Correcting False Memories: The Effect of Mnemonic Generalization on Original Memory Traces

Nathan M. Muncy[1] and C. Brock Kirwan[1,2*]

[1] Department of Psychology, Brigham Young University, Provo, Utah, USA and [2] Neuroscience Center, Brigham Young University, Provo, Utah, USA

* Corresponding author

E-mail: kirwan@byu.edu (CBK)

Abstract

False memories are a common occurrence but the impact of misremembering on the original memory trace is ill-described. While the original memory may be rewritten, it is also possible for a second false memory to exist concurrently with the original, and if a false memory exists concurrently then recovery of the original information should be possible. This study investigates first, whether false recognition overwrites the original memory representation using a mnemonic discrimination task, and second, which neural processes are involved in recovering the original memory following a false memory. Thirty-five healthy, young adults performed multiple recognition memory tests, where the design of the experiment induced participants to make memory errors in the first recognition memory test and then allowed us to determine whether the memory error would be corrected in the second test session. FMRI signal associated with the encoding and retrieval processes during the experiment were investigated in order to determine the important regions for false memory correction. We found that false memories do not overwrite the original trace in all instances, as recovery of the original information was possible. Critically, we determined that recovery of the original information was dependent on higher-order processes during the formation of the false memory during the first test, and not on processing at the time of encoding or the second test episode.

KEYWORDS

episodic memory, discrimination, generalization, functional magnetic resonance imaging

1 INTRODUCTION

The field of study on mnemonic discrimination in episodic memory has largely focused on a single aspect—the capacity of the participant to correctly identify lure stimuli. In animal models, where pattern separation processes can be targeted, previous work has demonstrated that the successful identification of lures is dependent on the orthogonalization processes of the dentate gyrus in the hippocampus as well as an intact hippocampal trisynaptic pathway (Gilbert et al., 1998, 2001; Gilbert & Kesner, 2006; Gold & Kesner, 2005; Hunsaker et al., 2008; Kesner, 2007a, 2007b; Leutgeb et al., 2007). Human studies have implicated the same processes for successful mnemonic discrimination performance: participants with healthy hippocampi perform well on tests of mnemonic discrimination while participants with hippocampal atrophy perform poorly and participants with hippocampal lesions perform worst of all (Bakker et al., 2008; Kirwan et al., 2012; Mueller et al., 2011; Small et al., 2011). Additionally, hippocampal CA3/DG volumes in humans are known to positively correlate with performance on tests of mnemonic discrimination (Doxey & Kirwan, 2015; Grady & Ryan, 2017; Leal & Yassa, 2014; Marshall et al., 2016; Reagh et al., 2014; Roberts et al., 2014; Sheppard et al., 2016). That said, while mnemonic discrimination processes appear to be pattern separation dependent, higher order processes such as semantic labeling, visual spatial processing, attention, and decisiveness (Goebel & Vincze, 2007; Hunsaker & Kesner, 2013; Lacy et al., 2011; Leal & Yassa, 2014; Pidgeon & Morcom, 2016; Reagh et al., 2016; Reagh & Yassa, 2014; Rolls & Kesner, 2016; Steemers et al., 2016; Wais et al., 2017) integrate with hippocampal pattern separation output to coordinate the overall behavior. But while the roles of the medial temporal lobe structures in mnemonic discrimination are well understood, higher-order processes are relatively understudied in this context.

In addition to these gaps in the field of mnemonic discrimination, the failure in detecting lures is also relatively unaddressed. The detection of lure stimuli in a recognition memory test is the result of both encoding and retrieval processes: the lure's representation must be successfully orthogonalized from the target's, and the target must be successfully retrieved. At the point of successful lure identification, then, two distinct representations exist (Target, Lure) between which the participant is able to discriminate. This process is sometimes known as "recall to reject" (J. Kim & Yassa, 2013; Kirwan & Stark, 2007; Lacy et al., 2011; van den Honert et al., 2016). A failure in lure detection, then, could have multiple antecedents: a failure to encode the lure's salient details, a failure to retrieve the salient target information, and/or lure interference may lead to mnemonic generalization rather than discrimination. Such failures may result from underperforming lower-order hippocampal processes, but may also be the result of deficient attention, semantic labelling, and/or decision making. Further, and perhaps more interestingly, the consequence of incorrectly identifying the Lure as Target (a False Alarm, FA) is unknown: as studies of mnemonic discrimination typically use a single study/test presentation, the effect of the FA response on the original memory trace has not been established. It is possible that either the lure representation could overwrite the target memory trace, or that a second representation could exist concurrently with the original. If the latter is true and multiple memory traces result from a FA response, then the two memory traces (one of the Target, and one of the Lure) may be dissociable in subsequent testing. And while certain studies, particularly in the false-memory literature, have specifically targeted the FA responses (rather than record it as a task failure in the mnemonic discrimination task) and the underlying mechanism at both the encoding and test phase of the trial (Cabeza, 2013; Hanczakowski & Mazzoni, 2011; H. Kim, 2011; Kubota et al., 2006; Moritz et al., 2006; Okado & Stark, 2005), the differential processing that could

potentially be involved with correcting versus perpetuating FA behavior has not been investigated.

In order to address these gaps in the literature, we developed a paradigm to investigate both the impact of a false alarm on the original memory trace as well as the relative contributions of hippocampal and cortical processes in promoting and recovering from a failed lure detection. We hypothesized that (a) FAs would result in multiple memory traces that exist concurrently, and as such, we predicted that participants would be able to distinguish target from lure when both were presented simultaneously. Further, we hypothesized that subsequent target-trace recovery would be predicted by (b) increased BOLD signal in CA3/DG during the orienting task and false alarm, and (c) increased BOLD signal in a priori regions associated with encoding and retrieval. Finally, we performed exploratory whole-brain analyses with the prediction that (d) we would observe regions associated with higher-order cognitive processes, such as attention and content processing, that were involved in subsequent target-trace recovery. In short, we predicted that not all FA responses are equal, but that some would occur during better processing of the Lure stimulus which would then precede a retrieval of the Target memory trace during a subsequent test.

2 METHODS

2.1 Participants

Thirty-five participants (15 female, mean age = 23.1, SD = 2.2) were recruited from the University and local community and were compensated for their participation with either payment or a 3D-printed, quarter-scale model of their brain. The university's Institutional Review Board approved the research, and all participants gave written informed consent prior to participation. Inclusion criteria consisted of speaking English as a native language, being right-

hand dominant, and having normal or corrected-to-normal vision. Exclusion criteria consisted of a history of psychologic disorder or brain injury, head dimensions that were too large for the 32-channel coil, the existence of non-surgical metal in the participant's body (including non-removable piercings), and color blindness. Participants were excluded from MRI analyses separately for excess motion within a particular phase of the experiment as noted in the MRI Data Preprocessing section below.

2.2 Mnemonic Discrimination Task

Stimuli for the task consisted of 200 pairs of perceptually and semantically similar Target-Lure pairs that were randomly selected from those hosted at https://github.com/celstark/MST. Each stimulus depicted an everyday object in the center of the image with a white background, and one member of each set was randomly selected to be a Target while the other was selected to be a Lure for each participant. Previous to this study, approximately 900 Target-Lure pairs were rated for similarity by 35 individuals (who were independent of the current study) on a 7-point Likert scale where a response of "1" indicated "extremely dissimilar" and "7" indicated "extremely similar". Only items that received an averaged similarity rating of 6.0 or higher were used in this task in order to ensure a sufficiently large number of FA responses for functional analyses.

The behavioral task consisted of three phases (Study, Test1, and Test2), all of which were performed while the participant was scanned using fMRI. Stimuli were displayed by means of an MR-compatible LCD monitor placed at the head of the scanner and viewed using a mirror mounted on the head coil. Responses were collected via a fiber optic button box. E-prime (v 2.0) was used to control stimulus display and record behavioral responses. While in the scanner but prior to engaging in the task, participants viewed a task instruction video; for consistency, the

same researcher (NMM) screened and debriefed all participants, responded to any questions about the task, and operated the scanner. Three different instruction videos were shown, one immediately prior to each respective phase of the task, in which explicit directions and examples were given to the participant.

Study phase: In the initial encoding phase, participants were shown 200 images of everyday objects and were instructed to indicate via button press whether each object is typically encountered indoors or outdoors (Figure 0, left). Each stimulus was preceded by a fixation cross for 500 ms, then presented for 3500 ms across one continuous block. For all phases of the experiment, 16 blank trials were added in a pseudo-random fashion to create jitter in stimulus timing, with care taken to ensure that each block contained an equal number of blank trials (when applicable). These blank trials, in addition to the inter-stimulus fixation crosses, served as the baseline condition in the single subject regression analysis (see MRI Data Pre-Processing).

Test1 phase: Participants were shown 100 Target stimuli (items identical to the original stimuli in the Study phase), and 100 Lure stimuli (items which were similar to, but differed visually from, the original items) one at a time in the center of the screen. Participants were instructed to indicate via button press whether each stimulus was the “same” (an exact repeat, or Target) or “different” (a related Lure) to the study item (Figure 0, middle). Stimuli were presented in a randomized order for both type (Target/Lure) and for level of Target-Lure similarity. The Test1 phase was conducted in two blocks of 100 stimuli each (50 Target, 50 Lure), where each stimulus was presented for 3500 ms and preceded by a fixation cross for 500 ms.

Test2 phase: In this phase, all stimuli from the Study phase (Targets), accompanied with the similar item (Lure), were presented side-by-side to the participant in a two-alternative forced-

choice format. The stimuli were counterbalanced for whether the Target appeared on the Left or Right (Figure 0, right). Participants were asked to indicate which of the two items they had seen during the Study phase (Left or Right) and their confidence in their decision (Very or Somewhat) via one of four potential response options (left very confident; left somewhat confident; right somewhat confident; right very confident). This phase was conducted in four blocks of 50 stimuli in an attempt to combat participant fatigue. Each stimulus was presented for 5000 ms and preceded by a fixation cross for 500 ms. During all phases of this experiment, participants were instructed to respond while the stimulus was presented on the screen, and no feedback was given relating to participant performance.

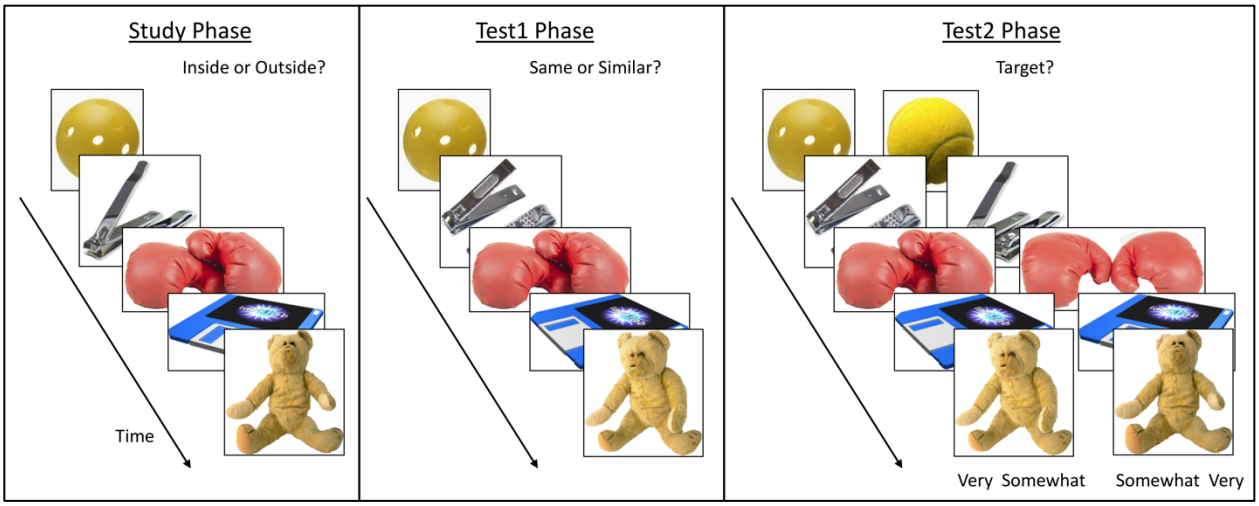

Figure 0. Mnemonic discrimination task. Left, during the study phase the participants made indoor/outdoor judgements about a series of objects. Middle, during the first test phase, participants were presented with either the same (Target) stimulus or one that was similar (Lure) and were asked to decide whether the stimulus was the same or similar. Right, during the second test phase, participants were presented with both Target and Lure stimuli simultaneously and asked to indicate which item was originally presented during the Study phase. Additionally, participants indicated the level of confidence (Very or Somewhat) in their decision and made both their target and confidence decision with a single button press. For instance, if they decided that the left stimulus was the Target and they were very confident, they pressed “1” whereas if they were only somewhat confident the right stimulus was the Target, they pressed “3”.

2.3 Statistical Analyses of Behavioral Data

Participant responses during the Study phase of the experiment were not analyzed for indoor-outdoor accuracy as this question was merely meant to orient participants' attention to the stimuli. During the second phase of the experiment, Test1, participants saw either Target or Lure stimuli and made a forced-dichotomous judgment of whether the item was "Same" or "Different". Correctly identifying the Target and Lure is termed Hit and Correct Rejection (CR), respectively, and the incorrect identification is termed Miss and False Alarm (FA), respectively. Here, we operationalize a false memory as the FA response to the lure stimulus. The third phase of the experiment, Test2, was a two-alternative forced-choice where both Target and Lure stimuli were presented and the participant was asked to identify the Target; consequently, only Hits and FAs were possible.

Behavioral analyses were conducted on response proportions, considering the number of items to which each participant responded. D-prime (d') scores were calculated for both Test1 and Test2 phases as an indicator of each participant's sensitivity to the stimuli. Test1 d' scores were calculated the standard way (Equation 1; Wickens, 2002) where Z = z-score, $pHit$ = proportion of Hits to the number of Targets, and pFA = proportion of FAs to the number of Lures. As the Test2 phase presented both stimuli simultaneously, a correction was added to the d' calculation (Equation 2; Wickelgren, 1968).

$$d'_{T1} = z(pHit) - z(pFA) \qquad d'_{T2} = \frac{z(pHit) - z(pFA)}{\sqrt{2}} \qquad \text{Equations (1, 2)}$$

Further, one-sample proportion testing using a continuity correction were used assess whether participants were more likely to respond one way versus another (e.g., if participants were more likely to Hit or Miss when presented with a Target, or more likely to respond Very versus Somewhat confident during CRs). Outliers were detected according to the 1.5 × IQR

method. Paired and one-sample *t*-testing was used to compare the d' scores to each other and against 0, respectively, utilizing an FDR correction to account for multiple comparisons. When the exclusion of outliers differentially impacted the statistics, tests including and excluding outliers are reported, otherwise reporting represents the exclusion of outliers.

2.4 MRI Acquisition

Functional and structural images were acquired on a Siemens TIM Trio 3T MRI scanner utilizing the 32-channel head coil. Participants contributed a T1-weighted scan, a high-resolution T2-weighted scan, and a series of functional T2*-weighted scans as described below. Standard-resolution structural images were acquired using a T1-weighted magnetization-prepared rapid acquisition with gradient echo (MP-RAGE) sequence with the following parameters: TR = 1900 ms, TE = 2.26 ms, flip angle = 9°, FoV = 218 × 250, voxel size = 0.97 × 0.97 × 1 mm, slices = 176 interleaved. High-resolution structural images were acquired using a T2-weighted sequence with the following parameters: TR = 8020 ms, TE = 80 ms, flip angle = 150°, FoV = 150 × 150 mm, voxel size = 0.4 × 0.4 × 2 mm, slices = 30 interleaved. The T2 data were acquired perpendicular to the long axis of the hippocampus. High-resolution multi-band echo-planar images were acquired with a T2*-weighted pulse sequence with the following parameters: Multi-band factor = 8, TR = 875 ms, TE = 43.6 ms, flip angle = 55°, FoV = 180 × 180 mm, voxel size = 1.8 mm^3, slices = 72 interleaved. This sequence was aligned parallel with the long axis of the hippocampus, and the first 11 TRs were discarded in order to allow for T1 equilibration.

2.5 MRI Data Pre-Processing

The pre-processing and visualization of structural and functional MRI data were accomplished with the following software: dcm2nii (Li et al., 2016), Convert 3D Medical Image Processing Tool (c3d) and ITK-SNAP (Yushkevich et al., 2006), Analysis of Functional

NeuroImages (AFNI; version 18.3.15; Cox, 2012), Advanced Normalization Tools software (ANTs; Avants et al., 2008, 2011), Automatic Segmentation of Hippocampal Subfields (ASHS; Yushkevich et al., 2015), and Multi-image Analysis GUI (Mango; Lancaster et al., 2012).

T1-weighted structural files were converted into 3D NIfTI files, rotated into functional space (referencing the volume with the minimum calculated outlier voxels) using a rigid transformation with a Local Pearson Correlation cost function. The rotated scan was then skull-stripped and warped into MNI space via a non-linear diffeomorphic transformation.

High-resolution T2-weighted files were converted to 3D NIfTI files, and both T1w and T2w files, in native space, were used to segment hippocampal subregions via ASHS, referencing the UPenn atlas (Yushkevich et al., 2015). Hippocampal subregion masks from each participant were then used to construct template priors via Joint Label Fusion (Figure 1; H. Wang et al., 2013), and study-specific masks for the subiculum, CA1, and the combined CA2-CA3-DG regions where constructed and resampled into functional dimensions. Voxels associated with overlapping subregions were excluded to combat partial-voluming effects when down-sampling, and the CA2, CA3, and DG masks were combined into a single mask due to functional resolution constraints.

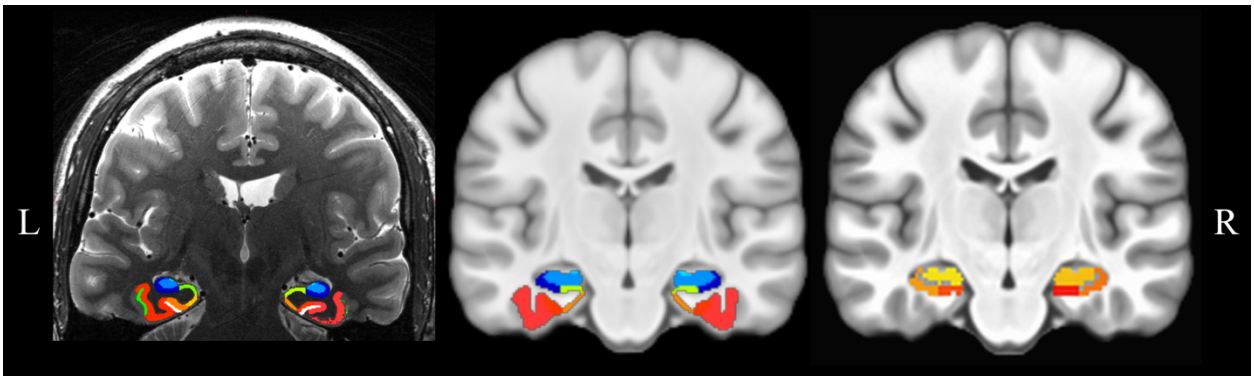

Figure 1. Hippocampal subregion masks. Left, a participant's T2-weighted scan segmented for the medial temporal lobe via ASHS. Middle, template medial temporal lobe masks constructed in MNI space via Joint Label Fusion. Right, the resampled masks, where overlapping labels were excluded.

241 T2*-weighted DICOMs were first converted to 4D NIfTI files, and the volume from the
242 experiment with the smallest percentage of outlier voxels was extracted to serve as a volume
243 registration base. The calculations for moving each volume into the same space as the volume
244 registration base were then performed, using cubic polynomial interpolations, and the T1-
245 weighted file was also aligned with the volume registration base using a rigid transformation that
246 utilized a Local Pearson Correlation cost function. This rotated structural scan was then used to
247 produce normalization calculations via a non-linear transformation into MNI space. Movement
248 of the functional data into template space was then accomplished by concatenating the volume
249 registration calculations with the MNI transformation calculations, thereby moving all volumes
250 from native space to MNI space via a single transformation, resulting in less partial-voluming
251 and blurring of the functional data than if each transformation was applied separately. A series of
252 masks were then generated in order to remove volumes with missing data as well as determine
253 the intersection between functional and structural data, and the functional data were then scaled
254 by the mean signal. Finally, centered (demeaned) motion files for six degrees of freedom were
255 generated for each run of the experiment.

256 Behavioral data during each of the three phases (Study, Test1, and Test2) were coded
257 according to behavioral responses during that and subsequent phases of the experiment. Trials
258 from the Study phase were coded according to subsequent Test1 performance (i.e., Study
259 predicting Test1 activation or SpT1): subsequent Hit, subsequent FA, subsequent Miss, or
260 subsequent CR. Additionally, Study-phase trials were coded according to behavioral responses in
261 Test1 and Test2 (SpT1pT2). For example, a stimulus presented during Study might subsequently
262 receive a FA in Test1 and a Hit in Test2 (subsequent FA-Hit). Coding Study trials in this way
263 allows us to potentially localize differences during the encoding process that result in Test1 FAs

264 that are perpetuated into Test2 (subsequent FA-FA) compared to FAs that are corrected in Test2
265 (subsequent FA-Hit). It should be noted, however, that a low number of certain behaviors
266 (particularly Test1 Misses preceding Test2 behaviors) necessitated collapsing across CR and
267 Miss response bins in the deconvolution phase in order to avoid producing vectors for some
268 participants consisting entirely of zeros, vectors which are incompatible with deconvolution.
269 Such behavioral bins were not included in group-level analyses. Test1 trials were coded
270 according to four outcomes (Hit, FA, CR, and Miss). Responses during Test1 that preceded
271 Test2 (T1pT2) produced 4×2 possible outcomes since Test2 had two possible behaviors (Hit,
272 FA). Test2 responses that followed Test1 (T2fT1) also produced 4×2 possible outcomes, and
273 Test2 had two outcomes. Dividing all Test2 behavioral data according to confidence interval as
274 well as response type resulted in bins with trial counts that were too small for analyses, and as
275 such the confidence interval was dropped from functional analyses. Finally, trials where
276 participants failed to respond were coded separately from each of the above.
277 Vectors coding for the above behavioral outcomes were used in single-subject regression
278 models (Generalized Least Square) in which the canonical hemodynamic response was
279 convolved with a boxcar function with a duration equal to stimulus presentation. Modelling
280 included polynomial regressors accounting for scanner drift and scan run and regressors coding
281 for movement. The baseline condition consisted of collapsing across inter-stimulus intervals and
282 jittered blank stimuli. Volumes consisting of movement greater than 0.3° and/or more than 0.3
283 mm in any direction relative to the previous volume were discarded along with the previous
284 volume. Each deconvolution was assessed for the number of volumes that necessitated removal
285 and any participant with more than 10% of volumes dropped *and* 1.5 × IQR of volumes dropped
286 were excluded from subsequent analyses: one subject was removed from the Study phase, two

287 subjects were removed from the Test1 phase, and three subjects were removed from the Test2
288 phase. As such 34 participants were used to investigate Study performance, 33 for Test1, and 32
289 for Test2 performance.

290 Region of interest (ROI) specific masks were constructed to test *a priori* hypotheses. In
291 addition to the hippocampal subregion masks described above, clusters associated with memory
292 tasks were constructed by using the key words "Memory Encoding" and "Memory Retrieval" as
293 search terms in the NeuroSynth.org algorithm, which conducts a meta-analysis of PubMed
294 articles reporting the same key words and constructs a set of clusters derived from the reported
295 coordinates in said papers. The search term "Memory Encoding" yielded 124 articles, while
296 "Memory Retrieval" yielded 183. Association clusters (corrected for meta-analysis multiple
297 comparisons at the $p = .01$ level) were downloaded and thresholded at $k=80$ for the encoding
298 clusters, and $k=100$ for the retrieval clusters, which produced a set of binary masks that were
299 then resampled into functional dimensions. The "encoding" set contained four clusters,
300 consisting of the left hippocampus, right collateral sulcus, left inferotemporal gyrus, and right
301 amygdala (Figure 2). The "retrieval" set contained 10 clusters, consisting of the left retrosplenial,
302 angular gyrus, hippocampus, medial PFC, temporal-parietal junction, dorsal medial PFC, and the
303 right hippocampus, parietal-occipital sulcus, and posterior hippocampus (Figure 3). NeuroSynth
304 masks were then rotated and resampled into functional space and used to extract region-of-
305 interest (ROI) specific mean β-coefficients, using "encoding" masks to investigate volumes from
306 the Study phase, and "retrieval" masks for the Test1 and Test2 phases. Two-factor repeated
307 measures ANOVAs were used to investigate an ROI × behavior interaction for hippocampal
308 subregion and NeuroSynth masks, individually, where a Greenhouse-Geisser (GG) correction
309 was conducted for violations of sphericity. Any significant interaction surviving FDR multiple-

comparison adjustments were investigated via a one-way repeated measures ANOVA to determine which ROIs had differential activation for the various behaviors, and subsequent pairwise *t*-testing to determine how the β-coefficients for the behaviors differed within the ROI.

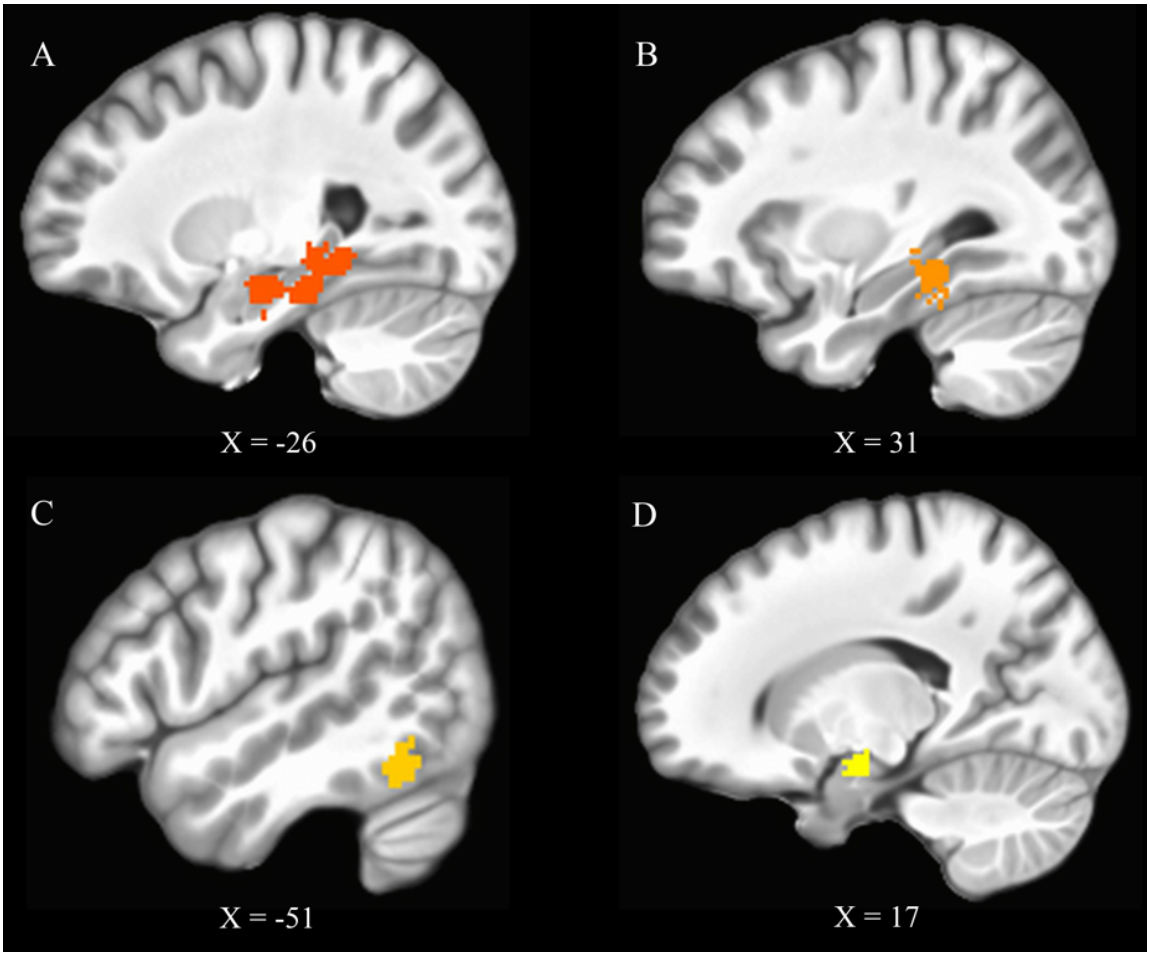

Figure 2. NeuroSynth encoding masks. A left hippocampus, B left inferotemporal gyrus, C right collateral sulcus, D right posterior amygdala.

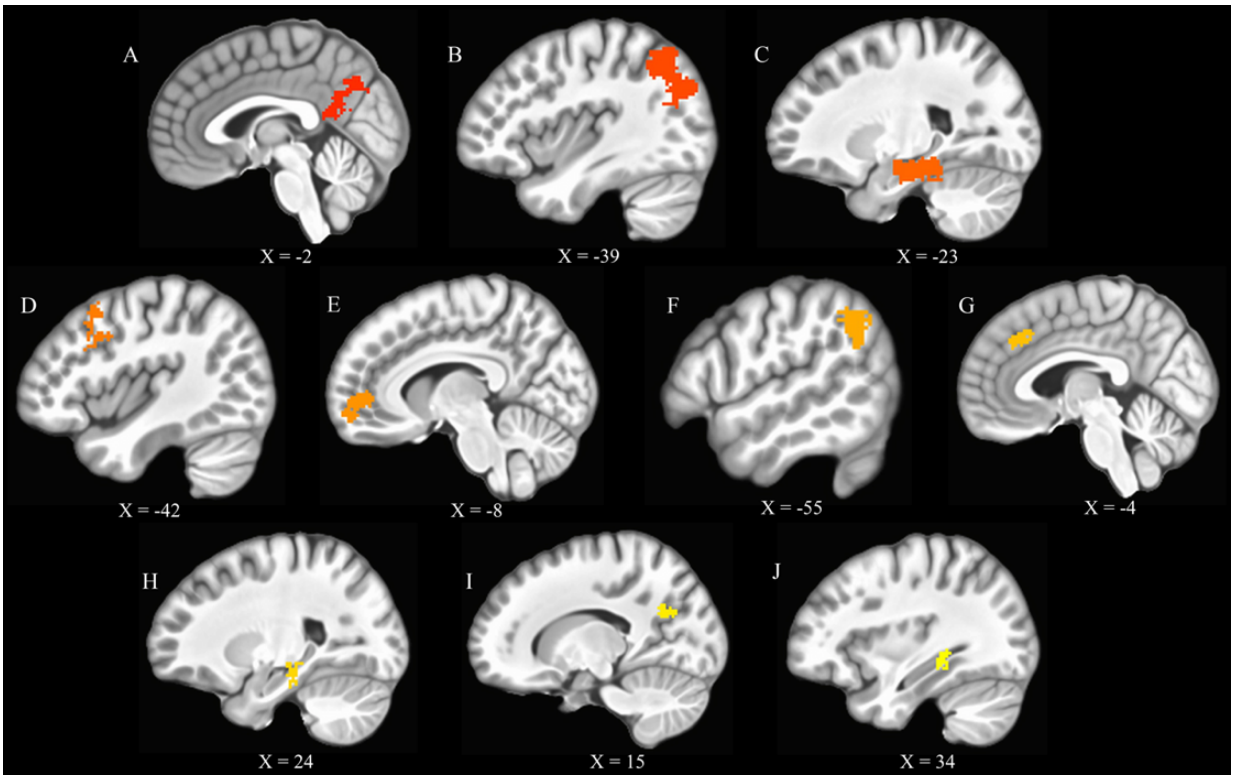

Figure 3. NeuroSynth retrieval masks. A bilateral posterior cingulate, B left posterior parietal, C left hippocampus, D left middle frontal gyrus, E left medial prefrontal, F left temporoparietal junction, G left dorsal medial prefrontal, H right hippocampus, I right parieto-occipital sulcus, and J right posterior hippocampus.

In preparation for exploratory group-level analyses, a gray matter mask was constructed in template space using Atropos priors (Tustison et al., 2014), which was multiplied by the intersection mask to produce a gray matter, intersection inclusion mask. Exploratory group-level analyses were then conducted on all voxels within this inclusion mask for each phase of the experiment using the Equitable Thresholding and Clustering method (ETAC; Cox, 2019) using the p-values 0.01, 0.005, and 0.001, blurs of 4, 6, and 8mm, nearest neighbors 1, and two-sided thresholding. Any surviving clusters were used to extract mean β-coefficients from participant data. All scripts used in this experiment can be found at https://github.com/nmuncy/STT.

3 RESULTS

3.1 Behavioral Analyses and Descriptive Statistics

Analyses of Test1 performance yield a mean d'_{T1} score of 0.66, indicating that successfully distinguishing between Targets and Lures was difficult, as the design of the experiment intended. However, participants were successful at detecting Targets and Lures in Test1, as the d'_{T1} differed significantly from zero ($t(34) = 11.3$, $p_{FDR} < .0001$, 95% CI [0.54, 0.78]). Response proportions were tested against chance (0.5) for each stimulus type; participants were more likely to Hit than Miss on Targets (estimate = .79, $\chi^2 = 32.18$, $p_{FDR} < .0001$, 95% CI [.69, .86]) while being equally likely to CR and FA on Lures (estimate = .43, $\chi^2 = 1.43$, $p_{FDR} = .41$, 95% CI [.33, .54]; Figure 1, left). On average, participants had 76 Hits, 54 FAs, 42 CRs, 20 Misses, and 6 non-responses in this phase. The 54 FA responses are of particular interest to the study: we designed the difficulty of the experiment to elicit approximately 60 FA responses in Test1 as two responses (Hit, FA) were possible in Test2, so subdividing Test1 FAs bins according to subsequent Test2 responses would yield sufficient trials to test statistically and model the HRF.

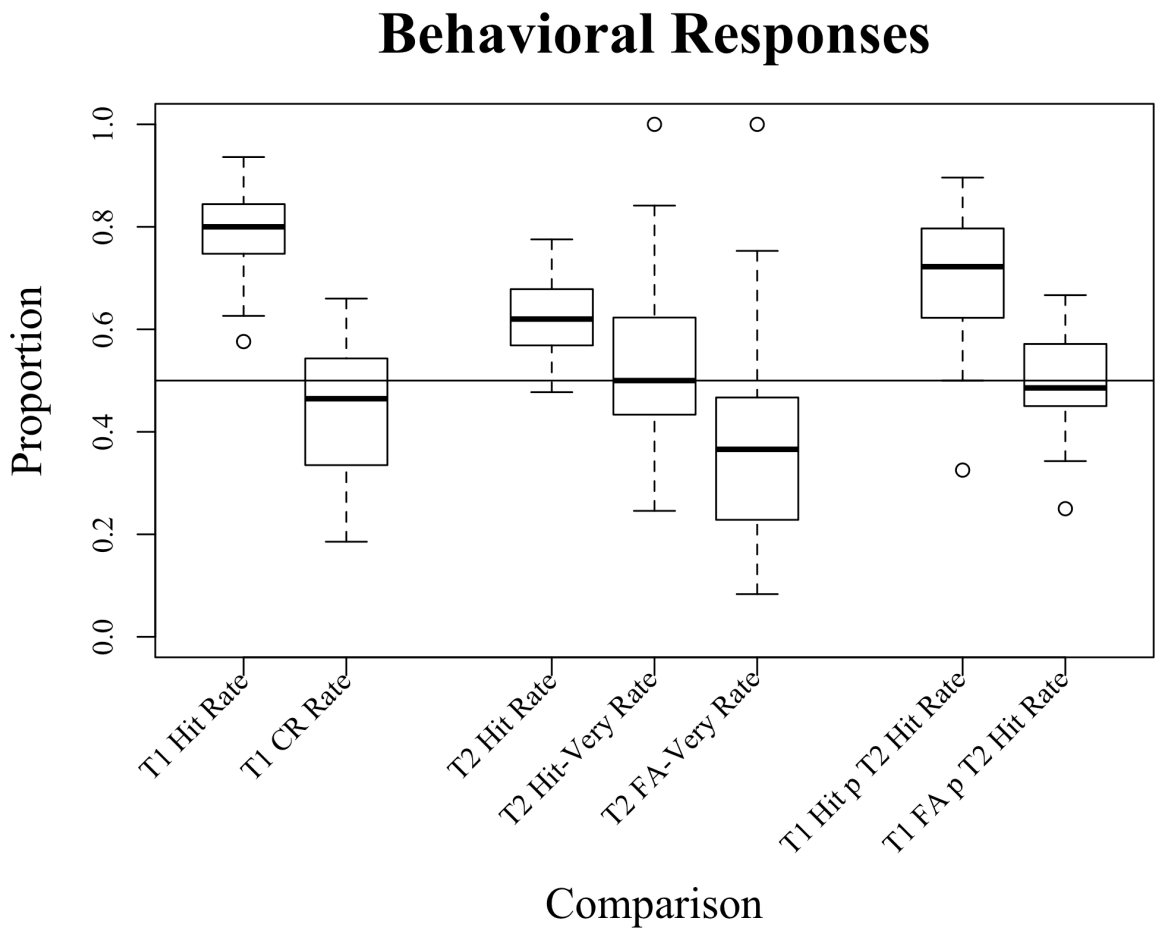

Figure 1. Boxplots of response proportions. Performance at chance results in a proportion of 0.5. Test1 (T1), Test2 (T2), and Test1 preceding Test2 (T1 p T2). "Hit-Very Rate" = likelihood of responding Very Confident versus Somewhat Confident during Hits. "T1 Hit p T2 Hit Rate" = proportion of Test1 Hits that precede a Test2 Hit versus Test2 FA. Outliers are included.

Analyses of Test2 performance yielded a d' $_{T2}$ of 0.45, which also differed significantly from zero ($t(34) = 9.12$, $p_{FDR} < .0001$, 95% CI [.35, .55]), and indicated that participants were more likely to Hit than FA (estimate = .62, $\chi^2 = 11.43$, $p_{FDR} = .002$, 95% CI [.55, .69]; Figure 1, middle). In terms of confidence, participants were equally likely to indicate Very and Somewhat confident during Test2 Hits (estimate = .53, $\chi^2 = 0.2$, $p_{FDR} = .7$, 95% CI [.43, .62]) and during Test2 FAs (estimate = .38, $\chi^2 = 4.1$, $p_{FDR} = .09$, 95% CI [.27, .49]). Finally, testing the d' scores for Test1 versus Test2 indicated that participants performed better during Test1 ($t(34) = 4.37$,

356 p_{FDR} = .001, 95% CI [.11, .31]). On average, participants had 121 Hits, 73 FAs, and 5 non-
357 responses.

358 Proportion testing of Test1 Hits and FAs which precede Test2 responses (T1pT2) reveal
359 that a Test1 Hit was more likely to result in a subsequent Test2 Hit than a FA (estimate = .72, χ^2
360 = 13.8, p_{FDR} < .001, 95% CI [.60, .82]), but that Test1 FAs were equally likely to result in either
361 a subsequent Test2 Hit or FA (estimate = .5, $\chi^2 \approx 0$, $p_{FDR} \approx 1$, 95% CI [.37, .64]; Figure 1, right).
362 These results, when viewed with respect to the Test2 results (d' which differs from 0 and a
363 tendency to detect the target), suggest that recovery from a Test1 FA is possible but not
364 guaranteed. It should be noted, though, that Test2 Target detection (particularly following a
365 Test1 Hit) may be driving some of the behavioral findings. Analyses of the BOLD response
366 during Test1 FAs that precede Hits versus FAs in Test2 time will help determine why some
367 Test1 FAs are perpetuated into Test2 while others are corrected, and if participants are indeed
368 performing differently from chance.

369 Test1 responses that preceded Test2 confidence responses consist of permutations of
370 Test1 behaviors, Test2 behaviors, and Test2 confidence responses. It was originally thought that
371 Test1-Test2 FA-Hit responses would yield higher confidence responses than FA-FA, which
372 could show interesting activation in the BOLD response. Paired *t*-testing revealed, however, an
373 equal number of "Very Confident" responses for both Hits and FAs in Test2 following a Test1
374 FA ($t(34) = 0.1$, $p = .9$, 95% CI [-2.3, 2.5]), indeed, participants were equally likely to select
375 "Very" and "Somewhat" during Test2 Hits and FAs following Test1 FAs (estimate = .4, χ^2 =
376 0.7, p_{FDR} = .53, 95% CI [.23, .61], estimate = .4, χ^2 = 0.8, p_{FDR} = .53, 95% CI [.22, .59],
377 respectively). Finally, parsing the data in such a fashion yielded bins that were too small to
378 reasonably model the hemodynamic response (e.g. Test1 FA preceding Test2 Hit-Very = 11 ±6

379 responses on average). As such, all subsequent analyses were collapsed across confidence
380 ratings.

381 **3.2 Functional MRI Analyses**

382 To address our research hypotheses, we conducted *a priori* analyses on both hippocampal
383 subregions and NeuroSynth memory regions, separately, as well as *a posteriori* exploratory,
384 whole-brain analyses. First, analyses investigated the effects of encoding on subsequent test
385 performance. The first analysis examined signal during the Study phase coded for subsequent
386 Test1 behavior, termed "SpT1" or "Study preceding Test1" (Figure 2, blue), while the second
387 analysis investigated the Study phase coded instead for trials which resulted in either a Test1 FA
388 that was then corrected (Test2 Hit) or perpetuated (Test2 FA), termed "SpT1pT2" (Figure 2,
389 purple). One participant was excluded from these analyses due to excessive movement during the
390 Study phase (n=34). Next, two analyses investigated the Test1 phase of the experiment. One
391 analysis investigated signal during Test1 associated with the various Test1 behaviors, termed
392 "T1", and the next analysis investigated Test1 FAs according to whether or not they would be
393 corrected in Test2, termed "T1pT2" (Figure 2, red). Two participants were excluded from these
394 analyses due to excessive movement in the Test1 phase (n=33). The final set of analyses
395 involved signal from the Test2 phase. First, we investigated signal during Test2 associated with
396 Test2 behaviors, termed "T2". The final analysis coded trials according to whether a Test2 Hit or
397 FA followed a Test1 FA, here termed "T2fT1" or "Test2 following Test1" (Figure 2, green).
398 Three participants were excluded from the last set of analyses due to excessive movement during
399 Test2 (n=32). In this way, the signal from each respective phase of the experiment is
400 investigated, but trials are coded for the current phase (T1 and T2), for future responses to
401 current stimuli (SpT1, SpT1pT2, T1pT2), or with respect to previous responses (T2fT1), thereby

allowing for the investigation of potential encoding and retrieval effects that differentially impact the memory trace and attempted recovery thereof.

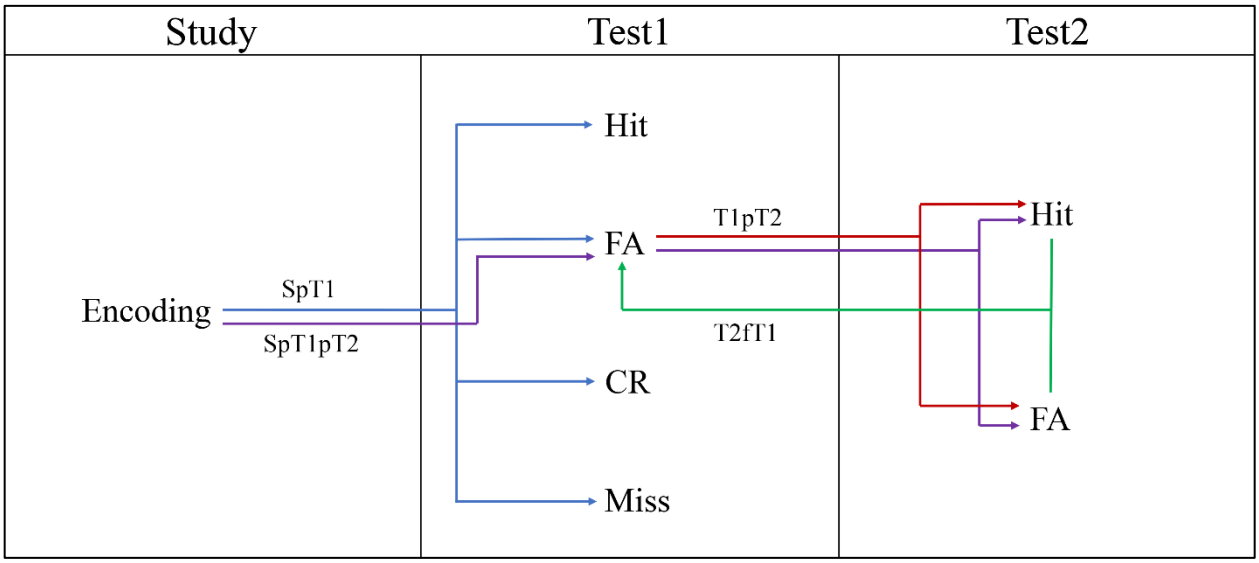

Figure 2. Schematic of analyses performed. The experiment consists of three phases, Study, Test1, Test2. Study trials were coded for which stimuli preceded Test1 responses (SpT1, blue) as well as the resulting response in Test2 following a Test1 response (SpT1pT2, purple). Test1 trials were analyzed both for the behavior which occurred in Test1 (not colored) as well as for responses which preceded a Test2 response (T1pT2, red). Similarly, Test2 trials were analyzed for Test2 behaviors (not colored) as well as for which Test1 behavior the Test2 response followed (T2fT1).

3.2.1 Study trials preceding Test1 responses (SpT1). Study stimulus responses were coded for subsequent Test1 behavior. For example, one stimulus in Study could elicit a Hit in Test1 while a different stimulus could elicit a FA response.

For *a priori* analyses, a two-factor repeated measures MANOVA, where hippocampal subregion ROIs (Left and Right Subiculum, CA1, and CA2-CA3-DG) served as within-subject factor A and subsequent Test1 behavior (Hit, FA, CR) served as within-subject factor B, failed to detect any significant interaction between the factors ($F(10, 330) = .89$, $\eta_G^2 < .001$, GGε = .45, $p(\mathrm{GG})_{\mathrm{FDR}} = .47$). A similar analysis using NeuroSynth ROIs associated with encoding (Left hippocampus, inferotemporal gyrus, and right amygdala, and collateral sulcus; see Figure 2) also

failed to detect a significant interaction between the factors ($F(6, 198) = 1.4$, $\eta_G^2 < .001$, GGε = .38, $p(\mathrm{GG})_{\mathrm{FDR}} = .24$).

Exploratory whole-brain analyses revealed seven clusters (Figure 3, Table 1) which significantly differed in Study volumes for subsequent Test1 Hits versus FAs (two subsequent behaviors were investigated given that ETAC only supports *t*-test equivalent analyses). Three clusters were more active preceding a Hit than a FA (left and right visual stream [Figure 3, A, B], right intraparietal sulcus [C]) which are consistent with visual and attention networks. Four clusters demonstrated the opposite pattern of activation (left and right angular gyrus [D, E], left precuneus [F], and left dorsal medial PFC [G]), where the angular gyri and precuneus regions are consistent with the default mode network.

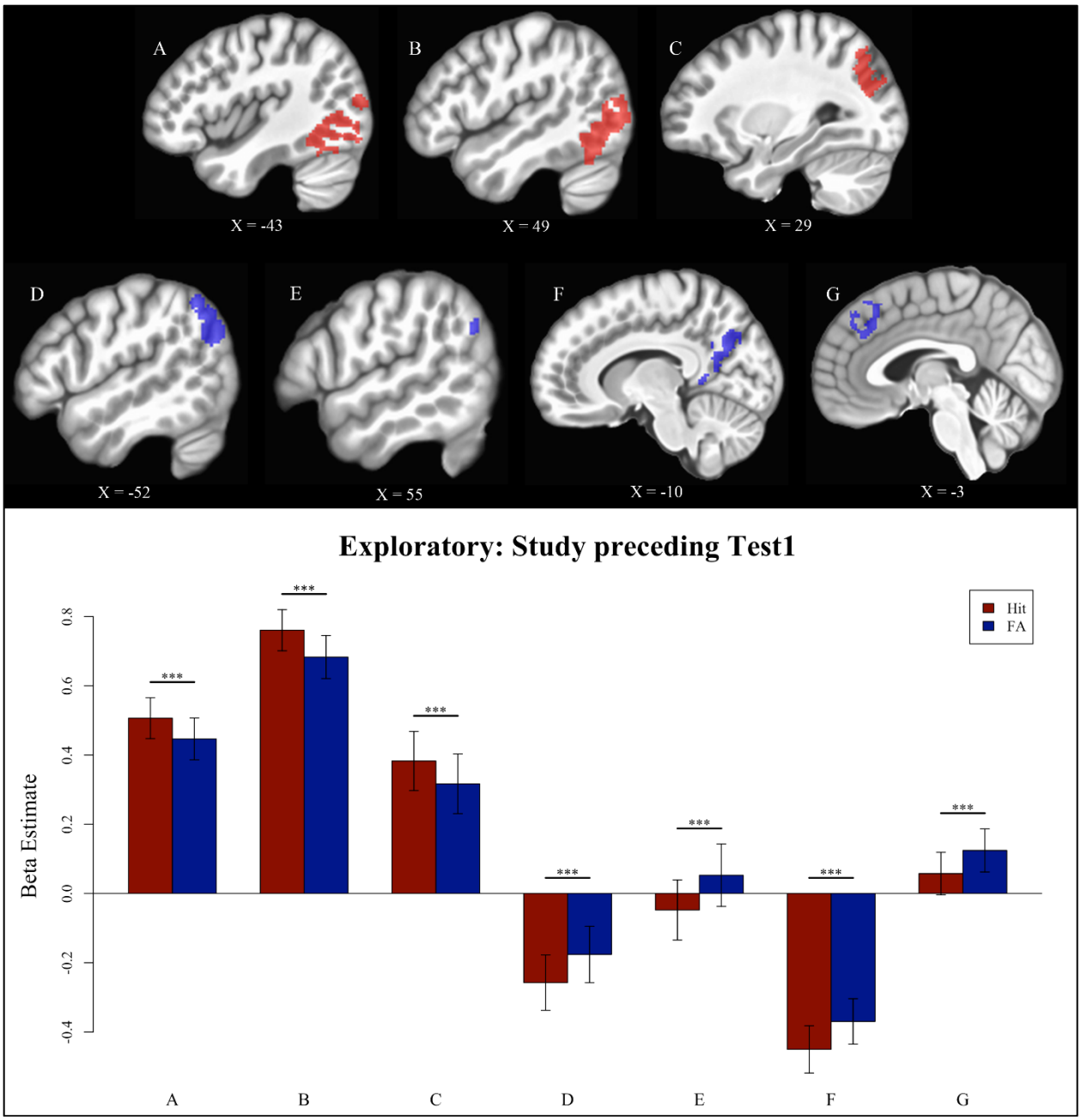

Figure 3. SpT1 whole-brain results. Regions which demonstrated differential signal during SpT1. Hits versus FAs during an exploratory whole-brain analysis. A left visual stream, B right visual stream, C right intraparietal sulcus, D left angular gyrus, E right angular gyrus, F left precuneus, G left dorsal medial PFC.

Table 1

Phase	ROI	K	X	Y	Z	T(33)	Sig	LB	UB
SpT1	LVS	1649	-42.6	-74.3	-1.2	6.8	***	0.04	0.08
	RVS	2054	42.1	-72.4	-2.4	7.1	***	0.06	0.10
	RIPS	561	30.1	-75.2	33.7	5.3	***	0.04	0.09
	LAG	825	-47.8	-64.7	39.1	-6.1	***	-0.11	-0.05

	RAG	56	53.6	-64	30.6	-5.1	***	-0.14	-0.06
	LPPCU	1334	-6.3	-60.4	27.1	-6.7	***	-0.11	-0.06
	LDMPFC	269	-4.7	34.6	39.6	-7.7	***	-0.08	-0.05
T1	ROI	K	X	Y	Z	T(32)	Sig	LB	UB
	RPUT	46	25	-9	-22	4.51	***	0.066	0.174
	RMPFC	984	6	-3	-2	6.27	***	0.056	0.11
	RAG	917	52	-14	-40	7.37	***	0.053	0.093
	RAMTG	646	59	22	-40	6.03	***	0.05	0.101
	RPCU	174	4	-41	0	6.48	***	0.068	0.13
	LDMPFC	18	-3	-48	5	-5.23	***	-0.22	-0.097
	LINS	103	-33	-3	38	-6.2	***	-0.141	-0.071
	RIPS	321	28	-40	-18	-6.26	***	-0.11	-0.056
	RIFG	47	47	-25	-43	-6.64	***	-0.149	-0.079
T1pT2	ROI	K	X	Y	Z	T(32)	Sig	LB	UB
	LINS	98	31	-26	-3	4.7	***	0.05	0.12
	RSMG	19	-63	30	41	6.8	***	0.08	0.15
T2fT1	ROI	K	X	Y	Z	T(31)	Sig	LB	UB
	RAG	193	45	-72	18	4.95	***	0.05	0.12
	RPCU	813	2	-56	16	4.45	***	0.05	0.13

Cluster sizes and test statistics for the exploratory whole-brain analyses. ROI = region of interest, K = cluster size, X-Z = coordinate (peak), T(xx) = t-test(df), Sig = significant, *** = p < .0001, LB = 95% confidence interval lower bound, UB = upper bound. See text for ROI abbreviations.

3.2.2 Study trials preceding Test1 FAs preceding Test2 responses (SpT1pT2). Study trials were coded for whether or not an encoding stimulus which resulted in a Test1 FA was then corrected in Test2. For example, while two stimuli both may have resulted in a Test1 FA, the participant may have successfully identified one stimulus as the Target but failed to do so for the other stimulus in Test2. Such coding allowed us to investigate the potential role of encoding in recovering from interference in Test2. On average there were an equal number of encoding trials (26) that resulted in a T1 FA that were and were not corrected in Test2.

For the a priori analyses, individual two-factor repeated measures MANOVAs were used to investigate the interaction of ROI and behaviors for hippocampal subregion as well as NeuroSynth ROIs associated with encoding. ROIs were used as factor A in the model and factor B consisted of whether a FA would be corrected (Test1 FA, Test2 Hit) or perpetuated (Test1 FA, Test2 FA). No ROI × behavior interaction was detected in this phase of the experiment for either the hippocampal or NeuroSynth ROIs ($F(5, 165) = 1.4$, $\eta_G^2 < .001$, GGε = .72, $p = .23$; $F(3, 99) =$

2.7, η_G^2 = .001, GGε = .58, $p(\mathrm{GG})_{\mathrm{FDR}}$ = .09, respectively). Further, exploratory whole-brain analyses failed to detect any regions during the encoding phase that discriminated for whether a subsequent Test1 FA would be corrected or perpetuated in Test2. Taken together, our data suggest that while the neural activation during encoding trials had a significant impact on subsequent Test1 performance, the activation during the encoding phase did not predict a recovery from interference, as indicated by a Test1 FA preceding a Test2 Hit.

3.2.3 Test1 responses. Test1 trials were coded for the relative success of target and lure detection, using the labels Hit, FA, CR, and Miss. A two-factor repeated measures MANOVA detected a significant interaction of hippocampal subregions (left and right subiculum, CA1, and CA2-CA3-DG) and Test1 behaviors (Hit, FA, CR; $F(10, 320) = 2.7$, η_G^2 = .001, GGε = .58, $p(\mathrm{GG})_{\mathrm{FDR}}$ = .03). One-way ANOVA testing revealed that left CA2-CA3-DG and CA1 were driving this effect ($F(2, 64) = 8.28$, η_G^2 = .01, GGε = .96, $p(\mathrm{GG}) < .001$; $F(2, 64) = 3.64$, η_G^2 < .01, GGε = .86, $p(\mathrm{GG}) < .05$, respectively). Further, post-hoc pairwise *t*-testing indicated that the left CA2-CA3-DG was significantly more active during Hits than FAs and CRs ($t(32) = 2.16$, p < .05, 95% CI [.004, .12]; $t(32) = 4.47$, p < .001, 95% CI [.06, .17], respectively) and that the left CA1 was also more active during Hits than CRs ($t(32) = 3.06$, p < .01, 95% CI [.01, .06]; Figure 4).

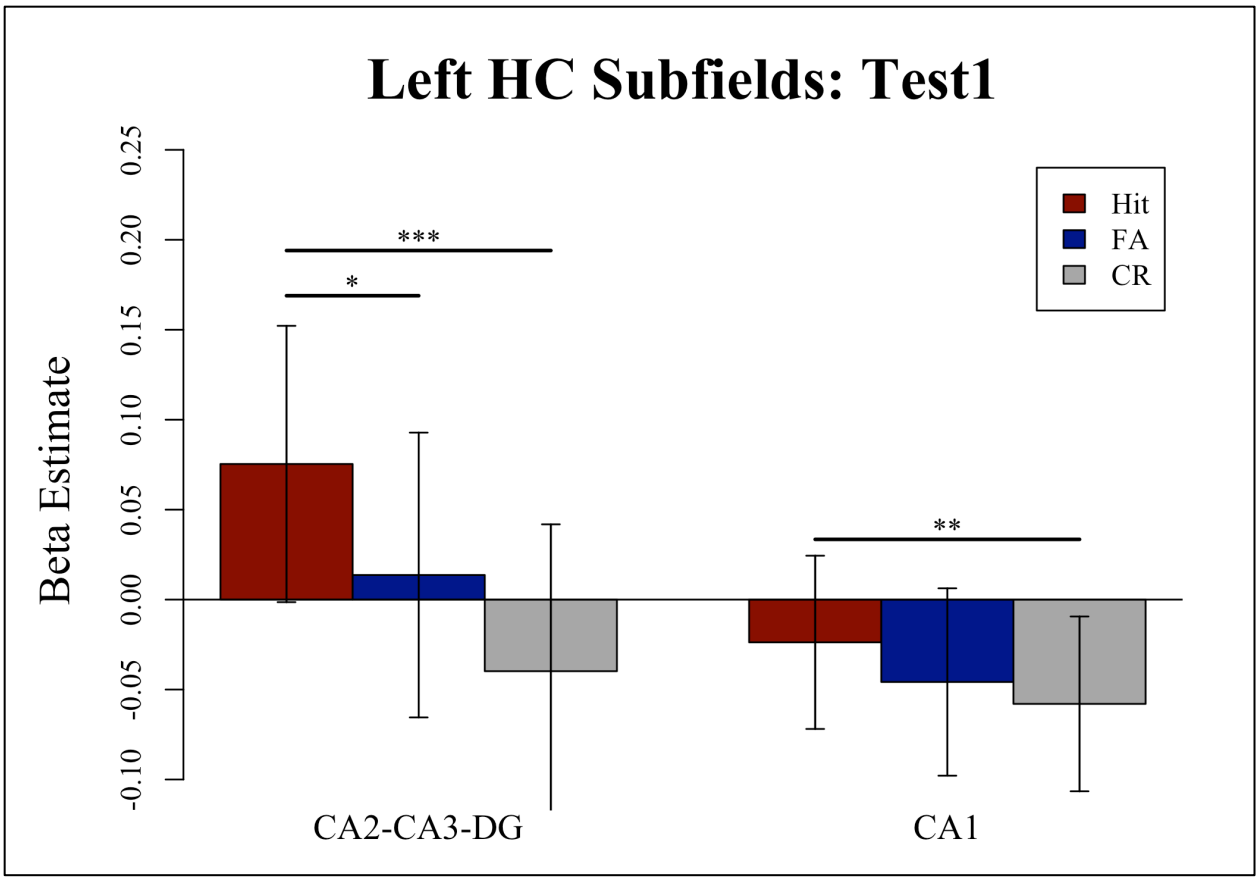

Figure 4. Test1 hippocampal subfield results. The CA2-CA3-DG region is more active during Hits than FAs or CRs. The CA1 is more active during Hits than CRs.

Using the same analysis as above to investigate NeuroSynth masks associated with retrieval (left angular gyrus, dorsal medial PFC, hippocampus, middle frontal gyrus, medial PFC, retrosplenial cortex, temporo-parietal junction, and right hippocampus, parieto-occipital sulcus, and posterior hippocampus; see Figure 3) revealed a significant mask × behavior interaction was detected ($F(18, 576) = 13.15$, $\eta_G^2 = .01$, GGε = .36, $p(\mathrm{GG})_{\mathrm{FDR}} < .0001$). Subsequent one-factor ANOVA testing revealed that five regions (Table 2, T1) were driving this effect: the left temporo-parietal junction (LTPJ), ventral medial PFC (LVMPFC), dorsal medial PFC (LDMPFC), middle frontal gyrus (LMFG), and angular gyrus (LAG). Post-hoc t-testing was conducted for these regions in order to determine for which behaviors a difference in BOLD signals occurred (Table 3, T1), which revealed two patterns of activity (Figure 5). First, the LTPJ

and LVMPFC were associated with greater BOLD signal during Hits than CRs; with such a pattern of activity it appears that these two regions are involved in recognition aspects of memory retrieval. Second, activity in the LDMPFC, LMFG, and LAG regions during CRs demonstrated greater BOLD signal than either Hits or FAs, consistent with novelty detection in memory systems. In sum, the effect found in these regions is driven largely by the differential BOLD response during CRs versus Hits and FAs (Table 3, T1), and the effects appear to be associated with recognition memory and novelty detection.

Table 2

	ROI	F(2,64)	η_G^2	GGe	Sig
T1	LTPJ	6.9	0.006	0.82	**
	LVMPFC	7.3	0.016	0.81	**
	LDMPFC	22.5	0.029	0.72	***
	LMFG	10.9	0.011	0.80	***
	LAG	13.9	0.012	0.81	***
T1pT2	ROI	F(2,64)	η_G^2	GGe	Sig
	LDMPFC	20.4	0.025	0.7	***
	LAG	17.9	0.02	0.9	***
	LMFG	9.6	0.01	0.86	***
T2fT1	ROI	F(2,62)	η_G^2	GGe	Sig
	LAG	16.2	0.023	0.96	***
	LDMPFC	12.2	0.019	0.95	***
	LMFG	10.7	0.017	0.91	**
	LPCU	11.8	0.019	0.88	***
	LTPJ	15.1	0.022	0.87	***
	RPOS	13.9	0.023	0.87	***

NeuroSynth ROI ANOVA statistics. T1 = Test1,T1pT2 = Test1 preceding Test2, T2fT1 = Test2 following Test1. GGe = Greenhouse-Geisser epsilon. ** = p < .01, *** = p < .001.

Table 3

	Contrast	Hit-FA				Hit-CR				FA-CR			
T1	ROI	T(32)	Sig	LB	UB	T(32)	Sig	LB	UB	T(32)	Sig	LB	UB
	LTPJ	3.65	***	0.03	0.10	3.37	**	0.03	0.13	0.72	n.s.	-0.04	0.07
	LVMPFC	2.22	*	0.00	0.08	3.60	**	0.04	0.15	1.87	n.s.	0.00	0.12
	LDMPFC	-0.21	n.s.	-0.04	0.03	-5.62	***	-0.21	-0.10	-4.72	***	-0.22	-0.09
	LMFG	-0.14	n.s.	-0.02	0.02	-3.98	***	-0.09	-0.03	-3.39	**	-0.09	-0.02
	LAG	-1.22	n.s.	-0.02	0.01	-4.44	***	-0.07	-0.03	-3.64	***	-0.07	-0.02
T1pT2	Contrast	FpH-FpF				FpH-CpH				FpF-CpH			
	ROI	T(32)	Sig	LB	UB	T(32)	Sig	LB	UB	T(32)	Sig	LB	UB
	LDMPFC	2.10	*	0.00	0.08	-4.28	***	-0.23	-0.08	-5.13	***	-0.27	-0.12
	LAG	0.99	n.s.	-0.01	0.03	-4.36	***	-0.08	-0.03	-5.21	***	-0.09	-0.04

	LMFG	0.22	n.s.	-0.02	0.03	-3.49	**	-0.09	-0.02	-3.45	**	-0.10	-0.02
T2fT1	Contrast	HfF-FfF				HfF-HfC				FfF-HfC			
	ROI	T(31)	Sig	LB	UB	T(31)	Sig	LB	UB	T(31)	Sig	LB	UB
	LAG	-1.00	n.s.	-0.08	0.03	-5.84	***	-0.14	-0.07	-3.33	**	-0.13	-0.03
	LDMPFC	-0.07	n.s.	-0.03	0.03	-5.42	***	-0.10	-0.04	-4.49	***	-0.10	-0.04
	LMFG	0.42	n.s.	-0.03	0.04	-4.14	***	-0.11	-0.04	-3.64	***	-0.13	-0.04
	LPCU	0.50	n.s.	-0.04	0.06	-4.91	***	-0.17	-0.07	-4.24	***	-0.19	-0.07
	LTPJ	0.86	n.s.	-0.03	0.07	-3.87	***	-0.15	-0.05	-4.21	***	-0.17	-0.06
	RPOS	-0.40	n.s.	-0.06	0.04	-5.02	***	-0.19	-0.08	-3.75	***	-0.20	-0.06

NeuroSynth ROI post-hoc t-tests corresponding to Table 2. * = p < .05, n.s. = p > .05.

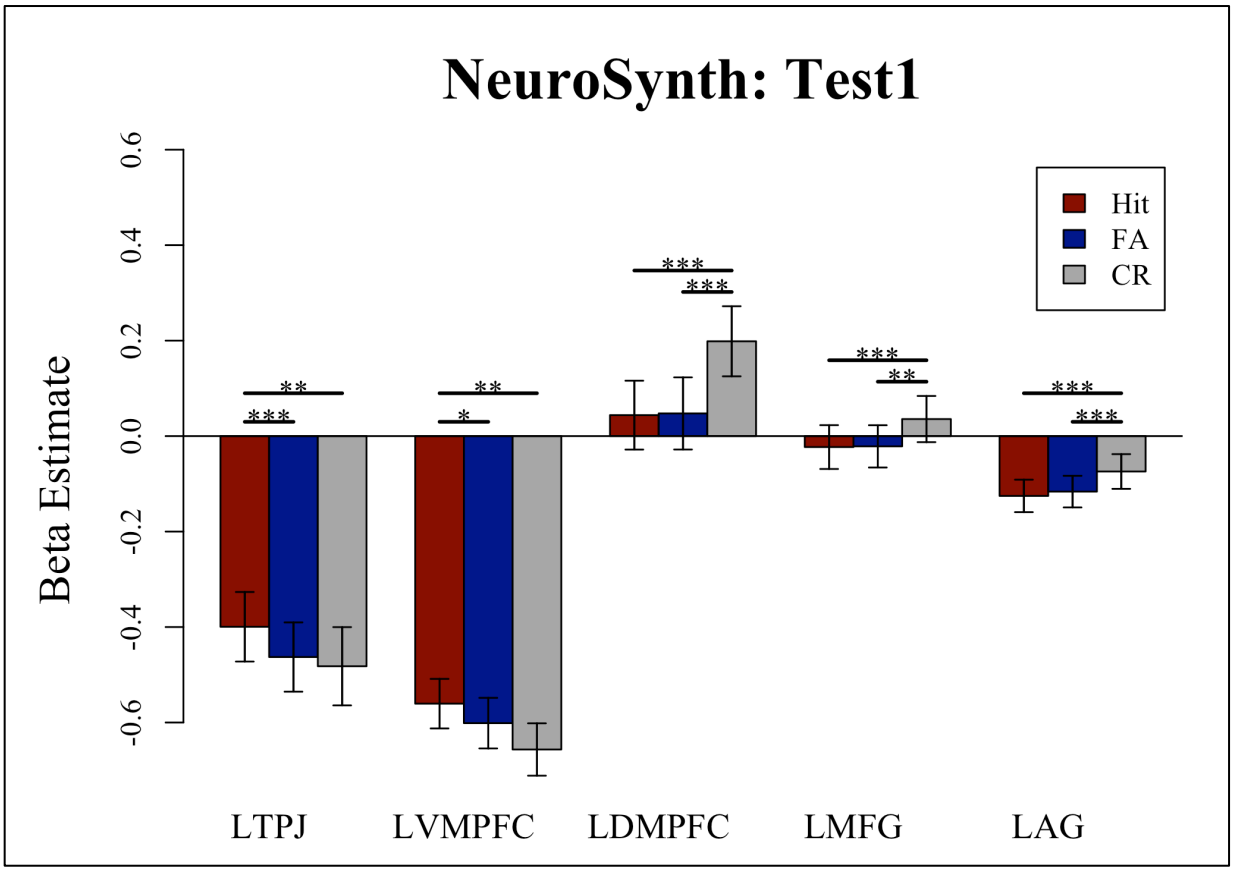

Figure 5. Test1 NeuroSynth results. Five regions (left temporo-parietal junction [LTPJ], ventral medial prefrontal cortex [LVMPFC], dorsal medial prefrontal cortex [LDMPFC], middle frontal gyrus [LMFG], angular gyrus [LAG]) differentiate between Hits, FAs, and CRs. The LTPJ and LVMPFC demonstrate significantly greater activity during Hits than CRs, while the LDMPFC, LMFG, and LAG show the opposite pattern of activity.

Exploratory analyses detected nine clusters (Figure 6) which demonstrated differential activation according to whether a Hit versus a FA was occurring (Table 1, Figure 6). Five regions had a greater BOLD response during Hits all of which were located in the right

510 hemisphere: the putamen (RPUT, A), medial PFC (RMPFC, B), angular gyrus (RAG, C),
511 anterior middle temporal gyrus (RAMTG, D), and precuneus (RPCU, E). These findings are
512 largely consistent with known patterns of activity in the default mode network associated with
513 successful retrieval during episodic memory tasks (Figure 6, B-E). Four regions demonstrated
514 the opposite pattern of BOLD response: left dorsal medial PFC (LDMPFC, F), insular cortex
515 (LINS, G), right intraparietal sulcus (RIPS, H), and inferior frontal gyrus (RIFG, I). These
516 regions have been associated with dorsal and ventral networks which attend to self-referential
517 and saliency processing, respectively, in episodic memory tasks (Figure 6, F-H) as well as
518 attention networks (Figure 6, I).
519

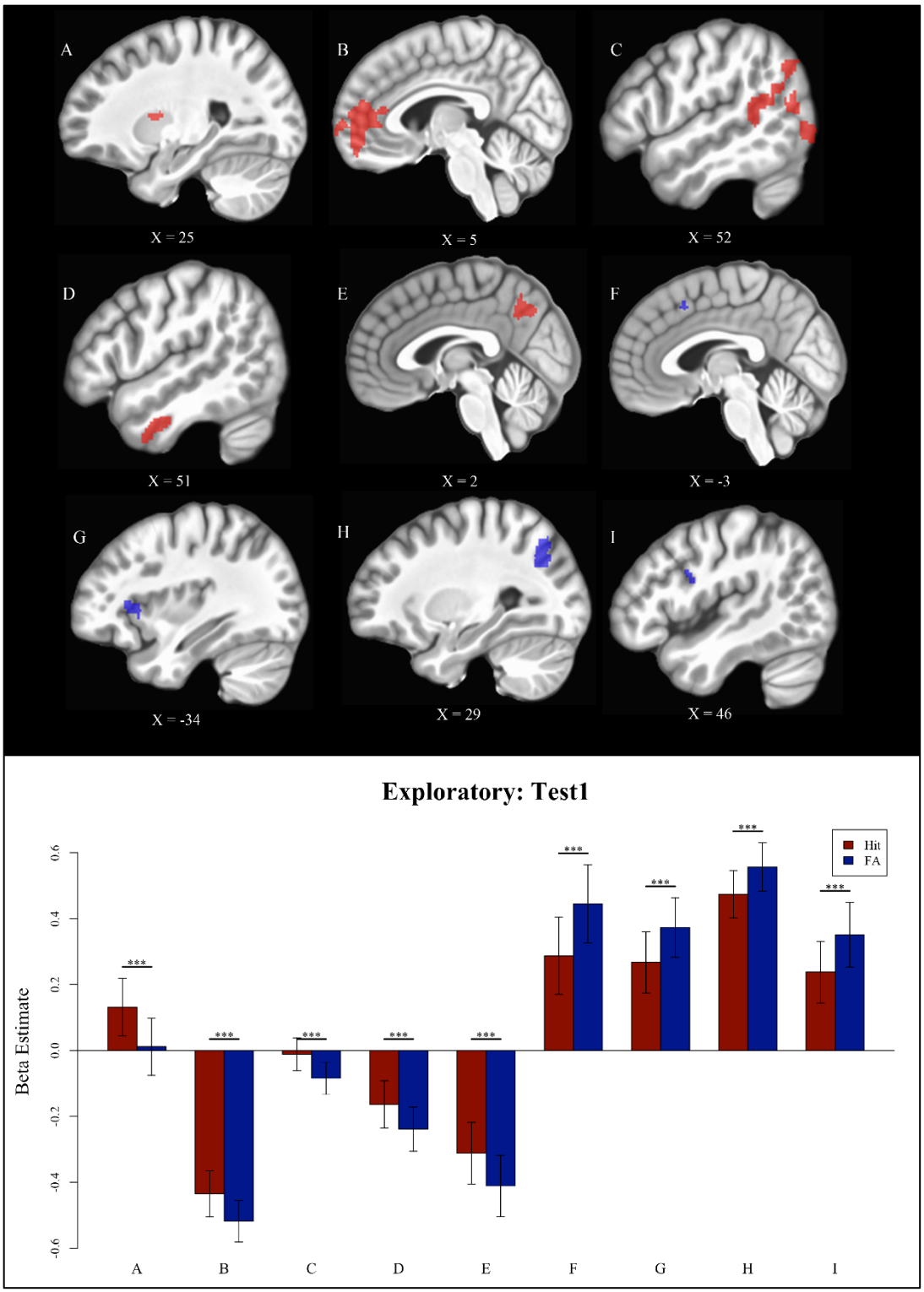

Figure 6. Test1 whole-brain results. Regions which demonstrated differential signal during Test1 Hits versus FAs during an exploratory whole-brain analysis. A right putamen, B right medial PFC, C right angular gyrus, D right anterior middle temporal gyrus, E right precuneus, F left dorsal medial PFC, G left insula, H right intraparietal sulcus, I right inferior frontal gyrus.

3.2.4 Test1 FA responses preceding Test2 responses (T1pT2). Test1 FA responses were coded according to whether or not they would be corrected in Test2 (Hit versus FA), resulting in the same number of trials per bin as in Section 3.2.2. This was done in order to investigate any potential processes occurring during the FA that may allow for subsequent recovery from such interference. Additionally, Test1 CR trials which preceded a Hit were included in the analysis in order to investigate processes which may differ during Test1 Lure trials (FA versus CR) that precede the same Test2 performance (Hit).

A two-factor repeated measures MANOVA failed to detect a significant interaction of hippocampal subregions and beta-coefficients ($F(10, 320) = 0.97$, $\eta_G^2 < .001$, GGε = .69, $p(\mathrm{GG})_{\mathrm{FDR}} = .45$). However, an investigation into NeuroSynth regions associated with retrieval showed a significant interaction ($F(18, 576) = 7.28$, $\eta_G^2 = .008$, GGε = .42, $p(\mathrm{GG})_{\mathrm{FDR}} < .0001$). Subsequent one-factor ANOVA testing revealed that three ROIs were driving this effect (Table 2, T1pT2; Figure 7): the left dorsal medial PFC (LDMPFC), angular gyrus (LAG), and middle frontal gyrus (LMFG). The LDMPFC differed significantly in signal across all behaviors, and of particular interest had stronger signal during a Test1 FA that preceded a Test2 Hit than a Test2 FA (Table 3, T1pT2). The LAG and LMFG did not differ among the Test1 FA responses but showed stronger signal for Test1 CR preceding a Hit than either Test1 FA response. These results suggest that while all three regions are particularly important for Test1 performance in terms of CRs versus FAs, only the LDMPFC had differential activity during Test1 FAs that predicted subsequent Test2 performance.

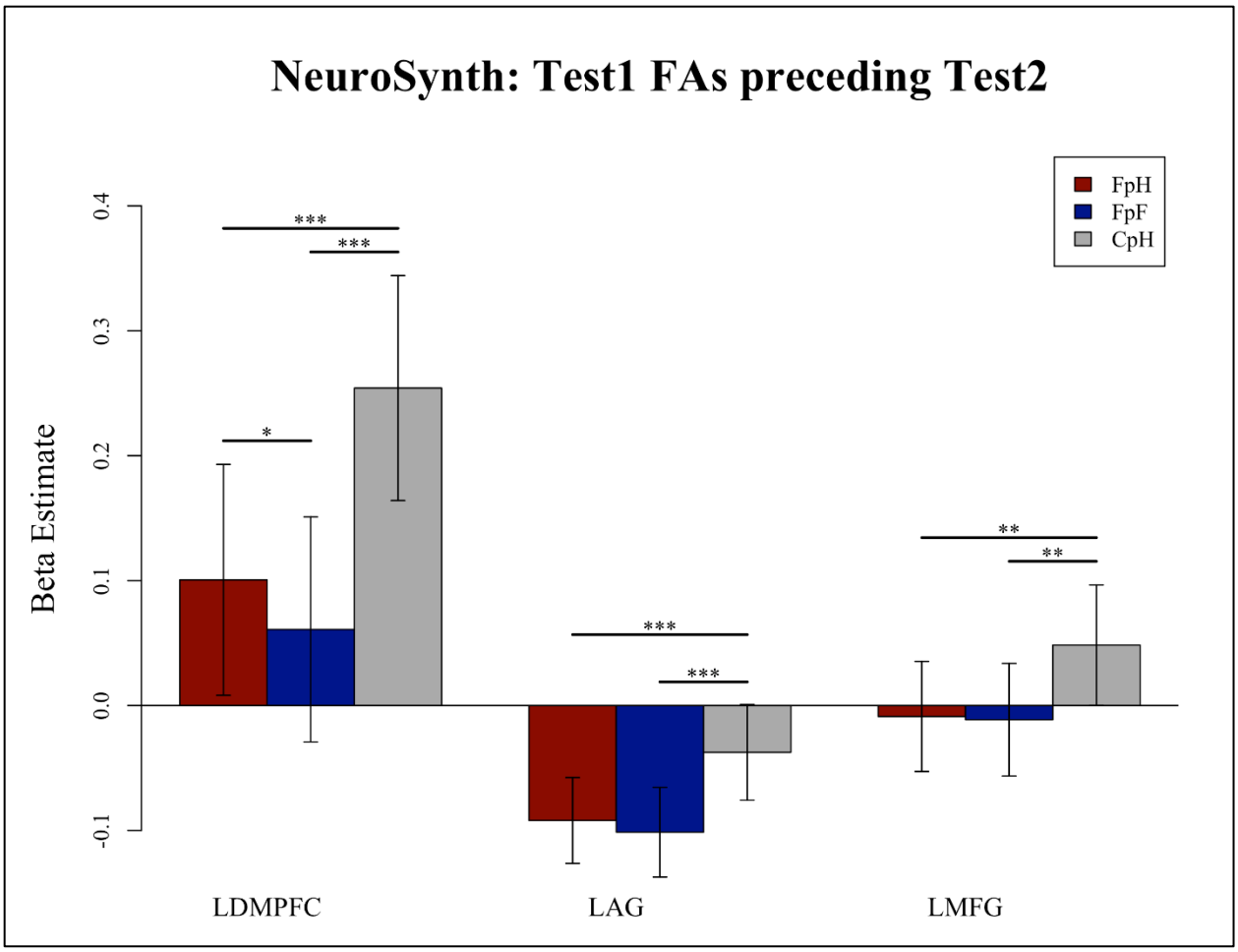

Figure 7. T1pT2 NeuroSynth results. Three regions (left dorsal medial prefrontal cortex [LDMPFC], angular gyrus [LAG], middle frontal gyrus [LMFG]) demonstrated differential activity during Test1 behaviors (FA, CR) preceding Test2 Hits and FAs. FpH = FA preceding Hit, FpF = FA preceding FA, CpH = CR preceding Hit.

Exploratory whole-brain analyses of Test1 FA trials that precede a Test2 Hit versus FA implicated two regions as being differentially engaged (Figure 8): the left insular cortex (LINS) and right supramarginal gryus (RSMG). Both regions had greater BOLD signal preceding a Hit than a FA (Figure 8, Table 3).

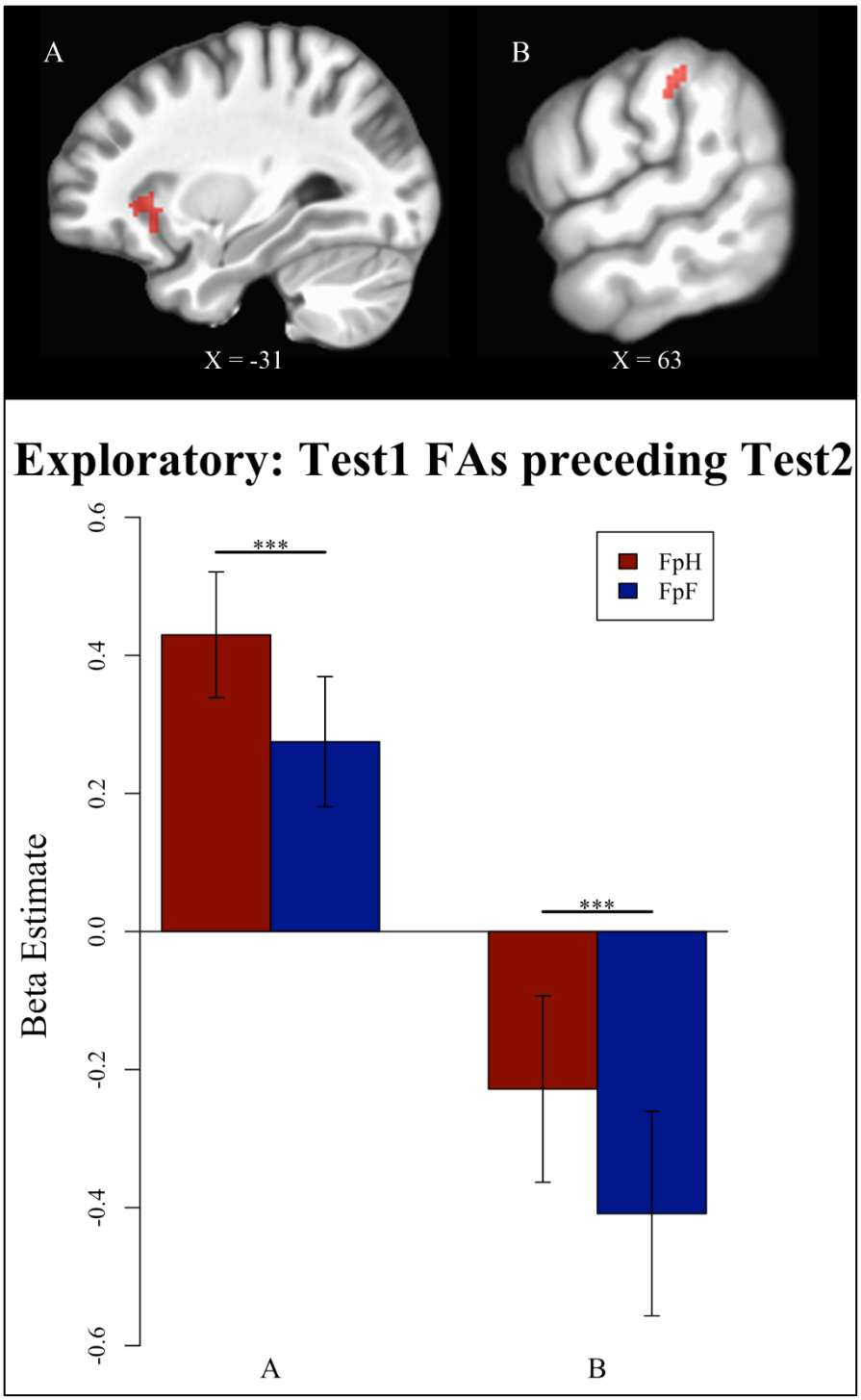

Figure 8. T1pT2 whole-brain results. Clusters which demonstrate differential signal during Test1 FAs that precede a Test2 Hit versus FA. A left insular cortex, B right supramarginal gyrus. FpH = FA preceding Hit, FpF = FA preceding FA.

3.2.5 Test2 responses. A two-factor repeated measures MANOVA failed to detect a significant interaction of hippocampal subregions and Test2 behaviors (Hit, FA; $F(5, 155) =$

561 1.04, η_G^2 < .001, GGε = .75, $p(\mathrm{GG})_{\mathrm{FDR}}$ = .39) or among NeuroSynth ROIs associated with

562 retrieval ($F(9, 279) = .19$, η_G^2 < .001, GGε = .51, $p(\mathrm{GG})_{\mathrm{FDR}}$ = .99). Further, and unexpectedly, an

563 exploratory whole-brain analysis failed to detect any regions which differentially engaged in Hits

564 versus FAs.

565 **3.2.6 Test2 responses following Test1 FAs (T2fT1).** Test2 responses (Hit, FA)

566 following a Test1 FA were coded in order to investigate any potential processes during the

567 second testing phase involved in overcoming interference resulting from a Test1 FA. Further,

568 Test2 Hits following a Test1 CRs were also investigated in order to determine whether any

569 processes were differentially involved in Test2 Hits with respect to whether the Hit was preceded

570 by an incorrect or correct Test1 Lure identification (FA, CR). On average there were 27 Hits

571 following a CR.

572 A two-factor repeated measures MANOVA failed to detect a significant interaction of

573 hippocampal subregions and behavioral outcomes during the behavioral coding described above

574 ($F(10, 310) = 1.1$, η_G^2 = .001, GGε = .53, $p(\mathrm{GG})_{\mathrm{FDR}}$ = .39). An investigation in NeuroSynth

575 masks associated with Retrieval, however, showed a significant interaction in the same analysis

576 ($F(18, 558) = 4.68$, η_G^2 = .004, GGε = .48, $p(\mathrm{GG})_{\mathrm{FDR}}$ < .0001), and subsequent one-way ANOVA

577 analyses indicated that six regions were driving this effect (left angular gyrus, dorsal medial

578 PFC, middle frontal gyrus, precuneus, temporoparietal junction, and right parietooccipital sulcus;

579 Table 3, T2fT1). Greater signal was detected during Test2 in all regions for Hits following a

580 Test1 CR than for Hits following a Test1 FA (Figure 9), while no difference was detected in the

581 same ROIs for Hits versus FAs following a Test1 FA (Table 3, T2fT1). As such, we do not

582 detect any compensatory processes in Test2 in these ROIs associated with overcoming

583 interference resulting from a Test1 FA, as would be implicated by differential signal in a region

during Test2 Hits versus FAs following a Test1 FA. Further, we find the difference between Test2 Hits according to whether they were preceded by a Test1 FA or CR to be interesting; as Test2 presents both the Target and Lure simultaneously, increased activity during a Hit following a CR but not a FA may be indicative of processes associated with selecting between two salient memory representations. Indeed, in Test2 Hits following a Test1 CR, the participant (a) encoded the Target sufficiently to be able to (b) correctly identify the Lure. When both are presented simultaneously and the participant is tasked with identifying the Target, the participant may have to resolve interference resulting from two competing representations. Finally, exploratory whole-brain analyses did not detect any regions that differed significantly between their BOLD signal associated with Test2 Hits versus FAs following a Test1 FA.

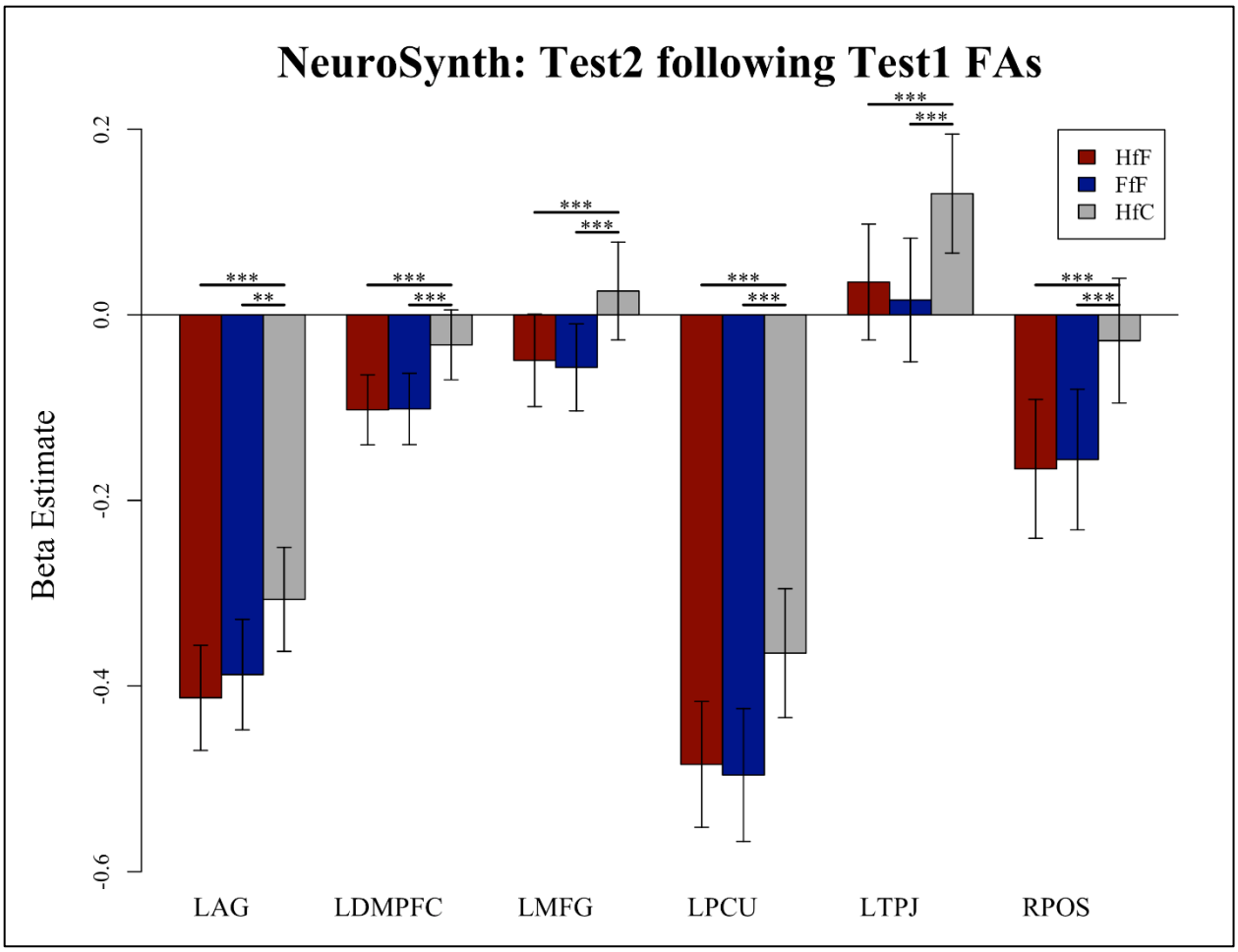

Figure 9. T2fT1 NeuroSynth results. Six regions (left angular gyrus [LAG], dorsal medial prefrontal cortex [LDMPFC], middle frontal gyrus [LMFG], precuneus [LPCU], temporoparietal junction [LTPJ], right parietooccipital sulcus [RPOS]) demonstrated differential activity during Test2 behaviors (Hit, FA) following Test1 responses (FA, CR). HfF = Hit following FA, FfF = FA following FA, HfC = Hit following CR.

4 DISCUSSION

This experiment set out to investigate the impact of a false memory response on the original memory representation, and to determine what, if any, neural processes were associated with recovering the original memory trace. To this end, we employed a Study-Test1-Test2 paradigm within an MRI scanner, in which participants were trained on a series of images (Study), tasked with identifying either Targets or similar Lures presented individually (Test1), and then tasked with identifying the Target when both Targets and Lures were presented

simultaneously (Test2). False Alarm (FA) responses in Test1 were operationalized as false memories and were used to investigate whether participants could correctly identify the Target (Hit) in Test2 following a Test1 FA. Briefly, we hypothesized (a) that a FA would write a memory trace of the Lure without necessarily overwriting the Target's trace such that the two representations would be dissociable, and that (b) the CA3/DG regions of the hippocampus as well as (c) encoding and retrieval nodes and (d) higher-order executive processes would be integral to the formation of a second, dissociable representation.

As the paradigm anticipated, participants made a large number of FA responses in Test1, and proportion testing revealed that participants were equally likely to correct (Hit) and perpetuate (FA) the mistake in Test2 (Figure 1). That is, a FA does not unambiguously overwrite the Target memory trace, as would be indicated by a greater likelihood to perpetuate the FA response in Test2, nor does the FA necessarily write multiple representations from which the Target is dissociable. Given that participants were indeed successful in Test2 (Section 3.1), the behavioral data is ambiguous as to the effect of the FA on the Target trace, and so from the behavioral analyses alone we are unable to reject the null of hypothesis (a). The functional data, however, shed some light on why some FA responses are corrected and others perpetuated.

Functional analyses both replicate and reveal several aspects of memory. First, consistent with literature on subsequent memory and forgetting (Cabeza et al., 2008; H. Kim, 2011; Shrager et al., 2008), seven regions were differentially involved during the encoding phase according to test performance (SpT1). Three regions, the bilateral visual stream and right intraparietal sulcus (Figure 3, A-C), showed greater BOLD signal preceding a Test1 Hit versus a Test1 FA, indicative of increased visual processing and attention, respectively. Conversely, three regions, the bilateral angular gyri and left precuneus (Figure 3, D-F), had the opposite pattern of activity,

631 which is consistent with increased default mode network activity during encoding preceding an
632 erroneous response during test. The final region, the dorsal medial PFC region (Figure 3, G), also
633 demonstrated greater BOLD signal preceding a FA versus Hit and appears to be associated with
634 recognition memory (Phillips et al., 2009; Suzuki et al., 2009); as our stimulus set consists of
635 everyday objects, increased familiarity with certain objects may result in top-down interference
636 thereby resulting in poorer encoding. As such, suppressed signal in this region during encoding
637 trials for stimuli that precede a Hit may be indicative of decreased familiarity with the stimulus
638 which allowed for greater encoding and subsequent successful target identification.

639 Second, in addition to confirming that hippocampal subfields and NeuroSynth retrieval
640 clusters demonstrated differential signal during this retrieval task (Figure 4, Figure 5), multiple
641 regions were detected in an exploratory analysis of the Test1 phase of the experiment that were
642 consistent with the retrieval literature: four regions in the right hemisphere, consisting of the
643 medial PFC, angular gyrus, anterior middle temporal gyrus, and precuneus (Figure 6, B-E), all of
644 which are again consistent with the default mode network, were differentially engaged in the task
645 such that they demonstrated greater BOLD signal during Hits relative to FAs. These findings
646 replicate previous results (H. Kim, 2010; Sestieri et al., 2011) which have demonstrated that the
647 default mode network is involved in self-referential processing, an aspect of which is episodic
648 memory retrieval. Additionally, regions associated with saliency processing and attention
649 networks (left anterior insula, right intraparietal sulcus, right inferior frontal gyrus; Figure 6, G-I)
650 had greater signal during FA responses relative to Hits. It appears that such patterns of activation
651 are the result of the unsuccessful attempt to overcome significant interference resulting from the
652 highly similar Lures used in this task. Indeed, the right intraparietal sulcus is particularly
653 interesting as it is implicated in post-retrieval processes, where hippocampal information is

integrated with executive orchestration (Cabeza et al., 2008; Gilmore et al., 2015; Hao et al., 2018; Sestieri et al., 2011; Uncapher & Wagner, 2009).

Third, and importantly, differential signal was detected during Test1 FAs according to whether or not the FA was perpetuated into Test2 (T1pT2). Three regions, the left dorsomedial PFC, left anterior insula, and right supramarginal gyrus (Table 3, Figure 7, Figure 8), demonstrated significantly greater BOLD signal during Test1 FAs the preceded a Test2 Hit versus a Test2 FA. The dorsomedial PFC region was derived from a NeuroSynth search term of "Memory Retrieval" and is often implicated in the retrieval of episodic memory while the left anterior insular cortex and right supramarginal gyrus are themselves associated with saliency and bottom-up attention networks (Cabeza et al., 2008; Euston et al., 2012; Menon & Uddin, 2010; Seeley et al., 2007; Shirer et al., 2012). These findings, then, suggest that increased activity in these regions during a Test1 FA may be associated with better retrieval of the original memory trace (LDMPFC) and attention to the relevant details of the stimulus (LINS, RSMG) which, while subthreshold for Lure detection, nevertheless led to the formation of a Lure representation such that the Target and Lure memory traces were subsequently dissociable in Test2. Correspondingly, no dissociable pattern of activity was detected during the encoding phase that predicted which Test1 FAs would be corrected in Test2 (SpT1pT2), nor is any such pattern detected in Test2 when comparing Hits to FAs for stimuli following a Test1 FA (T2fT1) as would be expected if such resolution of Target-Lure interference was only a function of the quality of Target encoding (SpT1pT2) or retrieval processes used to dissociate Target and Lure representations (T2fT1). Instead, our data suggest that the capacity to recover the Target trace following a FA is entirely dependent on processes occurring *during* the FA response, and not before or after, where greater attention and retrieval processes result in a dissociable Lure trace.

As such, and with respect to the behavioral data, the rejection of hypothesis (a) may or may not be warranted; it appears that multiple representations *can* occur during a FA but we do not demonstrated that multiple representations *will* occur, as the probability appears to dependent on retrieval and attention processes during Test1. We do, however, reject the null of hypotheses (c) and (d).

Some results of this study were surprising. First, participants were equally likely to Hit or FA in Test2 to stimuli that had previously resulted in a Test1 FA (Figure 1, right), which is to say that their Test2 behavior for these trials was not significantly different from chance. While the d' score of Test2 differed significantly from 0, this was likely driven by the large number of Test2 Hits following a Test1 Hit. Analyses of confidence ratings indicate that participants may have been less confident in the Test2 FA responses (Figure 1, middle); note, though, that while the confidence intervals of the proportion test do not include 0.5, indicating a difference from chance, the FDR-corrected p-value is not significant at the predetermined α. Having confidence intervals that do not include 0.5 gives some indication that participants were not responding with FAs and Hits in Test2 at random. Regardless of the responses at varying confidences, however, an equal likelihood to Hit or FA in Test2 following a Test1 Hit undermines subsequent fMRI analyses in that any differential signal detected may simply be the capitalization on noise and confirmation bias. That said, such a result may very well be a function of the experimental design itself: a large number of difficult Target-Lure pairs were utilized in order to elicit a sufficient number of Test1 FA responses such that dividing the number of Test1 FA responses by subsequent behavior would still have sufficient trials to model the HRF, resulting in a task that lasted 1.5 hours. Such an ambiguous result may very well therefore be the result of participant fatigue in a long, arduous task. It is interesting, however, that the response proportion did not

approach 0, which would be indicative of a greater tendency for the Test1 FA to overwrite the original representation. It is our interpretation, then, that the participants were indeed engaged in the task and not performing at random given the Test2 d' and a decreased confidence in Test2-FA responses. As such, a false alarm may or may not overwrite the original representation, according to processes occurring during the initial false alarm (Section 3.2.4). Somewhat relatedly, another minor limitation is the fact that participants performed better during the Test1 phase than Test2, even though it has been documented that a two-alternative forced-choice task is easier than a Yes/No (Jou et al., 2016). We interpret this fact as evidence of participant fatigue.

A second unexpected result was the lack of findings within the hippocampal subfields in all analyses save those of Test1 (Section 3.2.3). While we demonstrate the differential roles of left CA1 and CA3-CA3-DG in this task with respect to Hits and CRs, in which the hippocampus appeared to be biased towards Target detection in this task, the lack of subfield findings in other analyses of the task was surprising to us give the wealth of literature that has well established the roles of these regions in similar tasks (Bakker et al., 2008; J. Kim & Yassa, 2013; Kirwan & Stark, 2007; Lacy et al., 2011; Rolls & Kesner, 2016). Accordingly, we cannot reject the null of hypothesis (b) as we did not detect differential subregion processing during the FA that predicted subsequent performance. These null results are perhaps a type II or false negative error which resulted from extracting a single parameter estimate for the entire ROI, thereby collapsing across a significant amount of noise as well as signal, possibly washing out the effect. Indeed, it has been noted that the long-axis of the hippocampus contains a functional gradient and discrete domains (Strange et al., 2014), so collapsing across all variance along the anterior-posterior axis of the hippocampal subregions may be inappropriate. Similarly, our contrast of Test1 FA signal coded for subsequent Test2 performance may have simply failed to detect a difference between

723 those two conditions in terms of subregion BOLD signal, but this does not necessarily imply that
724 the hippocampus is not differentially processing the lures such that subsequent behavior results
725 in a Hit versus FA. Rather, the processes involved in such a computation could be rather similar
726 to one another and so a comparison of the mean subfield BOLD signal associated with each
727 process may again be simply too insensitive.

728 Third, we were surprised by the lack of significant differences for the fMRI data of the
729 Test2 phase. The behavioral data indicate that successful Target detection was possible, and that
730 participants were more likely to Hit in Test2 following a Test1 Hit, so performance was not at
731 chance. The lack of findings may not be unwarranted, however, as there are very few studies
732 reporting positive results from two-alternative, forced choice, recognition memory paradigms
733 using fMRI. One notable exception by Olsen and colleagues (2009) is somewhat similar to this
734 experiment in terms of participants and power. Olsen et al. (2009) demonstrated that the medial
735 temporal region was active while maintaining the memory of a set of faces during a delay period
736 prior to a two-alternative forced choice test, an area similar to where we expected activation to
737 occur. The lack of findings in the present experiment may be due to differences between the
738 experiments in fMRI processing methods. One stark difference is in the methods modelling both
739 the noise and auto-correlation function at the group-analysis step employed in this experiment, as
740 this improvement is more conservative than previous random-field based models in terms of
741 false discovery rates but was only recently developed (Cox, 2018, 2019; Eklund et al., 2016).
742 Furthermore, while significant differences were not detected within the fMRI data at the group
743 level, one should not assume that "[i]f the result is not statistically significant, then it proves that
744 no effect or difference exists" (Chen et al., 2017, p. 955). Indeed, as Chen et al. (2017) points

745 out, null results may very well be the result of sub-threshold signal and/or a lack of differential
746 signal between our chosen contrasts at our resolution.

747 In conclusion, our data suggest that a false alarm does not necessarily overwrite the
748 original memory trace of the target, but that subsequent detection of the target is possible,
749 suggesting that a false alarm results in multiple memory representations in some instances.
750 Further, the successful recovery of the target trace following a false alarm appears to be
751 dependent not on the initial encoding trial or on retrieval processes following the false alarm, but
752 on processes which occurred during the false alarm itself. Differential activation was detected in
753 the left dorsal medial prefrontal cortex, left insular cortex, and right supramarginal gyrus during
754 the false alarm that predicted whether or not the target would subsequently be retrieved, regions
755 which are associated with episodic memory, saliency, and bottom-up attention, respectively.

References

Avants, B. B., Epstein, C. L., Grossman, M., & Gee, J. C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*(1), 26–41.

Avants, B. B., Tustison, N. J., Song, G., Cook, P. A., Klein, A., & Gee, J. C. (2011). A reproducible evaluation of ANTs similarity metric performance in brain image registration. *NeuroImage*, *54*(3), 2033–2044.

Bakker, A., Kirwan, C. B., Miller, M., & Stark, C. E. (2008). Pattern separation in the human hippocampal CA3 and dentate gyrus. *Science*, *319*. https://doi.org/10.1126/science.1152882

Cabeza, R. (2013). Introduction to the special issue on functional neuroimaging of episodic memory. *Neuropsychologia*, *51*(12), 2319–2321.

Cabeza, R., Ciaramelli, E., Olson, I. R., & Moscovitch, M. (2008). The parietal cortex and episodic memory: An attentional account. *Nature Reviews Neuroscience*, *9*(8), 613.

Chen, G., Taylor, P. A., & Cox, R. W. (2017). Is the statistic value all we should care about in neuroimaging? *NeuroImage*, *147*, 952–959. https://doi.org/10.1016/j.neuroimage.2016.09.066

Cox, R. W. (2012). AFNI: What a long strange trip it's been. *NeuroImage*, *62*(2), 743–747.

Cox, R. W. (2018). Equitable thresholding and clustering. *BioRxiv*, 295931.

Cox, R. W. (2019). Equitable thresholding and clustering (ETAC): A novel method for fMRI clustering in AFNI. *Brain Connectivity*. https://doi.org/10.1089/brain.2019.0666

Doxey, C. R., & Kirwan, C. B. (2015). Structural and functional correlates of behavioral pattern separation in the hippocampus and medial temporal lobe. *Hippocampus*, *25*(4), 524–533.

Eklund, A., Nichols, T. E., & Knutsson, H. (2016). Cluster failure: Why fMRI inferences for spatial extent have inflated false-positive rates. *Proceedings of the National Academy of Sciences*, *113*(28), 7900–7905. https://doi.org/10.1073/pnas.1602413113

Euston, D. R., Gruber, A. J., & McNaughton, B. L. (2012). The role of medial prefrontal cortex in memory and decision making. *Neuron*, *76*(6), 1057–1070. https://doi.org/10.1016/j.neuron.2012.12.002

Gilbert, P. E., & Kesner, R. P. (2006). The role of the dorsal CA3 hippocampal subregion in spatial working memory and pattern separation. *Behavioural Brain Research*, *169*(1), 142–149.

Gilbert, P. E., Kesner, R. P., & DeCoteau, W. E. (1998). Memory for spatial location: Role of the hippocampus in mediating spatial pattern separation. *Journal of Neuroscience*, *18*(2), 804–810.

Gilbert, P. E., Kesner, R. P., & Lee, I. (2001). Dissociating hippocampal subregions: A double dissociation between dentate gyrus and CA1. *Hippocampus*, *11*(6), 626–636.

Gilmore, A. W., Nelson, S. M., & McDermott, K. B. (2015). A parietal memory network revealed by multiple MRI methods. *Trends in Cognitive Sciences*, *19*(9), 534–543. https://doi.org/10.1016/j.tics.2015.07.004

Goebel, P. M., & Vincze, M. (2007). Implicit modeling of object topology with guidance from temporal view attention. *Proceedings of the International Cognitive Vision Workshop, ICVW*.

Gold, A. E., & Kesner, R. P. (2005). The role of the CA3 subregion of the dorsal hippocampus in spatial pattern completion in the rat. *Hippocampus*, *15*(6), 808–814.

Grady, C. L., & Ryan, J. D. (2017). Age-related differences in the human hippocampus: Behavioral, structural and functional measures. In *The Hippocampus from Cells to Systems* (pp. 167–208). Springer.

Hanczakowski, M., & Mazzoni, G. (2011). Both differences in encoding processes and monitoring at retrieval reduce false alarms when distinctive information is studied. *Memory*, *19*(3), 280–289.

Hao, X., Wang, X., Song, Y., Kong, X., & Liu, J. (2018). Dual roles of the hippocampus and intraparietal sulcus in network integration and segregation support scene recognition. *Brain Structure & Function*, *223*(3), 1473–1485. https://doi.org/10.1007/s00429-017-1564-2

Hunsaker, M. R., & Kesner, R. P. (2013). The operation of pattern separation and pattern completion processes associated with different attributes or domains of memory. *Neuroscience & Biobehavioral Reviews*, *37*(1), 36–58.

Hunsaker, M. R., Rosenberg, J. S., & Kesner, R. P. (2008). The role of the dentate gyrus, CA3a, b, and CA3c for detecting spatial and environmental novelty. *Hippocampus*, *18*(10), 1064–1073.

Jou, J., Flores, S., Cortes, H. M., & Leka, B. G. (2016). The effects of weak versus strong relational judgments on response bias in two-alternative-forced-choice recognition: Is the test criterion-free? *Acta Psychologica*, *167*, 30–44. https://doi.org/10.1016/j.actpsy.2016.03.014

Kesner, R. P. (2007a). A behavioral analysis of dentate gyrus function. *Progress in Brain Research*, *163*, 567–576.

Kesner, R. P. (2007b). Behavioral functions of the CA3 subregion of the hippocampus. *Learning & Memory*, *14*(11), 771–781.

Kim, H. (2010). Dissociating the roles of the default-mode, dorsal, and ventral networks in episodic memory retrieval. *NeuroImage*, *50*(4), 1648–1657. https://doi.org/10.1016/j.neuroimage.2010.01.051

Kim, H. (2011). Neural activity that predicts subsequent memory and forgetting: A meta-analysis of 74 fMRI studies. *NeuroImage*, *54*(3), 2446–2461.

Kim, J., & Yassa, M. A. (2013). Assessing recollection and familiarity of similar lures in a behavioral pattern separation task. *Hippocampus*, *23*(4), 287–294.

Kirwan, C. B., Hartshorn, A., Stark, S. M., Goodrich-Hunsaker, N. J., Hopkins, R. O., & Stark, C. E. (2012). Pattern separation deficits following damage to the hippocampus. *Neuropsychologia*, *50*(10), 2408–2414.

Kirwan, C. B., & Stark, C. E. (2007). Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learn Mem*, *14*. https://doi.org/10.1101/lm.663507

Kubota, Y., Toichi, M., Shimizu, M., Mason, R. A., Findling, R. L., Yamamoto, K., & Calabrese, J. R. (2006). Prefrontal hemodynamic activity predicts false memory—A near-infrared spectroscopy study. *Neuroimage*, *31*(4), 1783–1789.

Lacy, J. W., Yassa, M. A., Stark, S. M., Muftuler, L. T., & Stark, C. E. (2011). Distinct pattern separation related transfer functions in human CA3/dentate and CA1 revealed using high-resolution fMRI and variable mnemonic similarity. *Learning & Memory*, *18*(1), 15–18.

Lancaster, J. L., Laird, A. R., Eickhoff, S. B., Martinez, M. J., Fox, P. M., & Fox, P. T. (2012). Automated regional behavioral analysis for human brain images. *Frontiers in Neuroinformatics*, *6*, 23.

Leal, S. L., & Yassa, M. A. (2014). Effects of aging on mnemonic discrimination of emotional information. *Behavioral Neuroscience*, *128*(5), 539.

Leutgeb, J. K., Leutgeb, S., Moser, M.-B., & Moser, E. I. (2007). Pattern separation in the dentate gyrus and CA3 of the hippocampus. *Science*, *315*(5814), 961–966.

Li, X., Morgan, P. S., Ashburner, J., Smith, J., & Rorden, C. (2016). The first step for neuroimaging data analysis: DICOM to NIfTI conversion. *Journal of Neuroscience Methods*, *264*, 47–56.

Marshall, A. C., Cooper, N. R., & Geeraert, N. (2016). The impact of experienced stress on aged spatial discrimination: Cortical overreliance as a result of hippocampal impairment. *Hippocampus*, *26*(3), 329–340.

Menon, V., & Uddin, L. Q. (2010). Saliency, switching, attention and control: A network model of insula function. *Brain Structure and Function*, *214*(5–6), 655–667.

Moritz, S., Gläscher, J., Sommer, T., Büchel, C., & Braus, D. F. (2006). Neural correlates of memory confidence. *Neuroimage*, *33*(4), 1188–1193.

Mueller, S. G., Chao, L. L., Berman, B., & Weiner, M. W. (2011). Evidence for functional specialization of hippocampal subfields detected by MR subfield volumetry on high resolution images at 4T. *NeuroImage*, *56*(3), 851–857.

Okado, Y., & Stark, C. E. (2005). Neural activity during encoding predicts false memories created by misinformation. *Learning & Memory*, *12*(1), 3–11.

Olsen, R. K., Nichols, E. A., Chen, J., Hunt, J. F., Glover, G. H., Gabrieli, J. D. E., & Wagner, A. D. (2009). Performance-related sustained and anticipatory activity in human medial temporal lobe during delayed match-to-sample. *Journal of Neuroscience*, *29*(38), 11880–11890. https://doi.org/10.1523/JNEUROSCI.2245-09.2009

Philippi, C. L., Tranel, D., Duff, M., & Rudrauf, D. (2015). Damage to the default mode network disrupts autobiographical memory retrieval. *Social Cognitive and Affective Neuroscience*, *10*(3), 318–326. https://doi.org/10.1093/scan/nsu070

Phillips, J. S., Velanova, K., Wolk, D. A., & Wheeler, M. E. (2009). Left posterior parietal cortex participates in both task preparation and episodic retrieval. *NeuroImage*, *46*(4), 1209–1221. https://doi.org/10.1016/j.neuroimage.2009.02.044

Pidgeon, L. M., & Morcom, A. M. (2016). Cortical pattern separation and item-specific memory encoding. *Neuropsychologia*, *85*, 256–271.

Reagh, Z. M., Ho, H. D., Leal, S. L., Noche, J. A., Chun, A., Murray, E. A., & Yassa, M. A. (2016). Greater loss of object than spatial mnemonic discrimination in aged adults. *Hippocampus*, *26*(4), 417–422.

Reagh, Z. M., Roberts, J. M., Ly, M., DiProspero, N., Murray, E., & Yassa, M. A. (2014). Spatial discrimination deficits as a function of mnemonic interference in aged adults with and without memory impairment. *Hippocampus*, *24*(3), 303–314.

Reagh, Z. M., & Yassa, M. A. (2014). Object and spatial mnemonic interference differentially engage lateral and medial entorhinal cortex in humans. *Proceedings of the National Academy of Sciences*, *111*(40), E4264–E4273.

Roberts, J. M., Ly, M., Murray, E., & Yassa, M. A. (2014). Temporal discrimination deficits as a function of lag interference in older adults. *Hippocampus*, *24*(10), 1189–1196.

Rolls, E. T., & Kesner, R. P. (2016). *Pattern separation and pattern completion in the hippocampal system. Introduction to the Special Issue*. Elsevier.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., Reiss, A. L., & Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, *27*(9), 2349–2356. https://doi.org/10.1523/JNEUROSCI.5587-06.2007

Sestieri, C., Corbetta, M., Romani, G. L., & Shulman, G. L. (2011). Episodic memory retrieval, parietal cortex, and the default mode network: Functional and topographic analyses. *Journal of Neuroscience*, *31*(12), 4407–4420. https://doi.org/10.1523/JNEUROSCI.3335-10.2011

Sheppard, D. P., Graves, L. V., Holden, H. M., Delano-Wood, L., Bondi, M. W., & Gilbert, P. E. (2016). Spatial pattern separation differences in older adult carriers and non-carriers for the apolipoprotein E epsilon 4 allele. *Neurobiology of Learning and Memory*, *129*, 113–119.

Shirer, W. R., Ryali, S., Rykhlevskaia, E., Menon, V., & Greicius, M. D. (2012). Decoding subject-driven cognitive states with whole-brain connectivity patterns. *Cerebral Cortex*, *22*(1), 158–165. https://doi.org/10.1093/cercor/bhr099

Shrager, Y., Kirwan, C. B., & Squire, L. R. (2008). Activity in both hippocampus and perirhinal cortex predicts the memory strength of subsequently remembered information. *Neuron*, *59*(4), 547–553.

Small, S. A., Schobel, S. A., Buxton, R. B., Witter, M. P., & Barnes, C. A. (2011). A pathophysiological framework of hippocampal dysfunction in ageing and disease. *Nature Reviews Neuroscience*, *12*(10), 585–601.

Steemers, B., Vicente-Grabovetsky, A., Barry, C., Smulders, P., Schröder, T. N., Burgess, N., & Doeller, C. F. (2016). Hippocampal attractor dynamics predict memory-based decision making. *Current Biology*, *26*(13), 1750–1757.

Strange, B. A., Witter, M. P., Lein, E. S., & Moser, E. I. (2014). Functional organization of the hippocampal longitudinal axis. *Nature Reviews. Neuroscience*, *15*(10), 655–669. https://doi.org/10.1038/nrn3785

Suzuki, C., Tsukiura, T., Mochizuki-Kawai, H., Shigemune, Y., & Iijima, T. (2009). Prefrontal and medial temporal contributions to episodic memory-based reasoning. *Neuroscience Research*, *63*(3), 177–183. https://doi.org/10.1016/j.neures.2008.11.010

Tustison, N. J., Cook, P. A., Klein, A., Song, G., Das, S. R., Duda, J. T., Kandel, B. M., van Strien, N., Stone, J. R., Gee, J. C., & Avants, B. B. (2014). Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. *NeuroImage*, *99*(0), 166–179. http://dx.doi.org/10.1016/j.neuroimage.2014.05.044

Uncapher, M. R., & Wagner, A. D. (2009). Posterior parietal cortex and episodic encoding: Insights from fMRI subsequent memory effects and dual-attention theory. *Neurobiology of Learning and Memory*, *91*(2), 139–154.

van den Honert, R. N., McCarthy, G., & Johnson, M. K. (2016). Reactivation during encoding supports the later discrimination of similar episodic memories. *Hippocampus*, *26*(9), 1168–1178.

Wais, P. E., Jahanikia, S., Steiner, D., Stark, C. E. L., & Gazzaley, A. (2017). Retrieval of high-fidelity memory arises from distributed cortical networks. *NeuroImage*, *149*, 178–189. https://doi.org/10.1016/j.neuroimage.2017.01.062

Wang, H., Suh, J. W., Das, S. R., Pluta, J., Craige, C., & Yushkevich, P. A. (2013). Multi-atlas segmentation with joint label fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(3), 611–623. https://doi.org/10.1109/TPAMI.2012.143

Wickelgren, W. A. (1968). Unidimensional strength theory and component analysis of noise in absolute and comparative judgments. *Journal of Mathematical Psychology*, *5*(1), 102–122.

Wickens, T. D. (2002). *Elementary signal detection theory*. Oxford University Press, USA.

Yushkevich, P. A., Piven, J., Hazlett, H. C., Smith, R. G., Ho, S., Gee, J. C., & Gerig, G. (2006). User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage*, *31*(3), 1116–1128.

Yushkevich, P. A., Pluta, J. B., Wang, H., Xie, L., Ding, S.-L., Gertje, E. C., Mancuso, L., Kliot, D., Das, S. R., & Wolk, D. A. (2015). Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. *Human Brain Mapping*, *36*(1), 258–287. https://doi.org/10.1002/hbm.22627