frontiers in
HUMAN NEUROSCIENCE

SUPPLEMENTARY MATERIAL
published: 10 August 2010
doi: 10.3389/fnhum.2010.00061

# Listening for recollection: a multi-voxel pattern analysis of recognition memory retrieval strategies

***Joel R. Quamme[1], David J. Weiss[2] and Kenneth A. Norman[3]****

[1] *Department of Psychology, Grand Valley State University, Allendale, MI, USA*
[2] *Department of Computer and Information Science, University of Pennsylvania, Philadelphia, PA, USA*
[3] *Department of Psychology and Princeton Neuroscience Institute, Princeton University, Princeton, NJ, USA*

Edited by: Neal J. Cohen, University of Illinois, USA

Reviewed by: Alison Preston, The University of Texas at Austin, USA
Craig Stark, University of California at Irvine, USA

*Correspondence: Kenneth A. Norman, Department of Psychology, Princeton University, Green Hall, Washington Road, Princeton, NJ 08540, USA. e-mail: knorman@princeton.edu

Received: 30 March 2010; paper pending published: 23 April 2010; accepted: 13 July 2010; published online: 10 August 2010.

Citation: Quamme JR, Weiss DJ and Norman KA (2010) Listening for recollection: a multi-voxel pattern analysis of recognition memory retrieval strategies. *Front. Hum. Neurosci.* **4**:61. doi: 10.3389/fnhum.2010.00061



## ADDITIONAL METHODOLOGICAL DETAILS RELATING TO THE CLASSIFIER ANALYSIS

This section provides additional methodological details regarding: (1) the regularized logistic regression classifier used in the classification analyses; (2) the alignment of data across subjects; (3) the computation of the across-subject reliability of the classifier results; and (4) the computation of the sphere density maps shown in **Figures 4 and 5** of the main paper.

### CLASSIFIER DETAILS

The L2-regularized logistic regression classifier described in the main paper was fit using a modified version of the iteratively reweighted least squares (IRLS) algorithm (Bishop, 2006). The L2 penalty term was incorporated into the standard IRLS algorithm by adding an identity matrix multiplied by a hyperparameter $\lambda$ to the weight matrix during the reweighting step of the algorithm. Specifically, we modified Equation 4.99 of Bishop (2006; see p. 208) by replacing the $\mathbf{\Phi^T R \Phi}$ term with $\mathbf{\Phi^T R \Phi + \lambda I}$. The rest of the algorithm remained identical to standard IRLS.

In this algorithm, greater values of $\lambda$ result in a higher degree of regularization. For our analyses, we used $\lambda = 10$. We did not systematically try to optimize $\lambda$. Midway through data collection, we investigated the effects of varying $\lambda$ on Phase 1 cross-validation accuracy in a subset of subjects, and we found that Phase 1 classification accuracy was not notably sensitive to the precise setting of $\lambda$. We chose $\lambda = 10$ arbitrarily from a range of $\lambda$ values that yielded roughly similar Phase 1 cross-validation accuracy scores.

### ALIGNMENT

In order to analyze the across-subject reliability of our effects, we first had to co-register all 24 subjects to the same template space; we did this using a combination of AFNI (Cox, 1996) and FreeSurfer (Segonne et al., 2004). First, the anatomical images were skull stripped automatically, using AFNI's 3dSkullStrip program or Freesurfer's mri_watershed program when AFNI's 3dSkullStrip program repeatedly failed to converge. Next, these skull-stripped anatomical images were aligned using AFNI's 3dAllineate program to the corresponding subject's mean EPI image, to account for subject motion in between structural and functional scans. This transform used six parameters (rotation and translation) to avoid distorting the structural image. Finally, AFNI's @auto_tlrc program was used to automatically generate a 12-parameter affine transform to the "Colin" N27 Talairach space template provided by AFNI, using the aligned skull-stripped brains as a base.

For each subject, maps of the nine sphere-based metrics (one cross-validation score, $D_{RL}$ for the four temporal windows, and $D_{STUDIED}$ for the four temporal windows) were aligned to the Talairach space template and resampled linearly to 3 mm cubic voxels. No additional smoothing was applied before or after alignment. Next, we generated a cross-subject mask by running AFNI's 3dAutomask on each of the individual subject Talairach space anatomical images, and then including voxels that were included in at least 75% of all of the individual subject automasks. Finally, all maps were loaded into Matlab using the Princeton MVPA Toolbox (Detre et al., 2006) for further analysis. Note that all maps were generated separately for the MVPA and AVG versions of the analysis.

### ASSESSING THE ACROSS-SUBJECT RELIABILITY OF THE CLASSIFIER RESULTS

As described in the main paper, there were two steps to our statistical analysis procedure. In the first step, we used a random effects analysis to find sphere locations where cross-validated classifier accuracy was significantly above chance ($p < 0.05$) in Phase 1 of the experiment; sphere locations that failed to show significantly reliable classification were discarded. The second step of the analysis focused on the relationship between classifier output and recognition behavior during Phase 2, as indexed by the $D_{RL}$ and $|D_{RL}| - |D_{STUDIED}|$ measures described in the main paper. We used a permutation test (Nichols and Holmes, 2002) to non-parametrically assess whether $D_{RL}$ and $|D_{RL}| - |D_{STUDIED}|$ showed reliably positive values across subjects (we used a non-parametric procedure here because of concerns that these novel dependent measures might

violate the normality assumption). **Figure S1** provides a step-by-step description of the non-parametric permutation procedure applied to the present data. The first part of the non-parametric resampling procedure (**Figure S1**, steps 1–3) involved generating samples from the null distributions of $D_{RL}$ and $|D_{RL}| - |D_{STUDIED}|$ for each sphere location. To generate a sample, we shuffled each subject's Phase 2 behavioral data by randomly permuting the subject's responses within the studied-item and related-lure conditions. This approach preserves the studied-item hit rate, the related-lure false alarm rate, and the classifier output time series for each subject, but disrupts the specific trial-by-trial relationship between classifier output and behavior. After permuting the behavioral data, we re-computed the $D_{RL}$ and $D_{STUDIED}$ measures (at each sphere location) for each subject. Finally, we aligned the re-computed $D_{RL}$ and $D_{STUDIED}$ values to Talairach space and averaged them across all 24 subjects. We generated 20,000 null samples in this fashion (**Figure S1**, step 4). We then computed *p*-values separately for the two dependent measures of interest ($D_{RL}$ and $|D_{RL}| - |D_{STUDIED}|$) for each sphere location. To compute *p*-values for a given sphere location, we counted the number of null samples where the sampled value for that sphere location exceeded the observed value, and divided by the total number of samples (**Figure S1**, step 5). Note that this procedure was repeated for both the MVPA and AVG variants of our analysis and for each of the four time windows of interest. We should emphasize that the resampling procedure was only applied to Phase 2 data; the Phase 1 data were not shuffled. As such, all voxels that were excluded from further analysis (due to poor Phase 1 classification) in the actual data were also excluded from further analysis in the resampled data.

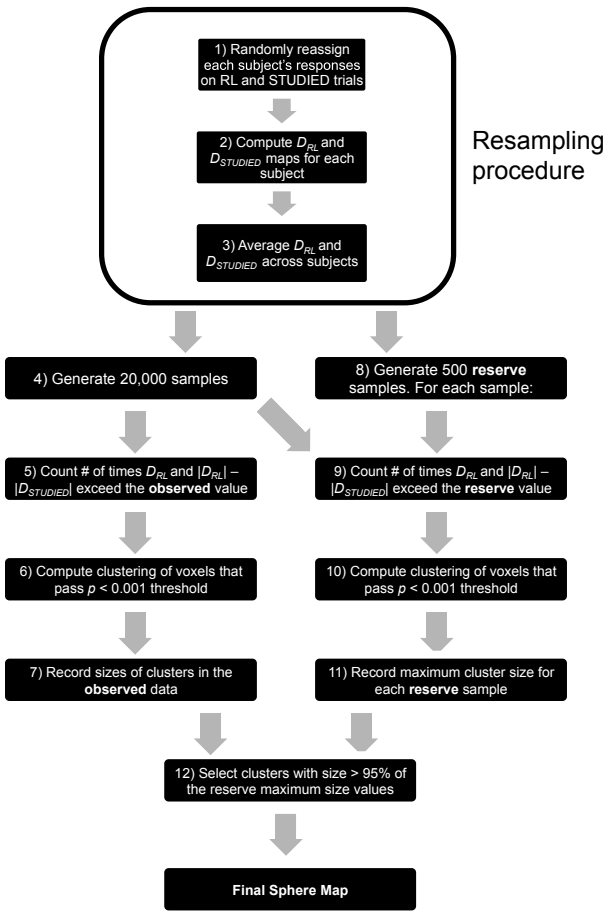


**FIGURE S1 | Flowchart indicating the steps of the non-parametric resampling procedure that was used to identify significantly informative spheres (see text for explanation).**

The above steps result in *p*-values for $D_{RL}$ and $|D_{RL}| - |D_{STUDIED}|$ for each of the 3 mm cubic Talairach voxels. To identify sphere locations where brain activity reliably predicted behavior, we applied a *p*-value threshold of 0.001 to both the $D_{RL}$ measure and the $|D_{RL}| - |D_{STUDIED}|$ measure (note that, at this point in the analysis, we were only considering sphere locations that passed the $p < 0.05$ threshold for the Phase 1 cross-validation measure).

Next, we corrected for multiple comparisons using the non-parametric cluster-based approach described in Nichols and Hayasaka (2003) – this procedure controls the family-wise error rate (FWER) by comparing cluster sizes in the observed data to the distribution of maximum cluster sizes observed under the null hypothesis. The first step in this procedure was to cluster the observed data: Sets of contiguous spheres (i.e., spheres with contiguous center voxels) that passed the *p*-value thresholds listed above were grouped into clusters (**Figure S1**, step 6), and the sizes of these clusters were recorded (**Figure S1**, step 7). Next, we computed the distribution of maximum cluster sizes observed under the null hypothesis. To do this, we first generated 500 *reserve samples* (**Figure S1**, step 8) using the same resampling procedure we used to generate the initial 20,000 samples in step 4. Conceptually, this procedure treats each of the 500 reserve samples as if it were the real data: For each reserve sample, we computed *p*-values for each sphere location using the same set of 20,000 samples used to compute the significance of the real data (**Figure S1**, step 9, which is analogous to step 5 for the real data); we then thresholded the resulting statistical maps at $p < 0.001$ for both $D_{RL}$ and $|D_{RL}| - |D_{STUDIED}|$ and clustered the voxels that survived this thresholding procedure (**Figure S1**, step 10, which is analogous to step 6 for the real data). Finally, for each of the 500 reserve samples, we recorded the size of the largest cluster (**Figure S1**, step 11). The net result of this process is a null distribution (based on the 500 reserve samples) of maximum cluster sizes (Nichols and Hayasaka, 2003). As the final step of the analysis, we computed cluster-wise *p*-values for each cluster in the actual data by computing (for a given cluster) the proportion of times that the cluster's size was exceeded by the 500 samples of the maximum cluster size statistic (**Figure S1**, Step 12). We then reported all clusters that passed a FWER threshold of $p < 0.05$ (i.e., clusters whose size exceeded 95% of the maximum cluster sizes in the null distribution). This cluster-wise FWER correction procedure was repeated for each variant of the analysis (MVPA, AVG) and each time window separately.

As a technical note: Our sampling procedure involves computing brain maps for $D_{RL}$ and $D_{STUDIED}$ for each of 20,000 samples. Storing all of these maps on disk at the same time would be difficult given current constraints on disk space. To get around this problem, we computed samples in batches of 200 samples at a time. After running each batch, we updated (for each sphere location

and dependent measure) a running tally of the number of samples whose value exceeded the actually observed value; after updating these counts, we discarded the sample values for that batch. Using this method, we only needed disk space for the running tally and for the current batch of 200 samples.

### COMPUTING SPHERE DENSITY MAPS

To represent the final results of the across-subject significance tests, we computed *sphere density maps*. These maps show the spatial extent and density of the significant sphere clusters (where "density" indicates the degree of overlap between spheres belonging to these clusters). One challenge in making these maps is that overlap between spheres in Talairach space does not exactly correspond to overlap between spheres in native space. To address this issue, we took the spheres belonging to significant clusters in Talairach space, and we used AFNI's 3dFractionize program to warp all of these significant spheres back to the native space of each subject. Then, for each voxel in each subject, we summed up the number of significant spheres (in that subject's native space) that encompassed that voxel; we will refer to these maps as *native-space sphere density maps*. We then warped each subject's native-space sphere density map back into Talairach space, and we averaged these maps across subjects. The net result of this process is a single Talairach-space map showing the average sphere density across subjects (see **Figures 4 and 5** of the main paper).

## PHASE 1-CLASSIFIER CROSS-VALIDATION RESULTS

As described in the Section "Materials and Methods" of the main text, we ran a cross-validation test on Phase 1 data to assess how well each sphere (in both the MVPA and AVG analyses) discriminated between scans from recollection test blocks and familiarity test blocks. **Figure S2** shows the results of this analysis. Points shown in color in the figure correspond to sphere locations where cross-validated classifier accuracy was significantly above chance at $p < 0.05$. Spheres that failed to pass this (relatively liberal) cross-validation threshold were excluded from the Phase 2 classifier analysis. Intuitively, any region that is involved in subjects' use of recollection should show some degree of differential responding when subjects are instructed to use recollection vs. familiarity in Phase 1 – this is why we excluded areas with poor Phase 1 cross-validation performance from the Phase 2 analysis.

However, it is also important to realize that "use of recollection" was not the only difference between recollection and familiarity blocks; for example, subjects showed more liberal responding and faster RTs in familiarity blocks. Regions that are sensitive to these other factors (but not subjects' use of recollection *per se*) might show above-chance classification of Phase 1 data but poor generalization to Phase 2 data. In keeping with this idea, there were several regions (e.g., medial frontal gyrus, precentral gyrus) that showed highly significant discrimination between recollection vs. familiarity blocks in Phase 1 ($p < 10^{-6}$, uncorrected) but did not significantly predict behavior in the Phase 2 analysis.

## SINGLE-VOXEL AND AVG-RADIUS-1 RESULTS

In the main paper, we showed that MVPA was able to detect a significantly informative region during the pre-trial (Window 1) time period, but the AVG analysis did not detect any significantly informative regions during this time window. There are two possible explanations for this result: One possibility is that pre-trial differences in subjects' cognitive state affected the pattern of activity but not the overall level of activity. Another possibility is that pre-trial differences in cognitive state affected the overall

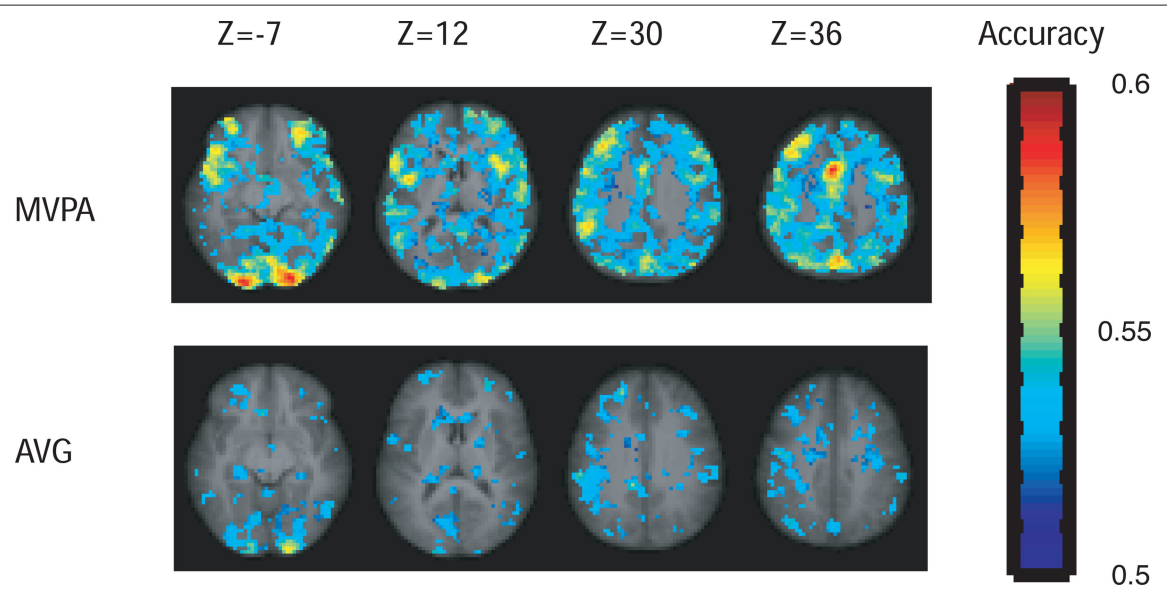


**FIGURE S2 | Brain maps showing the accuracy of local-pattern-mapping (searchlight) classifiers at discriminating between recollection vs. familiarity test blocks in Phase 1, for both the MVPA and AVG analyses.** The colors on the map denote cross-validated classifier accuracy (measured in terms of percent correct) when the classifier was applied to a sphere of voxels centered at that point. The figure only shows points where classifier accuracy was significantly above chance at $p < 0.05$.

activity of localized regions of interest, but the spatial scale of these ROIs was smaller than our two-voxel-radius spheres – this kind of spatial scale mismatch would be expected to affect the AVG analysis more than the MVPA analysis (since the MVPA analysis can learn to selectively weight a subset of voxels within the sphere, but the AVG analysis can not do this). To evaluate this idea, we re-ran the AVG analysis at a smaller spatial scale (using one-voxel-radius spheres encompassing seven voxels in total, and single, unsmoothed voxels). The results of these analyses are shown in **Table S1**. The AVG analysis with one-voxel-radius spheres only found one significantly informative region across all time windows: a left lingual gyrus area in Window 2 that was also found by the two-voxel-radius AVG analysis. The single-voxel analysis also only found one significantly informative region across all time windows: a very small right supramarginal region in Window 2. Importantly, neither analysis found any significantly informative regions during Window 1. This suggests that MVPA's ability to detect informative pre-trial activity during Window 1 (and AVG's failure to do so) is attributable to MVPA's superior ability to detect complex spatial patterns, and is not a simple artifact of the spatial scale that we chose for the AVG analysis in the main paper.

## GLM ANALYSIS

Both variants of our primary analysis (MVPA and AVG) are quite different from the standard sorts of recollection vs. familiarity comparisons that have been performed in existing studies (see, e.g., Vilberg and Rugg, 2008). To facilitate comparison with other fMRI studies of recollection vs. familiarity, we also ran standard general linear model (GLM) analyses. We describe our GLM methods and results in the following three subsections.

### GLM ANALYSIS METHODS

Preprocessing for the GLM analyses was performed in AFNI (Cox, 1996). The first three scans were removed from the beginning of each run and slice times were aligned to 1 s after the onset of each 2-s scan (i.e., the middle of the scan). AFNI's 3dDespike was used to remove signal spikes in the time course of each voxel. All functional volumes were co-registered to the first scan of the experiment, and were corrected for motion artifacts based on co-registration parameters. The co-registered volumes were smoothed using an 8 mm Gaussian kernel. Data from each run were then rescaled to a mean of 100 across the volumes.

Separate GLM analyses were performed on Phase 1 and Phase 2 data. The Phase 1 analysis looked for regions showing differential average activity during recollection vs. familiarity testing blocks. The Phase 2 analysis looked for regions showing differential average activity for hits, unrelated-lure correct rejections, related-lure correct rejections, and related-lure false alarms. For the Phase 1 GLM analysis, 12 regressors (six motion parameters and six regressors of interest) were entered into the GLM for each subject. The six regressors of interest corresponded to recollection test blocks, familiarity test blocks, the cue periods immediately preceding the recollection and familiarity test blocks, the study phase, and cues to fixate. Each of the regressors was convolved with a model hemodynamic response function (HRF). A group analysis was performed on the resulting coefficient values with subjects as a random effect. For the Phase 2 GLM analysis, 12 regressors were included. Six were motion parameters and were treated as regressors of no interest. The remaining six corresponded to the hits, misses, unrelated-lure correct rejections, unrelated-lure false alarms, related-lure correct rejections, and related-lure false alarms. Each of these regressors was modeled with an HRF, and the group analysis was performed with subjects as a random effect. For the Phase 2 analysis, we were particularly interested in three sets of contrasts: hits vs. unrelated-lure correct rejections; hits vs. related-lure false alarms; and related-lure correct rejections vs. related-lure false alarms. In addition to those three contrasts, we also examined the relationship between pre-trial activity and behavior during Phase 2. For these analyses, we focused on the two scans immediately prior to the onset of each test item (i.e., the same time period that we refer to as "Window 1" in our MVPA and AVG analyses). Separate regressors were created for the scans preceding related-lure correct rejections, related-lure false alarms, studied-item hits, and studied-item misses; note that these regressors were not convolved with an HRF function. The first contrast examined whether there was differential pre-trial activity associated with related-lure correct rejections vs. false alarms. The second contrast examined whether there was differential pre-trial activity associated with studied-item hits vs. misses.

For the GLM analyses described here, we used a threshold of $p < 0.001$ for 20 or more contiguous voxels. The cluster size threshold was set using AFNI's AlphaSim program, which uses Monte Carlo simulations of random (i.e., noise-only) data to determine the probability that a cluster of $n$ contiguous significant voxels would appear by chance. The threshold values used here correspond to a FWER of $p < 0.05$.

**Table S1 | Sphere clusters from single-voxel and AVG (radius = 1) analyses passing cross-validation, $D_{RL}$, and $|D_{RL}| - |D_{STUDIED}|$ thresholds at family-wise error rate < 0.05.**

| | | | | | | Center cluster | | |
|---|---|---|---|---|---|---|---|---|
| **Rad.** | **Win.** | **Region** | **BA** | **Envelope** | **Size** | ***x*** | ***y*** | ***z*** |
| 0 | 2 | R. supramarginal gyrus | 40 | 3 | 3 | 46 | –40 | 36 |
| 1 | 2 | L. lingual gyrus | 18 | 193 | 24 | –18 | –69 | –9 |

*Clusters were defined as sets of significant spheres with contiguous center voxels. Rad. = radius of sphere (0 = single-voxel, 1 = 7-voxel sphere); envelope = total volume (in voxels) covered by significantly informative spheres; Size = the contiguous volume covered by center voxels only; Center cluster = location of center voxels; Win = time window; BA = Brodmann area; x, y, and z coordinates refer to locations in Talairach space on the left–right, posterior–anterior, and inferior–superior axes respectively, with negative numbers indicating left, posterior and inferior (LPI).*

### PHASE 1 GLM RESULTS

A GLM analysis was performed to examine the relative BOLD response to recollection vs. familiarity blocks during Phase 1. Regions showing a significantly different response to recollection vs. familiarity blocks are shown in the top panel of **Figure S3** (see **Table S2** for a list of significant regions). There was a greater response to recollection blocks than familiarity blocks bilaterally in the occipital lobe, centered on the left lingual gyrus and right middle occipital gyrus. These effects may reflect more extensive visual processing during retrieval when subjects were trying to recollect the mental image formed at study. A larger number of regions showed the opposite pattern, responding more strongly during familiarity blocks than recollection blocks. Two distinct bilateral pairs of regions around the temporal–parietal junction showed this effect. One region of activity was more superior and posterior, over the intraparietal sulcus on both sides with peak activity in the inferior parietal lobule/BA40 on the left and superior parietal lobule/BA 7 on the on the right; the other region was more inferior and anterior, with peak activity on the boundary between the inferior parietal lobule (BA 40) and postcentral gyrus (BA 2). These regions are broadly consistent with familiarity-related dorsal parietal areas that have been identified in previous studies (for summaries, see Cabeza, 2008; Cabeza et al., 2008; Vilberg and Rugg, 2008). However, comparisons to previous studies are difficult to make for this analysis because the regressors used in this contrast (corresponding to instructions to use recollection vs. familiarity, keeping other aspects of the task constant) have not been used in previous studies, to the best of our knowledge. Additional regions showing greater response to familiarity blocks than recollection blocks were found bilaterally in the insula/superior temporal gyrus and middle occipital gyrus, and medially along the cingulate gyrus.

### PHASE 2 GLM RESULTS

Three primary contrasts were performed on the Phase 2 data. First, we compared the BOLD response associated with hits to that of unrelated-lure correct rejections. This comparison reflects the standard old–new effect commonly reported in the literature; it is sensitive to differences in both recollection and familiarity, as well as differences in the perception of oldness and yes vs. no responses (Wagner et al., 2005). Regions showing differential activity for hits vs. unrelated-lure correct rejections are shown in the second panel of **Figure S3** (see **Table S3** for a list of significant regions). Consistent with numerous previous studies (see Cabeza, 2008; Vilberg and Rugg, 2008), we found greater activity for hits in and around the region of the intraparietal sulcus bilaterally, with peak activity in the inferior

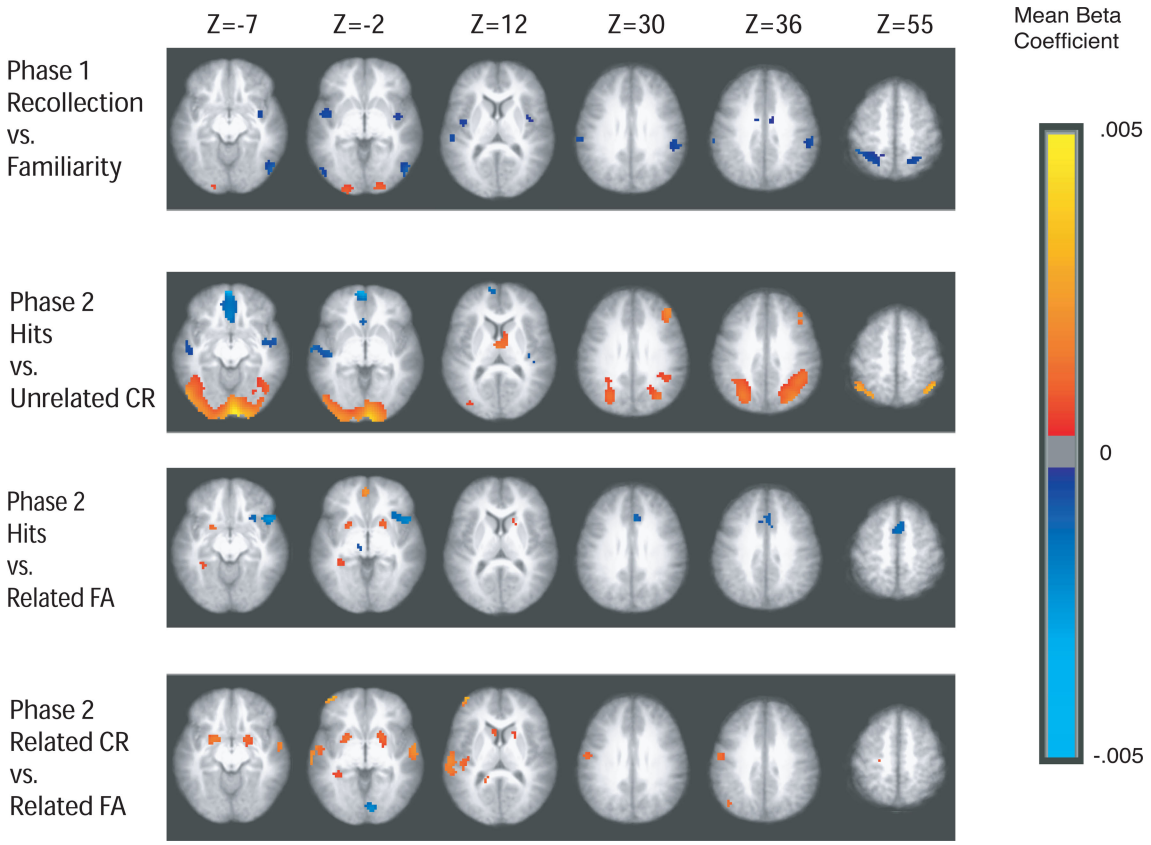


**FIGURE S3 | Areas of differential BOLD response to the experimental conditions, as indicated by our GLM analyses.** Blue regions indicate a negative difference for the given contrast, and orange-to-yellow regions indicate positive differences. The top panel contrasts Phase 1 recollection blocks with Phase 1 familiarity blocks. The second panel shows the standard old–new effect in the Phase 2 plurals task, contrasting hits with correct rejections of unrelated lures. The third panel contrasts hits with related-lure false alarms. The fourth panel contrasts related-lure correct rejections with related-lure false alarms.

**Table S2 | Brain regions showing significantly different BOLD signal (voxel-wise $p < 0.001$, cluster size ≥ 20, resulting in family-wise error rate < 0.05) for recollection blocks vs. familiarity blocks in Phase 1.**

| Region | Left/right | BA | No. of voxels | Talairach coordinates | | | *t*-Value |
|---|---|---|---|---|---|---|---|
| | | | | *x* | *y* | *z* | |
| **RECOLLECTION > FAMILIARITY** | | | | | | | |
| Middle occipital gyrus | R | 18 | 36 | 26 | –92 | 3 | 4.07 |
| Lingual gyrus | L | 17 | 35 | –19 | –95 | –4 | 5.00 |
| **FAMILIARITY > RECOLLECTION** | | | | | | | |
| Inferior parietal lobule | L | 40 | 128 | –44 | –44 | 57 | –4.06 |
| | L | 40/2 | 70 | –62 | –29 | 36 | –4.02 |
| | R | 40/2 | 118 | 59 | –29 | 42 | –4.19 |
| Insula/Sup. temporal gyrus | L | 13/22 | 85 | –47 | –2 | 3 | –6.00 |
| | R | 13 | 72 | 41 | 2 | –10 | –4.19 |
| Cingulate gyrus | L | 24 | 80 | –2 | –8 | 42 | –4.12 |
| Middle occipital gyrus | R | 19 | 62 | 53 | –71 | –4 | –5.53 |
| | L | 19 | 24 | –53 | –74 | 3 | –5.08 |
| Superior parietal lobule | R | 7 | 49 | 26 | –62 | 57 | –4.30 |

*Locations and t-values of maximal activity within clusters of significant voxels are listed; BA = Brodmann area; x, y, and z coordinates refer to locations in Talairach space on the left–right, posterior–anterior, and inferior–superior axes respectively, with negative numbers indicating left, posterior and inferior (LPI).*

**Table S3 | Brain regions showing significantly different BOLD signal (voxel-wise $p < 0.001$, cluster size ≥ 20, resulting in family-wise error rate < 0.05) for hits vs. unrelated-lure correct rejections in Phase 2.**

| Region | Left/right | BA | No. of voxels | Talairach coordinates | | | *t*-Value |
|---|---|---|---|---|---|---|---|
| | | | | *x* | *y* | *z* | |
| **HITS > CORRECT REJECTIONS** | | | | | | | |
| Inferior occipital gyrus | L/R | 18 | 2441 | –38 | –77 | –22 | 7.87 |
| Inferior parietal lobule | R | 40 | 613 | 44 | –56 | 51 | 5.23 |
| | L | 40 | 391 | –44 | –59 | 54 | 4.40 |
| Thalamus | R | | 165 | 2 | –11 | 12 | 4.49 |
| Middle frontal gyrus | R | 46/9 | 103 | 47 | 35 | 27 | 4.98 |
| **CORRECT REJECTIONS > HITS** | | | | | | | |
| Medial frontal gyrus | L | 10 | 290 | –2 | 56 | –7 | –4.42 |
| Superior temporal gyrus | R | 38 | 117 | 35 | 11 | –22 | –4.12 |
| | L | 22 | 81 | –62 | –8 | 3 | –4.31 |
| Insula | R | 13 | 31 | 41 | –20 | 15 | –4.00 |

*Locations and t-values of maximal activity within clusters of significant voxels are listed; BA = Brodmann area; x, y, and z coordinates refer to locations in Talairach space on the left–right, posterior–anterior, and inferior–superior axes respectively, with negative numbers indicating left, posterior and inferior (LPI).*

parietal lobule (BA 40) on both sides, extending also into the superior parietal lobule on both sides. Also, we found greater activity for hits than unrelated-lure correct rejections in dorsolateral prefrontal cortex, around middle frontal regions BA 46 and 9. Additionally, we found a larger posterior area that showed greater activity for hits than unrelated-lure correct rejections, encompassing the inferior occipital, middle occipital, and lingual gyri, and extending into the cerebellum; this posterior activation suggests greater visual processing of hits relative to correct rejections. The right thalamus also showed greater response to hits than unrelated-lure correct rejections. Conversely, parts of medial frontal gyrus, right insula, and the anterior portion of the superior temporal gyrus bilaterally showed a greater response to unrelated-lure correct rejections than to hits.

We also contrasted the BOLD response to hits vs. related-lure false alarms. This comparison should isolate recollection-based activity, since hits can be driven by either recollection or familiarity, but related-lure false alarms only occur when there is familiarity in the absence of plurality recollection (if subjects recollect the corresponding studied item, they will reject the related lure; Curran, 2000). Note that response-related activity is held constant across this comparison (since subjects responded "yes" for both hits and related-lure false alarms). Regions showing differential activity for hits vs. related-lure false alarms are shown in the third panel of **Figure S3** (see **Table S4** for a list of significant regions). The posterior left parahippocampus showed greater responding to hits than related-lure false alarms; this fits with results from Kirwan and Stark (2007), who found that

activity in distinct parts of the hippocampus and parahippocampus distinguished studied face and object stimuli from highly similar (viewpoint-shifted) lures. Other areas showing greater activity for hits compared to related-lure false alarms in our study were the left anterior cingulate gyrus, left thalamus, and bilateral lentiform nucleus. An area at the inferior frontal/superior temporal junction over Brodmann areas 47 and 38 showed greater response to false alarms than hits, as did a dorsal portion of the medial frontal gyrus.

Next, we looked for regions showing a differential response to related-lure correct rejections vs. related-lure false alarms. This comparison, like the previous comparison, should isolate recollection-related activity: Insofar as related-lure correct rejections are driven by recollection of plurality information, and false alarms result from failure to recollect plurality information, the overall level of recollection-related neural activity should be larger for related-lure correct rejections than false alarms. This comparison does not control for response-related activity (since subjects responded "no" on correct-rejection trials and "yes" on false-alarm trials) but it does control for the prior history of the item (i.e., in both cases the item was studied previously in switched-plural form). Regions showing differential activity for related-lure correct rejections vs. false alarms are shown in the fourth panel of **Figure S3** (see **Table S5** for a list of significant regions). We found a posterior left parahippocampal region that showed greater activity for correct rejections than related-lure false alarms. The left angular gyrus/BA39 also showed a greater response to related-lure correct rejections than false alarms; this area is slightly inferior and posterior to the peak of the old/new effect that we observed previously (in the comparison of hits vs. unrelated-lure correct rejections). Greater activity for related-lure correct rejections vs. false alarms was also found in the left middle frontal gyrus, precuneus/BA 7, left and right superior temporal gyrus, bilateral lentiform and left caudate nuclei, left precentral gyrus, right cingulate, and left posterior cingulate. The only region showing significantly greater activity for related-lure false alarms compared to related-lure correct rejections was an area of the right lingual gyrus. The left parahippocampal effect fits with prior results from Kirwan and Stark (2007), who (using pictorial materials) found a greater response to related-lure correct rejections than related-lure false alarms in a similar area. The angular gyrus/BA39 result is consistent with prior studies showing that more ventral portions of the lateral parietal region are involved in recollection-related responding (Cabeza, 2008; Vilberg and Rugg, 2008).

In addition to the three contrasts described above, we contrasted pre-trial activity (i.e., activity in the two scans preceding the appearance of the test item) for related-lure correct rejections vs. false alarms, and we also contrasted pre-trial activity for studied-item hits vs. misses. For both of these contrasts, there were no clusters that met our significance threshold.

### COMPARISON OF GLM WITH MVPA AND AVG ANALYSES

Overall, the GLM analyses tended to identify different areas of importance than the MVPA and AVG analyses. This can be seen by comparing **Figure S3** (GLM) to **Figure 4** (MVPA) and **Figure 5** (AVG) from the main paper, where four transverse slices are common across all three figures (at *z*-axis coordinates −7, 12, 30, and 36). There were also some interesting points of convergence between the results of the GLM and the results of the other analyses. Most notably, both the AVG analysis and the Phase 2 GLM analysis identified overlapping regions of the left posterior parahippocampus (for the most direct comparison, see Window 2 of **Figure 5** and the hits vs. related FA panel of **Figure S3** at $z = -7$). In the AVG analysis, we found that the average activity of spheres in this region during Window 2 (i.e., stimulus onset) was informative regarding how subjects would respond to related lures, but it was not informative regarding how they would respond to studied items. In the GLM, we observed greater activity for both hits and related-lure correct rejections (both of which should be associated with relatively high levels of recollection) relative to related-lure false alarms (which should be associated with relatively low levels of recollection). This is consistent with prior work (e.g., Kirwan and Stark, 2007) showing that posterior medial temporal activity differentiates studied items from similar lures (see the main paper for discussion of why this left parahippocampal region was not significant in the MVPA analysis).

**Table S4 | Brain regions showing significantly different BOLD signal (voxel-wise $p < 0.001$, cluster size ≥ 20, resulting in family-wise error rate < 0.05) for hits vs. related-lure false alarms in Phase 2.**

| Region | Left/right | BA | No. of voxels | Talairach coordinates | | | *t*-Value |
|---|---|---|---|---|---|---|---|
| | | | | *x* | *y* | *z* | |
| **HITS > RELATED-LURE FALSE ALARMS** | | | | | | | |
| Lentiform nucleus | R | | 50 | 23 | 8 | –1 | 5.43 |
| | L | | 21 | –17 | 5 | –7 | 4.05 |
| Thalamus | L | | 41 | –11 | –20 | 6 | 5.20 |
| Parahippocampal gyrus | L | 19 | 26 | –32 | –38 | –4 | 5.47 |
| Anterior cingulate gyrus | L | 32/10 | 20 | –2 | 50 | 1 | 4.54 |
| **RELATED-LURE FALSE ALARMS > HITS** | | | | | | | |
| Medial frontal gyrus | R | 6 | 191 | 8 | 5 | 63 | –3.95 |
| Superior temporal gyrus | R | 38 | 137 | 50 | 17 | –7 | –5.41 |

*Locations and t-values of maximal activity within clusters of significant voxels are listed; BA = Brodmann area; x, y, and z coordinates refer to locations in Talairach space on the left–right, posterior–anterior, and inferior–superior axes respectively, with negative numbers indicating left, posterior and inferior (LPI).*

**Table S5 | Brain regions showing significantly different BOLD signal (voxel-wise $p < 0.001$, cluster size ≥ 20, resulting in family-wise error rate < 0.05) for related-lure correct rejections vs. false alarms in Phase 2.**

| Region | Left/right | BA | No. of voxels | Talairach coordinates | | | *t*-Value |
|---|---|---|---|---|---|---|---|
| | | | | *x* | *y* | *z* | |
| **RELATED-LURE CORRECT REJECTIONS > FALSE ALARMS** | | | | | | | |
| Superior temporal gyrus | L | 22 | 246 | –62 | –17 | 6 | 6.06 |
| | R | 22 | 73 | 62 | –11 | 3 | 4.07 |
| Lentiform nucleus | R | | 109 | 20 | 8 | –4 | 5.88 |
| | L | | 72 | –17 | 2 | –10 | 4.20 |
| Middle frontal gyrus | L | 10 | 62 | –38 | 59 | 6 | 4.77 |
| | L | 8 | 40 | –32 | 17 | 45 | 4.34 |
| Caudate | L | | 51 | –2 | 11 | 9 | 4.10 |
| Precentral gyrus | L | 6 | 48 | –56 | –8 | 36 | 4.86 |
| | L | 4 | 31 | –14 | –29 | 69 | 3.87 |
| Cingulate gyrus | R | 31 | 42 | 20 | –44 | 21 | 4.98 |
| Precuneus | R | 7 | 42 | 2 | –35 | 48 | 4.19 |
| Posterior cingulate | L | 29 | 31 | –17 | –44 | –12 | 3.90 |
| Angular gyrus | L | 39 | 25 | –50 | –68 | 42 | 4.31 |
| Parahippocampal gyrus | L | 19 | 21 | –32 | –41 | –4 | 4.49 |
| **RELATED-LURE FALSE ALARMS > CORRECT REJECTIONS** | | | | | | | |
| Lingual gyrus | R | 18 | 32 | 8 | –80 | –1 | 4.59 |

*Locations and t-values of maximal activity within clusters of significant voxels are listed; BA = Brodmann area; x, y, and z coordinates refer to locations in Talairach space on the left–right, posterior–anterior, and inferior–superior axes respectively, with negative numbers indicating left, posterior and inferior (LPI).*

Both the MVPA and GLM analyses revealed task-relevant regions in parietal cortex. However, the specific parietal regions that were identified by MVPA differed from the regions identified with the GLM. The right-lateralized cluster of spheres that was informative during the Window 1 MVPA analysis is located posterior and inferior to the right-lateralized parietal regions that showed differential activity for recollection and familiarity blocks in the Phase 1 GLM analysis. This Window 1 MVPA cluster does overlap somewhat with the large right-parietal old–new effect (i.e., greater activity for hits than unrelated correct rejections) in the GLM analysis, but location of peak activity for the old–new effect is outside of the Window 1 MVPA cluster. The more inferior left-lateralized cluster of spheres in MVPA (centered on the middle-temporal gyrus) that was informative during Window 3 likewise does not overlap with the left-lateralized GLM effects around the temporal–parietal junction. The regions of the left parietal lobe identified by the GLM as differentiating both hits and unrelated-lure correct rejections (see second panel from top in **Figure S3**) and related-lure correct rejections from related-lure false alarms (see the bottom panel of **Figure S3**) are completely superior to the Window 3 MVPA sphere cluster in the left temporal–parietal area.

Why were there so many differences between the GLM and AVG results, given that the GLM and AVG analyses are both (nominally) univariate? It is important to keep in mind that the procedures used for the GLM analyses were very different from the procedures used for the AVG and MVPA analyses. The GLM analyses involved simple binary contrasts between trial types, and each of these contrasts was performed on data from either Phase 1 or Phase 2 (never both). By contrast, the spheres identified as informative for AVG and MVPA analyses were those that generalized across Phases 1 and 2 to predict a specific pattern of yes and no responses for both hits and related lures (preprocessing was also different for the GLM analyses vs. the MVPA and AVG analyses). The differences between the results of the GLM, MVPA, and AVG analyses highlight the importance of using multiple approaches when analyzing fMRI data.

## REFERENCES


Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. New York, NY: Springer.

Cabeza, R. (2008). Role of parietal regions in episodic memory retrieval: the dual attentional processes hypothesis. *Neuropsychologia* 7, 1813–1827.

Cabeza, R., Ciaramelli, E., Olson, I. R., and Moscovitch, M. (2008). The parietal cortex and episodic memory: an attentional account. *Nat. Rev. Neurosci.* 9, 613–625.

Cox, R. W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput. Biomed. Res.* 29, 162–173.

Curran, T. (2000). Brain potentials of recollection and familiarity. *Mem. Cogn.* 28, 923–938.

Detre, G. J., Polyn, S. M., Moore, C. D., Natu, V. S, Singer, B. D., Cohen, J. D., Haxby, J. V., and Norman, K. A. (2006). "The multi-voxel pattern analysis (MVPA) toolbox," in *Annual Meeting of the Organization of Human Brain Mapping*, June 11–15, 2006, Florence, Italy.

Kirwan, C. B., and Stark, C. E. L. (2007). Overcoming interference: an fMRI investigation of pattern separation in the medial temporal lobe. *Learn. Mem.* 14, 625–633.

Nichols, T., and Hayasaka, S. (2003). Controlling the familywise error rate in functional neuroimaging: a comparative review. *Stat. Methods Med. Res.* 12, 419–446.

Nichols, T., and Holmes, A. P. (2002). Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Hum. Brain Mapp.* 15, 1–25.

Segonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., and Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *Neuroimage* 22, 1060–1067.

Vilberg, K. L., and Rugg, M. D. (2008). Memory retrieval and the parietal cortex: a review of evidence from a dual-process perspective. *Neuropsychologia* 46, 1787–1799.

Wagner, A. D., Shannon, B. J., Kahn, I., and Buckner, R. L. (2005). Parietal lobe contributions to episodic memory retrieval. *Trends Cogn. Sci.* 9, 445–453.