the correlations between prefrontal and temporal ROIs were not, overall, different for Face vs. Scene trials. The main effect of prefrontal ROI was significant ($F(4,68) = 6.84$, $p < .001$), with MFG displaying the strongest correlations with temporal ROIs and OFC displaying the weakest. The main effect of temporal ROI was marginally significant ($F(2,34) = 2.75$, $p = .08$), with HIPP displaying somewhat stronger correlations with prefrontal ROIs relative to FG and PHG. Additionally, the interaction between prefrontal ROI and temporal ROI was significant ($F(8,136) = 4.36$, $p < .001$), indicating that the strength of correlations across temporal ROIs varied as a function of the prefrontal ROI with which it was correlated. Moreover, an interaction between image category and prefrontal ROI ($F(4,68) = 2.93$, $p < .05$) indicated that the strength of correlations with temporal structures differed across prefrontal ROIs as a function of the type of image being encoded. For example, during Scene encoding, MFG displayed the strongest correlations with temporal ROIs, whereas during Face encoding mPFC displayed the strongest correlations with temporal ROIs. The interaction between image category and temporal ROI was not significant ($F(2,34) = 2.51$, $p = .10$), nor was the three-way interaction between image category, prefrontal ROI, and temporal ROI ($F(8,136) = 1.63$, $p = .12$). Individual ANOVAs with factors of temporal ROI and image category were also applied to each prefrontal ROI to test whether any of the prefrontal ROIs displayed correlations that varied across temporal ROIs as a function of the category of image being encoded. For OFC, a robust interaction between image category and temporal ROI was observed ($F(2,34) = 7.22$, $p < .005$). This interaction reflected stronger correlations between OFC and HIPP/PHG during Scene encoding, relative to Face encoding, and stronger correlations between OFC and FG during Face encoding, relative to Scene encoding. No other prefrontal ROI displayed a significant interaction between temporal ROI and image category (all $ps > .23$).

3.5. Univariate analyses of subsequent memory effects

To assess the relationship between the preceding MVPA and more typical univariate subsequent memory analyses, we conducted two univariate analyses on the present data. First, we contrasted encoding trials that were subsequently Remembered vs. those subsequently Forgotten (collapsing across image category). At a standard threshold ($p < .001$, 5-voxel extent) we did not observe any clusters positively associated with subsequent memory that overlapped with the prefrontal or temporal ROIs. At a very liberal threshold ($p < .01$, 5-voxel extent) the only clusters that overlapped with the prefrontal ROIs were in bilateral IFG; no clusters overlapped with the temporal ROIs. To more closely parallel the subsequent memory analysis applied to the classifier data, we also tested for an interaction between the subsequent memory effects for Faces and Scenes [(Face Remembered > Face Forgotten) > (Scene Remembered > Scene Forgotten)]. At a standard threshold ($p < .001$, 5-voxel extent) there were no clusters, either from the positive or negative tail of the contrast that overlapped with the prefrontal or temporal ROIs. At a very liberal threshold ($p < .01$, 5-voxel extent) there were no clusters from the positive tail of the contrast that overlapped with the prefrontal or temporal ROIs; for the negative tail, a few small clusters of activation, bilaterally, overlapped with the PHG and FG ROIs.

3.6. Pattern similarity analysis

To complement the main classification analyses reported above, we also conducted a pattern similarity analysis (e.g., Kriegeskorte, Mur, & Bandettini, 2008) for which the pattern of activity elicited during each encoding trial was correlated with the pattern of activity elicited on every other encoding trial. This analysis allowed us to consider how correlations varied across trials as a function of subsequent memory (Remembered vs. Forgotten) and visual category (within-category vs. between-category). This was separately performed for each of the temporal and prefrontal ROIs. All correlations were transformed to z-scores and then averaged according to subsequent memory status and visual category. Subjects with five or fewer trials in one or more of the four relevant bins (Face-Remember, Face-Forget, Scene-Remember, Scene-Forget) were excluded to reduce the influence of small samples on the correlations.

Consistent with the general success of our pattern classifier in discriminating Face vs. Scene trials, within-category correlations (e.g., Face trials correlated with other Face trials) were significantly higher than between-category correlations, both in prefrontal ($t(10) = 4.99$, $p < .001$) and temporal regions ($t(10) = 10.74$, $p < .001$) (Fig. 6). Notably, for the temporal ROIs, within-category correlations were greater among Remembered items than among Forgotten items ($t(10) = 2.89$, $p < .05$). Indeed, Forgotten items were more positively correlated with within-category Remembered items than other within-category Forgotten items ($t(10) = 3.10$, $p < .05$). Between-category correlations were numerically, but not significantly more negative for Remembered items (e.g., Face-Remember to Scene-Remember) than Forgotten items (e.g., Face-Forget to Scene-Forget) ($t(10) = -1.03$, $p = .32$). There was, however, a significant interaction between subsequent memory group (Remembered–Remembered vs. Forgotten–Forgotten) and category (within vs. between) ($F(1,10) = 7.98$, $p < .05$), reflecting the tendency for Remembered items to be associated with greater within-category similarity and greater between-category dissimilarity than Forgotten items.

For the prefrontal ROIs, correlations did not differ among Remembered items and Forgotten items either within categories ($p = .58$) or between categories ($p = .83$). The correlations in prefrontal ROIs were, however, much weaker than the temporal ROIs
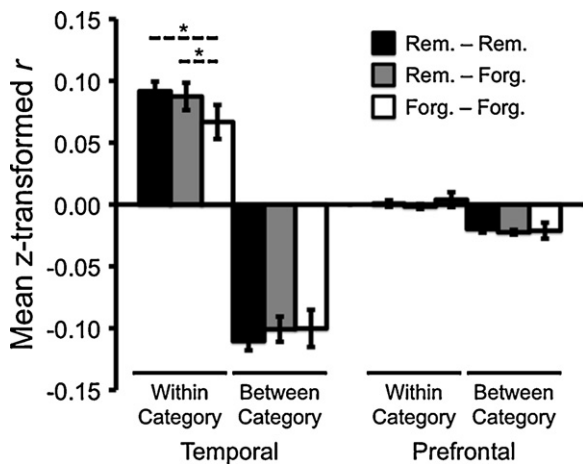
Fig. 6. Pattern similarity analysis. Correlation coefficients were computed for all pairs of encoding trials, reflecting the similarity of the neural response across voxels for each pair of trials. The resulting *r* values were *z*-transformed and averaged according to whether they represented within-category correlations (e.g., Face–Face), between-category correlations (Face–Scene) and according to subsequent memory status (e.g., Remembered–Remembered). The similarity analysis was separately performed for each prefrontal and temporal ROI and data were then averaged across the three temporal ROIs and the five prefrontal ROIs. Within-category similarity was greater than between-category similarity for both temporal and prefrontal regions. For the temporal regions, within-category similarity was greater among Remembered trials (Rem.–Rem.) than among Forgotten trials (Forg.–Forg.). Additionally, Forgotten items were more similar to within-category Remembered items (Rem.–Forg.) than to other within-category Forgotten items (Forg.–Forg.). Error bars represent standard error of the mean. **p < .05.

(see Fig. 6), likely reflecting a lower proportion of category-selective voxels in the prefrontal ROIs.

4. Discussion

The present study yielded three main findings. First, during encoding of words paired with images of Faces or Scenes, MVPA revealed that information highly diagnostic of visual category was present in distributed patterns of activity in temporal and prefrontal structures. Second, representation of category information during encoding was positively associated with subsequent memory outcomes. Notably, in prefrontal cortex, both trial-by-trial variation and individual differences in classifier-based measures of category information were predictive of subsequent memory. Third, classifier-based measures of category information within temporal cortex were correlated with prefrontal information, with the strength of these correlations varying across specific temporal and prefrontal pairings and as a function of the category of encoded material. Below, we first consider some basic issues regarding the use of pattern classification to infer representational strength during episodic encoding and then consider the specific implications and significance of each of our main findings.

4.1. Representational strength at encoding as measured by pattern classification

Central to the present study is the idea that pattern classification can be used to measure the strength of category representation during episodic encoding. As there have been relatively few studies to date that have addressed this topic with similar methods, it is important to consider some of the advantages and caveats inherent to this analysis approach.

Perhaps the most important consideration with respect to the present methodology is that our measure of category representation was, in fact, a measure of category differentiation: that is, our classifier was trained to discriminate Faces vs. Scenes. As such, a

given trial should have been more likely to be successfully classified to the extent that its neural representation was (a) similar to the prototypical exemplar of its category, and (b) dissimilar to the prototypical exemplar of the other category. While our pattern classification approach does not, on its own, allow for separating the relative importance of these two factors, the pattern similarity analysis described in Section 3.6 provides some support for each of these ideas, at least within temporal regions. Namely, Remembered items tended to be associated with greater within-category similarity and greater between-category dissimilarity (though, there was clearer evidence for the former). Our finding of greater within-category pattern similarity for Remembered items vs. Forgotten items may suggest a benefit of prototypicality (Posner & Keele, 1968) and may conceptually relate to recent evidence that higher pattern similarity across repeated acts of encoding of an item is associated with better subsequent memory for that item (Xue et al., 2010). While we did not observe strong evidence for greater between-category dissimilarity for Remembered items, another recent study found that pattern dissimilarity across temporally adjacent encoding events is associated with better subsequent memory for context (Jenkins & Ranganath, 2010).

The pattern similarity analysis described here also addresses another important question raised by the pattern classification approach. Namely, because our pattern classifier was trained to discriminate Face vs. Scene trials by using data from the encoding phase, and because most of the encoded trials were subsequently remembered, it is theoretically possible that the observed relationship between classifier performance and subsequent memory was influenced by a subtle bias in classifier training. That is, it is possible that Remembered and Forgotten items were, in fact, associated with comparably 'strong' representations, but that these representations were simply distinct. If so, the fact that the training data used by the classifier was, on average, comprised more Remembered than Forgotten items might have led to better classification of Remembered items than Forgotten items simply because the classifier was trained on more Remembered items. However, the results of our pattern similarity analyses argue against this interpretation. Specifically, the within-category correlations indicated that Forgotten items were: (a) less correlated with other Forgotten items than Remembered items were with other Remembered items, and (b) *more* correlated with Remembered items than other Forgotten items. The weaker correlation among Forgotten items argues against a distinct but comparably strong representation for Forgotten items. The greater correlation of Forgotten items with Remembered items, relative to other Forgotten items, is consistent with the idea that Forgotten items tended to be weaker or noisier versions of the representations for Remembered items in the same way that two copies of an original, each subject to independent influences of noise, will each tend to be more correlated with the original than with each other.

While the pattern similarity results provide a compelling argument against concerns about bias in the training data leading to better classification of Remembered items, we also addressed this concern in a second way. Namely, as an alternative to training the pattern classifier using the encoding data, we ran a separate classification analysis for which data from the face/scene localizer were used to train the classifier and the classifier was then tested on all encoding trials. While the localizer was comprised fewer trials—and thus potentially underpowered relative to our first approach—the advantages of training the classifier on the localizer data are (a) that the localizer did not involve intentional episodic encoding, (b) stimuli in the localizer task were not accompanied by words, and (c) all stimuli in the localizer were non-famous, novel images, thus reducing the contribution of semantic representations. Critically, data from this classifier strongly replicated our main findings: classifier evidence was greater for subsequently

Remembered items both in temporal ROIs ($F(1,17) = 4.72$, $p < .05$; no interaction with ROI ($F < 1$)) and prefrontal ROIs ($F(1,17) = 9.77$, $p < .01$; no interaction with ROI ($F < 1$)). Thus, these data indicate that our subsequent memory results cannot be fully attributed to the fact that our classifier was: (a) trained on data from an episodic encoding task with an imbalance in Remembered vs. Forgotten items, (b) trained on data where words were paired with images, or (c) trained on well-known and semantically rich images.

A final issue related to our approach is whether pattern classification analyses, or multi-voxel pattern analyses more generally, represent a more sensitive means for assessing encoding success. At first pass, the present results appear consistent with this idea as we observed reliable subsequent memory effects across prefrontal and temporal regions using our pattern classification analysis, but we did not observe univariate results that were significant at a standard threshold ($P < .001$). However, the comparison of univariate vs. classification results is not straightforward. For example, whereas a fast event-related design is often optimal for univariate analyses, a slow event-related design of the type used here is typically better-suited to pattern classification analyses. Additionally, because univariate analyses typically involve applying thousands or even tens of thousands of statistical tests (one test per voxel), statistical thresholds are typically much stricter to protect against false positives, whereas with classification analyses, data from thousands or tens of thousands of voxels can be aggregated so that a key analysis may reduce to a single statistical test. Despite the caution that is warranted in comparing univariate analyses to MVPA, it is likely that MVPA will offer increased sensitivity in many contexts (Norman et al., 2006), potentially including the study of episodic encoding success[1] (Watanabe et al., 2011).

4.2. Prefrontal category information and subsequent memory

In the present study, Face and Scene trials were associated with differential patterns of activity within prefrontal cortex during episodic encoding, as reflected by the success of our pattern classification analyses. Somewhat surprisingly, classification accuracy was robust across all prefrontal ROIs, suggesting widespread representation of category information across prefrontal cortex. However, the success of classification across prefrontal ROIs does not indicate that the information content was equivalent across ROIs. Rather, as the univariate contrast of Face vs. Scene encoding indicated, Face- and Scene-sensitive voxels were differentially distributed across prefrontal ROIs, with Face-sensitive voxels most prevalent in inferior frontal and medial prefrontal cortex and Scene-sensitive voxels most prevalent in middle frontal regions (Fig. 1B). Consideration of image sub-category classification (i.e., decoding the Male/Female

status of Faces or the Natural/Manmade status of Scenes) provided further evidence for category representation in prefrontal cortex, as we observed reliable sub-category classification in prefrontal cortex—particularly in lateral prefrontal regions (IFG, MFG, SFG) (Kaul, Rees, & Ishai, 2011).

The observation of category sensitivity in prefrontal cortex raises a fundamental question: what is the nature of prefrontal representations of category? Despite considerable interest in the topic, a definitive specification of the functional organization of prefrontal cortex has proven to be elusive (e.g., Wilson, Gaffan, Browning, & Baxter, 2010; Wood & Grafman, 2003). At a first level of analysis, the dissociable patterns of activity for Faces and Scenes observed here appear consistent with the idea of content-sensitivity within prefrontal cortex during encoding (e.g., Golby et al., 2001; Grady, McIntosh, Rajah, & Craik, 1998; Johnson, Raye, Mitchell, Greene, & Anderson, 2003; McDermott, Buckner, Petersen, Kelley, & Sanders, 1999; Wagner, Poldrack, et al., 1998). In particular, the localization of Face-sensitive voxels to IFG is consistent with prior fMRI studies of Face processing in humans (e.g., Courtney, Petit, Maisog, Ungerleider, & Haxby, 1998; Ishai, Schmidt, & Boesiger, 2005) and with studies of monkey prefrontal cortex which have demonstrated Face-sensitive responses in the inferior frontal convexity using both fMRI (Tsao, Schweers, Moeller, & Freiwald, 2008) and recordings from individual neurons (Scalaidhe, Wilson, Goldman-Rakic, 1997). Similarly, the localization of Scene-sensitive voxels to more dorsal aspects of prefrontal cortex in the present study is potentially consistent with evidence that dorsal prefrontal cortex represents spatial information, both in humans (e.g., Courtney et al., 1998) and monkeys (e.g., Wilson, Scalaidhe, & Goldman-Rakic, 1993).

While at least some degree of content-sensitivity in prefrontal cortex seems likely, prefrontal content-sensitivity is thought to fundamentally differ from content-sensitivity in posterior sites. For example, a hallmark of prefrontal representations of perceptual information is that they are modulated by behavioral relevance—that is, prefrontal cortex preferentially represents relevant or diagnostic features of an event (e.g., Duncan, 2001; Freedman et al., 2001; Freedman, Riesenhuber, Poggio, & Miller, 2003; Li, Ostwald, Giese, & Kourtzi, 2007; Rainer, Asaad, & Miller, 1998). In the present study, image category (Face vs. Scene) was of clear relevance to subjects during encoding, as they were aware of the forthcoming retrieval phase. Notably, however, we also observed reliable classification of image sub-category in prefrontal cortex even though the distinctions between sub-categories were not explicitly relevant. Thus, in future work it may be of interest to specifically consider how prefrontal representations of information during encoding vary as a function of perceived behavioral relevance, and whether the relationship between representation and subsequent memory is modulated by perceived relevance. Additionally, prefrontal representations may differ from posterior representations in the degree to which they allow for integration across distinct types of information, particularly when conjunctions of information are behaviorally relevant (e.g., Prabhakaran, Narayanan, Zhao, & Gabrieli, 2000; Rao, Rainer, & Miller, 1997). Thus, in contrast to Face and Scene-sensitive regions in temporal lobe structures that may predominantly reflect visual features of stimuli, prefrontal regions may integrate visual, semantic, or other forms of information.

An alternative to this content-representation account is the possibility that the observed prefrontal sensitivity to visual categories reflects not the representation of visual stimuli, per se, but the engagement of distinct control processes engaged during Face vs. Scene encoding. On the one hand, separable processes could be engaged precisely because prefrontal cortex exhibits content-sensitivity—that is, analogous processes may be supported by distinct structures according to the type of information being

---

[1] In light of a recent study that used MVPA to show that individual encoding trials can be successfully classified as subsequently Remembered vs. Forgotten (Watanabe et al., 2010), we considered whether such an approach could be applied to the present data. However, whereas our classification of Faces vs. Scenes involved a balanced set of 48 trials per condition, classification of Remembered vs. Forgotten items required artificially balancing the Remembered and Forgotten bins (to avoid biased classification) and doing so within each image category (to avoid confounds of image type). Accordingly, four subjects were excluded from this analysis because they had fewer than 5 trials in either the Face–Forget or Scene–Forget bins. For each of the remaining 14 subjects, we included 10 iterations where, for each iteration, a different random set of trials was excluded to artificially balance the conditions. Averaging across temporal ROIs, classification accuracy did not significantly differ from chance ($M = 48.8\%$, $t(13) = -.77$, $p = .45$); likewise for prefrontal ROIs ($M = 51.7\%$, $t(13) = 1.25$, $p = .23$). Considering performance for individual ROIs, however, SFG was significantly above chance ($M = 54.7\%$, $t(13) = 3.42$, $p < .005$, significant following Bonferroni correction). While it is of theoretical interest to determine whether this approach of directly classifying items as subsequently Remembered vs. Forgotten will yield fundamentally different conclusions than the approach employed in the present paper, the lack of adequate power for this analysis precludes such a discussion here.

processed (Johnson et al., 2003). Alternatively, different visual categories may tend to differentially engage domain-general processes. For example, in the present study, Face and Scene trials may have differed in the degree to which they elicited semantic analysis, sub-vocal rehearsal, attention to spatial information, or any number of processes that might be reflected in differential prefrontal activation (e.g., Baker, Sanders, Maccotta, & Buckner, 2001; Demb et al., 1995; Johnson et al., 2003; Otten & Rugg, 2001; Race, Shanker, & Wagner, 2009; Wig, Miller, Kingstone, & Kelley, 2004). The potential for spontaneous variation in encoding strategy is particularly plausible in the present study, where we did not prescribe a specific strategy (Kirchhoff & Buckner, 2006). Ultimately, while accounts of visual category sensitivity in prefrontal cortex based on content-specificity vs. type of processing are theoretically dissociable, these accounts are not mutually exclusive (Wood & Grafman, 2003).

The content representation and control process account of prefrontal category representation can also be extended to account for the relationship between classifier performance and subsequent memory. For example, one account of the subsequent memory results is that encoding success was a function of the degree to which control or strategic processes were engaged during encoding. Prefrontal cortex is known to be particularly necessary when mnemonic tasks require strategic processing (e.g., Shimamura, 1995) and the encoding task employed here, which required the formation of novel associations between words and images, was likely to engage such processes. Thus, to the extent that Faces and Scenes engaged distinct processes, the observed relationship between classifier performance and subsequent memory could reflect the success with which these mechanisms were engaged. Alternatively, to the extent that prefrontal classification of visual category was driven by content representation, the present results may reflect a relationship between the fidelity with which stimuli were represented at encoding and the likelihood that they were later retrieved. To the extent that prefrontal representations involve integration of various features, this relationship could also reflect the degree to which features were successfully integrated. These competing accounts can potentially be addressed by considering how prefrontal category representation—and its relation to subsequent memory—is modulated by task demands: for example, is category information in prefrontal cortex weaker when images are incidentally encoded and strategic processing is not invoked or are these representations more obligatory and independent of control processing?

A separate question concerning the present results is why the relationship between category information and subsequent memory did not significantly differ across prefrontal ROIs? While the general point that prefrontal cortex supports successful episodic encoding is reflected in an extensive literature (e.g., Blumenfeld & Ranganath, 2007; Brewer et al., 1998; Clark & Wagner, 2003; Kim, 2011; Spaniol et al., 2009), this literature has typically described subsequent memory effects in IFG and less frequently in more dorsal prefrontal cortex (Blumenfeld & Ranganath, 2007; Kim, 2011). At present, it is not clear whether this apparent dissociation between the present results and the broader literature considering univariate subsequent memory analyses is simply attributable to the increased sensitivity of classification analyses (Norman et al., 2006; Watanabe et al., 2011), or whether it reflects a fundamentally different relationship that was captured by the present analyses (e.g., being due to the associative demands of the subsequent memory test; Blumenfeld & Ranganath, 2007).

4.3. Temporal lobe category information and subsequent memory

Category-selectivity within temporal lobe structures has been extensively studied with respect to face vs. scene processing,

with faces known to elicit activation in fusiform gyrus (which includes multiple distinct patches that differentially respond to faces; e.g., Weiner & Grill-Spector, 2010) and scenes eliciting activation in parahippocampal cortex (e.g., Epstein & Kanwisher, 1998). Responses in these regions have been shown to differentiate successful vs. unsuccessful Face or Scene encoding. For example, during Scene encoding, greater activation in parahippocampal cortex is associated with superior subsequent memory (e.g., Awipi & Davachi, 2008; Brewer et al., 1998; Hayes, Nadel, & Ryan, 2007; Kirchhoff et al., 2000; Preston et al., 2010; Prince et al., 2009; Turk-Browne, Yi, & Chun, 2006). A similar relationship is even observed when considering pre-trial parahippocampal activation (Turk-Browne et al., 2006), suggesting a relationship between attentional variance and memory formation. Likewise, responses within fusiform gyrus during Face encoding are predictive of subsequent face memory (Nichols, Kao, Verfaellie, & Gabrieli, 2006; Prince et al., 2009; Sergerie, Lepage, & Armony, 2005;).

In the present study, the relationship between category representation in temporal lobe structures and subsequent memory was subtle: while subsequently Forgotten items were not associated with lower classification accuracy during encoding, they were associated with reliably weaker classifier evidence. In other words, subsequently Forgotten items were clearly processed to a degree that allowed the vast majority of trials to be correctly classified (e.g., 98.9% accuracy for FG), but there was nonetheless evidence that representations in temporal structures were stronger for subsequently Remembered vs. Forgotten images.

The present results relating temporal lobe category representation during encoding to subsequent memory outcomes complement prior work demonstrating that memory outcomes are closely related to the strength with which temporal lobe category information is *reactivated* at retrieval (e.g., Kuhl et al., 2011; for review, see Rissman & Wagner, in press). Together, these findings indicate that strong category representation in temporal lobe regions is diagnostic of both successful encoding and successful retrieval. While the application of MVPA to studying neural reactivation at retrieval has, to date, received more attention (e.g., Johnson, McDuff, Rugg, & Norman, 2009; Lewis-Peacock & Postle, 2008; McDuff et al., 2009; Polyn, Natu, Cohen & Norman, 2005), the present findings suggest that MVPA may prove a useful tool for relating encoding operations to subsequent retrieval or reactivation. While the present study focused on relatively coarse levels of representation (faces vs. scenes), a particularly interesting question for future research is whether the relative strength with which individual features of an event are represented during encoding, as measured by MVPA, is predictive of the degree to which these features are later remembered and/or reactivated. As such, MVPA may constitute a very sensitive and unique tool for measuring how attention is oriented during event encoding and how attentional allocation relates to memory outcomes.

4.4. Correlations between prefrontal and temporal regions

In the present study, we separately considered category information in prefrontal and temporal structures. However, successful encoding likely depends on interactions between prefrontal and posterior sites (e.g., Simons & Spiers, 2003; Summerfield et al., 2006). While functional connectivity (Friston, Frith, Liddle, & Frackowiak, 1993) has often been considered in terms of inter-regional correlations in the fMRI time series data, connectivity can also be indexed by correlated fluctuations in trial-by-trial activity estimates across regions (e.g., Rissman, Gazzaley, & D'Esposito, 2004). Here, rather than correlating univariate activity measures across distinct regions, we assessed whether MVPA-based measures of category information derived from distinct ROIs were correlated. Specifically, we

tested whether trial-by-trial variance in the strength of posterior representations of visual category was correlated with variance in the strength of prefrontal representations of visual category. We assessed this relationship across all pairings of prefrontal and temporal ROIs and as a function of image category (Face vs. Scene encoding).

Overall, the correlations between prefrontal and temporal regions were positive, indicating that the strength of classifier evidence in temporal ROIs was positively related to the strength of classifier evidence in prefrontal ROIs. The robust positive relationship between information in temporal and prefrontal ROIs indicates that these representations were not independent, consistent with the idea of frontal–temporal interactions. However, we did observe a significant interaction between prefrontal ROI and image category, indicating that correlations with temporal lobe structures varied across prefrontal ROIs according to whether a Face or Scene was being encoded. Considering individual prefrontal ROIs, OFC was selectively associated with an interaction between temporal ROI and image category—that is, the strength of OFC's correlation with HIPP, PHG, and FG was strongly modulated by image category. Although not predicted a priori, the selective interaction for OFC is intriguing and may suggest a sensitivity of OFC to visual category representations in temporal lobe structures (Bar et al., 2006).

The present region-to-region correlation analyses also suggest an interesting alternative to traditional functional connectivity analyses. That is, by leveraging the sensitivity of MVPA, we were able to characterize correlations in information representation across distinct neural sites. A related approach is to test for individual voxels whose activation correlates with the output of a classifier applied in some region of interest (e.g., Kuhl et al., 2011; Li, Mayhew, & Kourtzi, 2009). In the context of episodic encoding, using MVPA to test for correlations between neural sites is particularly appealing for consideration of how perceptual information propagates from early visual regions to higher-level visual regions and, ultimately, to putatively higher-order prefrontal regions.

4.5. Conclusion

Here, we employed a novel approach to examine encoding factors that support successful memory formation. By using MVPA, we were able to decode the visual category of information currently being encoded and to assess (a) how these representations were distributed across prefrontal and temporal regions, and (b) how the strength of these representations related to later memory outcomes. We observed strong evidence for category representation both in prefrontal and temporal lobe structures, with the strength of prefrontal information predictive of later memory success both on a trial-by-trial basis and across subjects. The relationship between information strength and subsequent memory was more subtle—but still robust—in temporal regions, with small reductions in strength reflecting a lower likelihood of subsequent remembering. Notably, these reductions in information strength in temporal regions were not evident in trial-level measures of classification accuracy as the reductions were too subtle to substantially lower the probability that an individual trial was successfully classified. This dissociation between measures of classifier performance provides important evidence for graded representations of information during encoding. Finally, consideration of trial-by-trial variance in classifier-based evidence derived from prefrontal vs. temporal regions revealed robust positive correlations between information strength in these regions, suggesting their functional interactivity. Together, by characterizing the nature and consequences of neural representations during event encoding, these results further our understanding of how visual experiences are translated into lasting memories.

References

Awipi, T., & Davachi, L. (2008). Content-specific source encoding in the human medial temporal lobe. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 769–779.

Baker, J. T., Sanders, A. L., Maccotta, L., & Buckner, R. L. (2001). Neural correlates of verbal memory encoding during semantic and structural processing tasks. *NeuroReport*, 12, 1251–1256.

Bar, M., Kassam, K. S., Ghuman, A. S., Boshyan, J., Schmid, A. M., Dale, A. M., et al. (2006). Top-down facilitation of visual recognition. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 449–454.

Blumenfeld, R. S., & Ranganath, C. (2007). Prefrontal cortex and long-term memory encoding: An integrative review of findings from neuropsychology and neuroimaging. *Neuroscientist*, 13, 280–291.

Brewer, J. B., Zhao, Z., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Making memories: Brain activity that predicts how well visual experience will be remembered. *Science*, 281, 1185–1187.

Clark, D., & Wagner, A. D. (2003). Assembling and encoding word representations: fMRI subsequent memory effects implicate a role for phonological control. *Neuropsychologia*, 41, 304–317.

Cohen, N. J., & Eichenbaum, H. (1994). *Memory, amnesia, and the hippocampal system*. Cambridge, MA: MIT Press.

Courtney, S. M., Petit, L., Maisog, J. M., Ungerleider, L. G., & Haxby, J. V. (1998). An area specialized for spatial working memory in human frontal cortex. *Science*, 279, 1347–1351.

Demb, J. B., Desmond, J. E., Wagner, A. D., Vaidya, C. J., Glover, G. H., & Gabrieli, J. D. (1995). Semantic encoding and retrieval in the left inferior prefrontal cortex: A functional MRI study of task difficulty and process specificity. *The Journal of Neuroscience*, 15, 5870–5878.

Duncan, J. (2001). An adaptive coding model of neural function in prefrontal cortex. *Nature Reviews: Neuroscience*, 2, 820–829.

Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, 392, 598–601.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2001). Categorical representation of visual stimuli in the primate prefrontal cortex. *Science*, 291, 312–316.

Freedman, D. J., Riesenhuber, M., Poggio, T., & Miller, E. K. (2003). A comparison of primate prefrontal and inferior temporal cortices during visual categorization. *The Journal of Neuroscience*, 23, 5235–5246.

Friston, K. J., Frith, C. D., Liddle, P. F., & Frackowiak, R. S. (1993). Functional connectivity: The principal–component analysis of large (PET) data sets. *Journal of Cerebral Blood Flow and Metabolism*, 13, 5–14.

Golby, A. J., Poldrack, R. A., Brewer, J. B., Spencer, D., Desmond, J. E., Aron, A. P., et al. (2001). Material-specific lateralization in the medial temporal lobe and prefrontal cortex during memory encoding. *Brain*, 124, 1841–1854.

Grady, C. L., McIntosh, A. R., Rajah, M. N., & Craik, F. I. (1998). Neural correlates of the episodic encoding of pictures and words. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 2703–2708.

Haxby, J. V., Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*, 293, 2425–2430.

Hayes, S. M., Nadel, L., & Ryan, L. (2007). The effect of scene context on episodic object recognition: Parahippocampal cortex mediates memory encoding and retrieval success. *Hippocampus*, 17, 873–889.

Ishai, A., Schmidt, C. F., & Boesiger, P. (2005). Face perception is mediated by a distributed cortical network. *Brain Research Bulletin*, 67, 87–93.

Jenkins, L. J., & Ranganath, C. (2010). Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. *The Journal of Neuroscience*, 30, 15558–15565.

Johnson, J. D., McDuff, S. G. R., Rugg, M. D., & Norman, K. A. (2009). Recollection, familiarity, and cortical reinstatement: A multivoxel pattern analysis. *Neuron*, 63, 697–708.

Johnson, M. K., Raye, C. L., Mitchell, K. J., Greene, E. J., & Anderson, A. W. (2003). fMRI evidence for an organization of prefrontal cortex by both type of process and type of information. *Cerebral Cortex*, 13, 265–273.

Kanwisher, N., McDermott, J., & Chun, M. M. (1997). The fusiform face area: A module in human extrastriate cortex specialized for face perception. *The Journal of Neuroscience*, 17, 4302–4311.

Kaul, C., Rees, G., & Ishai, A. (2011). The gender of face stimuli is represented in multiple regions in the human brain. *Frontiers in Human Neuroscience*, 4, 238.

Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: A meta-analysis of 74 fMRI studies. *NeuroImage*, 54, 2446–2461.

Kirchhoff, B. A., & Buckner, R. L. (2006). Functional–anatomic correlates of individual differences in memory. *Neuron*, 51, 263–274.

Kirchhoff, B. A., Wagner, A. D., Maril, A., & Stern, C. E. (2000). Prefrontal–temporal circuitry for episodic encoding and subsequent memory. *The Journal of Neuroscience*, 20, 6173–6180.

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience*, 2, 1–28.

Kuhl, B. A., Rissman, J., Chun, M. M., & Wagner, A. D. (2011). Fidelity of neural reactivation reveals competition between memories. *Proceedings of the National Academy of Sciences of the United States of America*, 108, 5903–5908.

Lewis-Peacock, J. A., & Postle, B. R. (2008). Temporary activation of long-term memory supports working memory. *The Journal of Neuroscience*, 28, 8765–8771.

Li, S., Mayhew, S. D., & Kourtzi, Z. (2009). Learning shapes the representation of behavioral choice in the human brain. *Neuron*, 62, 441–452.

Li, S., Ostwald, D., Giese, M., & Kourtzi, Z. (2007). Flexible coding for categorical decisions in the human brain. *The Journal of Neuroscience*, 27, 12321–12330.

McDermott, K. B., Buckner, R. L., Petersen, S. E., Kelley, W. M., & Sanders, A. L. (1999). Set- and code-specific activation in frontal cortex: An fMRI study of encoding and retrieval of faces and words. *Journal of Cognitive Neuroscience*, 11, 631–640.

McDuff, S. G. R., Frankel, H. C., & Norman, K. A. (2009). Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during single-agenda source monitoring. *The Journal of Neuroscience*, 29, 508–516.

Miller, B. T., Vytlacil, J., Fegen, D., Pradhan, S., & D'Esposito, M. (2010). The prefrontal cortex modulates category selectivity in human extrastriate cortex. *Journal of Cognitive Neuroscience*, 23, 1–10.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, 24, 167–202.

Nichols, E. A., Kao, Y. C., Verfaellie, M., & Gabrieli, J. D. (2006). Working memory and long-term memory for faces: Evidence from fMRI and global amnesia for involvement of the medial temporal lobes. *Hippocampus*, 16, 604–616.

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: Multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10, 424–430.

Otten, L. J., & Rugg, M. D. (2001). Task-dependency of the neural correlates of episodic encoding as measured by fMRI. *Cerebral Cortex*, 11, 1150–1160.

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, 6, 93–102.

Polyn, S. M., Natu, V. S., Cohen, J. D., & Norman, K. A. (2005). Category-specific cortical activity precedes retrieval during memory search. *Science*, 310, 1963–1966.

Posner, M. I., & Keele, S. W. (1968). On the genesis of abstract ideas. *Journal of Experimental Psychology*, 77, 353–363.

Prabhakaran, V., Narayanan, K., Zhao, Z., & Gabrieli, J. D. E. (2000). Integration of diverse information in working memory within the frontal lobe. *Nature Neuroscience*, 3, 85–90.

Preston, A. R., Bornstein, A. M., Hutchinson, J. B., Gaare, M. E., Glover, G. H., & Wagner, A. D. (2010). High-resolution fMRI of content-sensitive subsequent memory responses in human medial temporal lobe. *Journal of Cognitive Neuroscience*, 22, 156–173.

Prince, S. E., Dennis, N. A., & Cabeza, R. (2009). Encoding and retrieving faces and places: Distinguishing process- and stimulus-specific differences in brain activity. *Neuropsychologia*, 47, 2282–2289.

Puce, A., Allison, T., Asgari, M., Gore, J. C., & McCarthy, G. (1996). Differential sensitivity of human visual cortex to faces, letterstrings, and textures: A functional magnetic resonance imaging study. *The Journal of Neuroscience*, 16, 5205–5215.

Race, E. A., Shanker, S., & Wagner, A. D. (2009). Neural priming in human frontal cortex: Multiple forms of learning reduce demands on the prefrontal executive system. *Journal of Cognitive Neuroscience*, 21, 1766–1781.

Rainer, G., Asaad, W. F., & Miller, E. K. (1998). Selective representation of relevant information by neurons in the primate prefrontal cortex. *Nature*, 393, 577–579.

Rao, S. C., Rainer, G., & Miller, E. K. (1997). Integration of what and where in the primate prefrontal cortex. *Science*, 276, 821–824.

Rissman, J., Gazzaley, A., & D'Esposito, M. (2004). Measuring functional connectivity during distinct stages of a cognitive task. *Neuroimage*, 23, 752–763.

Rissman, J., Greely, H. T., & Wagner, A. D. (2010). Detecting individual memories through the neural decoding of memory states and past experience. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 9849–9854.

Rissman, J., & Wagner, A. D. Distributed representations in memory: Insights from functional brain imaging. *Annual Review of Psychology*, in press.

Scalaidhe, S. P., Wilson, F. A. W., & Goldman-Rakic, P. S. (1997). Areal segregation of face-processing neurons in prefrontal cortex. *Science*, 278, 1135–1138.

Scoville, W. B., & Milner, B. (1957). Loss of recent memory after bilateral hippocampal lesions. *Journal of Neurology, Neurosurgery, and Psychiatry*, 20, 11–21.

Sergerie, K., Lepage, M., & Armony, J. L. (2005). A face to remember: Emotional expression modulates prefrontal activity during memory formation. *Neuroimage*, 24, 580–585.

Shimamura, A. P. (1995). Memory and the prefrontal cortex. *Annals of the New York Academy of Sciences*, 769, 151–159.

Simons, J. S., & Spiers, H. J. (2003). Prefrontal and medial temporal lobe interactions in long-term memory. *Nature Reviews: Neuroscience*, 4, 637–648.

Spaniol, J., Davidson, P. S., Kim, A. S., Han, H., Moscovitch, M., & Grady, C. L. (2009). Event-related fMRI studies of episodic encoding and retrieval: Meta-analyses using activation likelihood estimation. *Neuropsychologia*, 47, 1765–1779.

Squire, L. R. (1992). Declarative and nondeclarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience*, 4, 232–243.

Summerfield, C., Greene, M., Wager, T., Egner, T., Hirsch, J., & Mangels, J. (2006). Neocortical connectivity during episodic memory formation. *PLoS Biology*, 4, e128.

Tomita, H., Ohbayashi, M., Nakahara, K., Hasegawa, I., & Miyashita, Y. (1999). Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature*, 401, 699–703.

Tsao, D. Y., Schweers, N., Moeller, S., & Freiwald, W. A. (2008). Patches of face-selective cortex in the macaque frontal lobe. *Nature Neuroscience*, 11, 877–879.

Turk-Browne, N. B., Yi, D. J., & Chun, M. M. (2006). Linking implicit and explicit memory: Common encoding factors and shared representations. *Neuron*, 49, 917–927.

Wagner, A. D., Maril, A., & Schacter, D. L. (2000). Interactions between forms of memory: When priming hinders new episodic learning. *Journal of Cognitive Neuroscience*, 12(S2), 52–60.

Wagner, A. D., Poldrack, R. A., Eldridge, L. L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D. (1998). Material-specific lateralization of prefrontal activation during episodic encoding and retrieval. *NeuroReport*, 9, 3711–3717.

Wagner, A. D., Schacter, D. L., Rotte, M., Koutstaal, W., Maril, A., Dale, A. M., et al. (1998). Building memories: Remembering and forgetting of verbal experiences as predicted by brain activity. *Science*, 281, 1188–1191.

Watanabe, T., Hirose, S., Wada, H., Katsura, M., Chikazoe, J., Jimura, K., et al. (2011). Prediction of subsequent recognition performance using brain activity in the medial temporal lobe. *Neuroimage*, 54, 3085–3092.

Weiner, K. S., & Grill-Spector, K. (2010). Sparsely-distributed organization of face and limb activations in human ventral temporal cortex. *Neuroimage*, 52, 1559–1573.

Wheeler, M. A., Stuss, D. T., & Tulving, E. (1995). Frontal lobe damage produces episodic memory impairment. *Journal of the International Neuropsychological Society*, 1, 525–536.

Wig, G. S., Miller, M. B., Kingstone, A., & Kelley, W. M. (2004). Separable routes to human memory formation: Dissociating task and material contributions in the prefrontal cortex. *Journal of Cognitive Neuroscience*, 16, 139–148.

Wilson, C. R., Gaffan, D., Browning, P. G., & Baxter, M. G. (2010). Functional localization within the prefrontal cortex: Missing the forest for the trees? *Trends in Neurosciences*, 33, 533–540.

Wilson, F. A., Scalaidhe, S. P., & Goldman-Rakic, P. S. (1993). Dissociation of object and spatial processing domains in primate prefrontal cortex. *Science*, 260, 1955–1958.

Wood, J. N., & Grafman, J. (2003). Human prefrontal cortex: Processing and representational perspectives. *Nature Reviews: Neuroscience*, 4, 139–147.

Xue, G., Dong, Q., Chen, C., Lu, Z., Mumford, J. A., & Poldrack, R. A. (2010). Greater neural pattern similarity across repetitions is associated with better memory. *Science*, 330, 97–101.

Zanto, T. P., Rubens, M. T., Thangavel, A., & Gazzaley, A. (2011). Causal role of the prefrontal cortex in top-down modulation of visual processing and working memory. *Nature Neuroscience*, 14, 656–661.