Effects of age on across-participant variability of cortical reinstatement effects

Preston P. Thakral[a,*], **Tracy H. Wang**[b], and **Michael D. Rugg**[c]

[a]Department of Psychology, Harvard University, USA

[b]Department of Psychology, University of Texas at Austin, USA

[c]Center for Vital Longevity and School of Behavioral and Brain Sciences, University of Texas at Dallas, USA

Abstract

Using functional magnetic resonance imaging data, we assessed whether across-participant variability of content-selective retrieval-related neural activity differs with age. We addressed this question by employing across-participant multi-voxel pattern analysis (MVPA), predicting that increasing age would be associated with reduced variability of retrieval-related cortical reinstatement across participants. During study, 24 young and 24 older participants viewed objects and concrete words. Test items comprised studied words, names of studied objects, and unstudied words. Participants judged whether the items were recollected, familiar, or new by making 'Remember', 'Know' and 'New' responses, respectively. MVPA was conducted on each region belonging to the 'core recollection network', dorsolateral prefrontal cortex, and a previously identified content-selective voxel set. A leave-one-participant-out classification approach was employed whereby a classifier was trained on a subset of participants and tested on the data from a yoked pair of held-out participants. Classifiers were trained on the study phase data to discriminate the study trials as a function of content (picture or word). The classifiers were then applied to the test phase data to discriminate studied test words according to their study condition. In all of the examined regions, classifier performance demonstrated little or no sensitivity to age and, for the test data, was robustly above chance. Thus, there was little evidence to support the hypothesis that across-participant variability of retrieval-related cortical reinstatement differs with age. The findings extend prior evidence by demonstrating that content-selective cortical reinstatement is sufficiently invariant to support across-participant multi-voxel classification across the healthy adult lifespan.

Keywords

Episodic memory; Recollection; Familiarity; Aging; MVPA; fMRI

---

*Corresponding author. Department of Psychology, Harvard University, Cambridge, MA, 02138, USA. prestonthakral@fas.harvard.edu (P.P. Thakral).

1. Introduction

Episodic memory, the ability to recollect specific details about a unique event (Tulving, 1985), is widely held to depend on the reinstatement of neural processes that were active when the event was experienced (e.g., Alvarez and Squire, 1994; Marr, 1971; Norman and O'Reilly, 2003). A wealth of functional magnetic resonance imaging (fMRI) studies have provided support for this idea by demonstrating retrieval-related 'cortical reinstatement' (for reviews, see Danker and Anderson, 2010; Rissman and Wagner, 2012; Rugg et al., 2015; Xue, 2018). For example, fMRI studies employing univariate analyses have demonstrated that the regions selectively engaged during the encoding of specific episodic information overlap regions engaged during subsequent recollection of that information (e.g., Kahn et al., 2004; Johnson and Rugg, 2007; Thakral et al., 2015). Other studies have demonstrated reinstatement with multi-voxel pattern analyses (MVPA), reporting similarity between the patterning of blood oxygenation level-dependent (BOLD) activity occurring during the encoding and subsequent retrieval of episodic information (e.g., Johnson et al., 2009; Staresina et al., 2012; Gordon et al., 2014; Kuhl and Chun, 2014; Wing et al., 2015).

A sizeable number of fMRI studies have examined age-related differences in the selectivity with which different cortical regions respond to different stimulus categories (e.g., Voss et al., 2008; Carp et al., 2011a, 2011b; Grady et al., 2011; Park et al., 2004, 2010, 2012; Du et al., 2016; Kleemeyer et al., 2017; Koen et al., 2019). A consistent finding from these studies is that, for some stimulus categories and cortical regions (e.g., activity elicited by visual scenes in the 'parahippocampal place area', but not by visual objects in the 'lateral occipital complex', Koen et al., 2019; see also Voss et al., 2008 for additional evidence of material and regional specificity of age-related differences in neural selectivity), neural activity becomes less category-selective with increasing age (a phenomenon sometimes referred to as 'age-related neural dedifferentiation'). These findings have been interpreted as evidence that the precision with which information is represented in the cortex declines with age, and that this may play a role in age-related cognitive decline (e.g., Li et al., 2001; Li and Rieckmann, 2014).

In contrast with the substantial body of research examining age-related neural dedifferentiation, only a handful of prior fMRI studies have examined the related question of whether retrieval-related cortical reinstatement - operationalized as strength of overlap between content-selective activity at study and test (see above) - varies across the lifespan. This question is important, since it might shed light on the well-documented finding that episodic memory performance declines with advancing age (for reviews, see Nilsson, 2003; Dodson, 2017). The hypothesis motivating these prior studies (e.g. Wang et al., 2016) is that one cause (or consequence) of age-related memory decline is a weakening in the fidelity of reinstatement, leading to a reduction in the accuracy of recollection-related memory judgments. Findings from three studies (McDonough et al., 2014; St-Laurent et al., 2014; Abdulrahman et al., 2017) are arguably consistent with this hypothesis (but see Wang et al., 2016 for critiques of the first two studies). By contrast, findings from a prior experiment in our laboratory (Wang et al., 2016; Thakral et al., 2017) failed to find supporting evidence. In this experiment, older and younger participants first encoded a set of objects and concrete words. In the subsequent retrieval phase, test items comprised unstudied words, studied

words and the names of the studied objects, with the requirement to signal whether each test item was recollected, familiar, or new (Tulving, 1985). Regardless of whether reinstatement was assessed with univariate analysis or classifier-based MVPA, there was no evidence of weaker reinstatement in older relative to young adults.

One potential caveat to the interpretation of this finding is that null effects of age were reported not only with respect to retrieval-related reinstatement, but also for the selectivity of the neural activity elicited by the two classes of study trials. If there is a correspondence between the fidelity with which study events are represented at the time of encoding, and the fidelity with the events when 're-represented' if they are later successfully retrieved from memory, age differences in the strength of reinstatement would be evident only for study items that demonstrated age-related dedifferentiation at the time of encoding, and then, only in those cortical regions where the dedifferentiation was evident. The sparse evidence relevant to this issue does not support this assumption, however. St. Laurent et al. (2014) required young and older participants to repeatedly view and recall a small set of short movie clips. Employing whole brain 'searchlight' MVPA, they found essentially no evidence of age-related differences in neural selectivity for the study events, along with widespread evidence of age-related decline in the strength of retrieval-related reinstatement. As the authors acknowledged, interpretation of these findings is complicated by the employment of repeated study-test cycles. Nonetheless, the findings suggest that age-related differences in retrieval-related cortical reinstatement do not depend upon corresponding differences in neural selectivity at the time of encoding.

Here, we report the results of a further analysis of the Wang et al. (2016) data-set that examines age-related differences in retrieval-related cortical reinstatement from a quite different perspective. In our prior reports, classifier-based MVPA was conducted within-participants, as is usual for this approach (for a review, see Rissman and Wagner, 2012). In the analysis presented here, however, we adopt an across-participants approach (cf., Rissman et al., 2010, 2016; Richter et al., 2016), assessing the extent to which reinstatement can be identified in held-out participants using a classifier trained on study data from the remainder of the sample. In this case, classifier performance is dependent on the consistency with which reinstatement effects are manifest at the voxel level not only across trials, but across participants also. Thus, other things being equal, the accuracy of a classifier trained on across-participant data will be inversely proportional to the variability across participants in the patterning of neural activity within the population of voxels in question. The across-participant approach therefore provides a direct assessment of the similarity (or, equivalently, the idiosyncrasy) of the distributed activity elicited by stimulus events in the brains of a sample of individuals.

Our goal here in employing the across-participants approach to MVPA classification was to examine a different prediction about age-related differences in cortical reinstatement effects to that examined in the studies cited above. The prediction is motivated both by theoretical proposals, alluded to above, that brains become less 'complex' and 'differentiated' with increasing age, resulting in neural representations with lower dimensionality and specificity (e.g., Li et al., 2001; Li and Sikstrom, 2002; Moran et al., 2014), and by empirical findings indicating that there is an increase with age in the level of abstraction ('gist') of retrieved

episodic memories, with a concomitant loss of specific episodic detail (e.g., Norman and Schacter, 1997; Pierce et al., 2005; Brainerd and Reyna, 2015). These proposals and findings converge to suggest that retrieved episodic information, and hence any associated cortical reinstatement effects, should become less idiosyncratic, and therefore less variable across participants, with increasing age. For the reasons outlined above, this leads to the prediction that the accuracy of across participant classification of retrieved (reinstated) content should be greater among older than among young participants. We tested this prediction by examining across-participant classification accuracy of retrieved information both in regions of interest (ROIs) belonging to the 'core recollection network' (Rugg and Vilberg, 2013) (where we recently reported robust, age-invariant reinstatement effects; Thakral et al., 2017), and in the voxel set employed in our original report of null effects of age on retrieval-related cortical reinstatement (Wang et al., 2016).

2. Materials and methods

Data from the experiment described below have been reported in three prior papers (King et al., 2015; Wang et al., 2016; Thakral et al., 2017). The outcomes of the principal analyses reported here have not been described previously.

2.1. Participants

The experiment was approved by the Institutional Review Boards of the University of Texas at Dallas and the University of Texas, Southwestern Medical Center. Informed consent was obtained prior to participation. A total of 48 cognitively healthy participants were included in the analyses (24 older participants (13 female; mean age 68 years) and 24 younger participants (13 female; mean age 24 years). All participants reported themselves to have normal or corrected-to-normal vision and to be right-handed. Participants were administered a neuropsychological test battery no more than six months in advance of the scanning session. A description of the inclusion and exclusion criteria employed for participant recruitment, the number of and the reasons for exclusions of data-sets can be found in Wang et al. (2016), and the demographic and neuropsychological characteristics of the participants in each age group can be found in Supplemental Table 1.

2.2. Stimulus materials

Stimuli comprised 216 colored pictures of common objects and their names. The objects were drawn from Hemera Photo Objects 50,000 vol 2 (http://www.hemera.com/index.html). The names of each picture were 3–12 letters in length and had a mean frequency of 14.3 counts/million (Kucera and Francis, 1967). For each participant, the 216 picture-name pairs were randomly sorted. A study list was created using 144 picture-name pairings. The study list was then subdivided into three sub-lists of equal length, one for each fMRI study session. In each sub-list, half of the items were presented as pictures and half were shown as names (24 pictures and 24 names). A test list was created using the words from the 144 picture-name pairings presented during study and 72 new words. This test list was subdivided into three sub-lists, one for each fMRI test session. Each test sub-list contained a pseudorandom ordering of 48 old words (24 words previously presented as pictures and 24 old words) and 24 new words.

2.3. Experimental procedures

During study, participants performed one of two tasks (Fig. 1A). For pictures, participants judged whether the object would fit into a shoebox. For words, participants judged whether the object denoted by the word would more likely be found inside or outside a house. The rationale for employing two different study tasks for each class of studied content was to elicit maximally distinctive patterns of cortical activity associated with each class of studied content, making them suitable for the identification of retrieval-related cortical reinstatement effects (see also, Johnson et al., 2009; Leiker and Johnson, 2015; Wang et al., 2016). Study items were presented in 'mini-blocks' which comprised three consecutive word or three consecutive picture trials (cf., Johnson et al., 2009; McDuff et al., 2009). We presented study items as mini-blocks to maximally segregate BOLD activity associated with each class of studied content (i.e., by taking advantage of the summing of the BOLD signal across successive trials within a mini-block; for similar procedures see, Johnson et al., 2009; McDuff et al., 2009). This procedure stands in contrast to a slow-event-related design which would have entailed a prolonged scanning session. Each mini-block began with the presentation of a red fixation cross presented at the center of the screen for 500 ms. This was replaced by a task cue for 2 s ('Used inside of a house?' or 'Fits inside a shoebox?'). Pictures or words were then presented for 2 s each with an inter-stimulus interval of 2 s. Responses were made with the left and right index fingers. Finger assignment was counterbalanced across participants.

After the three study sessions and before beginning the first test session, participants practiced the test task while remaining in the scanner (Fig. 1B). The practice test task lasted approximately 5 min. Each test trial began with the presentation of a red fixation cross for 500 ms followed by the presentation of a word for 3 s. Following word offset, a black fixation cross was presented for a variable duration (~66% of trials at 4.5 s, ~25% of trials at 6.5 s, and ~9% of trials at 8.5 s). A Remember/ Know/New task was employed (Tulving, 1985). A 'Remember' response was to be made when recognition of the test word was accompanied by retrieval of at least one specific detail from the study episode. A 'Know' response was to be made when they were confident that a test item was old but could not recollect any specific details from the study episode. A 'New' response was to be made when a test item was judged to not be part of the study list or when participants were unconfident with respect to the study status of the test item. 'Remember', 'Know', and 'New' responses were made with the index and middle fingers of one hand, and the index finger of the opposite hand. Finger assignment was counterbalanced across participants.

2.4. Image acquisition and analysis

Functional and anatomic images were acquired with a 3 T Phillips Achieva MRI scanner (Philips Medical System, Andover, MA, USA) using a 32-channel head coil. Anatomic images were acquired with a magnetization-prepared rapid gradient echo sequence (matrix size of 220×193, 150 slices, voxel size of 1 mm^3). Functional images were acquired with an echo-planar imaging sequence (SENSE factor of 2, TR = 2 s, TE = 30 ms, flip angle of 70°, field-of-view of 240×240, matrix size of 80×79, 30 slices, 3 mm slice thickness, 1 mm gap). Slices were acquired in ascending order and oriented parallel to the anteriorposterior

commissure plane. For each study and test session, 170 and 338 vol were acquired, respectively.

Univariate analysis was conducted using Statistical Parametric Mapping (SPM8, Wellcome Department of Cognitive Neurology, London, UK). MVPA was conducted using the Princeton MVPA Toolbox (https://code.google.com/p/princeton-mvpa-toolbox/) and custom MATLAB scripts. Functional image preprocessing included two-stage spatial realignment, slice-time correction, and normalization to an age group-specific template (for fuller description, see De Chastelaine et al., 2011; Mattson et al., 2014). The normalized images were smoothed with an 8 mm Gaussian kernel. The time series in each voxel was high-pass filtered at 1/128 Hz and scaled to a constant mean within session. Anatomic images were normalized using an analogous procedure as that employed for the functional images.

2.5. Univariate analysis

Univariate analysis was conducted in a two stage mixed effects model. In the first stage, neural activity was modeled at stimulus onset by a delta function (i.e., picture or word onset at study and word onset at test). The associated BOLD response was modeled by convolving the delta functions with a canonical hemodynamic response function and its temporal and dispersion derivatives (Friston et al., 1998). Events of interest for the analysis of the study data were picture and word trials (an additional trial of no interest included failures to respond and filler trials). Events of interest for the analysis of the test data were Remember-picture trials (i.e., 'Remember' responses given to test words previously presented as pictures), Remember-word trials (i.e., 'Remember' responses given to test words previously presented as words), and Know trials ('Know' responses were collapsed across study content due to low trial numbers). Three further categories of test events comprised correct rejections (new words correctly judged as 'New'), misses (studied items incorrectly judged to be 'New'), and an event of no-interest (false alarms, filler trials, and failures to respond). Six regressors representing movement related variance (three for rotation and three for rigid-body translation) and regressors modeling each scan session were also entered into the design matrix.

In the second stage of the analysis, linear contrasts were performed on the aforementioned parameter estimates treating participants as a random effect. To identify recollection effects common to both classes of study content, we inclusively masked the contrast between all test items (i.e., regardless of study history) endorsed Remember versus those endorsed Know (Remember > Know, $p < 0.01$, after FWE, with a 22 voxel cluster extent threshold) with the separate contrasts for each class of study content (Remember-word > Know and Remember-pic-ture > Know, each contrast at $p < 0.01$, uncorrected). The subsidiary contrasts played no role in the identification of the primary recollection effect, and were employed solely for the purpose of inclusive masking to insure that in no ROI was the main effect of recollection driven by recollection of only one of the two study conditions but reliable for both classes of study content.

2.6.   Feature selection for MVPA

MVPA was conducted on each region belonging to the 'core recollection network', and dorsolateral prefrontal cortex (i.e., 6 ROIs; Thakral et al., 2017), and a previously identified content-selective voxel set (Wang et al., 2016). Four regions were identified by the univariate analysis described above to identify peak clusters in regions of the 'core recollection network' demonstrating recollection effects. The regions comprised the left angular gyrus (AG; peak MNI coordinates: 45, −76, and 28; 620 voxels), medial prefrontal cortex (MPFC; peak MNI coordinates: 6, 65 and 13; 775 voxels), left middle temporal gyrus (MTG; peak MNI coordinates: 63, −46, and −14; 279 voxels), and left retro-splenial/posterior cingulate cortex (RSP/PCC; peak MNI coordinates: 15, −55, and 34; 337 voxels). An 85 voxel cluster was also identified in the right hippocampus (HIPP; peak MNI coordinates: 21, −7, and −23). As in our original MVPA of these ROIs (Thakral et al., 2017), given the small size of the hippocampal cluster, we opted to employ an anatomically defined bilateral anterior HIPP mask (Frisoni et al., 2015) comprising 283 voxels (we note that findings for the functionally defined hippocampal ROI did not differ from those reported below for the anatomically defined region). In contrast to our prior analyses of these data-sets, which utilized within-participant MVPA and ROIs that were separately identified for each participant (Wang et al., 2016; Thakral et al., 2017), here ROIs were identified by univariate analyses conducted across all participants, giving a single voxel set for each ROI. We also included an additional ROI, the left dorsolateral prefrontal cortex (DLPFC), to assess whether we could extend our original within-participant MVPA findings, which revealed robust cortical reinstatement effects in the region (see, Thakral et al., 2017). The DLPFC ROI was identified using the WFU PickAtlas v3.0 (Maldjian et al., 2003). Voxels in each ROI were restricted to gray matter by inclusively masking them with the default SPM gray matter probability map thresholded at $p > 0.2$.

In addition to the 6 ROIs described above, we employed the univariate approach described in Wang et al. (2016; see also Johnson and Rugg, 2007) to identify a feature set that manifested content-selective retrieval effects (Fig. 2B). The voxel set was identified by a series of masked contrasts. To identify picture-selective retrieval, the directional contrast of picture > word trials at study (thresholded at $p < 0.001$) was inclusively masked with the contrast identifying picture recollection effects (Remember-picture > Know; $p < 0.01$), and then exclusively masked with the contrast identifying word recollection effects (Remember--word > Know, $p < 0.05$), leaving a voxel set manifesting picture-selective, univariate retrieval effects. The analogous procedure was employed to identify univariate word-selective effects. A 1000 voxel ROI was formed by combining the top ranked (as ordered by Z value) 500 voxels derived from each of the two recollection contrasts that survived the inclusive and exclusive masking procedures. The rationale for combining the two classes of univariate cortical reinstatement effects (i.e., picture- and word-selective) into a single voxel set derived from our aim to examine whether MVPA classifiers applied to voxels manifesting content-selective univariate recollection effects differed in their accuracy as a function of age group or test judgment (Remember vs. Know), as well as to directly extend our prior findings for the same feature set (Wang et al., 2016). It should be noted that the approach of combining differentially-selective voxels into a single feature set was adopted not only in the initial analysis of the current data-set (Wang et al., 2016), but also in several other prior

studies of retrieval-related reinstatement (e.g. Johnson et al., 2009; Gordon et al., 2014; Koen and Rugg, 2016). The resulting content-selective ROI is illustrated in Fig. 2B. For a list of the peaks of clusters associated with each class of content-selective reinstatement effect see Supplemental Table 2.

2.7. MVPA

Functional image preprocessing prior to MVPA was conducted as described above with the exception of the spatial smoothing step. For each ROI, the data were then de-trended to remove linear and quadratic trends, and z-scored across volumes within each session. Estimates of the BOLD signal for each trial were obtained by averaging the signal from TRs 4–5 following stimulus onset (this time range corresponds to the period encompassing the peak of the evoked hemodynamic response; see Wang et al., 2016). To reduce possible reaction time (RT)-related confounds across trials, the single-trial BOLD signals for each voxel were regressed on the corresponding RTs separately for the study and test data (Todd et al., 2013). The resulting single trial BOLD estimates were z-scored across the trials belonging to the same study or response category, and then across voxels, eliminating any condition-dependent differences in mean BOLD signal within the ROI (for similar procedures, see Kuhl et al., 2013; Kuhl and Chun, 2014; Koen and Rugg, 2016).

Linear classifiers were employed to examine whether studied content could be decoded from patterns of across participant neural activity within the seven aforementioned ROIs (AG, MPFC, MTG, RSP/PCC, HIPP, DLPFC, and the 'distributed', content-sensitive ROI). We employed a leave-one-participant-out cross validation approach in which the classifier was trained on a subset of the participants, and then tested on the data from held-out participants (see, Rissman et al., 2010, 2016; Richter et al., 2016 for examples of this approach). An iterative procedure was employed whereby each participant's data was 'held-out' once. Three separate sets of classifiers were employed: one set was trained on data pooled across age-groups (across-group classifiers), a second set was trained on data from young participants only (young-only classifiers), and the final set was trained on data from older participants only (old-only classifiers). The across-group classifier was employed to assess the ability to decode content across participants (i.e., accuracy of the across-group classifier reflects consistency of content-dependent differences in patterns of activity across all participants). The age-specific classifiers allowed us to directly assess whether the two age groups differed in the extent to which the patterns of activity demonstrated similarity across participants. Classification was implemented with regularized logistic regression (L2) using a penalty parameter (λ) of 0.05[1] (cf., Wang et al., 2016; Thakral et al., 2017).

2.7.1. Study-study classification—For study-study classification, the across-group classifiers were trained on a total of 2208 study trials from 23 older and 23 younger participants (1104 picture trials and 1104 word trials) and tested on 48 study trials (24

---

[1]The regularization parameter of 0.05 was chosen to be consistent with our two prior MVPA analyses of the same data (Wang et al., 2016; Thakral et al.,2017). However, a reviewer noted that our findings could be idiosyncratic to this specific parameter. To examine this possibility, we re-ran the age-specific classification analyses for the AG and content-selective ROIs using a range of parameters (0.01, 0.05, and 0.5). The results of these analyses, which are available from the first author upon request, revealed that the only reliable effect of this manipulation was on overall study-test classifier accuracy (with accuracy slightly but significantly co-varying with the magnitude of the parameter). Crucially, in no analyses did the factor of age group interact with the parameter manipulation

picture and 24 word) from the two held-out participants. Note that each study trial corresponds to a single mini-block (i.e., a miniblock of 3 study items belonging to the same category, picture or word; see *2.3 Experimental procedures*). The leave-two-participant-out approach allowed us to identify age-related differences in classifier performance where the classifier for each yoked pair of younger and older participants had been trained on the same data (i.e., the classifier was trained on each of the 23 yoked pairs and iteratively tested on the left-out pair). The young-only classifier was trained on 1104 trials from 23 young participants (552 picture trials and 552 word trials), and iteratively tested on each held-out younger participant and a yoked older participant. Analogously, the old-only classifier was trained on 1104 trials from 23 older participants (552 picture trials and 552 word trials), and iteratively tested on each held-out older participant and a yoked younger participant. Classification was deemed accurate if the returned classifier evidence for the correct study category was greater than 0.5 (chance). Accuracy was binarized to give a score of 1 for correct and 0 for incorrect classification. Equivalent findings were obtained when the continuous variable of classifier evidence, rather than binarized accuracy, was analyzed, alleviating concerns about possible classifier bias (results available from the first author upon request).

2.7.2.   Study-test classification—For study-test classification, the across-group, young-only, and old-only study phase classifiers were separately tested for the ability to discriminate Remember-picture from Remember-word trials, and Know-picture from Know-word trials (cf., Wang et al., 2016; Thakral et al., 2017). Classifier accuracy was compared as a function of response category (Remember vs. Know). The mean (standard deviation) number of trials contributing to the classification of Remember-picture and Remember-word trials were 44 (12.6) and 39 (14.3), respectively. The mean (standard deviation) number of trials contributing to the classification of Know-picture and Know-word trials were 13 (7.5) and 22 (11.1), respectively. As in the original MVPA analysis of the same data (Thakral et al., 2017), to ensure that the performance of the study phase classifier was not biased by unequal numbers of test trials from each study condition, we randomly subsampled from the response category with the largest number of trials to equate trial numbers, using 10 iterations of the sub-sampling procedures. Therefore, for each participant, the number of Remember-word and Remember-picture trials was matched, as was the number of Know-word and Know-picture trials. The 10 sets of matched trials were used to assess performance of the three study phase classifiers (i.e., across-group, young-only, and old-only study phase classifiers). For each participant, mean classifier accuracy was estimated across all 10 iterations.

2.8.   Statistical analyses

The principal statistical analyses were conducted using mixed effects analyses of variance (ANOVA), followed up as necessary with additional repeated measures ANOVAs and pairwise t-tests. Although we report all results from each ANOVA, we focus on interactions involving the factors of age group and classifier, as these motivated the subsidiary ANOVAs and t-tests. For all significant results, we report the relevant effect sizes (d_z in the case of within-participant t-tests and partial η^2 for F-tests; see, Lakens, 2013). Nonsphericity between the levels of repeated-measures factors in the ANOVAs was corrected with the

Greenhouse-Geisser procedure. For any non-significant ANOVA results, Bayes factors were computed using JASP software (http://jasp-stats.org/; Ly et al., 2018; Wagenmakers et al., 2018) to estimate the strength of evidence favoring the null hypothesis (Jeffreys, 1961; Dienes, 2014). We report these values as BF_{10}, with values < 1 indicating substantial evidence for the null hypothesis and values near 1 indicating little or no evidence in favor of either the alternative or the null hypothesis. For multi-voxel analyses, one-tailed t-tests were employed to assess whether classifier accuracy exceeded the chance value of 0.5 (permutation analyses conducted for the AG ROI data revealed no evidence of classifier bias, with chance accuracy for test trial classification differing minimally from 0.5; the results of these analyses are available from the first author upon request). Consistent with our prior reporting procedures (Thakral et al., 2017), we report results for classification accuracy both before and after correction for multiple comparisons.

3. Results

3.1. Behavioral results

The behavioral results were first reported in Wang et al. (2016). Table 1 lists the strength of recollection and familiarity estimated from the proportion of Remember and Know responses (i.e., pR and pF, respectively). Strength of recollection (pR) was calculated as the probability of a Remember response to a studied item minus the probability of a Remember response to a new item. Familiarity strength (pF) was calculated in an analogous fashion to pR but corrected assuming independence between Remember and Know responses (Yonelinas and Jacoby, 1995). An ANOVA on the recollection estimates with factors of content and age revealed main effects of content (F(1, 46) = 5.78, p = 0.02, partial η^2 = 0.11) and age (F(1, 46) = 4.75, p = 0.04, partial η^2 = 0.09), with no significant interaction (F < 1, BF_{10} = 0.30). These results reflected higher pR for picture relative to word trials and for young relative to old participants, respectively. An ANOVA on the familiarity estimates revealed a main effect of content (F(1, 46) = 74.53, p = 3.51×10^{-11}, partial η^2 = 0.62) with no significant main effect of age or content by age interaction (Fs < 1.71, ps > 0.20, BFs_{10} < 0.58). These results reflected higher pF for word relative to picture trials.

Table 1 also lists the reaction times for each age group, memory response, and content. An ANOVA on the RTs with factors, content, age, and memory response revealed significant main effects of content (F(1, 46) = 18.51, p = 8.70×10^{-5}, partial η^2 = 0.29) and memory response (F(1, 46) = 90.98, p = 1.79×10^{-12}, partial η^2 = 0.66), as well as a content by response interaction (F(1,46) = 9.25, p = 3.88×10^{-3}, partial η^2 = 0.16). The interaction was driven by the longer reaction times for picture than word trials given a Know response (p = 2.08×10^{-4}, d_z = 0.58), with no significant difference between picture and word trials given a Remember response (p = 0.59, BF_{10} = 0.18). The ANOVA failed to reveal a main effect of age, a response by age interaction or a threeway interaction (Fs < 1, BFs_{10} < 0.41).

3.2. fMRI results

3.2.1. Study-study classification, across-group classifier—We first examined study-study classification collapsed across age group to determine whether the across-participants MVPA would replicate our earlier within-participants findings (the outcome of

each test was considered significant if the p value survived Bonferroni correction for multiple comparisons; 0.05/12 tests (i.e., 6 regions by 2 age groups), yielding a corrected $p = 4.17 \times 10^{-3}$; Thakral et al., 2017). Study-study classification was robustly above chance in every region (Table 2A, Fig. 3A, top). When examining each age group separately, as is evident from Table 2A and Fig. 3A (bottom), study-study classification in young participants was significantly above chance in every region. Study-study classification in older participants was significantly above chance in every region other than the HIPP (accuracy was above chance before correction in this region).

To examine age-related differences in classification accuracy, we compared study-study classifier accuracy as a function of age. An ANOVA with factors of ROI and age revealed a main effect of region (F(4.44, 204.04) = 31.55, $p = 6.28 \times 10^{-22}$, partial $\eta^2 = 0.41$), while the age by region interaction and main effect of age both failed to reach significance (Fs < 1; $BFs_{10} < 0.42$).

3.2.2. Study-study classification, young-only classifier

When the classifier was trained on young participants only, study- study classification in both young and old participants (Table 2B and Fig. 3B) was above chance in every region other than the HIPP. Here, classifier accuracy was above chance before correction in younger participants. Bold denotes corrected p < 0.05 and italics denotes uncorrected p < 0.05 (across- region results are uncorrected).

3.2.3. Study-study classification, old-only classifier

When the study phase classifier was trained on old participants only (Table 2C and Fig. 3C), study-study classification in young participants was above chance in the AG, MTG, RSP/PCC, and DLPFC and above chance before correction in the MPFC. Study-study classification in old participants was above chance in every region except the HIPP, where once again accuracy was above chance before correction.

3.2.4. Study-study classification, young versus old classifier

To assess differences in classifier accuracy as a function of the age-specific classifiers, we conducted an ANOVA with factors of classifier (young-only, old-only), ROI, and age (compare Fig. 3B and C). The ANOVA revealed a significant classifier by age interaction (F(1, 46) = 8.39, $p = 5.75 \times 10^{-3}$, partial $\eta^2 = 0.15$) and a main effect of region (F(3.98, 189.03) = 38.43, $p = 3.02 \times 10^{-23}$, partial $\eta^2 = 0.46$). All other ANOVA results were non-significant (Fs < 1.47, ps > 0.23, $BFs_{10} < 0.31$).

Two sets of subsidiary ANOVAs were conducted on the study-study accuracy values to probe the classifier by age interaction. The first set of ANOVAs was conducted on data pertaining to each classifier (with factors of age and region). The ANOVA conducted on the accuracy values for the young-only classifier (Fig. 3B) revealed only a main effect of region (F(4.23, 194.63) = 25.11, $p = 1.97 \times 10^{-17}$, partial $\eta^2 = 0.35$) with no effect of age or region by age interaction (Fs < 1.46, ps > 0.23, $BFs_{10} < 0.52$). The ANOVA conducted on the accuracy values for the old-only classifier (Fig. 3C) revealed a main effect of region (F(4.29, 197.12) = 26.23, $p = 2.79 \times 10^{-18}$, partial $\eta^2 = 0.36$), with no region by age interaction (F < 1, $BF_{10} = 0.31$) or main effect of age (F(1, 46) = 2.25, p = 0.14, $BF_{10} = 0.71$). The next set

of ANOVAs was conducted on data pertaining to each age group (with factors of classifier and region). The ANOVA conducted on the accuracy values for the younger participants (Fig. 3B-C, dark grey bars) revealed only a main effect of region (F(3.50, 80.41) = 18.57, p = 4.54×10^{-10}, partial η^2 = 0.45) with no main effect of classifier or region by classifier interaction (Fs < 1.22, ps > 0.28, BF_{10} = 0.28). In contrast, the ANOVA conducted on the accuracy values for the older participants (Fig. 3B-C, light grey bars) revealed both a main effect of region (F(3.90, 89.63) = 20.42, p = 8.26×10^{-12}, partial η^2 = 0.47) and a main effect of classifier, reflecting greater accuracy for the old-only relative to the young-only classifier (F(1, 23) = 10.05, p = 4.27×10^{-3}, partial η^2 = 0.30). The region by classifier interaction was not significant (F < 1, BF_{10} = 0.17).

3.2.5. Study-test classification, across-group classifier—The findings for the across-group classifier are summarized in Table 3A and Fig. 4A. To parallel our original within-participants MVPA (Thakral et al., 2017), we first examined study-test classifier accuracy collapsed across age group (Table 3A, Fig. 4A, top). When collapsed across age, classifier accuracy was above chance for Remember responses in the AG, MTG, RSP/PCC, and before correction in the HIPP and DLPFC. Classifier accuracy was above chance for Know responses in the MTG, and before correction in the AG, MPFC and RSP/PCC. When examining young participants separately (Table 3A, Fig. 4A, bottom), study-test classifier accuracy was above chance for Remember responses in the AG and MTG. Accuracy was also above chance in the MTG for Know responses, but only before correction. For older participants (Table 3A, Fig. 4A, bottom), accuracy was significantly above chance for Remember responses in the AG, MTG, RSP/PCC, and, before correction, in the DLPFC. Accuracy was significantly greater than chance for Know responses before correction in the AG, MTG, RSP/PCC and DLPFC. To examine whether study-test accuracy differed as a function of response or age group we employed an ANOVA with factors of region, response, and age group (Fig. 4A, bottom). The ANOVA failed to reveal any significant effects (Fs < 2.96, ps > 0.09, BF_{10} = 0.95).

3.2.6. Study-test classification, young-only classifier—Findings for the young-only classifier are summarized in Table 3B and Fig. 4B. Accuracy for young participants was above chance for Remember responses in the MTG, and before correction in the AG and HIPP. Accuracy did not differ from chance in any region for Know responses. Study-test classifier accuracy for older participants was above chance before correction for Remember responses in the AG, MTG, and RSP/ PCC. Accuracy was also above chance for Know responses in the AG.

3.2.7. Study-test classification, old-only classifier—Performance of the old-only classifier is summarized in Table 3C and Fig. 4C. Accuracy for young participants was above chance for Remember responses in the AG and MTG and before correction, in the RSP/PCC and HIPP. Accuracy for young participants was above chance for Know responses before correction in the MTG, HIPP, and DLPFC. For older participants, accuracy was above chance for Remember responses before correction in the AG, RSP/PCC, and DLPFC. Accuracy was also above chance for Know responses in the AG, MPFC, and DLPFC, but only before correction.

3.2.8. Study-test classification, young versus old classifier—We assessed whether study-test classification differed as a function of age group, classifier (young-only, old-only), or response category by conducting an ANOVA with factors of region, response, classifier, and age (compare Fig. 4B and C). The ANOVA revealed only a main effect of classifier ($F(1, 46) = 6.00$, $p = 0.02$, partial $\eta^2 = 0.12$), reflecting greater accuracy for the old- than the young-only classifier. All other effects were non-significant ($Fs < 2.82$ $ps > 0.10$, $BFs_{10} < 0.89$)[2].

3.2.9. Spatial extent of the across-participant MVPA effects—The above analyses provide evidence that cortical reinstatement effects in regions belonging to the core recollection network generalize across participants. To gain a sense of the extent of the across-participant consistency in the neural patterns underpinning these findings, we mapped the single trial estimates of the BOLD signal from the study and test phases that were entered into the MVPA for each ROI. For each participant, and for each voxel within an ROI, we subtracted the mean BOLD signal for picture trials from the mean BOLD signal for word trials. We separately examined the study and test phase data (e.g., a difference score was computed between picture and word study trials, Remember-picture and Remember-word test trials, and Know-picture and Know-word test trials). Across-participant voxel-wise two-tailed t-tests were then conducted to assess whether the score in each voxel was significantly different from zero (uncorrected for multiple comparisons). Fig. 5 depicts the results of this analysis. For illustrative purposes, we plot only those voxels where the across-participant difference between the BOLD signal for picture and words trials (and vice versa) exceeded a t-value of ± 2.00. We collapsed across age group as our MVPA analysis failed to find any evidence that cortical reinstatement effects differed as a function of age (see above). The analysis revealed sizeable, spatially contiguous voxel sets demonstrating reliable content-related differences in BOLD signal in each ROI for both the study and test phases.

3.2.10. Across-participant classification within content-selective ROIs— Consistent with our previously reported findings from the same dataset, where MVPA was conducted within-participants (Wang et al., 2016; Thakral et al., 2017), we found no hint in any of the analyses reported above of differences in study-test classifier accuracy according to whether test items were endorsed as Remember or Know. In contrast to these findings, Wang et al. (2016) reported greater study-test classifier accuracy for Remember relative to Know responses. In this earlier report the analysis of reinstatement effects was not targeted at the core recollection network, but focused exclusively on content-selective cortical regions. Therefore, we conducted a follow-up analysis employing the same feature set used by Wang et al. (2016) to examine whether across-participant classification analyses would reveal greater cortical reinstatement effects for recollection-relative to familiarity-based responses (see *2.6 Feature selection for MVPA*).

As is evident in Fig. 6A and Table 4, study-study accuracy for the across-group classifier and for each of the age-specific classifiers was significantly above chance (the outcome of each test was considered significant if the p value survived Bonferroni correction for

[2]At the request of a reviewer, we examined whether classifier accuracy for the hippocampal ROI differed according to hemisphere. An ANOVA revealed no evidence of any such laterality effects.

multiple comparisons; 0.05/2 tests (i.e., 2 age groups), yielding a corrected p = 0.025). For the across-group classifier, study-study accuracy did not differ as a function of age (t(46) = 0.01, p = 0.99, BF_{10} = 0.29; Fig. 6A, bars 2 and 3). To examine differences in study-study classifier accuracy as a function of the age-specific classifiers, we conducted an ANOVA with factors of classifier (young-only, old-only) and age (Fig. 6A, last 4 bars). The ANOVA failed to reveal any significant effects (Fs < 1.85, ps > 0.18, BFs_{10} < 0.61).

Fig. 6B illustrates study-test classifier accuracy for the across-group classifier (see Table 5A). Study-test classifier accuracy collapsed across age group was above chance for both Remember and Know responses. When examining each age group separately, classifier accuracy was above chance in younger and older participants for Remember responses, but was only above chance for Know responses in young participants. An ANOVA with factors of response and age group (Fig. 6B, last 4 bars) revealed a main effect of response with greater accuracy for Remember relative to Know responses (F(1, 46) = 8.16, p = 6.42 × 10^{-3}, partial η^2 = 0.15). All other effects were non-significant (Fs < 1.47, ps > 0.23, BFs_{10} < 0.52).

Fig. 6C illustrates study-test classifier accuracy for each of the age-specific classifiers. For the young-only classifier (see Table 5B), accuracy for young participants was above chance for Remember and Know responses, whereas accuracy for old participants was above chance only for Remember responses. For the old-only classifier (see Table 5C), accuracy was above chance for Remember and Know responses for both young and old participants. We assessed whether study-test classification differed as a function of age group, classifier (young-only, old-only) or response category by conducting an ANOVA with factors of response, classifier, and age (Fig. 6C). Consistent with the findings from the across-group classifier, the ANOVA revealed a main effect of response, with study-test accuracy significantly greater for Remember than Know responses (F(1, 46) = 8.79, p = 4.80 × 10^{-3}, partial η^2 = 0.16). All other effects were non-significant (Fs < 3.08, ps > 0.09, BFs_{10} < 0.65).

4. Discussion

We examined whether prior MVPA findings demonstrating age-invariant reinstatement of encoding-related activity would be replicated when MVPA classification was performed across-rather than within-participants. For the reasons outlined in the Introduction, we predicted that, in contrast to the prior findings, across-participant classification would be more accurate for older than for young adults. We also asked whether, when using an across participants approach, there would be evidence of differential reinstatement effects in core recollection regions according to whether or not memory test items elicited a subjective sense of recollection (Remember vs. Know responses). As well as examining reinstatement in the core recollection network and the DLPFC, we extended our original MVPA of this data-set (Wang et al., 2016) to include an across-participant analysis of the content-sensitive voxel set identified in that report. In agreement with prior reports demonstrating cortical reinstatement with across participant analysis approaches (e.g., Richter et al., 2016; Chen et al., 2017; Oedekoven et al., 2017), we found robust evidence that across participant classifiers could detect retrieval-related reinstatement in some regions of the core

recollection network, as well in the voxel set demonstrating 'content-selective' recollection effects identified by Wang et al. (2016). Contrary to our prediction, however, but in line with our previously reported results (Wang et al., 2016; Thakral et al., 2017), we could find no evidence that reinstatement differed according to age group. Also in agreement with our prior findings, reinstatement within core recollection regions did not differ according to whether test items elicited a sense of recollection or were judged old on the basis of familiarity only, but reinstatement was robustly stronger for recollected than for familiar-only items in the voxel set employed by Wang et al. (2016).

As was just noted, we found no evidence for a moderating effect of age on across participant classification accuracy of retrieval-related neural activity and, hence, no evidence for less idiosyncratic patterns of reinstatement in older than in young individuals. Indeed, in both age groups, and with the two minor exceptions noted below, classification accuracy for held-out participants during study and test was remarkably similar whether classifiers were trained on data from all participants, or from same- or different-age participant samples. Thus, there was little in the way of systematic differences between the age groups in across participant consistency of the patterns of content-dependent cortical activity elicited during either the study or test tasks. This implies that, at least at the spatial scale and sensitivity afforded by the present analyses, there were minimal differences between the two age groups in the spatial distributions of the voxels that were differentially responsive to the two classes of retrieved study content. Clearly, the present findings offer no support for our prediction that older subjects would demonstrate more 'stereotyped' reinstatement effects than their young counterparts. Whether these null findings merely reflect limitations of experimental design (e.g., employment of highly distinct classes of study trials) or the analysis approach (use of across trial classifiers that identified reinstatement at the 'category-' rather than the 'item-level'; cf., Thakral et al., 2017) are important questions for future research.

That being said, the present analyses did uncover two subtle effects of age. First, mean across region classification of older participants' study data from core recollection regions and DLPFC was slightly but significantly more accurate when classifiers were trained on other older participants' data than when they were trained on young participants' data (a similar but non-significant trend was evident for the content-sensitive voxel set). The second age effect was found for the classification of retrieval data: for both age groups and classes of memory judgment, accuracy was slightly higher for the classifier trained on the older rather than the young participants' study data. Together, these two findings raise the possibility that differences in the patterns of activity elicited by the two classes of study trial were more stable across the older than the young participants, perhaps offering partial support for our prediction of lower across-participant variability in the neural responses of older individuals. The findings do not, however, modify the conclusion that the present analyses offer no evidence in support of the prediction that retrieval-related reinstatement would be less variable across older than across young participants.

In keeping with common practice (see, for example, Johnson et al., 2009; Kuhl et al., 2013; Kuhl and Chun, 2014), in our prior analyses of the present data-set that employed within-participant classifier-based MVPA we assessed whether classifier accuracy exceeded chance

with across participant one-sample t-tests against the null hypothesis of an accuracy of 0.5 (chance). It has been demonstrated (Allefeld et al., 2016) that this and related approaches constitute a fixed rather than a random effects test of the null hypothesis; thus, a significant outcome licenses the conclusion that at least one, and perhaps more, participants demonstrated above chance classification, but it provides no basis for generalizing the result to the population from which the participants were sampled. Critically, Allefeld et al. (2016) argued that this limitation does not apply when the same statistical approach is applied to accuracy rates derived from *across*, rather than within participant classification, when the test reverts to a random effects analysis. From this perspective, the present findings take on additional significance, in that they largely replicate our prior findings for both the core recollection and content-selective feature sets that were based on within participant classifier training. Hence, with the exceptions discussed below, the present analysis approach provides grounds for concluding that our prior findings can indeed be generalized to the young and older adultpopulations from which our participants were recruited (see Rugg, 2016, for discussion of more general issues concerning the representativeness of the participants recruited into typical fMRI studies of aging).

In light of this issue, it is relevant to note one specific difference between the present and prior findings that pertains to the core recollection and DLPFC ROIs (Thakral et al., 2017). The present results of the study-study classification analyses were remarkably similar to what was previously reported: accuracy was highest for the AG, equivocal in the HIPP, and robustly above chance in the remaining ROIs. By contrast, regions where above-chance classification of recollected items was observed (i.e. recollection-related reinstatement) were less numerous than previously: whereas within participant classification identified reinstatement in all regions other than MPFC and HIPP, evidence for reinstatement in the present analyses was confined largely to the AG, MTG, and RSP/PCC (Fig. 4A top). Perhaps most strikingly, whereas classification accuracy of recollected test trials was robustly above chance in the DLPFC in the within participant analyses (Thakral et al., 2017), here it barely, if at all, exceeded chance levels. This finding cannot be attributed to a general insensitivity of the DLPFC to the across participant classification approach: classification accuracy of study trials in the region was around or above 60%, depending on the analysis (Fig. 3A-C). The present findings might indicate that, unlike in the cases of the AG, MTG, and RSP/PCC, the spatial distribution of reinstatement effects in the DLPFC is too variable across individuals to support across participant classification. This raises the possibility that our prior findings for this region are a reflection of the limitations of the statistical approach that was adopted (see above), and that DLPFC does not, in fact, manifest robust reinstatement effects at the population level (see Bhan-dari et al., 2018, for evidence that, along with other PFC regions, MVPA in the DLPFC is less sensitive than MVPA in other cortical regions across a wide variety of cognitive domains).

In addition to the examination of recollection-related reinstatement, the present analyses also provided an opportunity to investigate whether test trials endorsed as familiar-only (Know responses) elicited reinstatement effects. Consistent with prior findings using within participant classification (Johnson et al., 2009; Wang et al., 2016; Thakral et al., 2017), we found evidence of reinstatement in association with these test trials in both core recollection ROIs (albeit, most robustly when the data were collapsed across ROIs; Fig. 4A-C), as well

as in the content-selective feature set (Fig. 6B-C). Thus, belying some theoretical perspectives (see Thakral et al., 2017, and Rugg and King, 2018, for recent discussions), there seems little doubt that 'unrecollected' test items can elicit retrieval-related activity sufficient to allow accurate classification of their study histories, at least at the category level (cf., Wang et al., 2016; Thakral et al., 2017). Also consistent with prior results, we found no evidence of differences in classification accuracy for Remember and Know test trials in the core recollection and DLPFC ROIs, whereas accuracy, while above chance, was significantly lower for Know trials in the voxel set of Wang et al. (2016). These findings reinforce our prior conclusion that univariate and multi-voxel classifier analysis approaches can lead to highly disparate conclusions about whether regions such as the AG play a selective role in recollection, or whether they are engaged during familiarity-driven recognition also (cf. Jimura and Poldrack, 2012). The findings also suggest that retrieval-related cortical reinstatement in content-selective cortical regions (as these are identified by univariate analyses) is more sensitive to the distinction between recollection- and familiarity-driven recognition memory than is reinstatement in regions that manifest generic recollection effects.

5. Conclusion

In conclusion, the present findings extend prior results in two main ways. They indicate that patterns of content-selective retrieval-related activity can be no more variable in young than in older individuals and that, regardless of age, the patterns are sufficiently spatially invariant to support across participant multi-voxel classification. In addition, the findings add weight to the proposal that, at least as operationalized by classification accuracy, cortical reinstatement is not limited to test trials associated with a phenomenal sense of recollection but is also present on trials where recognition memory is associated only with a sense of familiarity. It remains to be seen whether this finding extends to other operationalizations of the recollection/familiarity distinction (e.g., objective indices of recollection, such as source memory; cf., Gordon et al., 2014; Leiker and Johnson, 2015), and to item-rather than category-level measures of retrieval-related reinstatement (e.g., Ritchey et al., 2013; Wing et al., 2015; Oedekoven et al., 2017).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

Abdulrahman H, Fletcher PC, Bullmore E, Morcom AM, 2017 Dopamine and memory dedifferentiation in aging. Neuroimage 153, 211–220. [PubMed: 25800211]

Allefeld C, Gorgen K, Haynes JD, 2016 Valid population inference for information-based imaging: from the second-level t-test to prevalence inference. Neuroimage 141, 378–392. [PubMed: 27450073]

Alvarez P, Squire LR, 1994 Memory consolidation and the medial temporal lobe: a simple network model. Proc. Natl. Acad. Sci. U. S. A. 91, 7041–7045. [PubMed: 8041742]
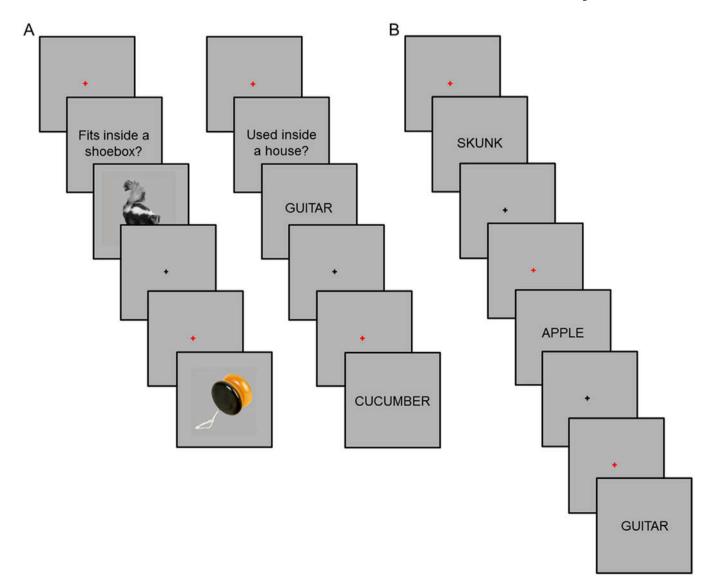
Bhandari A, Gagne C, Badre D, 2018 Just above chance: is it harder to decode information from prefrontal cortex hemodynamic activity patterns? J. Cognit. Neurosci. 1, 26.

Brainerd CI, Reyna VF, 2015 Fuzzy-trace theory and lifespan. Dev. Rev. 38, 89–121. [PubMed: 26644632]

Carp J, Park J, Polk TA, Park DC, 2011a Age differences in neural distinctiveness revealed by multi-voxel pattern analysis. Neuroimage 56, 736–743. [PubMed: 20451629]

Carp J, Park J, Hebrank A, Park DC, Polk TA, 2011b Age-related neural dedifferentiation in the motor system. PLoS One 6, 1–6.

Chen J, Leong YC, Honey CJ, Yong CH, Norman KA, Hasson U, 2017 Shared memories reveal shared structure in neural activity across individuals. Nat. Neurosci. 20, 115–125. [PubMed: 27918531]

Danker JF, Anderson JR, 2010 The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. Psychol. Bull. 136, 87–102. [PubMed: 20063927]

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. Cerebr. Cortex 21, 2166–2176.

Dienes Z, 2014 Using Bayes to get the most out of non-significant results. Front. Psychol. 5, 1–17. [PubMed: 24474945]

Dodson CS, 2017 Aging and memory In: Byrne JH (Ed.), Learning and Memory: A Comprehensive Reference, second ed Academic Press, Oxford, pp. 403–421.

Du Y, Buchsbaum BR, Grady CL, Alain C, 2016 Increased activity in frontal motor cortex compensates impaired speech perception in older adults. Nat. Commun. 7, 1–12.

Frisoni GB, et al., 2015 The EADC-ADNI harmonized protocol for manual hippocampal segmentation on magnetic resonance: evidence of validity. Alzheimer's Dementia 11, 111–125.

Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg MD, Turner R, 1998 Event-related fMRI: characterizing differential responses. Neuroimage 7, 30–40. [PubMed: 9500830]

Gordon AM, Rissman J, Kiani R, Wagner AD, 2014 Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. Cerebr. Cortex 24, 3350–3364.

Grady CL, Charlton R, He Y, Alain C, 2011 Age differences in fMRI adaptation for sound identity and location. Front. Hum. Neurosci. 5, 1–12. [PubMed: 21283556]

Jeffreys H, 1961 Theory of Probability. Oxford University Press, Oxford, UK.

Jimura K, Poldrack RA, 2012 Analyses of regional-average activation and multivoxel pattern information tell complementary stories. Neuropsychologia 50, 544–552. [PubMed: 22100534]

Johnson JD, Rugg MD, 2007 Recollection and the reinstatement of encoding-related cortical activity. Cerebr. Cortex 17, 2507–2515.

Johnson JD, McDuff SGR, Rugg MD, Norman KA, 2009 Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis. Neuron 63, 697–708. [PubMed: 19755111]

Kahn I, Davachi L, Wagner AD, 2004 Functional-neuroanatomic correlates of recollection: implications for models of recognition memory. J. Neurosci. 24, 4172–4180. [PubMed: 15115812]

King DR, de Chastelaine M, Elward RL, Wang TH, Rugg MD, 2015 Recollection-related increases in functional connectivity predict individual differences in memory accuracy. J. Neurosci. 35, 1763–1772. [PubMed: 25632149]

Kleemeyer MM, Polk TA, Schaefer S, Bodammer NC, Brechtel L, Lindenberger U, 2017 Exercise-induced fitness changes correlate with changes in neural specificity in older adults. Front. Neurosci. 11, 1–8.

Koen JD, Rugg MD, 2016 Memory reactivation predicts resistance to retroactive interference: evidence from multivariate classification and pattern similarity analyses. J. Neurosci. 36, 4389–4399. [PubMed: 27076433]

Koen JD, Hauck N, Rugg MD, 2019 The relationship between age, neural differentiation, and memory performance. J. Neurosci. 39, 149–162. [PubMed: 30389841]

Kucera H, Francis WN, 1967 Computational Analysis of Present Day American English. Brown University Press, Rhode Island.

Kuhl BA, Chun MM, 2014 Successful remembering elicits event-specific activity patterns in lateral parietal cortex. J. Neurosci. 34, 8051–8060. [PubMed: 24899726]

Kuhl BA, Johnson MK, Chun MM, 2013 Dissociable neural mechanisms for goal- directed versus incidental memory reactivation. J. Neurosci. 33, 16099–16109. [PubMed: 24107943]

Lakens D, 2013 Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Front. Psychol. 4, 1–12. [PubMed: 23382719]

Leiker EK, Johnson JD, 2015 Pattern reactivation co-varies with activity in the core recollection network during source memory. Neuropsychologia 75, 88–98. [PubMed: 26004057]

Li S-C, Rieckmann A, 2014 Neuromodulation and aging: implications of aging neuronal gain control on cognition. Curr. Opin. Neurobiol. 29, 148–158. [PubMed: 25064177]

Li SC, Lindenberger U, Sikstrom S, 2001 Aging cognition: from neuromodulation to representation. Trends Cognit. Sci. 5, 479–486. [PubMed: 11684480]

Li SC, Sikstrom S, 2002 Integrative neurocomputational perspectives on cognitive aging, neuromodulation, and representation. Neurosci. Biobehav. Rev. 26, 795–808. [PubMed: 12470691]

Ly A, Raj A, Etz A, Marsman M, Gronau QF, Wagenmakers E-J, 2018 Bayesian reanalyses from summary statistics: a guide for academic consumers. Adv. Meth. Pract. Psychol. Sci. 1, 367–374.

Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH, 2003 An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 19, 1233–1239. [PubMed: 12880848]

Marr D, 1971 Simple memory: a theory for archicortex. Philos. Trans. R. Soc. Lond. B Biol. Sci. 262, 23–81. [PubMed: 4399412]

Mattson JT, Wang TH, de Chastelaine M, Rugg MD, 2014 Effects of age on negative subsequent memory effects associated with the encoding of item and item-context information. Cerebr. Cortex 24, 3322–3333.

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. [PubMed: 24828546]

McDuff SGR, Frankel HC, Norman KA, 2009 Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during singleagenda source monitoring. J. Neurosci. 29, 508–516. [PubMed: 19144851]

Moran RJ, Symmonds M, Dolan RJ, Friston KJ, 2014 The brain ages optimally to model its environment: evidence from sensory learning over the adult lifespan. PLoS Comput. Biol. 10, 1–8.

Nilsson LG, 2003 Memory function in normal aging. Acta Neurol. Scand. Suppl. 179, 7–13. [PubMed: 12603244]

Norman KA, Schacter DL, 1997 False recognition in younger and older adults: exploring the characteristics of illusory memories. Mem. Cognit. 25, 838–848.

Norman KA, O'Reilly RC, 2003 Modeling hippocampal and neocortical contributions to recognition memory: a complementary-learning-systems approach. Psychol. Rev. 110, 611–646. [PubMed: 14599236]

Oedekoven CSH, Keidel JL, Berens SC, Bird CM, 2017 Reinstatement of memory representations for lifelike events over the course of a week. Sci. Rep. 7, 1–12. [PubMed: 28127051]

Park DC, Polk TA, Park R, Minear M, Savage A, Smith MR, 2004 Aging reduces neural specialization in ventral visual cortex. Proc. Natl. Acad. Sci. U. S. A. 101, 13091–13095. [PubMed: 15322270]

Park J, Carp J, Hebrank A, Park DC, Polk TA, 2010 Neural specificity predicts fluid processing ability in older adults. J. Neurosci. 30, 9253–9259. [PubMed: 20610760]

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. J. Neurosci. 32, 2154–2158. [PubMed: 22323727]

Pierce BH, Schacter DL, Sullivan AL, Budson AE, 2005 Comparing source-based and gist-based false recognition in aging and Alzheimer's disease. Neuropsychology 19, 411–419. [PubMed: 16060815]

Richter FR, Chanales AJH, Kuhl BA, 2016 Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. Neuroimage 124, 323–335. [PubMed: 26327243]

Ritchey M, Wing EA, LaBar KS, Cabeza R, 2013 Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. Cerebr. Cortex 23, 2818–2828.

Rissman J, Wagner AD, 2012 Distributed representations in memory: insights from functional brain imaging. Annu. Rev. Psychol. 63, 101–128. [PubMed: 21943171]

Rissman J, Greely HT, Wagner AD, 2010 Detecting individual memories through the neural decoding of memory states and past experience. Proc. Natl. Acad. Sci. 107, 9849–9854. [PubMed: 20457911]

Rissman J, Chow TE, Reggente N, Wagner AD, 2016 Decoding fMRI signatures of real-world autobiographical memory retrieval. J. Cognit. Neurosci. 28, 604–620. [PubMed: 26741803]

Rugg MD, 2016 Interpreting age-related differences in memory-related neural activity In: Cabeza R, Nyberg L, Park DC (Eds.), Cognitive Neuroscience of Aging. Oxford University Press, Oxford, pp. 183–204.

Rugg MD, King DR, 2018 Ventral lateral parietal cortex and episodic memory retrieval. Cortex 107, 238–250. [PubMed: 28802589]

Rugg MD, Johnson JD, Uncapher MR, 2015 Encoding and retrieval in episodic memory: insights from fMRI In: Addis DR, Barense MD, Duarte A (Eds.), Handbook on the Cognitive Neuroscience of Memory. Wiley-Blackwell, Oxford, pp. 84–107.

Rugg MD, Vilberg KL, 2013 Brain networks underlying episodic memory retrieval. Curr. Opin. Neurobiol. 23, 255–260. [PubMed: 23206590]

St-Laurent M, Abdi H, Bondad A, Buchsbaum BR, 2014 Memory reactivation in healthy aging: evidence of stimulus-specific dedifferentiation. J. Neurosci. 34, 4175–4186. [PubMed: 24647939]

Staresina BP, Henson RNA, Kriegeskorte N, Alink A, 2012 Episodic reinstatement in the medial temporal lobe. J. Neurosci. 32, 18150–18156. [PubMed: 23238729]

Thakral PP, Wang TH, Rugg MD, 2015 Cortical reinstatement and the confidence and accuracy of source memory. Neuroimage 109, 118–129. [PubMed: 25583615]

Thakral PP, Wang TH, Rugg MD, 2017 Decoding the content of recollection within the core recollection network and beyond. Cortex 91, 101–113. [PubMed: 28077212]

Todd MT, Nystrom LE, Cohen JD, 2013 Confounds in multivariate pattern analysis: theory and rule representation case study. Neuroimage 77, 157–165. [PubMed: 23558095]

Tulving E, 1985 Memory and consciousness. Can. Psychol. Can. 26, 1–12.

Van Essen DC, 2005 A population-average landmark- and surface-based PALS atlas of human cerebral cortex. Neuroimage 28, 635–662. [PubMed: 16172003]

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 Res. 1244, 121–131. [PubMed: 18848823]

Wagenmakers E-J, et al., 2018 Bayesian inference for psychology. Part II: example applications with JASP. Psychon. Bull. Rev. 25, 58–76. [PubMed: 28685272]

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. Cerebr. Cortex 26, 1698–1714.

Wing EA, Ritchey M, Cabeza R, 2015 Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. J. Cognit. Neurosci. 27, 679–691. [PubMed: 25313659]

Xue G, 2018 The neural representations underlying human episodic memory. Trends Cognit. Sci. 22, 544–561. [PubMed: 29625850]

Yonelinas AP, Jacoby LL, 1995 The relation between remembering and knowing as bases for recognition: effects of size congruency. J. Mem. Lang. 34, 622–643.

Fig. 1.
A. Study tasks (two representative study trials from each task shown to the left and right). B. Test task. For each test word cue, participants were asked to make a 'Remember', 'Know', or 'New' response.
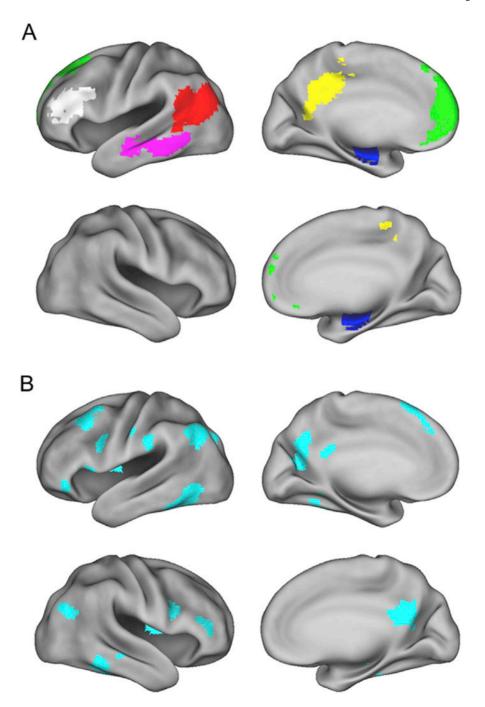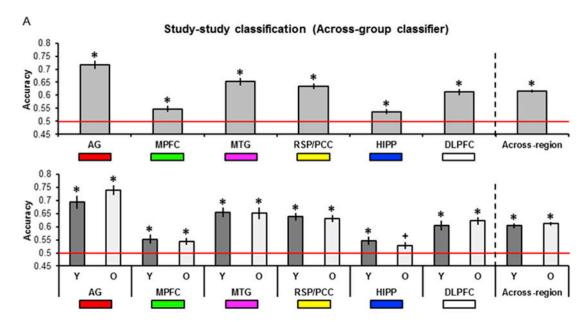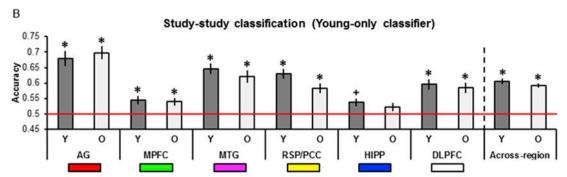
Fig. 2.
A. Features employed in the multi-voxel pattern analyses overlaid on the standardized brain of the PALS-B12 atlas implemented in Caret5 (Van Essen, 2005). Regions analyzed include the left angular gyrus (AG, red), left middle temporal gyrus (MTG, magenta), left retroslpenial/posterior cingulate cortex (RSP/PCC, yellow), left medial prefrontal cortex (MPFC, green), bilateral anterior hippocampus (HIPP, blue), and left dorsolateral prefrontal cortex (DLPFC, white). B. Features employed in the multi-voxel pattern analysis of Wang et

al. (2016) that manifested content-selective univariate reinstatement effects. Left hemisphere is on the top row of each panel.
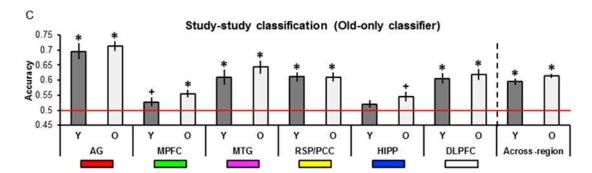
Fig. 3.
A. Across-group study-study classifier accuracy as a function of age group and ROI. B. Young-only study-study classifier accuracy as a function of age group and ROI. C. Old-only study-study classifier accuracy as a function of age group and ROI. In this and subsequent figures, * denotes corrected $p < 0.05$, + denotes uncorrected $p < 0.05$, and red line denotes chance classifier performance. Also illustrated is the mean classifier accuracy across regions (* denotes uncorrected $p < 0.05$).
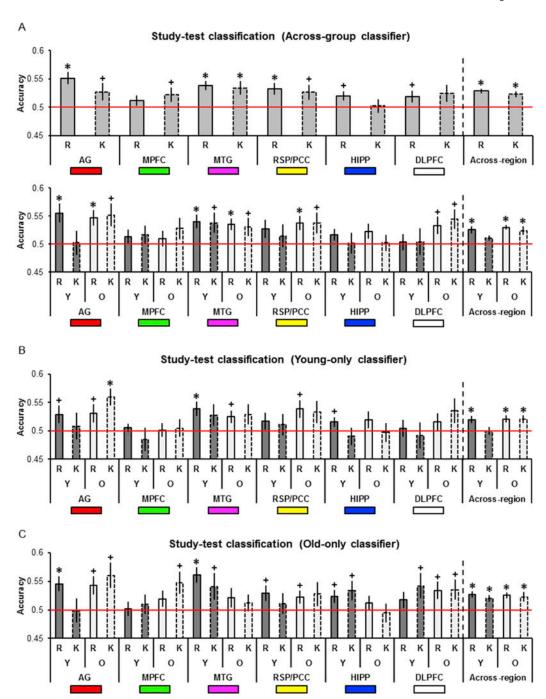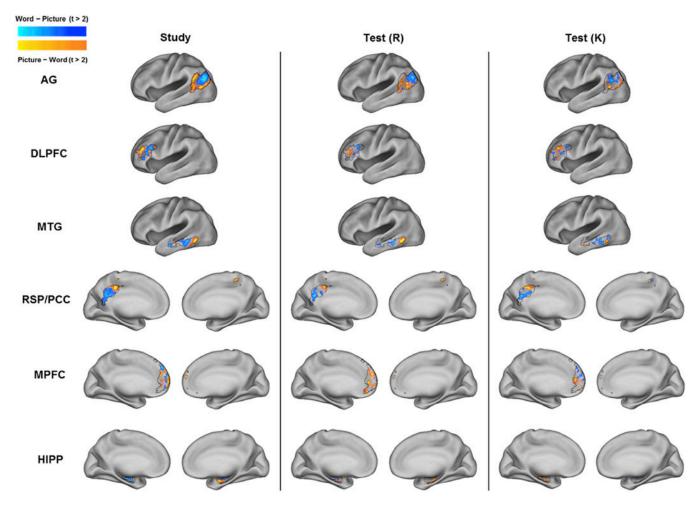
Fig. 4.
A. Across-group study-test classifier accuracy as a function of age group, response, and ROI. B. Young-only study-test classifier accuracy as a function of age group, response, and ROI. C. Old-only study-test classifier accuracy as a function of age group, response, and ROI.

Fig. 5.
Mean BOLD signal difference map between picture and word trials for each ROI, session (study and test), and response category (Remember and Know). Black lines depict ROI borders.

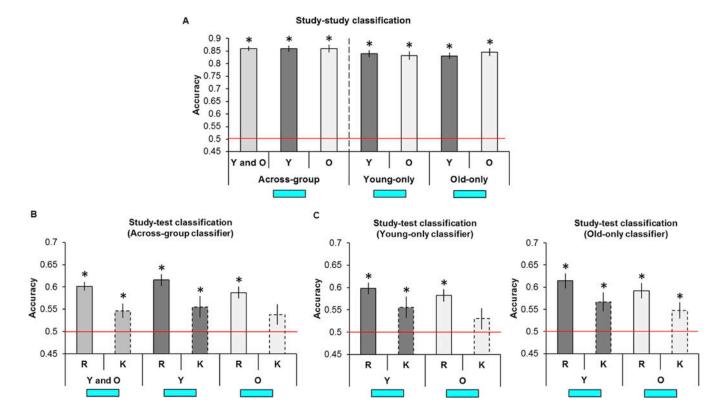**Fig. 6.**
A. Across-group, young-only, and old-only study-study classifier accuracy as a function of age group in the feature set employed in Wang et al. (2016, see Fig. 2B). B. Across-group study-test classifier accuracy collapsed across age groups and split as a function of age group and response. C. Young-only and old-only study-test classifier accuracy as a function of age group and response.

Author Manuscript

Table 1

Mean (±1 standard error) estimates of recollection (pR) and familiarity (pF) computed for Remember and Know responses, respectively, as a function of study category and age group. Reaction times for Remember and Know responses listed below.

	Memory performance			
	Words		**Pictures**	
	pR	**pF**	**pR**	**pF**
Young	0.56 (0.03)	0.61 (0.04)	0.62 (0.03)	0.40 (0.04)
Old	0.46 (0.05)	0.53 (0.03)	0.53 (0.04)	0.36 (0.03)
	Reaction times			
	Remember	**Know**	**Remember**	**Know**
Young	2.13 (0.16)	2.96 (0.22)	2.15 (0.15)	3.15 (0.23)
Old	2.10 (0.14)	2.80 (0.21)	2.10 (0.13)	3.09 (0.19)

Table 2

Study-study classifier results as a function of classifier ((A) across-group, (B) young-only (B), and (C) old-only), age group, and ROI.

A. Across-group

	Young and Old		Young		Old	
	p	d_z	p	d_z	p	d_z
AG	3.00×10^{-18}	1.97	4.09×10^{-8}	1.57	3.34×10^{-12}	2.60
MPFC	6.71×10^{-5}	0.60	3.86×10^{-3}	0.60	3.67×10^{-7}	0.60
MTG	3.66×10^{-14}	1.51	1.67×10^{-8}	1.66	3.88×10^{-7}	1.37
RSP/PCC	4.00×10^{-18}	1.95	1.91×10^{-9}	1.87	5.02×10^{-10}	2.01
HIPP	2.56×10^{-4}	0.54	2.67×10^{-3}	0.63	*0.02*	*0.47*
DLPFC	3.28×10^{-13}	1.41	4.10×10^{-6}	1.17	7.52×10^{-9}	1.80
Across-region	4.23×10^{-25}	2.91	1.23×10^{-11}	2.44	2.69×10^{-15}	3.65

B. Young-only

	Young		Old	
	p	d_z	p	d_z
AG	1.43×10^{-7}	1.46	1.00×10^{-9}	1.94
MPFC	3.10×10^{-3}	0.62	2.34×10^{-3}	0.64
MTG	3.96×10^{-9}	1.80	4.15×10^{-6}	1.17
RSP/PCC	1.60×10^{-8}	1.66	2.87×10^{-5}	1.00
HIPP	*6.78×10^{-3}*	*0.55*	0.06	0.32
DLPFC	8.18×10^{-6}	1.11	2.41×10^{-5}	1.02
Across-region	2.94×10^{-11}	2.33	3.28×10^{-12}	2.60

C. Old-only

	Young		Old	
	p	d_z	p	d_z
AG	7.11×10^{-8}	1.52	2.90×10^{-12}	2.62
MPFC	*0.04*	*0.38*	1.31×10^{-4}	0.88
MTG	1.36×10^{-4}	0.88	4.88×10^{-7}	1.35
RSP/PCC	1.00×10^{-7}	1.49	1.20×10^{-7}	1.47
HIPP	0.07	0.30	*6.58×18^{-3}*	*0.55*
DLPFC	7.87×10^{-7}	1.31	8.99×10^{-7}	1.30
Across-region	1.63×10^{-9}	1.89	1.05×10^{-13}	3.07

Bold denotes corrected p < 0.05 and italics denotes uncorrected p < 0.05 (acrossregion results are uncorrected).

Table 3

Study-test classifier results as a function of classifier ((A) across-group, (B) young-only (B), and (C) old-only), age group, response, and ROI.

A. Across-group

	Young and Old Remember		Know		Young Remember		Know		Old Remember		Know	
	p	d_z	p	d_z	p	d_z	p	d_z	p	d_z	p	d_z
AG	1.04×10^{-5}	0.68	0.04	0.26	1.51×10^{-3}	0.68	0.04	0.02	1.42×10^{-3}	0.68	9.55×10^{-3}	0.51
MPFC	0.10	0.19	0.04	0.25	0.14	0.23	0.17	0.20	0.24	0.15	0.08	0.29
MTG	1.26×10^{-5}	0.67	3.49×10^{-3}	0.41	1.92×10^{-3}	0.66	0.02	0.43	1.34×10^{-3}	0.69	0.04	0.38
RSP/PCC	1.34×10^{-3}	0.46	0.03	0.27	0.05	0.34	0.24	0.15	3.59×10^{-3}	0.60	0.03	0.41
HIPP	0.01	0.33	0.43	0.02	0.07	0.32	0.46	0.02	0.05	0.34	0.44	0.03
DLPFC	0.04	0.25	0.05	0.24	0.40	0.05	0.43	0.04	0.02	0.45	7.88×10^{-3}	0.53
Across-region	1.63×10^{-8}	0.95	9.21×10^{-5}	0.59	7.25×10^{-4}	0.74	0.05	0.35	8.77×10^{-7}	1.30	1.94×10^{-4}	0.85

B. Young-only

	Young Remember		Know		Old Remember		Know	
	p	d_z	p	d_z	p	d_z	p	d_z
AG	0.03	0.39	0.36	0.08	0.03	0.03	3.29×10^{-4}	0.80
MPFC	0.26	0.13	0.24	0.14	0.44	0.03	0.37	0.07
MTG	2.76×10^{-3}	0.63	0.07	0.31	0.02	0.46	0.05	0.34
RSP/PCC	0.13	0.23	0.28	0.12	0.01	0.50	0.05	0.34
HIPP	0.04	0.38	0.27	0.13	0.10	0.28	0.44	0.03
DLPFC	0.38	0.06	0.36	0.07	0.16	0.21	0.06	0.33
Across-region	3.15×10^{-3}	0.61	0.36	0.07	9.18×10^{-4}	0.72	1.55×10^{-3}	0.73

C. Old-only

	Young Remember		Know		Old Remember		Know	
	p	d_z	p	d_z	p	d_z	p	d_z
AG	1.28×10^{-3}	0.69	0.46	0.02	7.25×10^{-3}	0.54	6.24×10^{-3}	0.55
MPFC	0.44	0.03	0.29	0.11	0.10	0.27	0.01	0.49

C. Old-only

	Young Remember		Know		Old Remember		Know	
	p	d_z	**p**	d_z	**p**	d_z	**p**	d_z
MTG	$\mathbf{7.63 \times 10^{-5}}$	**0.92**	*0.05*	*0.36*	0.11	0.25	0.23	0.16
RSP/PCC	*0.03*	*0.42*	0.27	0.13	*0.03*	*0.39*	0.10	0.27
HIPP	*0.03*	*0.40*	*0.03*	0.40	0.19	0.19	0.37	0.07
DLPFC	0.12	0.24	*0.04*	*0.38*	*0.02*	*0.44*	*0.03*	*0.40*
Across-region	$\mathbf{1.05 \times 10^{-4}}$	**0.90**	$\mathbf{9.74 \times 10^{-4}}$	**0.71**	$\mathbf{8.01 \times 10^{-4}}$	**0.73**	$\mathbf{2.50 \times 10^{-3}}$	**0.63**

Bold denotes corrected $p < 0.05$ and italics denotes uncorrected

Table 4

Study-study classifier results as a function of classifier ((A) across-group, (B) young-only, and (C) old-only) and age group in the feature set employed in Wang et al. (2016, see Fig. 2B).

A. Across-group					
Young and Old		**Young**		**Old**	
p	**d_z**	**p**	**d_z**	**p**	**d_z**
2.30×10^{-37}	5.52	3.81×10^{-20}	6.06	2.68×10^{-18}	5.01

B. Young-only			
Young		**Old**	
p	**d_z**	**p**	**d_z**
1.24×10^{-18}	5.19	4.60×10^{-16}	3.96

C. Old-only			
Young		**Old**	
p	**d_z**	**p**	**d_z**
1.71×10^{-18}	5.11	1.15×10^{-17}	4.69

Bold denotes corrected $p < 0.05$ and italics denotes uncorrected $p < 0.05$.

Table 5

Study-test classifier results as a function of classifier ((A) across-group, (B) young-only, and (C) old-only), age group, and response in the feature set employed in Wang et al. (2016, see Fig. 2B).

A. Across-group

Young and Old Remember		Young Remember			Know		Old Remember			Know	
p	**d_z**	**p**	**d_z**		**p**	**d_z**	**p**	**d_z**		**p**	**d_z**
2.93×10^{-14}	**1.52**	**5.39×10^{-9}**	**1.77**		**0.01**	**0.48**	**7.14×10^{-7}**	**1.31**		0.05	0.34
Know											
p	**d_z**										
3.06×10^{-3}	**0.41**										

B. Young-only

Young Remember		Old Remember			Know	
p	**d_z**	**p**	**d_z**		**p**	**d_z**
7.57×10^{-8}	**1.52**	**2.46×10^{-6}**	**1.21**		0.11	0.26
Know						
p	**d_z**					
0.01	**0.47**					

C. Old-only

Young Remember		Old Remember			Know	
p	**d_z**	**p**	**d_z**		**p**	**d_z**
3.43×10^{-7}	**1.38**	**1.25×10^{-5}**	**1.07**		**7.71×10^{-3}**	**0.53**
Know						
p	**d_z**					
2.21×10^{-3}	**0.64**					

Bold denotes corrected p < 0.05 and italics denotes uncorrected p < 0.05.