Age differences in retrieval-related reinstatement reflect age-related dedifferentiation at encoding

Paul F. Hill[a,b]
Danielle R. King[a,b]
Michael D. Rugg[a,b,c,d]

[a]Center for Vital Longevity, University of Texas at Dallas
1600 Viceroy Dr. #800
Dallas, TX 75235

[b]School of Behavioral and Brain Sciences, University of Texas at Dallas
800 W Campbell Rd
Richardson, TX 75080

[c]Department of Psychiatry, University of Texas Southwestern Medical Center
6363 Forest Park Rd 7th floor suite 749
Dallas TX 75235

[d]School of Psychology, University of East Anglia
Norwich NR4 7TJ, UK

Correspondence:
Paul Hill
1600 Viceroy Dr. #800
Dallas, TX 75235
paul.hill@utdallas.edu
972-883-3771

Abstract

Age-related reductions in neural selectivity have been linked to cognitive decline. We examined whether age differences in the strength of retrieval-related cortical reinstatement could be explained by analogous differences in neural selectivity at encoding, and whether reinstatement was associated with memory performance in an age-dependent or an age-independent manner. Young and older adults underwent fMRI as they encoded words paired with images of faces or scenes. During a subsequent scanned memory test participants judged whether test words were studied or unstudied and, for words judged studied, also made a source memory judgment about the associated image category. Using multi-voxel pattern similarity analyses, we identified robust evidence for reduced scene reinstatement in older relative to younger adults. This decline was however largely explained by age differences in neural differentiation at encoding; moreover, a similar relationship between neural selectivity at encoding and retrieval was evident in young participants. The results suggest that, regardless of age, the selectivity with which events are neurally processed at the time of encoding can determine the strength of retrieval-related cortical reinstatement.

Keywords: cognitive aging, cortical reinstatement, episodic memory, neural dedifferentiation, pattern similarity

The ability to recollect information about a past event declines with advancing age (Koen & Yonelinas, 2014; Nilsson, 2003; Old & Naveh-Benjamin, 2008; Park et al., 1996). Here, we examined whether age-related differences in recollection are associated with a tendency for older adults to retrieve less detailed or differentiated information about a prior experience than younger individuals. To address this question, we used functional magnetic resonance imaging (fMRI) to examine age-related differences in retrieval-related *cortical reinstatement* effects. Cortical reinstatement refers to the finding that successful recollection is associated with patterns of cortical activity that partially overlap the patterns elicited when the recollected information was initially experienced (e.g. Johnson & Rugg, 2007; Nyberg et al., 2000; Wheeler et al., 2000; for reviews see Danker & Anderson, 2010; Rissman & Wagner, 2012; Rugg et al., 2015; Xue, 2018). These findings have led to the widely held view that reinstated cortical activity reflects retrieved content (Kuhl & Chun, 2014; Rugg et al., 2015; Xue, 2018). From this perspective, if there is a tendency to retrieve less detailed or lower fidelity information with increasing age, cortical reinstatement effects should be weaker in older adults relative to younger individuals.

Prior work examining age differences in cortical reinstatement have reported mixed results, with some studies reporting reduced reinstatement in older adults (Abdulrahman et al., 2017; Bowman et al., 2019; Folville et al., 2020; McDonough et al., 2014; St-Laurent et al., 2014; 2019) and others reporting null effects of age (Thakral et al., 2017; Wang et al., 2016). However, and of relevance to the present experiment, the analytical approaches employed in the two studies that reported null age effects for reinstatement (Thakral et al., 2017; Wang et al., 2016) characterized reinstatement at the categorical rather than the item level (that is, in terms of patterns of neural activity common to an entire class of encoding events). Thus, these studies leave open the possibility that the fidelity with which finer-grained, item-level information is

reinstated differs between young and older adults (see Folville et al., 2020; St-Laurent et al., 2014; 2019).

Of relevance to the question of age differences in retrieval-related reinstatement, it has consistently been reported that patterns of neural activity elicited by some stimulus categories become less selective in older age. This phenomenon - *age-related neural dedifferentiation* - has been conjectured to play a role in age-related cognitive decline (Koen & Rugg, 2019; Li et al., 2001; Li & Rieckmann, 2014). Notably, lower selectivity of neural responses to perceptual events has been reported to predict poorer subsequent memory for the events in both young and older adults (Berron et al., 2018; Koen et al., 2019; Maass et al., 2019). These findings raise the possibility that the precision or selectivity with which the perceptual features of an event are neurally represented at the time of encoding (that is, its level of neural differentiation) are a determinant of the strength with which the features are reinstated at retrieval.

In apparent contradiction to this possibility, St-Laurent and colleagues (2014) reported that while the viewing of multimodal video clips was associated with only minimal evidence for age-related dedifferentiation, there were pronounced age differences in cortical reinstatement effects as participants mentally 'replayed' the videos from memory. These findings were however based on analyses of data that had been pooled across numerous repeated viewings and retrieval attempts, raising the possibility that measures of neural selectivity in the young and older adults were differentially influenced by repetition. Indeed, further analysis of the same data-set revealed that neural differentiation of the video clips was greatest on their initial presentation, and declined with subsequent viewings (St-Laurent & Buchsbaum, 2019). The authors did not report, however, whether neural differentiation during the initial viewing was associated with the strength of neural reinstatement. In a study by Abdulrahman et al. (2017), the

authors were also unable to identify any evidence for age-related neural differentiation (operationalized as the accuracy of an MVPA classifier) during the encoding of blocks of words subjected to phonological vs. semantic processing (see also Dennis et al., 2019). Robust evidence of age-related reductions in reinstatement was nonetheless obtained at retrieval. However, in the retrieval phase of this experiment test items were blocked according to their study context, precluding the analysis of retrieval-related activity according to the accuracy of the associated memory judgment. Thus, it is unclear whether the reported age differences in reinstatement reflected differences in neural correlates of pre- or post-retrieval processing; by definition, reinstatement effects are a reflection only of the latter class of processes.

It has been proposed that cortical reinstatement depends critically on the hippocampus. According to several neurobiological models of memory retrieval (e.g. Alvarez & Squire, 1994; Norman & O'Reilly, 2003; Rugg et al., 2008), patterns of brain activity engendered during the original experience of an event are stored as a sparse hippocampal representation. At retrieval, reactivation of an encoded memory trace through the mechanism of hippocampal pattern completion leads to reinstatement of the original patterns of brain activity. Consistent with these theoretical accounts, fMRI studies in younger adults have demonstrated that trial-wise fluctuations in hippocampal activity co-vary with reinstatement strength, and that both of these variables predict within-subject variability in memory performance (Gordon et al., 2014; Ritchey et al., 2013; Staresina et al., 2012; Wing et al., 2015). Despite accumulating evidence from younger adults, however, the relationships between hippocampal activity, reinstatement strength, and memory performance in older adults have received relatively little attention (Trelle et al., 2020).

In the present study, healthy young and older adults undertook a study task, and subsequently a test of source memory, as they underwent fMRI. During study, words paired with images of faces or scenes were elaboratively encoded to encourage the formation of item-context associations. Face and scene images were selected as associated stimuli as they have been previously reported to give rise to robust age-related neural dedifferentiation effects (e.g., Koen et al., 2019; Park et al., 2004; Trelle et al., 2019; Voss et al., 2008; Zheng et al., 2018). Studied and unstudied words were subsequently presented in the scanned memory test in which, for words judged as studied, a source memory judgment for the corresponding image category was required. The primary reinstatement analyses focused on multivariate pattern similarity analyses (PSA) in a voxel set that demonstrated standard univariate reinstatement effects (Johnson & Rugg, 2007; for a similar approach to MVPA feature selection, see Wang et al., 2016). PSA was employed to quantify the strength of category- and item-level reinstatement within the voxel set. In a series of secondary analyses, also employing PSA, we supplemented this 'top-down' feature selection strategy with an exploratory bottom-up approach (see Methods).

We addressed three principal questions: (1) whether scene and face reinstatement effects associated with successful source memory judgments for the respective stimulus classes differed according to age (cf. Abdulrahman et al., 2017 vs. Wang et al., 2016), (2) whether age differentially impacts reinstatement of category- vs. item-level information, and (3) whether age differences in retrieval-related reinstatement could be accounted for by analogous differences in neural differentiation at the time of encoding. In further trial-wise analyses, we built on prior findings (Gordon et al., 2014; Ritchey et al., 2013) to ask whether strength of retrieval-related reinstatement co-varied within-participants with retrieval-related hippocampal activity, whether

either or both of these variables co-varied with memory performance, and whether any such relationships differed according to age.

Materials and Methods

The study phase data from this experiment, albeit subjected to a quite different analytic approach, were the topic of a prior report (Srokova et al., 2020). The present descriptions of the experimental design and procedures, the participant samples, and the neuropsychological and behavioral data, overlap heavily with the descriptions given in that report and are provided here for the convenience of the reader, and not because they convey new information. The fMRI findings described below have not been reported previously.

Participants.

Participants were 27 younger and 33 older adult volunteers recruited from the University of Texas at Dallas and surrounding community. All participants were right-handed and fluent English speakers before the age of five. No participants had a history of neurological or psychiatric disease or reported taking any prescription medications affecting the central nervous system. All participants gave informed consent in accordance with the UT Dallas and University of Texas Southwestern Institutional Review Boards and were compensated at the rate of $30 an hour.

Data from three younger and three older adult participants were excluded from subsequent analyses for the following reasons: voluntary withdrawal (n=2), behavioral performance resulting in memory bins with too few trials (n=2), technical malfunction (n=1), and an incidental MRI abnormality (n=1). An additional six older adult participants were excluded due to near-chance source memory performance according to a pre-determined cut-off score (pSR < .1). Data from the remaining 24 younger (18-28 years, M = 22.4 years, SD = 3.2 yrs; 15

females) and 24 older (65-75 years, *M* = 70.1 yrs, *SD* = 3.4 yrs; 14 females) participants were used in the analyses reported here.

Neuropsychological Testing.

All participants completed a neuropsychological test battery consisting of the Mini-Mental State Examination (MMSE), the California Verbal Learning Test-II (CVLT; Delis et al., 2000), Wechsler Logical Memory Tests 1 and 2 (Wechsler, 2009), Trail Making tests A and B (Reitan & Wolfson, 1985), the Symbol Digit Modalities test (SDMT; Smith, 1982), the F-A-S subtest of the Neurosensory Center Comprehensive Evaluation for Aphasia (Spreen & Benton, 1977), the Wechsler Adult Intelligence Scale–Revised subtests of forward and backward digit span (Wechsler, 1981), category fluency (Benton, 1968), Raven's Progressive Matrices (List 1, Raven et al., 2000) and the Wechsler Test of Adult Reading (WTAR; Wechsler, 1981). Participants were excluded from entry into the study if they scored < 27 on the MMSE, < 1.5 SDs below age-appropriate norms on any memory test, < 1.5 SDs below age norms on any two other tests, or if their estimated full-scale IQ was less than 100. These criteria were employed to minimize the likelihood of including individuals with mild cognitive impairment. Results from the neuropsychological test battery are presented in Supplementary Table 1.

Experimental Materials and Procedure.

The critical experimental stimuli consisted of 288 concrete words and 96 color images depicting faces (50% male) and a further 96 images depicting scenes (50% urban, 50% rural). An additional 68 words and 40 images were used as filler or practice stimuli.

Separate stimulus sets were created for yoked younger-older adult pairs. For the study phase, each stimulus set comprised 192 randomly selected critical image-word pairs interspersed with 96 null trials (white fixation cross). Study sets were divided evenly into four sub-lists (one

per scanning session) of 48 critical and 24 null trials. Stimulus sets for the test phase comprised the 192 'old' words encountered during the study task interspersed with 96 unstudied 'new' words and 96 null trials. Test stimuli were subdivided evenly into four sub-lists (one per scanning session), each comprising 48 'old' items, 24 'new' items, and 24 null trials. Trial order for both tasks was pseudo-randomized under the constraint that no more than three critical trials from the same encoding category (faces, scenes) or two null trials occurred consecutively.

The experiment consisted of two study-test cycles completed inside the scanner (Figure 1A). Each study-test cycle consisted of two consecutive scanning runs of the study phase (~8 minutes each) followed by two consecutive runs of the test phase (~10 minutes each). A brief break separated the study and test phases as participants were reminded of the task instructions while remaining in the scanner. The sequence and timing for each trial was identical in the study and test phases. Each trial commenced with a red fixation cross (500 ms) followed by the presentation of the study (word-image pair) or test (word only) items for 2000 ms. Each item was followed by a white fixation cross displayed for an additional 2000 ms, resulting in a stimulus-onset-asynchrony (excluding null trials) of 4.5 s. For both tasks, participants were instructed to make their response(s) within the 4000 ms period during which the stimulus items and then the fixation cross were presented on screen. A 30 s rest break occurred midway through each scan session. Two filler trials were presented at the beginning of each scan session and after each rest break. Instructions and practice for the study and test tasks were given outside of the scanner.

For each study trial, participants viewed a concrete noun paired with an image of a face or a scene. All face and scene images were scaled and cropped to 256 x 256 pixels. On face trials, participants were instructed to imagine the person depicted in the image interacting with or using the object denoted by the word. On scene trials, participants were instructed to imagine the

object denoted by the word interacting or moving about within the depicted scene. To encourage compliance with the instructions, participants rated the vividness of the imagined scenario on a 3-point scale ('Not Very Vivid', 'Somewhat Vivid', 'Very Vivid'). Responses were made on a scanner-compatible button box with the index, middle, and ring fingers of the right hand. Responses made within 500-4500 ms of the presentation of the study items were included in the analyses. Trials associated with responses falling outside this temporal window were discarded and included as events of no interest.

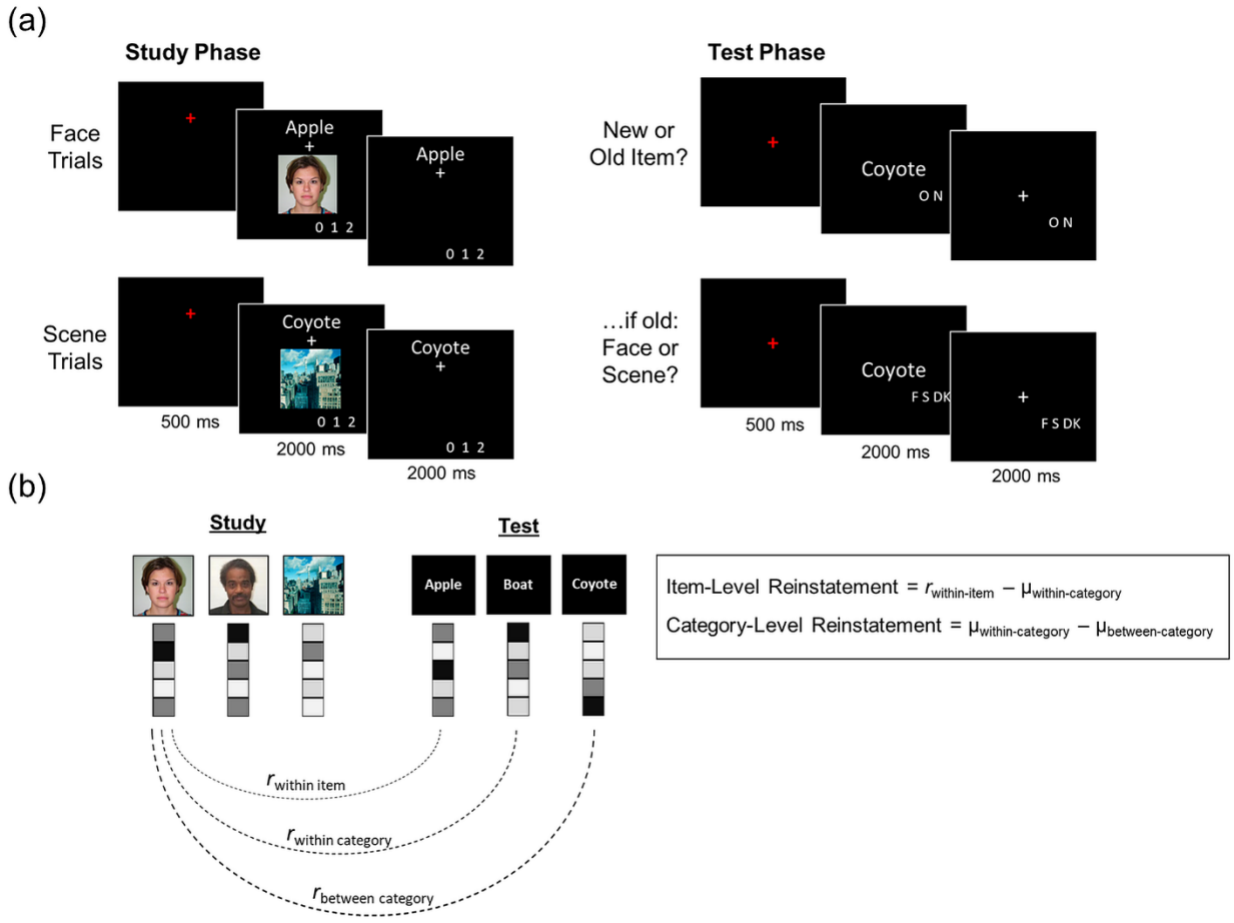

Figure 1. (a) Structure and timing of the study and test phases. (b) Item- and category-level PSA schematic. Item-level similarity was computed as the difference between the mean correlation between study-test pairs involving the same experimental items and the mean correlation between study-test pairs belonging to the same category. Category-level similarity was computed as the difference between the mean correlation between study-

test pairs belonging to the same stimulus categories and study-test pairs belonging to the opposite stimulus categories.

During the test phase, participants viewed old and new words presented one at a time. Each test item required either one or two memory judgments, both of which were made within the single 4000 ms interval occupied by the test item and subsequent white fixation cross. Participants were required to first judge whether a test item had been encountered during the study phase by making an "Old" or "New" response. They were instructed to refrain from guessing and to respond "Old" only when confident that a test item had been previously studied. Test items receiving a 'New' response were followed by the white fixation cross for the remainder of the trial. For items attracting an "Old" response, participants were required to make a source memory judgment about the category of the test word's studied associate (i.e., was the word studied with a face or a scene?). A "Don't Know" response option was also available to discourage guessing. Responses were made on a scanner compatible button box with the index, middle and ring fingers of the right hand. Old/new recognition judgments were always made with the index and middle fingers and counterbalanced across participants. Source memory responses were also fully counterbalanced across participants with the constraint that the "Don't Know" response was never assigned to the middle finger. As with the study phase, trials associated with responses made between 500ms and 4500 ms following the presentation of the test items were included in the analyses. Trials falling outside this time window were discarded and included as events of no interest in the design matrix.

MRI Data Acquisition and Preprocessing.

Functional and anatomical images were acquired with a 3T Philips Achieva MRI scanner (Philips Medical Systems, Andover, MA, USA) equipped with a 32-channel receiver head coil. Functional images were acquired using a T2*-weighted, blood-oxygen level-dependent

echoplanar (EPI) sequence (sensitivity encoding [SENSE] factor 2, flip angle 70 deg., 80 x 78 matrix, field of view [FOV] = 24 cm, repetition time [TR] = 2000 ms, and echo time [TE] = 30 ms). EPI volumes consisted of 34 slices (1-mm interslice gap) with a voxel size of 3x3x3 mm. Slices were acquired in ascending order oriented parallel to the anterior commissure-posterior commissure plane. Each functional run included 194 (study phase) or 248 (test phase) EPI volumes. T1-weighted anatomical images were acquired with a magnetization-prepared rapid gradient echo (MPRAGE) pulse sequence (FOV = 240 x 240, 1 x 1 x 1 mm isotropic voxels, 34 slices, sagittal acquisition.

fMRI preprocessing and analyses were conducted using a combination of Statistical Parametric Mapping (SPM12, Wellcome Department of Cognitive Neurology, London, UK) and custom scripts, run under Matlab R2017a (MathWorks). Functional images were realigned to the mean EPI image and slice-time corrected using sinc interpolation to the 17th slice. The images were then reoriented and spatially normalized to a sample-specific EPI template following previously published procedures(de Chastelaine et al., 2011; 2016). Normalized volumes were resampled into 3 mm isotropic voxels and smoothed with an isotropic 8 mm full-width half-maximum Gaussian kernel. Anatomical images were spatially normalized to a sample-specific T1 template following procedures analogous to those applied to the functional images.

MRI Data Analysis

fMRI data were analyzed using both mass univariate analysis and MVPA to, respectively, localize and quantitate reinstatement effects. The univariate analyses were performed in two stages. In the first stage, separate GLMs were constructed for each participant. Parameter estimates from events of interest were then carried forward to second-level random effects

factorial ANOVAs to test for group level effects. Separate GLMs were employed to analyze the study and test phases.

Study trials were binned according to encoding category (i.e., face trials vs. scene trials), giving two events of interest. These were modeled with a 2 s duration boxcar regressor convolved with SPM's canonical hemodynamic response function (HRF) and an orthogonalized delayed HRF (Andrade et al., 1999) generated by shifting the orthogonalized canonical HRF one TR (2 s) later in time. The results obtained for the late HRF added little of theoretical significance to the findings obtained with the canonical function and are not reported here. Filler trials, trials with missing or multiple responses, and trials falling outside of the aforementioned temporal response window were modeled as covariates of no interest, along with the 30 s rest periods and six regressors representing motion-related variance (three for rigid-body-translation and three for rotation). Data from volumes showing transient displacement > 1 mm or rotation > 1° in any direction were censored by their inclusion as additional covariates of no interest. Parameter estimates from the two events of interest were carried forward to a second-level random effects 2 x 2 factorial ANOVA treating age (younger, older) as a between subjects factor and category (faces, scenes) as a within subjects factor.

For the test phase, five events of interest were included in the design matrix: correct source memory for faces (SC_{faces}), correct source memory for scenes (SC_{scenes}), successfully recognized items eliciting an incorrect source memory endorsement (including 'DK' responses; SIDK), old items erroneously endorsed as new (item misses), and new items correctly endorsed as new (correct rejections). To ensure an adequate number of trials per memory bin, source incorrect and item misses were collapsed across the two encoding categories. Each event of interest was modeled with a delta function convolved with SPM's canonical HRF and an

orthogonalized delayed HRF. As before, the results obtained for the late HRF added little of theoretical significance to the findings obtained with the canonical function and are not reported here. Filler trials, false alarms, trials with missing or multiple responses, and trials falling outside of the aforementioned temporal response window were modeled as covariates of no interest, along with the 30s rest periods and six regressors representing motion-related variance (three for rigid-body translation and three for rotation). Data from volumes showing a transient displacement of > 1 mm or 1° in any direction were eliminated by their inclusion as covariates of no interest. Parameter estimates from the five events of interest were carried over to a second-level random effects 2 x 5 factorial ANOVA treating age (younger, older) as a between subjects factor and subsequent memory (source correct faces, source correct scenes, source incorrect, item miss, correct rejection) as a within subjects factor.

Univariate Reinstatement Effects.

Using univariate analyses of mean signal change, cortical reinstatement effects were operationalized as regions where category-selective encoding and recollection effects overlapped (cf. Johnson & Rugg, 2007; Wang et al., 2016). The regions were identified using a sequence of masking procedures similar to those reported by Wang et al. (2016). Recollection effects were operationalized by separately contrasting source correct trials from each image category with source incorrect trials pooled across the two encoding conditions (SC_{faces} > SIDK, SC_{scenes} > SIDK; height threshold $p < .001$). We then exclusively masked the two recollection contrasts with one another in order to identify recollection effects that were unique to each category (mask threshold $p < .05$). The resulting category-selective recollection contrasts were then inclusively masked with the corresponding category-selective encoding contrast (Faces > Scenes or Scenes

> Faces, inclusive mask threshold $p < .001$). Clusters that survived family-wise error (FWE) corrected extent thresholds in the masked recollection contrasts were retained.

Multivariate Pattern Similarity Feature Selection.

Pattern similarity analysis (PSA) was performed on the 151 voxels showing the largest z-scores for each of the two category-selective reinstatement effects identified by the mass univariate analyses described above. The size of the voxel set was selected to coincide with the spatial extent of the smaller of the two mass univariate reinstatement effects (Table 2). Thus, for both stimulus categories, PSA was performed only on those voxels that demonstrated reliable suprathreshold reinstatement effects at our pre-experimental statistical threshold (it should be noted, however, that the results were unchanged when the voxel sets were varied to include the top 50, 100, 250, or 500 voxels demonstrating univariate reinstatement effects for each category; see Supplementary Materials and Table S2). To avoid bias, the voxel sets were empirically defined at the single participant level by using an iterative 'leave-one-out' approach (Wang et al., 2016). For each of 24 randomly yoked young-older adult pairs, voxel selection was determined by a group-level analysis that identified face- and scene-selective univariate reinstatement effects on the data from the remaining 46 participants. This approach ensured that the data used to define the voxel sets for each participant were independent of the data subjected to PSA.

Pattern Similarity Analyses.

Unsmoothed data from the four study sessions were concatenated and subjected to a 'least-squares-all' GLM (Mumford et al., 2014; Rissman et al., 2004) to estimate the BOLD response for each trial separately. Each study event was modeled as a 2 s duration boxcar convolved with the canonical HRF. Unsmoothed data from the four test sessions were analyzed in a similar manner with the exception that, as in prior similar studies (e.g., Folville et al., 2020;

Wang et al., 2016; Wing et al., 2015), test events were modeled with a delta function convolved with the canonical HRF. (The employment of the boxcar and delta functions to model, respectively, the study and test data was motivated by the sustained presentation duration and attendant cognitive processing of study items, and the likely much more short-lived nature of the processing associated with the retrieval cue. The PSA results were unchanged, however, when the study events were modeled with a delta function). PSA was performed on the resulting single-trial β weights and was based on Fisher *z*-transformed Pearson correlation coefficients. A pattern similarity metric was computed separately for each image category within the respective category-selective 151-voxel sets.

Category-level reinstatement effects were defined as the difference between the mean across-voxel correlation between a given study trial and all test trials involving the same image type, and the mean correlation between that same study trial and all test trials involving the alternate image type (i.e., a within-between *category* similarity metric; see Carp et al., 2011; Haxby et al., 2001). This procedure was performed separately for correct (SC) and incorrect (SIDK + item miss) trials. Note that due to insufficient source incorrect trials numbers, 'incorrect' memory bins were expanded to include 'Don't Know' responses and item misses. The number of SC trials ranged from 27-171 (M = 109) and the number of combined incorrect trials ranged from 20-145 (M = 78). A summary measure of category-level pattern similarity was computed for each participant by averaging across all of the trial-wise within-between similarity estimates for a given image category and source memory outcome.

Item-level reinstatement effects were defined as the difference between the across-voxel correlation between a given study trial and its corresponding test trial, and the mean correlation between that same study trial and all other test trials involving the same image type (giving an

item-wise similarity metric). As for the category-level similarity metric, item-level pattern similarity was computed separately for correct and incorrect memory trials, and a summary measure was computed for each participant as the average trial-wise within-between similarity estimates corresponding to a given image category and source memory outcome.

In addition to assessing encoding-retrieval overlap, we also employed PSA to quantify encoding-related neural selectivity in the face and scene voxel sets used to identify category- and item-level reinstatement effects (c.f., Koen et al., 2019; Srokova et al., 2020). For the face-selective voxel set, the within-category measure was the average across-voxel similarity between a given face trial and all other face trials, and the between-category measure was the average correlation between a given face trial and all scene trials. The same approach was used for the scene-selective voxel set, except that scene trials were used for the within-category measures, and face trials were used for the between-category measures. A measure of neural selectivity for each voxel set was computed as the difference between the respective within- and between-category similarity effects, averaged across all trials. To avoid biased similarity estimates stemming from temporal autocorrelation between trials occurring within the same scanner run, PSA was computed between trials occurring in separate runs (Mumford et al., 2014). Note that this was only relevant when computing neural selectivity for the study phase, as study and test phases were always associated with separate scanning runs.

Exploratory Whole Brain PSA

The univariate approach described above to identify voxel-sets for PSA is arguably most appropriate when employed to delineate feature sets relevant to the reinstatement of category-level information (cf. Wang et al., 2016). It is less clear whether the approach is also appropriate for identification of robust item-level effects. Moreover, reliable category-level pattern similarity

effects can exist in voxel sets in which univariate effects are undetectable (e.g., Davis et al., 2014). In light of these potential limitations of our feature selection strategy, we supplemented the analysis of the feature set defined by our univariate approach by performing exploratory PSAs in functionally defined ROIs derived from the Atlas of Intrinsic Connectivity of Homotopic Areas brain atlas (Joliot et al., 2015). The atlas defines a set of spatially contiguous cortical parcels on the basis of the commonality of the resting state functional connectivity profiles of their constituent voxels, an approach held to confer a high degree of functional homogeneity to each parcel (see also Wig et al., 2014). For the purposes of the present analyses, the atlas was resampled to 3mm isotropic voxels, matching the resolution of our functional data. The resampling led to 16 of the original 384 regions containing fewer than five voxels. These regions were dropped from the analyses and PSA was performed on data extracted from the remaining 368 ROIs (mean ROI voxel count = 83, range 5-315). For each ROI, we computed item- and category-level pattern similarity metrics using the methods described in the preceding section. For regions where significant age differences in category-level reinstatement were identified, we went on to also compute metrics of neural selectivity from the study phase data, again using the same approach as described above. All *p*-values were FDR adjusted (q = .05) to correct for multiple comparisons.

Trial-wise Mixed Effects Analyses.

Bilateral hippocampal masks were manually traced on an anatomical T1 template derived from a large cross-sectional dataset from our lab (36 younger, 36 middle-aged, and 64 older adults (de Chastelaine et al., 2017). For each participant, we generated two unilateral vectors comprising single trial β weights averaged across all voxels falling within left and right hippocampal masks, respectively. Each vector was then *z*-transformed across trials separately for

each participant, and the correlation (Fisher-z transformed) between left and right trial-wise hippocampal activity was computed. The mean across-participant correlation between left and right hippocampal activity was highly significant (mean $r = .67$, $p = 2.20 \times 10^{-16}$). Motivated by this result, and lacking any a priori hypotheses regarding hippocampal lateralization, we computed bilateral trial-wise hippocampal activity by averaging the parameter weights across the two hemispheres.

To examine the link between trial-wise estimates of retrieval-related hippocampal activity, cortical reinstatement, and source memory accuracy, we performed a set of generalized linear mixed-effects models separately for each stimulus category. In each model, trial-wise binary source memory outcomes (correct, incorrect) were entered as the dependent variable. Age, hippocampal activity, category-level pattern similarity, and all two-way interaction terms involving age were entered as fixed effects predictors. Subject-wise intercept and slope terms were entered into the model as random-effects factors. R syntax for this model was as follows: Memory ~ Age + Similarity + Hipp + Age*Similarity + Age*Hipp + (1 + Similarity + Hipp | Subject). A separate set of models were specified as above but with item-level pattern similarity in place of category-level similarity. Models were fit using restricted maximum likelihood.

Statistical Analyses.

All statistical analyses were conducted with R software (R Core Team, 2017). All *t*-tests were two-tailed and performed using the t.test function in the base R package. Welch's unequal variance *t*-tests were performed when assumptions of equal variance were not met. ANOVAs were conducted using the *afex* package (Singmann et al., 2016) and the Greenhouse-Geisser procedure (Greenhouse & Geisser, 1959) was used to correct degrees of freedom for non-sphericity when necessary. Post-hoc tests on significant effects from the ANOVAs were

conducted using the *emmeans* package (Lenth, 2018) and corrected for multiple comparisons using the Holm-Bonferroni procedure where appropriate. Generalized linear mixed-effects models were performed using the *glmer* function in the lme4 package (Bates et al., 2015).

All data and custom code are available upon request from the first author.

Results

Behavioral Results

Means and standard deviations of behavioral measures are presented in Table 1. Vividness ratings and median response times (RTs) from the study phase were sorted according to subsequent memory status into source correct (SC) and incorrect (including source incorrect, 'Don't Know', and item misses) bins and submitted to separate 2 (age) x 2 (memory) x 2 (category) mixed factorial ANOVAs. The analysis of vividness ratings revealed a significant main effect of memory ($F_{(1,46)} = 53.33$, $p = 3.13 \times 10^{-9}$, partial-$\eta^2 = .54$), which was driven by reduced vividness for incorrect memory trials. The main effects of age ($F_{(1,46)} = 3.12$, $p = .084$, partial-$\eta^2 = .06$) and category ($F_{(1,46)} = 0.70$, $p = .409$, partial-$\eta^2 = .01$) were not significant, nor were there any significant interactions involving age or category (all *ps* > .1). The analysis of RTs revealed a significant main effect of category ($F_{(1,46)} = 5.35$, $p = .025$, partial-$\eta^2 = .10$), which reflected faster response times for face trials relative to scenes. The remaining main effects and interactions were not significant (all *ps* > .1).

Item recognition – operationalized as the difference between the proportion of old items correctly endorsed as 'old' (hit rate) and the proportion of new items erroneously endorsed as 'old' (false alarm rate) – was computed separately for each image category and submitted to a 2 x 2 mixed factorial ANOVA with factors of age (young, older) and stimulus category (faces, scenes). This analysis produced significant main effects of age ($F_{(1,46)} = 10.11$, $p = .003$, partial-

η^2 = .18) and category ($F_{(1,46)}$ = 5.46, p = .024, partial-η^2 = .11). The interaction between age and category was not significant ($F_{(1,46)}$ = 0.74, p = .393, partial-η^2 = .02). Post-hoc contrasts revealed that, across both categories, recognition accuracy was reduced in older relative to younger adults and that, across both age groups, accuracy was higher for items studied with faces.

Source memory accuracy was estimated using a single high-threshold model (Snodgrass & Corwin, 1988) corrected for guessing (Park et al., 2008) using the formula *pSR* = [*pHit* - .5 * (1 - *pDK*)] / [1 - .5 * (1 - *pDK*)], where pSR refers to the probability of source recollection, and pHit and pDK refer to the proportion of correct old responses attracting an accurate or a 'Don't Know' source memory endorsement, respectively. A *t*-test revealed that source memory accuracy was significantly lower in older (*M* = .51, *SD* = .16) than in younger (*M* = .68, *SD* = .18) adults ($t_{(45.51)}$ = -3.44, p = .001). To further unpack the effect of stimulus category on source accuracy, we computed proportions of SC trials (SC/SC+SIDK) and submitted these to a 2 (age) x 2 (stimulus category) mixed-factorial ANOVA. This revealed significant main effects of age ($F_{(1,46)}$ = 6.26, p = .016, partial-η^2 = .12 - consistent with the foregoing analysis of the pSR metric) and category ($F_{(1,46)}$ = 12.04, p = .001, partial-η^2 = .21). The interaction between age and category was not significant ($F_{(1,46)}$ = 0.36, p = .553, partial-η^2 = .01). Post-hoc tests revealed that, across both age groups, correct source judgments were more likely for faces than scenes.

Table 1. Means (SD) for behavioral performance measures

	Younger		Older	
	Faces	Scenes	Faces	Scenes
Study Vividness[Ŧ]				
Source Correct	2.42 (.32)	2.44 (.32)	2.24 (.39)	2.18 (.43)
Incorrect	2.23 (.42)	2.13 (.51)	2.06 (.46)	2.01 (.49)
Study RTs (ms)[t]				
Source Correct	2369 (678)	2398 (628)	2130 (570)	2266 (524)
Incorrect	2351 (656)	2350 (633)	2285 (605)	2327 (579)
Recognition Accuracy*[t]	.69 (.18)	.67 (.17)	.56 (.14)	.52 (.13)
Proportion Source Correct*[t]	.83 (.14)	.79 (.16)	.75 (.13)	.68 (.13)

Proportion Source Don't Know[ŧ]	.12 (.13)	.16 (.13)	.12 (.12)	.14 (.13)
Recognition RTs (ms)[#]				
Source Correct	1150 (210)	1182 (189)	1306 (219)	1348 (177)
Incorrect	1273 (202)	1546 (248)	1426 (193)	1505 (218)

Note: Incorrect memory bin was collapsed across source incorrect, 'Don't Know', and item misses.
*significant main effect of age; [ŧ]significant main effect of category; [Ŧ]significant main effect of memory;
[#]significant age x memory x category interaction

For test trials, median RTs for correct item memory judgments on the studied test words were sorted into SC and source incorrect/Don't Know (SIDK) memory judgment bins and submitted to a 2 (age group: younger, older) x 2 (stimulus category: faces, scenes) x 2 (source memory: SC, SIDK) mixed factorial ANOVA. This gave rise to a significant three-way interaction ($F_{(1,46)} = 8.77$, $p = .005$, partial-$\eta^2 = .16$). Post-hoc tests revealed that RTs were significantly slower for older adults relative to younger adults on face trials attracting SC ($t_{(86.8)} = 2.60$, $p = .011$) and SIDK ($t_{(86.8)} = 2.55$, $p = .013$) memory judgments, as well as on correct scene trials ($t_{(86.8)} = 2.77$, $p = .007$). The two age groups did not differ on SIDK scene trials ($t_{(86.8)} = -0.68$, $p = .501$). Correct rejection RTs in young ($M = 1384$ ms, $SD = 295$ ms) and older ($M = 1414$ ms, $SD = 213$ ms) adults did not significantly differ ($t_{(41.85)} = 0.41$, $p = .688$).

Univariate Reinstatement Effects

The results of the univariate reinstatement contrasts are described in Table 2 and illustrated in Figure 2A. Face-selective effects for study trials were identified in a bilateral cluster falling along the border of the cuneus and precuneus (and extending into the posterior cingulate), medial prefrontal cortex, left middle temporal gyrus, and left angular gyrus. Enhanced face-related activity was also evident in a large cluster spanning the bilateral anterior medial temporal lobes (overlapping amygdala and anterior hippocampus) and extending into right fusiform and middle temporal gyri. Scene-selective effects at study were evident in bilateral parahippocampal cortex (extending into retrosplenial cortex and occipital cortex), right precuneus (extending into posterior cingulate), as well as left middle and inferior frontal gyri.

Table 2. Loci of Univariate Encoding and Reinstatement Effects

Contrast	Region	MNI			Peak z	Cluster size
		x	y	z		
Face Encoding	**Precuneus**	**3**	**-58**	**23**	**Inf**	**587**
	Cuneus	0	-88	20	3.91	
	WM	-18	-46	17	3.62	
	Amygdala	**21**	**-7**	**-22**	**Inf**	**2518**
	Hippocampus	-24	-7	-22	7.04	
	Fusiform Gyrus	45	-79	-16	6.64	
	Superior Medial Gyrus	**-3**	**59**	**17**	**6.51**	**837**
	Mid Orbital Gyrus	3	38	-7	5.62	
	Superior Frontal Gyrus	-12	44	44	3.88	
	Angular Gyrus	**-57**	**-67**	**23**	**4.58**	**128**
	Angular Gyrus	-48	-64	26	4.48	
	Angular Gyrus	-54	-70	32	4.45	
	Middle Temporal Gyrus	**-57**	**-10**	**-16**	**4.15**	**154**
	Middle Temporal Gyrus	-66	-13	-13	4.13	
	Middle Temporal Gyrus	-54	-31	-7	4.11	
Scene Encoding	**Parahippocampal Cortex**	**30**	**-40**	**-19**	**Inf**	**6802**
	Fusiform Gyrus	-27	-46	-16	Inf	
	Middle Occipital Gyrus	42	-82	17	Inf	
	Precuneus	**9**	**-43**	**41**	**6.40**	**316**
	Precuneus	-9	-40	44	6.25	
Face Reinstatement	**Precuneus**	**0**	**-55**	**26**	**6.31**	**151**
	Precuneus	9	-55	26	5.33	
Scene Reinstatement	**Parahippocampal Cortex**	**27**	**-34**	**-25**	**7.31**	**405**
	Retrosplenial Cortex	15	-55	11	5.26	
	Parahippocampal Gyrus	21	-19	-31	4.06	
	Parahippocampal Cortex	**-24**	**-40**	**-22**	**6.92**	**224**
	Fusiform Gyrus	-33	-55	-25	3.45	
	Retrosplenial Cortex	**-12**	**-58**	**5**	**5.32**	**170**

Note that for reinstatement loci, peak values were derived from the category-selective recollection contrast (inclusively masked with the corresponding encoding contrast).

Cortical reinstatement effects were operationalized as regions where the above-described study effects overlapped with the corresponding category-selective recollection effects. The regions were identified with a series of masking procedures similar to those reported by Wang et al. (2016). These analyses identified face reinstatement effects in the precuneus, and scene reinstatement effects in bilateral parahippocampal cortex, extending into retrosplenial cortex. The effects are detailed in Table 2 and illustrated in Figure 2B. Replicating Wang et al. (2016), exclusively masking these reinstatement effects with a whole-brain age x recollection interaction

contrast (mask threshold $p < .05$) failed to identify any voxels where reinstatement effects were moderated by age.

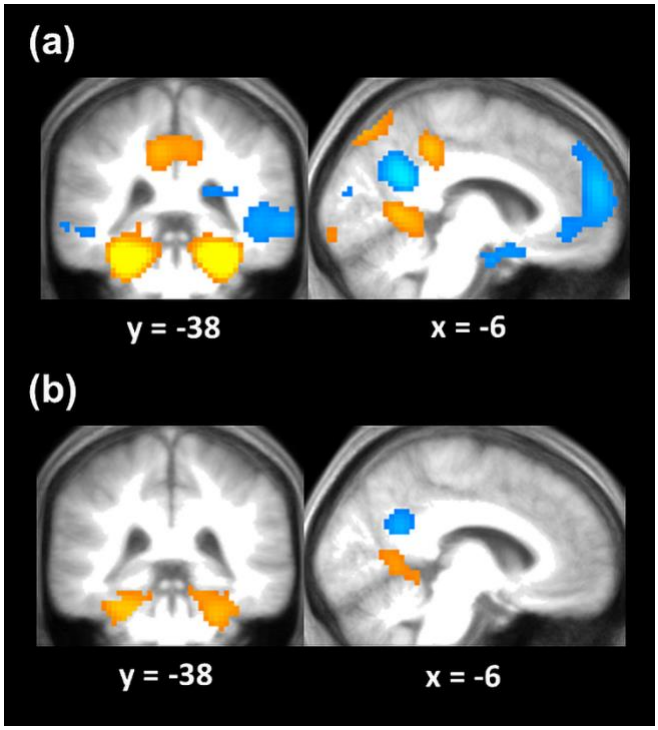

Figure 2. Univariate Reinstatement Effects. (a) category-selective encoding effects operationalized as Face > Scene (cool colors) and Scene > Face (warm colors), irrespective of subsequent memory status. (b) category-selective reinstatement effects, operationalized by inclusively masking each category-selective recollection contrast with the corresponding encoding contrast. All contrasts represent main effects collapsed across age-group.

Multi-Voxel Pattern Similarity Analyses

We performed PSA to quantify the strength of cortical reinstatement in the two category-selective voxel-sets identified by the foregoing univariate analyses. Note that unless otherwise noted, excluding the outliers identified in Figure 3 (> 3 SDs from the group means) did not alter the results reported below.

Category-Level Pattern Similarity

Category-level pattern similarity data were submitted to a three-way mixed-factorial ANOVA with factors of age (younger, older), category (faces, scenes) and memory (correct,

incorrect). The ANOVA gave rise to a significant three-way interaction ($F_{(1,46)} = 7.62$, $p = .008$, partial-$\eta^2 = .14$). To unpack these results, we performed subsidiary 2 (age) x 2 (memory) mixed-factorial ANOVAs on scene and face pattern similarity, respectively. The analysis of scene pattern similarity revealed a significant age by memory interaction ($F_{(1,46)} = 7.82$, $p = .008$, partial-$\eta^2 = .15$) which was driven by greater pattern similarity in younger relative to older adults for correct ($t_{(89.2)} = -5.34$, $p < .0001$) but not incorrect ($t_{(89.2)} = -1.04$, $p = .300$) memory trials. Additionally, scene pattern similarity was significantly greater for correct relative to incorrect memory trials in both young ($t_{(46)} = 6.14$, $p < .0001$) and older ($t_{(46)} = 2.19$, $p = .034$) adults. An analogous analysis of face pattern similarity revealed nonsignificant main effects of memory and age, as well as a nonsignificant age by memory interaction (all *ps* > .07). However, after excluding the two outlying data points evident in Figure 3A, the ANOVA gave rise to a marginally significant main effect of memory accuracy ($F_{(1,44)} = 4.05$, $p = .050$, partial-$\eta^2 = .08$). In both age groups category-level pattern similarity for the two stimulus categories was significantly greater than zero (two-tailed one-sample *t*-tests) on trials attracting a correct source memory judgment (*ps* > .02). By contrast, incorrect memory trials did not reliably differ from zero for either stimulus category regardless of age group (*ps* > .15).

In light of the age differences in retrieval RTs reported above, we performed a follow-up ANCOVA to contrast source correct scene reinstatement effects in young and older adults while controlling for RT. This analysis gave rise to identical results to the original ANOVA.

The above-described category-level pattern similarity effects were operationalized as the difference in the mean across-voxel correlation between study-test pairs belonging to the same stimulus category and the mean correlation between study-test pairs belonging to the opposite stimulus category. Thus the influence of age on these effects is ambiguous with respect to

whether the age differences were driven by reduced within-category neural similarity or increased between-category similarity. To examine this issue, the within- and between-category similarity measures elicited on source correct trials by the preferred stimulus categories for each voxel set were submitted to separate 2 (age) x 2 (similarity: within, between) mixed-factorial ANOVAs. (Note that this analysis was restricted to source correct trials as no age differences were evident on incorrect memory trials). The relevant data are presented in Supplementary Figure S1. The analysis of scene similarity measures revealed a significant main effect of similarity ($F_{(1,46)} = 55.13$, $p = 2.11 \times 10^{-9}$, partial-$\eta^2 = .55$) as well as a significant interaction between age and similarity ($F_{(1,46)} = 21.61$, $p = 2.83 \times 10^{-5}$, partial-$\eta^2 = .32$). The main effect of age was not significant ($F_{(1,46)} = 0.68$, $p = .414$, partial-$\eta^2 = .01$). Post-hoc tests revealed that the between-category similarity metric was significantly greater in older relative to younger adults ($t_{(54)} = 2.11$, $p = .040$) while the within-category similarity metric did not significantly differ between the two age groups ($t_{(54)} = -0.53$, $p = .599$). Turning to faces, there was a significant main effect of similarity ($F_{(1,46)} = 11.99$, $p = .001$, partial-$\eta^2 = .21$) which, unsurprisingly given the data illustrated in Figure 3A, was driven by greater within- relative to between-category similarity. The main effect of age was not significant, and nor was the interaction between age and similarity type ($ps > .1$).

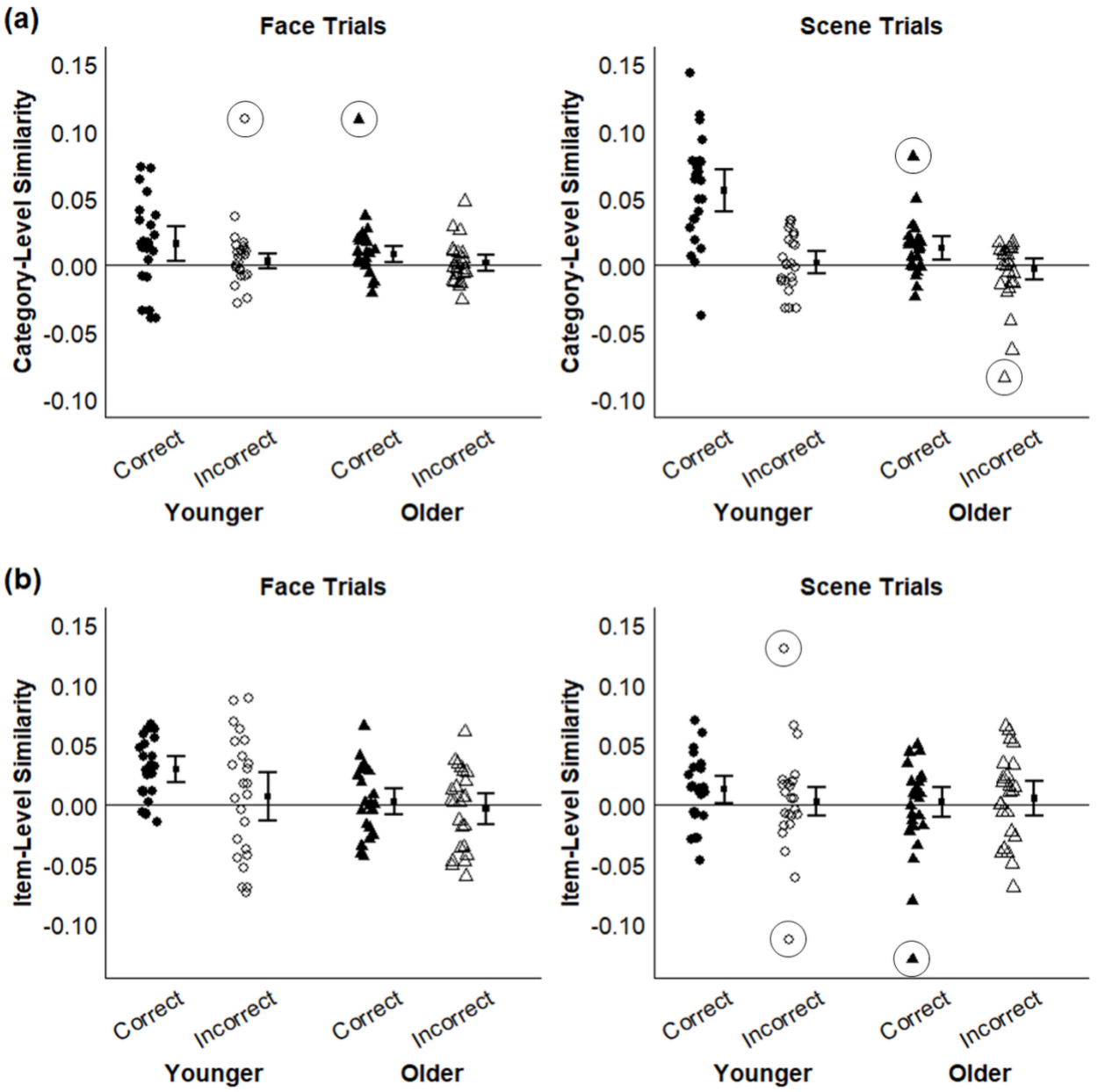

Figure 3. **Category- and item-level reinstatement strength.** Plots illustrating across-voxel (a) category-level and (b) item-level encoding-retrieval pattern similarity. Error bars reflect 95% confidence intervals which were computed after excluding the flagged outliers.

Item-Level Pattern Similarity

Turning to item-level pattern similarity, a mixed-factorial ANOVA with factors of age, category, and memory revealed a significant main effect of age ($F_{(1,46)} = 6.67$, $p = .010$, partial-η^2

= .13), which was driven by greater item-level pattern similarity in younger relative to older adults ($t_{(46)}$ = -2.58, p = .013). The main effects of memory and category were not significant, and nor were any of the two- and three-way interactions (ps > .1). Item-level pattern similarity in younger adults was significantly greater than zero (two-tailed one-sample t-test) for both face and scene trials attracting a correct source memory judgement (ps < .03). Item-level pattern similarity in younger adults did not significantly vary from zero on incorrect memory trials, and nor did it significantly differ from zero in any of the older adult trial bins (ps > .4).

Neural selectivity at encoding co-varies with strength of reinstatement at retrieval

We next asked whether the above-described effects of age on the strength of category-level scene reinstatement effects were related to the selectivity with which study events were processed at the time of encoding (i.e., neural differentiation). We computed the category-level selectivity of the neural responses elicited at study by the preferred vs. the non-preferred category in each of the voxel sets employed for the analysis of reinstatement effects (see Methods). We limited this analysis to source correct trials as it was these trials where age differences were observed. As is evident in Figure 4A, one young and one older adult had neural scene selectivity estimates that were extreme outliers (> 3 SDs from the group mean) and therefore these participants were dropped from the following analyses.

Scene pattern similarity at encoding was significantly greater than zero in younger (M = .13, SD = .06, $t_{(22)}$ = 9.93, p = 1.37 x 10^{-9}) and older (M = .05, SD = .03, $t_{(22)}$ = 6.43, p = 1.83 x 10^{-6}) adults, and these effects remained significant after correcting for multiple comparisons. Pattern similarity for face trials was similarly greater than zero in younger (M = .02, SD = .04, $t_{(23)}$ = 2.07, p = .050) and older (M = .02, SD = .03, $t_{(23)}$ = 3.46, p = .002) adults, although only the effect for older adults remained significant after correcting for multiple comparisons. When

combined across age groups, however, pattern similarity for face trials was robustly greater than zero ($t_{(47)} = 3.73$, $p = .001$). A 2 (age) x 2 (category) factorial ANOVA revealed a significant age group x category interaction ($F_{(1,46)} = 10.97$, $p = .002$, partial-$\eta^2 = .19$). Neural selectivity was significantly lower for scenes in older relative to younger adults ($t_{(91.7)} = -4.63$, $p < .0001$) but, echoing the category-level reinstatement effects illustrated in Figure 3A, no age differences were observed for face selectivity ($t_{(91.7)} = 0.18$, $p = .855$).

For each stimulus category, we computed the partial correlation between neural differentiation at encoding and category-level reinstatement, controlling for age. The resulting partial correlations were highly significant for both faces ($r_{partial} = .58$, $p = 2.26 \times 10^{-5}$) and scenes ($r_{partial} = .53$, $p = 1.63 \times 10^{-4}$) (Figure 4B) and multiple regression analyses indicated that age did not significantly moderate either of these relationships (*ps* > .2 for both interaction coefficients). In light of these findings, we conducted an ANCOVA to contrast category-level scene reinstatement in young and older adults while controlling for neural selectivity at study. The analysis revealed a non-significant age effect ($F_{(1,43)} = 3.82$, $p = .057$, partial-$\eta^2 = .08$). When the analysis was conducted after excluding the covariate, the main effect of age was of course highly significant ($F_{(1,44)} = 26.11$, $p = 6.71 \times 10^{-6}$), and the effect size increased substantially ($\eta^2 = .37$). In short, the inclusion of neural selectivity at study as a covariate led to a nearly five-fold decrease in the proportion of variance in scene reinstatement explained by the factor of age group.

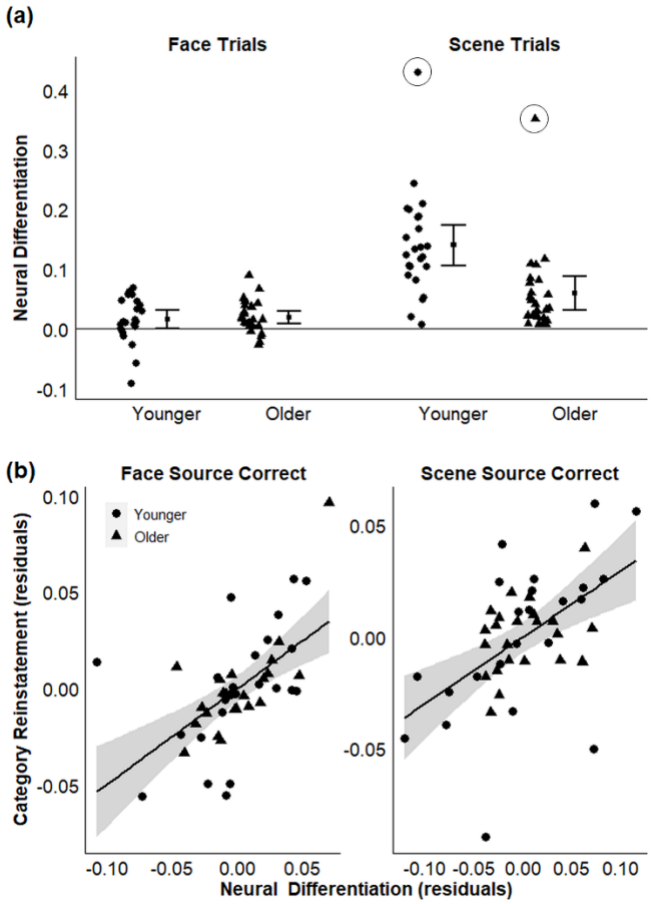

Figure 4. (a) Category-level similarity at encoding (i.e., neural differentiation). Error bars reflect 95% confidence intervals which were computed after excluding the flagged outliers. (b) Partial correlations between neural differentiation at encoding and category-level reinstatement strength, controlling for age.

The PSA measures used in the foregoing analyses to quantify neural selectivity at encoding and category-level reinstatement strength were based partially on the same study trial data. This raises the possibility that the correlations reported between these two measures reflect the employment of a shared metric. To address this possibility, we computed a measure of neural selectivity at retrieval using a PSA approach directly analogous to that used to quantify neural selectivity at encoding (see supplemental materials). This approach provided a measure of reinstatement strength conceptually similar to that obtained from the encoding-retrieval PSA metric, but now computed exclusively from test phase data. As is reported in the supplemental

materials, when the foregoing analyses were repeated using this additional reinstatement metric the results were qualitatively identical to those reported above, thus obviating concerns about non-independence of the study and test PSA estimates.

Trial-wise Mixed-Effects Analyses

Retrieval-related reinstatement and hippocampal retrieval effects independently co-vary with memory accuracy

We performed a series of generalized linear mixed-effects analyses to examine the relationship between trial-wise estimates of retrieval-related hippocampal activity, category-level pattern similarity, and source memory accuracy (Table 3 and Figure 5). We elected to run separate models for scene and face trials to ease interpretation and to avoid model overfitting. As is evident from Figure 5, the odds of making a correct source memory judgment significantly co-varied with increasing hippocampal activity for both face (Logit Odds = 0.22, 95% CI = 0.08-0.35) and scene (Logit Odds = 0.18, 95% CI = 0.07-0.30) trials. The odds of correct source memory judgments also significantly co-varied with increasing category-level pattern similarity for both face (Logit Odds = 0.28, 95% CI = 0.02-0.55) and scene (Logit Odds = 0.52, 95% CI = 0.05-1.00) trials. For scene trials, there was a significant age x category-level reinstatement interaction (Logit Odds = 0.80, 95% CI = 0.12-1.48) which was driven by a stronger within-subject relationship between reinstatement of category-level scene information and source memory performance in younger adults. As is evident in Figure 5, however, this relationship was reliably greater than chance in both age groups.

Table 3. Results of generalized linear mixed effects analyses predicting source memory accuracy

Fixed Effects Predictors	Face Trials		Scene Trials	
	Logit Odds	95% CI	Logit Odds	95% CI
Category-Level Models				
Age	0.85*****	0.31 - 1.39	0.82*****	0.28 - 1.36
Hippocampal Activity	0.22*****	0.08 - 0.35	0.18*****	0.07 - 0.30

Category-Level Similarity	0.28*	0.02 - 0.55	0.52*	0.05 - 1.00
Age x Hippocampus	-0.01	-0.21 - 0.19	0.02	-0.15 - 0.19
Age x Category Similarity	-0.04	-0.42 - 0.33	0.80*	0.12 - 1.48
Item-Level Models				
Age	0.79*	0.30 - 1.29	0.91*	0.47 - 1.36
Hippocampal Activity	0.21*	0.07 - 0.35	0.19*	0.08 - 0.30
Item-Level Similarity	0.05	-0.05 - 0.14	-0.02	-0.11 - 0.07
Age x Hippocampus	-0.01	-0.21 - 0.19	-0.02	-0.18 - 0.14
Age x Item Similarity	0.08	-0.06 - 0.22	0.06	-0.07 - 0.19

Notes: Incorrect trials and older adults treated as reference groups. *denotes significant fixed effect coefficients

Performing the same set of analyses with a model employing item-level pattern similarity as a predictor again revealed that the odds of a correct source memory judgment co-varied with increasing hippocampal activity for both face (Logit Odds = 0.21, 95% CI = 0.07-0.35) and scene (Logit Odds = 0.19, 95% CI = 0.08-0.30) trials. By contrast, item-level pattern similarity did not predict memory accuracy independently of hippocampal activity (Table 3).

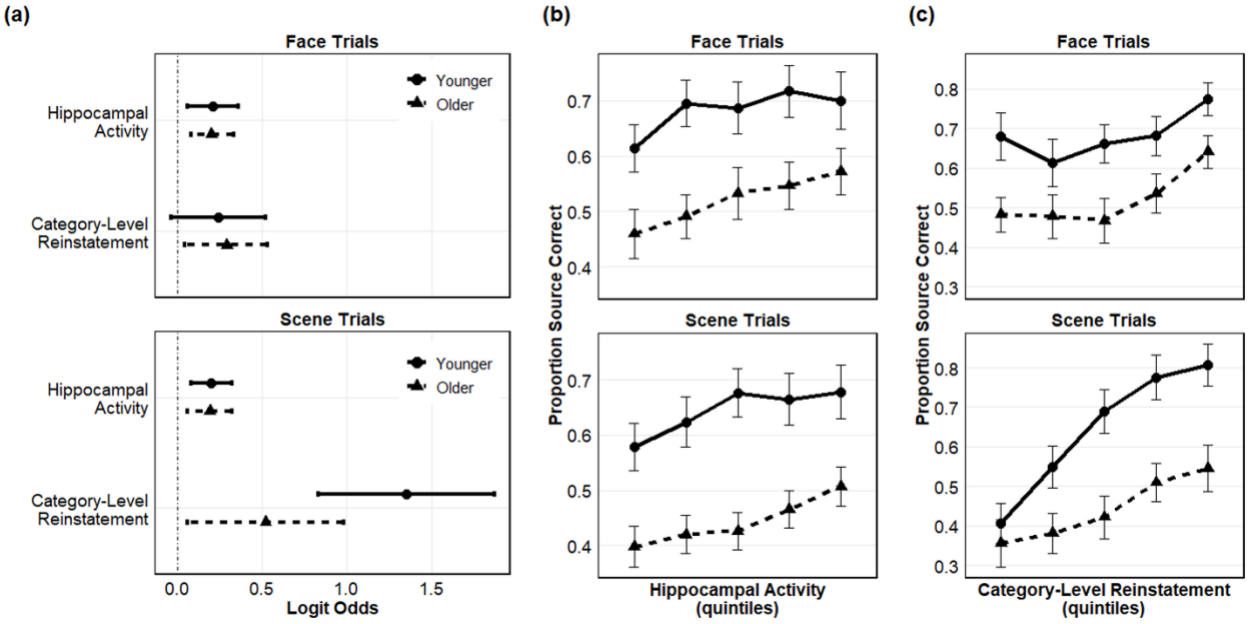

Figure 5. (a) Logit odds and 95% confidence intervals are plotted for hippocampal activity and category-level pattern similarity separated by age and category. (b) Increasing trial-wise retrieval-related hippocampal activity predicts an increased proportion of correct face (top panel) and scene (bottom panel) source memory judgments. (c) Increasing trial-wise category-level reinstatement strength predicts an increased proportion of correct face (top panel) and scene (bottom panel) source memory judgments. Error bars in (b) and (c) reflect ± 1 SE. Trial-wise hippocampal activity (b)

and category-level pattern similarity (c) were sorted within-subject into quintiles for display purposes only.

Retrieval-related hippocampal activity co-varies with item-level pattern similarity

In a final analysis, we used linear mixed effects analyses to examine relationships between trial-wise estimates of retrieval-related hippocampal activity and reinstatement. As previously, these analyses were performed separately for each stimulus category. In each model, trial-wise estimates of item- or category-level reinstatement were entered as the dependent variable. Age, hippocampal activity, and the age x hippocampal activity interaction term were entered as fixed effects predictors, and subject-wise intercept and slope terms were entered into the models as random-effects factors [$\text{Similarity}_{(\text{Item/Category})}$ ~ Age + Hipp + Age*Hipp + (1 + Hipp | Subject)]. The analyses revealed that hippocampal activity was a significant predictor of item-level reinstatement of scene information (b = .12, $p = 2.23 \times 10^{-4}$). The age x hippocampal interaction was not significant, and nor did hippocampal activity significantly co-vary with category-level scene reinstatement (b = .01, p = .836). Hippocampal activity did not significantly co-vary with either item- (b = -.05, p = .119) or category-level (b =.04, p = .145) reinstatement of face information. We note that, as reported in the foregoing section, trial-wise fluctuations in item-level pattern similarity did not predict memory accuracy, nor were across-subject item-level estimates moderated by memory accuracy. Accordingly, we express caution when interpreting item-level pattern similarity as representing reinstatement of trial-unique information.

Exploratory Whole-Brain Analyses

Category-Level Pattern Similarity

The findings reported above suggest that age effects on category-level reinstatement are confined to scene regions. To examine whether this pattern of results generalizes to regions outside of the empirically defined voxel sets, we performed category-level PSA separately for

face and scene trials in a set of 368 spatially contiguous ROIs offering whole brain coverage (see Methods). For each region, we performed one-sample *t*-tests (one-tailed) to identify ROIs where pattern similarity on trials attracting a correct source memory judgment was reliably greater than zero. All *p*-values were FDR corrected (q = .05) for multiple comparisons across the 368 ROIs. In young adults, pattern similarity was significantly greater than zero for face trials in three regions, for scene trials in 53 regions, and for both face and scene trials in seven regions (Supplementary Figure S2 and Table S3). In older adults, there were no regions where pattern similarity for face trials were significantly greater than zero. For scene trials, pattern similarity in older adults was significantly greater than zero in 13 regions, all of which were also significant in younger adults. There were no regions where pattern similarity effects were significant in older adults but not younger adults.

We next performed 2 (age) x 2 (memory) mixed-factorial ANOVAs on pattern similarity for faces in each of the 10 regions where category-level face effects were significant in at least one of the age groups, and for scenes in each of the 60 regions where scene effects were significant in at least one group. All *p*-values for the main effects of age and memory and their interaction were FDR corrected (q = .05) for multiple comparisons across the 10 (face trials) or 60 (scene trials) ROIs. We focus on those regions where pattern similarity was significantly moderated not only by age group but also by memory accuracy (given that there is no obvious functional interpretation of similarity effects that are unmoderated by memory accuracy). Co-occurring main effects of age and memory on face trials were evident in the left precuneus and right cuneus. For scene trials, co-occurring main effects of age and memory were evident in left and right parahippocampal gyrus, right retrosplenial cortex, and right fusiform gyrus (Supplementary Figure S3).

Paralleling the approach employed for the analysis of the feature sets that demonstrated univariate reinstatement effects, we went on to conduct ANCOVAs contrasting category-level pattern similarity in young and older adults in each of the six regions where age and memory differences were evident, controlling for neural selectivity at encoding. These analyses were restricted to source correct trials and were performed separately for face and scene trials in the respective category-specific regions. For face trials, age differences were eliminated in the cuneus when individual differences in face selectivity at study were controlled for ($F_{(1,45)}$ = 2.30, p = .136, partial-η^2= .05), but remained significant in the left precuneus ($F_{(1,45)}$ = 5.72, p = .021, partial-η^2= .11). For scene trials, age differences were eliminated in the right parahippocampal gyrus ($F_{(1,45)}$ = 1.68, p = .202, partial-η^2= .04), and right fusiform gyrus ($F_{(1,45)}$ = 1.36, p = .250, partial-η^2= .03) when scene selectivity at study were controlled for, but they remained significant in a sub-region of the left parahippocampal gyrus ($F_{(1,45)}$ = 5.72, p = .021, partial-η^2= .11) and right retrosplenial cortex ($F_{(1,45)}$ = 6.54, p = .014, partial-η^2= .13) (Supplemental Figure S3).

Item-Level Pattern Similarity

In young adults, item-level pattern similarity was significantly greater than zero for scene trials in 12 regions, for face trials in 48 regions, and for both face and scene trials in eight regions (Supplementary Table S4 and Figure S4). There were no regions where item-level pattern similarity was significantly greater than zero in older adults. As with the category-level effects, we performed 2 (age) x 2 (memory) ANOVAs on pattern similarity for faces in the 56 regions where item-level face effects were significant, and for scenes in the 18 regions where scene effects were significant. These analyses failed to identify any regions where there were co-occurring main effects of age and memory, and nor did any regions demonstrate a significant age by memory interaction.

Discussion

We examined whether the influence of age on retrieval-related cortical reinstatement was moderated by age differences in neural selectivity during encoding, whether reinstatement predicted trial-wise source memory performance, and whether any such relationship might explain age-related variance in source memory performance. Univariate reinstatement effects for scene information were identified in bilateral parahippocampal and retrosplenial cortex, regions ubiquitously linked to visual scene processing (Epstein & Baker, 2019). Univariate reinstatement effects for face information were evident in bilateral precuneus, a prominent member of the 'extended' face processing network (Haxby & Gobbini, 2011). Using multi-voxel PSA in ROIs defined by these univariate analyses, we identified robust age differences in the strength of recollection-related reinstatement of scene information which were largely explained by analogous differences in neural differentiation at the time of encoding. In addition, there was a significant relationship between the strength of trial-wise face and scene reinstatement and memory accuracy in both age groups, although in the case of scenes, the relationship was significantly stronger in the younger group.

Turning first to the behavioral results, consistent with prior reports (Jackson & Schacter, 2004) vividness ratings at study predicted memory performance, but this relationship did not vary with age. The age differences we observed in neural differentiation and reinstatement are therefore unlikely to reflect the confounding effects of this variable. At test, source memory performance was lower in older adults (and was accompanied by an analogous effect for item memory), as would be expected given the extensive prior literature documenting age-related episodic memory decline (for review, see Koen & Yonelinas, 2014). Regardless of age, both item and source memory were higher for test words paired with faces than with scenes. While

these memory benefits for faces are not without precedent (Lin et al., 2019; Sato & Yoshikawa, 2013; Trelle et al., 2020; but see Gordon et al., 2014; Keightley et al., 2011) they currently lack an explanation. Since the effects did not interact with age group, we do not discuss them further here.

We used multi-voxel PSA to examine whether face and scene reinstatement effects associated with successful source memory judgments differed in strength according to age. Category-level reinstatement of scene information was robustly weaker in older adults relative to younger individuals, a finding consistent with some prior reports (Abdulrahman et al., 2017; Bowman et al., 2019; McDonough et al., 2014) but inconsistent with others (Thakral et al., 2017, 2019; Wang et al., 2016). Weakened scene reinstatement in older adults was driven not by a reduction in levels of within-category (scene-scene) similarity, but rather, by elevated between-category (scene-face) similarity. This finding is arguably analogous to the 'neural broadening' effects which have been reported as a possible source of age-related neural dedifferentiation in studies employing univariate analysis methods (Park et al., 2012; but see Carp et al, 2011 and Koen et al., 2019 for examples of age-related dedifferentiation driven by neural attenuation).

In line with prior findings (e.g., Koen et al., 2019; Srokova et al., 2020; Zheng et al., 2018), we identified age differences in scene selectivity not only at retrieval, but at encoding also. Thus, echoing the findings for retrieval, neural selectivity in the parahippocampal and retrosplenial cortex was robustly lower in the older group at encoding. These and similar findings have been interpreted as evidence for age-related neural dedifferentiation, a reduction in neural selectivity with age that has been proposed to contribute to age-related cognitive decline (for review, see Koen & Rugg, 2019). Possible reasons why parahippocampal and retrosplenial cortex should be so sensitive to the effects of age are discussed in detail by Srokova et al. (2020).

We note here that it remains to be established whether these findings reflect a particular vulnerability of these regions to the effects of age (see below), as opposed to differences in how scenes and related perceptual categories (e.g. houses) are processed by young and older adults.

Crucially, the age differences that we identified in recollection-related reinstatement effects in scene sensitive regions failed to achieve statistical significance when the scene similarity metric derived from the same regions at encoding was employed as a covariate. Thus, the present findings suggest that age differences in retrieval-related reinstatement are largely, if not fully, attributable to the selectivity with which the retrieved information was represented in category-selective cortex when it was initially experienced (these conclusions are buttressed by the findings from the whole brain, exploratory analyses of category-level reinstatement discussed below). In other words, the present findings offer scant support for proposals that age-related memory decline reflects an impairment in retrieval processes that support recollection of detailed information (c.f., Abdulrahman et al., 2017; McDonough et al., 2014). Also of importance, for both stimulus categories, the relationship between neural differentiation at encoding and the strength of retrieval-related reinstatement was age-invariant (Figure 4B), suggesting that the relationships identified here between neural selectivity at encoding and retrieval reflect a general principle of brain function that operates across much of the healthy adult lifespan.

In addition to the aforementioned scene reinstatement effects, we also identified reliable category-level reinstatement of face information in both young and older adults within the precuneus. Why this region – which, as noted previously, is a member of the 'extended' face network – demonstrated face reinstatement effects while more canonical face-sensitive regions such as the 'fusiform face area' (Kanwisher et al., 1997) did not is unclear. One possibility is that the nature of the study task led participants to encode relatively abstract aspects of the faces that

depend on processing supported by regions, such as the precuneus, downstream of the fusiform face area (Silson et al., 2019). Regardless, and echoing the findings from the encoding phase, there was no evidence that these reinstatement effects differed with age. Age differences in face reinstatement were evident however in sub-regions of the left precuneus and right cuneus in the exploratory whole brain analyses. As with the majority of the scene reinstatement effects identified in these analyses that were also moderated by age, the age differences in face reinstatement in the cuneus were absent after controlling for neural selectivity at encoding. Taken together, these findings are consistent with the proposal advanced above that age differences in retrieval-related reinstatement largely depend on the existence of analogous differences in neural selectivity at the time of encoding.

Across both stimulus categories, reinstatement of item-level information in the voxel sets identified by univariate analyses was reduced in older adults relative to their younger counterparts. However, this effect did not interact with source memory accuracy and therefore is hard to interpret in terms of its implications for the age differences that were evident in both face- and scene-related memory. These findings are consistent with the outcome of the whole brain exploratory analyses, which likewise failed to identify any regions where item-level reinstatement effects were moderated by memory accuracy. The insensitivity of item-level reinstatement to memory accuracy might reflect our use of a memory test that did not require retrieval of trial-unique information: perfect performance on the test would have been possible simply by recalling which stimulus category a recognized test word had been associated with at study. Regardless of the validity of this account, the present findings offer little support for the proposal (e.g. Wang et al., 2016) that age differences in cortical reinstatement effects might be more prominent when the effects are examined at the item- rather than the category-level.

Turning to the trial-wise analyses, the strength of category-level reinstatement of face and scene information predicted trial-by-trial memory accuracy in both age groups. These results build on prior findings in young adults (Gordon et al., 2014), demonstrating that within-subjects relationships between reinstatement strength and memory success extend to healthy older adults (see Trelle et al., 2020, for similar findings). For face trials, the strength of this relationship did not differ between the two age groups. This was not the case however for scene trials. As is evident in Figures 5a and 5c, the relationship between scene reinstatement and memory performance in older adults, although reliably greater than chance, was significantly attenuated relative to that in the young group. Although it comes from a different analysis approach to that employed in prior aging studies (within- rather than between-participants correlations between neural and behavioral measures), this finding marks a departure from prior observations that relationships between neural differentiation and memory performance are not moderated by age (for review, see Koen & Rugg, 2019). It is not clear, however, why the impact of age on the trial-wise relationship between reinstatement strength and memory success should differ for face and scene categories. One possibility is that this finding is a further example of the particular vulnerability of 'scene-selective' cortical regions to advancing age (Koen et al., 2019), reflecting perhaps the sensitivity of neural selectivity in these regions to accumulation of pathological Tau (Maass et al., 2019). By this argument, either advancing age or age-associated pathology act to limit the 'dynamic range' of reinstatement in these scene-selective regions, attenuating the relationship with memory performance that is strongly evident in younger adults. Of course, this and related accounts of the present findings offer no explanation for why neural selectivity in parahippocampal and retrosplenial cortex, but seemingly not in other category-selective regions, should be so sensitive to effects of age. Notably, it remains to be established whether this reflects

an anatomical dissociation in the effects of age on cortical integrity, or a functional dissociation in the effects of age on the processing of scenes as opposed to other categories of visual stimuli.

Consistent with prior reports (Gordon et al., 2014; Richey et al. 2013; Trelle et al., 2020) trial-wise estimates of retrieval-related hippocampal activity co-varied within-subjects with source memory accuracy for both faces and scenes. Of importance, these effects were invariant with respect to age, consistent with prior proposals that recollection-related hippocampal activity does not differ in magnitude with age and demonstrates an age-invariant relationship with memory performance (at least for cognitively unimpaired older adults) (de Chastelaine et al., 2016; Wang et al., 2016; but see Daselaar et al., 2006 for conflicting findings). On their face, the present hippocampal retrieval effects are consistent with the widely accepted notion that successful episodic retrieval depends on the hippocampally-mediated 'reactivation' of patterns of cortical activity encoded in the hippocampus as an episode was initially experienced (e.g., Alvarez & Squire, 1994; Norman & O'Reilly, 2003; Rugg et al., 2015). Nevertheless, for both stimulus categories, hippocampal activity and the strength of category-level reinstatement independently predicted memory performance. Moreover, retrieval-related hippocampal activity did not explain a significant fraction of the variance in category-level reinstatement. Thus, we found no evidence that cortical reinstatement of face or scene information mediated (even partially) the relationship between retrieval-related hippocampal activity and memory performance (cf., Gordon et al., 2014; Ritchey et al., 2013; Trelle et al., 2020). The reasons for the disparity between the present and prior findings are unclear. While it is possible that we may have identified a relationship with reinstatement had we employed more nuanced measures of retrieval-related hippocampal activity (e.g. at the level of individual cell fields) we note that

Gordon et al. (2014) and Trelle et al. (2020) both employed a very similar approach to that used here to define a whole hippocampus ROI.

We did however identify a robust age-invariant relationship between hippocampal activity and reinstatement of trial-unique scene information. One possibility is that retrieval-related hippocampal activity is preferentially involved with reinstatement of trial-unique information, rather than with reinstatement of category-level patterns of activity (Ritchey et al., 2013; Staresina & Wimber, 2019; but see also Gordon et al., 2014). This possibility is undermined, however, by our failure to identify any evidence that item-level reinstatement of scene (or face) information co-varied with memory performance. The relationship between retrieval-related hippocampal activity and item-level pattern similarity should therefore be interpreted with caution.

In conclusion, we identified robust age differences in retrieval-related reinstatement of face and scene information which could be explained by analogous differences in neural selectivity at encoding. Importantly, for both faces and scenes, the relationship between neural selectivity at encoding and strength of retrieval-related reinstatement did not differ with age. These findings suggest that, regardless of age, the selectivity with which events are neurally processed at the time of encoding contributes significantly to the strength of cortical reinstatement at retrieval. Future research will be required to determine whether the findings generalize to memory tests that operationalize successful retrieval of episodic memory differently than in the source memory procedures adopted in the present study.

Acknowledgments: We are grateful to Jake Braunwalder, Marianne de Chastelaine, and Saad Alghamdi for their assistance with data collection and our colleagues at the University of Texas Southwestern Medical Center Advanced Imaging Research Center for their assistance with fMRI data acquisition.

Funding: This research was supported by the National Science Foundation (grant number 1633873) and the National Institute on Aging (grant number RF1AG039103).

Declaration of Interest: none

References

Abdulrahman H, Fletcher PC, Bullmore E, Morcom AM. 2017. Dopamine and memory dedifferentiation in aging. NeuroImage. 153:211–220. https://doi.org/10.1016/j.neuroimage.2015.03.031

Alvarez P, Squire LR. 1994. Memory consolidation and the medial temporal lobe: A simple network model. Proceedings of the National Academy of Sciences. 91:7041–7045. https://doi.org/10.1073/pnas.91.15.7041

Andrade A, Paradis AL, Rouquette S, Poline JB. 1999. Ambiguous Results in Functional Neuroimaging Data Analysis Due to Covariate Correlation. NeuroImage. 10:483–486. https://doi.org/10.1006/nimg.1999.0479

Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using **lme4**. Journal of Statistical Software, 67. https://doi.org/10.18637/jss.v067.i01

Berron D, Neumann K, Maass A, Schütze H, Fliessbach K, Kiven V, Jessen F, Sauvage M, Kumaran D, Düzel E. 2018. Age-related functional changes in domain-specific medial temporal lobe pathways. Neurobiology of Aging. 65:86–97. https://doi.org/10.1016/j.neurobiolaging.2017.12.030

Bowman CR, Chamberlain JD, Dennis NA. 2019. Sensory Representations Supporting Memory Specificity: Age Effects on Behavioral and Neural Discriminability. The Journal of Neuroscience. 39:2265–2275. https://doi.org/10.1523/JNEUROSCI.2022-18.2019

Carp J, Park J, Polk TA, Park, DC. 2011. Age differences in neural distinctiveness revealed by multi-voxel pattern analysis. NeuroImage. 56:736–743. https://doi.org/10.1016/j.neuroimage.2010.04.267

Danker JF, Anderson JR. 2010. The ghosts of brain states past: Remembering reactivates the brain regions engaged during encoding. Psychological Bulletin. 136:87–102. https://doi.org/10.1037/a0017937

Davis T, LaRocque KF, Mumford JA, Norman KA, Wagner AD, Poldrack RA. 2014. What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. NeuroImage. 97:271–283. https://doi.org/10.1016/j.neuroimage.2014.04.037

de Chastelaine M, Wang TH, Minton B, Muftuler LT, Rugg MD. 2011. The Effects of Age, Memory Performance, and Callosal Integrity on the Neural Correlates of Successful Associative Encoding. Cerebral Cortex. 21:2166–2176. https://doi.org/10.1093/cercor/bhq294

de Chastelaine M, Mattson JT, Wang TH, Donley BE, Rugg MD. 2016. The neural correlates of recollection and retrieval monitoring: Relationships with age and recollection performance. NeuroImage. 138:164–175. https://doi.org/10.1016/j.neuroimage.2016.04.071

de Chastelaine M, Mattson JT, Wang TH, Donley BE, Rugg MD. 2017. Independent contributions of fMRI familiarity and novelty effects to recognition memory and their stability across the adult lifespan. NeuroImage. 156:340–351. https://doi.org/10.1016/j.neuroimage.2017.05.039

Dennis NA, Overman AA, Gerver CR, McGraw KE, Rowley MA, Salerno JM. 2019. Different types of associative encoding evoke differential processing in both younger and older adults: Evidence from univariate and multivariate analyses. Neuropsychologia. 135:107240. https://doi.org/10.1016/j.neuropsychologia.2019.107240

Epstein RA, Baker CI. 2019. Scene Perception in the Human Brain. Annual Review of Vision Science. 5:373–397. https://doi.org/10.1146/annurev-vision-091718-014809

Folville A, Bahri MA, Delhaye E, Salmon E, D'Argembeau A, Bastin C. 2020. Age-related differences in the neural correlates of vivid remembering. NeuroImage. 206:116336. https://doi.org/10.1016/j.neuroimage.2019.116336

Gordon AM, Rissman J, Kiani R, Wagner AD. 2014. Cortical Reinstatement Mediates the Relationship Between Content-Specific Encoding Activity and Subsequent Recollection Decisions. Cerebral Cortex. 24:3350–3364. https://doi.org/10.1093/cercor/bht194

Greenhouse SW, Geisser S. 1959. On methods in the analysis of profile data. Psychometrika. 24:95–112. https://doi.org/10.1007/BF02289823

Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. Science 293:2425–2430. https://doi.org/10.1126/science.1063736

Haxby JV, Gobbini MI. 2011. Distributed Neural Systems for Face Perception. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199559053.013.0006

Jackson O, Schacter DL. 2004. Encoding activity in anterior medial temporal lobe supports subsequent associative recognition. NeuroImage. 21:456–462. https://doi.org/10.1016/j.neuroimage.2003.09.050

Johnson JD, Rugg MD. 2007. Recollection and the Reinstatement of Encoding-Related Cortical Activity. Cerebral Cortex. 17:2507–2515. https://doi.org/10.1093/cercor/bhl156

Joliot M, Jobard G, Naveau M, Delcroix N, Petit L, Zago L, Crivello F, Mellet E, Mazoyer B, Tzourio-Mazoyer N. 2015. AICHA: An atlas of intrinsic connectivity of homotopic areas.

Journal of Neuroscience Methods. 254:46–59. https://doi.org/10.1016/j.jneumeth.2015.07.013

Kanwisher N, McDermott J, Chun MM. 1997. The fusiform face area: A module in human extrastriate cortex specialized for face perception. The Journal of Neuroscience. 17:4302-4311. https://doi.org/10.1523/JNEUROSCI.17-11-04302.1997

Keightley ML, Chiew KS, Anderson JAE, Grady CL. 2011. Neural correlates of recognition memory for emotional faces and scenes. Social Cognitive and Affective Neuroscience. 6:24–37. https://doi.org/10.1093/scan/nsq003

Koen JD, Hauck N, Rugg MD. 2019. The Relationship between Age, Neural Differentiation, and Memory Performance. The Journal of Neuroscience. 39:149–162. https://doi.org/10.1523/JNEUROSCI.1498-18.2018

Koen JD, Rugg MD. 2019. Neural Dedifferentiation in the Aging Brain. Trends in Cognitive Sciences. 23:547–559. https://doi.org/10.1016/j.tics.2019.04.012

Koen JD, Yonelinas AP. 2014. The Effects of Healthy Aging, Amnestic Mild Cognitive Impairment, and Alzheimer's Disease on Recollection and Familiarity: A Meta-Analytic Review. Neuropsychology Review. 24:332–354. https://doi.org/10.1007/s11065-014-9266-5

Kuhl BA, Chun MM. 2014. Successful Remembering Elicits Event-Specific Activity Patterns in Lateral Parietal Cortex. Journal of Neuroscience. 34:8051–8060. https://doi.org/10.1523/JNEUROSCI.4328-13.2014

Lenth R. 2018. Emmeans: Estimated marginal means, aka least-squares means. R package version 1.3.0. https://CRAN.R-project.org/package=emmeans

Li SC, Lindenberger U, Sikström S. 2001. Aging cognition: From neuromodulation to representation. Trends in Cognitive Sciences. 5:479–486. https://doi.org/10.1016/S1364-6613(00)01769-1

Li SC, Rieckmann A. 2014. Neuromodulation and aging: Implications of aging neuronal gain control on cognition. Current Opinion in Neurobiology. 29:148–158. https://doi.org/10.1016/j.conb.2014.07.009

Lin H, Li W, Carlson S. 2019. A Privileged Working Memory State and Potential Top-Down Modulation for Faces, Not Scenes. Frontiers in Human Neuroscience. 13:2. https://doi.org/10.3389/fnhum.2019.00002

Maass A, Berron D, Harrison TM, Adams JN, La Joie R, Baker S, Mellinger T, Bell RK, Swinnerton K, Inglis B, Rabinovici GD, Düzel E, Jagust WJ. 2019. Alzheimer's pathology targets distinct memory networks in the ageing brain. Brain. 142:2492–2509. https://doi.org/10.1093/brain/awz154

McDonough IM, Cervantes SN, Gray SJ, Gallo DA. 2014. Memory's aging echo: Age-related decline in neural reactivation of perceptual details during recollection. NeuroImage. 98:346–358. https://doi.org/10.1016/j.neuroimage.2014.05.012

Mumford JA, Davis T, Poldrack RA. 2014. The impact of study design on pattern estimation for single-trial multivariate pattern analysis. NeuroImage. 103:130–138. https://doi.org/10.1016/j.neuroimage.2014.09.026

Nilsson LG. 2003. Memory function in normal aging. Acta Neurologica Scandinavica. 107:7-13. https://doi.org/10.1034/j.1600-0404.107.s179.5.x

Norman KA, O'Reilly RC. 2003. Modeling hippocampal and neocortical contributions to recognition memory: A complementary-learning-systems approach. Psychological Review. 110:611–646. https://doi.org/10.1037/0033-295X.110.4.611

Nyberg L, Habib R, McIntosh AR, Tulving E. 2000. Reactivation of encoding-related brain activity during memory retrieval. Proceedings of the National Academy of Sciences. 97:11120–11124. https://doi.org/10.1073/pnas.97.20.11120

Old SR, Naveh-Benjamin M. 2008. Differential effects of age on item and associative measures of memory: A meta-analysis. Psychology and Aging. 23:104–118. https://doi.org/10.1037/0882-7974.23.1.104

Park DC, Polk TA, Park R, Minear M, Savage A, Smith MR. 2004. Aging reduces neural specialization in ventral visual cortex. Proceedings of the National Academy of Sciences. 101:13091-13095. doi: 10.1073/pnas.0405148101

Park DC, Smith AD, Lautenschlager G, Earles JL, Frieske D, Zwahr M, Gaines CL. 1996. Mediators of Long-Term Memory Performance Across the Life Span. Psychology of Aging. 11:621-637. doi: 10.1037//0882-7974.11.4.621.

Park H, Uncapher MR, Rugg MD. 2008. Effects of study task on the neural correlates of source encoding. Learning & Memory. 15:417–425. https://doi.org/10.1101/lm.878908

Park J, Carp J, Kennedy KM, Rodrigue KM, Bischof GN, Huang CM, Rieck JR, Polk TA, Park DC. 2012. Neural Broadening or Neural Attenuation? Investigating Age-Related Dedifferentiation in the Face Network in a Large Lifespan Sample. Journal of Neuroscience. 32:2154–2158. https://doi.org/10.1523/JNEUROSCI.4494-11.2012

Rissman J, Gazzaley A, D'Esposito M. 2004. Measuring functional connectivity during distinct stages of a cognitive task. NeuroImage. 23:752–763. https://doi.org/10.1016/j.neuroimage.2004.06.035

Rissman J, Wagner AD. 2012. Distributed Representations in Memory: Insights from Functional Brain Imaging. Annual Review of Psychology. 63:101–128. https://doi.org/10.1146/annurev-psych-120710-100344

Ritchey M, Wing EA, LaBar KS, Cabeza R. 2013. Neural Similarity Between Encoding and Retrieval is Related to Memory Via Hippocampal Interactions. Cerebral Cortex. 23:2818–2828. https://doi.org/10.1093/cercor/bhs258

Rugg MD, Johnson JD, Uncapher MR. 2015. Encoding and Retrieval in Episodic Memory: Insights from fMRI. In D. R. Addis, M. Barense, & A. Duarte (Eds.), The Wiley Handbook on the Cognitive Neuroscience of Memory (pp. 84–107). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118332634.ch5

Sato W, Yoshikawa S. 2013. Recognition Memory for Faces and Scenes. The Journal of General Psychology. 140(1):1–15. https://doi.org/10.1080/00221309.2012.710275

Silson EH, Steel A, Kidder A, Gilmore AW, Baker CI. 2019. Distinct subdivisions of human medial parietal cortex support recollection of people and places, eLife. 8:e47391. https://doi.org/10.7554/eLife.47391.001

Singmann H, Bolker B, Westfall J, Aust F. 2016. Afex: Analysis of factorial experiments. R package version 0.22-1. https://CRAN.R-project.org/package=afex

Snodgrass JG, Corwin J. 1988. Pragmatics of Measuring Recognition Memory: Applications to Dementia and Amnesia. J Exp Psychol Gen. 117:34-50. doi: 10.1037//0096-3445.117.1.34.

Srokova S, Hill PF, Koen JD, King DR, Rugg MD. 2020. Neural differentiation is moderated by age in scene-selective but not face-selective cortical regions. eNeuro. 7:ENEURO.0142-20.2020 doi: 10.1523/ENEURO.0142-20.2020

Staresina BP, Wimber M. 2019. A Neural Chronometry of Memory Recall. Trends in Cognitive Sciences. 23:1071–1085. https://doi.org/10.1016/j.tics.2019.09.011

St-Laurent M, Abdi H, Bondad A, Buchsbaum BR. 2014. Memory Reactivation in Healthy Aging: Evidence of Stimulus-Specific Dedifferentiation. Journal of Neuroscience. 34:4175–4186. https://doi.org/10.1523/JNEUROSCI.3054-13.2014

St-Laurent M, Buchsbaum BR. 2019. How Multiple Retrievals Affect Neural Reactivation in Young and Older Adults. The Journals of Gerontology: Series B. 74:1086–1100. https://doi.org/10.1093/geronb/gbz075

Thakral PP, Wang TH, Rugg MD. 2017. Decoding the content of recollection within the core recollection network and beyond. Cortex. 91:101–113. https://doi.org/10.1016/j.cortex.2016.12.011

Thakral PP, Wang TH, Rugg MD. 2019. Effects of age on across-participant variability of cortical reinstatement effects. NeuroImage. 191:162–175. https://doi.org/10.1016/j.neuroimage.2019.02.005

Trelle AN, Carr VA, Guerin SA, Thieu MK, Jayakumar M, Guo W, Nadiadwala A, Corso NK, Hunt MP, Litovsky CP, Tanner NJ, Deutsch GK, Bernstein JD, Harrison MB, Khazenzon AM, Jiang J, Sha SJ, Fredericks CA, Rutt B K, … Wagner AD. 2020. Hippocampal and cortical mechanisms at retrieval explain variability in episodic remembering in older adults. eLife. 9:e55335. https://doi.org/10.7554/eLife.55335

Trelle AN, Henson RN, Simons JS. 2019. Neural evidence for age-related differences in representational quality and strategic retrieval processes. Neurobiology of Aging. 84:50–60. https://doi.org/10.1016/j.neurobiolaging.2019.07.012

Voss MW, Erickson KI, Chaddock L, Prakash RS, Colcombe SJ, Morris KS, Doerksen S, Hu L, McAuley E, Kramer AF. 2008. Dedifferentiation in the visual cortex: An fMRI investigation of individual differences in older adults. Brain Research. 1244:121–131. https://doi.org/10.1016/j.brainres.2008.09.051

Wang TH, Johnson JD, de Chastelaine M, Donley BE, Rugg MD. 2016. The Effects of Age on the Neural Correlates of Recollection Success, Recollection-Related Cortical Reinstatement, and Post-Retrieval Monitoring. Cerebral Cortex. 26:1698–1714. https://doi.org/10.1093/cercor/bhu333

Wheeler ME, Petersen SE, Buckner RL. 2000. Memory's echo: Vivid remembering reactivates sensory-specific cortex. Proceedings of the National Academy of Sciences. 97:11125–11129. https://doi.org/10.1073/pnas.97.20.11125

Wig GS, Laumann TO, Petersen SE. 2014. An approach for parcellating human cortical areas using resting-state correlations. NeuroImage. 93:276–291. https://doi.org/10.1016/j.neuroimage.2013.07.035

Wing EA, Ritchey M, Cabeza R. 2015. Reinstatement of Individual Past Events Revealed by the Similarity of Distributed Activation Patterns during Encoding and Retrieval. Journal of Cognitive Neuroscience. 27:679–691. https://doi.org/10.1162/jocn_a_00740

Xue G. 2018. The Neural Representations Underlying Human Episodic Memory. Trends in Cognitive Sciences. 22:544–561. https://doi.org/10.1016/j.tics.2018.03.004

Zheng L, Gao Z, Xiao X, Ye Z, Chen C, Xue G. 2018. Reduced Fidelity of Neural Representation Underlies Episodic Memory Decline in Normal Aging. Cerebral Cortex. 28:2283–2296. https://doi.org/10.1093/cercor/bhx130