Dissociating refreshing and elaboration and their impacts on memory

Lea M. Bartsch
University of Zurich and University Priority Program Dynamics of Healthy Aging

Vanessa M. Loaiza
University of Essex

Lutz Jäncke
University of Zurich and University Priority Program Dynamics of Healthy Aging

Klaus Oberauer
University of Zurich and University Priority Program Dynamics of Healthy Aging

Jarrod A. Lewis-Peacock
University of Texas at Austin

Author Note

We thank Raffael Schmitt and Danielle Pessach for helping with data collection. The behavioral data and corresponding analysis scripts can be accessed in the Open Science Framework (https:// osf.io/p2h8b/).

Correspondence should be addressed to Lea M. Bartsch, Department of Psychology, Cognitive Psychology Unit, University of Zurich, Binzmuehlestrasse 14/22, 8050 Zurich, Switzerland. E-mail: l.bartsch@psychologie.uzh.ch. Tel: +41 44 635 74 61

Abstract

Maintenance of information in working memory (WM) is assumed to rely on *refreshing* and *elaboration*, but clear mechanistic descriptions of these cognitive processes are lacking, and it is unclear whether they are simply two labels for the same process. This fMRI study investigated the extent to which refreshing, elaboration, and repeating of items in WM are distinct neural processes with dissociable behavioral outcomes in WM and long-term memory (LTM). Multivariate pattern analyses of fMRI data revealed differentiable neural signatures for these processes, which we also replicated in an independent sample of older adults. In some cases, the degree of neural separation within an individual predicted their memory performance. Elaboration improved LTM, but not WM, and this benefit increased as its neural signature became more distinct from repetition. Refreshing had no impact on LTM, but did improve WM, although the neural discrimination of this process was not predictive of the degree of improvement. These results demonstrate that refreshing and elaboration are separate processes that differently contribute to memory performance.

Keywords: Refreshing, Elaboration, Working memory, Long-term memory, MVPA

Introduction

Working memory (WM) is a system for holding a limited amount of information available for processing (Baddeley, 1986), whereas episodic long-term memory (LTM) stores information permanently with presumably unlimited capacity (Tulving, 1972). WM and LTM are highly correlated constructs, and models of their relation suggest that how information is processed in WM strongly affects how well it is maintained in LTM (D'Esposito & Postle, 2015; Eriksson, Vogel, Lansner, Bergström, & Nyberg, 2015; Lewis-Peacock & Postle, 2008; Ranganath, 2006; Ranganath & Blumenfeld, 2005; Ranganath, Cohen, & Brozinsky, 2005; Crowder, 1982; Melton, 1963; Nairne, 1990, 2002; Cowan, 1995; Oberauer, 2002). Two control processes on information in WM have been argued to contribute to encoding in episodic LTM: *refreshing* and *elaboration*. The aim of the present study is to investigate (1) whether refreshing and elaboration are neurally and behaviorally distinguishable processes, (2) how they affect WM and episodic LTM performance, and (3) to what extent age differences in these processes are responsible for memory deficits in older adults.

Refreshing and elaboration: Definitions and behavioral impacts on WM and LTM

Refreshing is understood as briefly thinking of a stimulus just after it is no longer physically present but while its representation is still active (Johnson, Reeder, Raye, & Mitchell, 2002). The process was introduced as a general attention-based mechanism for enhancing and prolonging the activation of memoranda in WM (see Camos et al., 2018 for a review), thereby improving WM (Souza & Oberauer, 2017; Souza, Vergauwe, & Oberauer, 2018); but see for contradicting evidence Bartsch, Singmann, & Oberauer, 2018) and episodic LTM (Johnson et al., 2002). Currently, there are two lines of evidence for effects of refreshing on memory: (a)

Instructions to "think of" an item in WM, and (b) variations of free time that could be used for refreshing. Both of these approaches are subject to an alternative interpretation: When instructed to "think of" a word, people are likely to think of its meaning, perhaps also form an image, and relate it to other WM contents. When given free time, people could use it to elaborate rather than refresh the WM contents. Elaboration refers the act of deeply processing information, especially with regard to its semantic and meaningful characteristics (Craik & Lockhart, 1972; Greene, 1987; Klatsky, 1988), and, as consequence, reliably improves episodic LTM (e.g., Craik & Tulving, 1975; Gallo, Meadow, Johnson, & Foster, 2008). Evidence for a benefit of elaboration for WM is more mixed: Correlational studies show a positive relationship between elaborative strategies and verbal WM recall (Bailey, Dunlosky, & Kane, 2008, 2011; Dunlosky & Kane, 2007; Kaakinen & Hyönä, 2007) and some experimental work has shown that semantic compared to shallow processing of the memoranda yields greater WM recall (Loaiza, McCabe, Youngblood, Rose, & Myerson, 2011; Rose, Buchsbaum, & Craik, 2014; Rose, Craik, & Buchsbaum, 2014). Conversely, other work has shown benefits of elaboration or semantic processing only for episodic LTM but not WM (Bartsch, Singmann, & Oberauer, 2018; Loaiza & Camos, 2016; Rose & Craik, 2012; Rose et al., 2010). Bartsch and colleagues showed that elaboration benefited LTM, but refreshing did not, and neither elaboration nor refreshing benefited WM.

Refreshing vs. elaboration: Neural correlates

Table 1 shows an overview of the reported regions associated with refreshing and/or elaboration. Refreshing has been associated with activity in the left dorsolateral prefrontal cortex (dlPFC, BA 8/9), and activity in the dlPFC during refreshing predicted subsequent LTM for the refreshed information (Johnson, Raye, Mitchell, Greene, & Adam, 2003; Raye, Johnson,

Mitchell, Greene, & Johnson, 2007; Raye, Johnson, Mitchell, Reeder, & Greene, 2002). A meta-analysis (Johnson et al., 2005) identified frontal regions, specifically left dlPFC (BA 9/46), ventrolateral PFC (vlPFC, BA 44/45/47), and the left anterior PFC (BA 10) as associates of refreshing various stimulus materials.

Although the dlPFC has been suggested to underlie refreshing, its activation has also been shown to predict subsequent LTM in studies of elaboration (or "relational encoding") wherein the semantic relationship between two items is elaborated upon (e.g. Blumenfeld & Ranganath, 2007). For the ease of the reader, we will refer to *relational encoding* as *elaboration* from now on. The neural correlates of elaboration have not always been that specific or limited to the dlPFC: earlier studies have more generally associated the lateral PFC with semantic elaboration (e.g., Kapur et al., 1994; Wagner et al., 1998) and relational elaboration (e.g., Addis & McAndrews, 2006; Fletcher, Shallice, & Dolan, 2000; Murray & Ranganath, 2007). Yet, numerous studies have associated the dlPFC with elaboration and subsequent memory effects (Blumenfeld, Parks, & Yonelinas, 2010; Blumenfeld & Ranganath, 2007; Davachi, Maril, & Wagner, 2001; Ragland et al., 2012). Collectively, this evidence suggests that elaboration of the memoranda in WM is what makes the dlPFC important for LTM.

Despite the neural similarities observed for refreshing and elaboration, there are also differences, which could be due to dissimilarities in the methods used to study these processes. First, the neural correlates of refreshing have been studied for single items only, with no instructed elaboration (e.g. Johnson et al., 2005; Raye et al., 2007, 2002), and this item-specific neural processing was localized almost exclusively to left lateral dlPFC. Conversely, elaboration studies have used multiple items, such as pairs (Blumenfeld et al., 2010) or triplets of words (e.g. Blumenfeld, 2006; Davachi, Maril, & Wagner, 2001), and localized the associated activity to the

bilateral dlPFC. Second, the refreshing studies have relied on incidental encoding, wherein participants are not informed of the upcoming memory test, whereas the elaboration studies employ intentional encoding. Therefore, clarifying the underlying neural processes of refreshing and elaboration requires greater consistency between the methods used to investigate them.

Refreshing and elaboration: Age effects

Past research has provided extensive evidence that episodic LTM declines with age (e.g., Hoyer & Verhaeghen, 2006; Naveh-Benjamin & Old, 2008; Zacks, Hasher, & Li, 2000), but the source of the deficit is still under debate. One view is that WM maintenance processes and recruitment of corresponding brain areas decline in older age (Hoareau, Lemaire, Portrat, & Plancher, 2016; Plancher, Boyer, Lemaire, & Portrat, 2017; Smith, 1980). For instance, it has been shown that older adults exhibit reduced refreshing-related brain activity in the left dlPFC and reduced refreshing benefits for episodic LTM relative to young adults (Johnson, Mitchell, Raye, & Greene, 2004; Raye, Mitchell, Reeder, Greene, & Johnson, 2008). Another possibility is that older adults are less likely than younger adults to engage in elaboration, thereby resulting in deficient retention (Smith, 1980). For example, some work has shown older adults are able to capitalize on experiment-administered elaborative strategies but show deficiencies in generating elaborative strategies themselves (Rankin & Collins, 1985, see also Kamp & Zimmer, 2015). A meta-analysis reported that age-related differences in subsequent memory are associated with under-recruitment of the occipital and fusiform cortex in older adults, as well as an over-recruitment of medial and lateral regions of PFC and parietal lobe (Maillet & Rajah, 2014). These findings suggest inefficient recruitment of brain regions that are important for elaboration, thereby leading to age-related memory deficits.

The present study

The goal of the present study was to investigate to what extent elaboration and refreshing are separable processes, given their neural overlap as well as their similar proposed beneficial effects for memory. So far, only one study has investigated both processes in one experiment, and the behavioral results demonstrated that the processes have divergent contributions to LTM (Bartsch et al., 2018). We aimed at extending this previous study by not only investigating whether refreshing and elaboration are distinct in their contribution to WM and LTM formation, but also whether they are supported by separable neural activation patterns. Furthermore, we aimed to investigate their contribution to age-related memory deficits.

We applied multivariate pattern analyses (MVPA; e.g., Haxby, Connolly, & Guntupalli, 2014; Haxby et al., 2001; Haynes & Rees, 2006; Lewis-Peacock & Norman, 2014; Norman, Polyn, Detre, & Haxby, 2006) to fMRI data of young adults and older adults performing the word list encoding task of Bartsch et al. (2018). This analysis approach allowed us to evaluate whether brain activity patterns associated with refreshing items and with elaborating items in WM could be differentiated. These neural measures were then linked to behavioral outcomes on tests of both WM and LTM. MVPA is especially sensitive to detecting fine-grained differences between neural activation patterns that are not detectable using conventional analyses (Kriegeskorte & Bandettini, 2007).

If refreshing and elaboration are two labels for the same process, then the pattern of behavioral effects should be similar for WM and on LTM, and the patterns of brain activity supporting these processes should be indistinguishable. If refreshing and elaboration are distinct

processes, they should have different behavioral effects and separable patterns of neural activation.

Method

Subjects and general procedure

We recruited 30 healthy, right-handed young adults (15 females; mean age = 24.2, SD = 2.97 years) from the student population of the University of Zurich as well as 27 healthy, right-handed older adults from the community (13 females; mean age = 69, SD = 3.47 years). Handedness was measured through observation of the writing hand. Subjects were screened for their ability to undergo a magnetic resonance imaging session. Furthermore, they completed the Digit–Symbol Substitution test (DSS; Wechsler, 1982), serving as an indicator of processing speed, and the mini-mental-status examination (MMSE; Folstein, Folstein, & McHugh, 1975) to screen for cognitive impairment. All subjects performed a WM task while being scanned with a 3-T MRI scanner, and subsequently an LTM task outside the scanner. The session ended with a computerized version of a vocabulary test (Lehrl, 2005), a marker test for crystallized intelligence. The study was approved by the ethical review board of the canton of Zurich. The participants were compensated with either 60 Swiss Francs (about 60 USD) or partial course credit for the two-hour session.

Paradigm

The paradigm is the same as reported in a recent study (Bartsch et al., 2018), adapted for use in the MRI scanner. We asked participants to remember six nouns in serial order (see Figure 1). After list presentation, either the first three words or the last three words were to be processed

again in one of four ways, depending on the experimental condition. During encoding it was not predictable which half of the items would have to be processed. In the *repeat* condition, the three words appeared again sequentially on the screen, and the subjects had to simply re-read them silently. In the *refreshing* condition, the to-be-processed words were replaced by refreshing prompts appearing at the same location. The subjects were instructed to "think of" the corresponding words as soon as the prompts were shown. In the *elaboration* condition, the three to-be-processed words were shown again sequentially on the screen, and subjects were instructed to generate a vivid mental image of the three objects interacting. The stimuli appearing on the screen in that condition did not differ from the repeat condition, leaving the instruction to form a vivid mental image as the only difference between these conditions. Finally, in the combined *refreshing with elaboration* condition the participants had to "think of" the words replaced by the prompts, and additionally form a vivid mental image of those items. Again, the event sequence of this condition does not differ from the refreshing condition apart from the instruction to form a mental image. Memory was tested with a four-alternatives forced-choice task, which we describe in detail below (see Section Procedure: Working memory task).

The experiment used a 2 x 2 x 2 x 2 (repeat/refresh [repeat, refresh] x elaboration [with elaboration, without elaboration] x processing [processed triplet, unprocessed triplet] x age [young adults, older adults]) within-subject, between-age group design. As in the studies of Ray and colleagues, we compared instructed refreshing to a repeating (re-reading) baseline during the maintenance phase of a WM task (e.g. Raye et al., 2002). In the two additional conditions we instructed participants to elaborate a subset of the items they held in memory. Elaboration logically entails attending to the words, either in memory or in the environment. When elaboration is applied to words just encoded into WM, but no longer presented, it entails

refreshing, whereas when elaboration is applied to words while they are presented, it entails (re-)reading, as in the *repeat* condition. Therefore, we implemented two elaboration conditions: One in which words were repeated and elaborated, and one in which they were refreshed and elaborated. Contrasting these two conditions allowed us to gauge any unique effects of elaboration. In addition, this design allowed us to evaluate whether combining elaboration with refreshing is more effective than either of them alone.

How can we measure the effect of refreshing in our paradigm? The Johnson et al. studies – testing the effect of refreshing on LTM – used *repeat* as the baseline, and therefore we followed their precedent for assessing the effect of refreshing on LTM. For assessing the effect of refreshing on WM, the *repeat* condition is not a suitable baseline because it provided a second chance for encoding the word into WM. Therefore, we assessed the effect of refreshing against the baseline used in Souza et al. (2016; i.e., comparison within the memory set between items refreshed more vs. less) by comparing the items that were processed in refreshing trials to the items within the same trial that were not processed further after initial encoding.

Materials

The stimuli were nouns randomly drawn from a pool of 863 German abstract and concrete nouns for each subject. The nouns were between three and 15 letters long and had a mean normalized lemma frequency of 30.81/million (drawn from the dlexdb.de lexical database).

Procedure: Working memory task

The sequence of an experimental trial is illustrated in Figure 1. The experiment was performed using Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley,

CA, www.neurobs.com). The six to-be-remembered words in each trial were sequentially presented in boxes from top to bottom on the screen, each for 500 ms. A processing cue was presented 1000 ms after the last memory item, indicating whether the first half or the second half of the list had to be processed again, and in which way. In the repeat and elaboration conditions, each word in the to-be-processed triplet was shown again for 1400 ms, followed by a 600 ms inter-stimulus interval. In the refreshing and refreshing-with-elaboration conditions, each to-be-processed word of a triplet was replaced by a refreshing prompt (*#?#?#*) in its corresponding box, and participants were instructed to "think of" the word in that box. In the elaboration and refreshing-with-elaboration conditions, participants were additionally instructed to form a vivid mental image of the three words interacting with each other.[1]

After processing the words in the cued triplet, participants' memory for each list item was tested in their order of presentation using a 4-alternative forced-choice procedure. For each tested item, four words were presented from which the subject could choose the correct word in the currently tested list position with a button press. All test sets included the following four response options: the target (i.e., correct) word, one lure from the same triplet of words within the present list, one lure from the other triplet of the present list, and one new word. This choice had to be made for each of the serial positions successively and with a time limit of 2500 ms for the young and 3500 ms for the older adults per serial position to ensure controlled timing for the fMRI image acquisition. We applied this 4-alternative forced-choice recognition task in order to

[1] The timing parameters where chosen based on a pilot experiment with young adults, which allowed participants to process the items in each of the 4 experimental conditions in a self-paced mode. The mean processing times (PT) where PT = 1419 ms in the repeat condition, PT = 1491 ms in the elaboration condition, PT = 1197 ms in the refreshing, and PT = 1198 ms in the refreshing with elaboration condition.

test both memory for items (i.e., discriminating between items that have been presented in the current memory list and new items) and for serial order (i.e., discriminating between the item in the tested position and other list items).

Within each block of four trials, the same type of processing was instructed throughout, and a screen repeating the instructions of the particular condition was shown prior to the beginning of each block. The order of the condition blocks was randomized between subjects. Each of the four fMRI runs consisted of four blocks, one for each condition (with 4 trials per block as described above).

Procedure: Long-term memory task

After leaving the scanner participants were brought into a separate room, where they performed the computerized LTM task. We assessed participants' LTM for the words they had encoded for the WM tests throughout the experiment. To this end, we presented in each trial the first word of a triplet from one of the studied memory lists. We asked participants to choose, from four different options, the word that had followed the target word in that triplet. The probe words included the correct word (i.e., which could be either the word in the second or third position of the target triplet for the first prompt, and the fifth or sixth word for the second prompt), two words from another list, and a new word. This allowed us to keep the format of the LTM test very similar to the WM test, and furthermore to compare in each trial the memory performance for words from the processed and from the unprocessed triplets. As in the WM test, the LTM test also provided information about both item memory (i.e., which words have been presented in the experiment) and relational memory (i.e., which words have been together in a triplet). The participants were made aware of the LTM test before the start of the experiment.

fMRI Data Acquisition and Preprocessing

Whole brain images were acquired with the 3 T Philips Ingenia MRI scanner with a 32-channel head coil, located at the University Hospital Zurich, Switzerland. High-resolution T1-weighted images were acquired for all subjects with a Turbo field echo (TFE) sequence (8ms time repetition (TR), 3.7ms time echo (TE), 8° flip angle, 160 sagittal slices, 240 × 240 inplane, 1.0mm isotropic). Blood oxygen level-dependent (BOLD)-sensitive functional MRI data were acquired using a gradient-echo, echo planar sequence (2 s TR, 35ms TE) within a 72 × 70 matrix (32 transverse slices, 3 mm isotropic).

Following the acquisition of the structural images, four MRI acquisition runs were collected for each subject, in which they performed a 10-min block of a six-item WM task with a processing delay. fMRI data preprocessing (slice-time correction and realignment) was performed with SPM12 (Penny, Friston, Ashburner, Kiebel, & Nichols, 2011). Subjects' functional scans were aligned by realigning the first volume in each run to the first volume of the first run, and then registering each image in each run to the first volume of that run. The middle functional slice served as a reference for slice-time correction. Further, the functional volumes were co-registered to the T1 anatomical image.

Analysis of Behavioral Data

All data and analysis scripts can be assessed on the Open Science Framework (osf.io/p2h8b/).We analyzed the behavioral data using a Bayesian generalized linear mixed model (BGLMM) implemented in the R package *rstanarm* (Stan Development Team, 2018) following the exact analysis pipeline reported by Bartsch and colleagues (2018). The dependent variable was the number of correct and incorrect responses in each cell of the design per

participant. Correct responses were defined as choosing the target item from the four alternatives. Bayesian procedures provide posterior probability distributions of the model parameters (i.e., the regression weights) that express uncertainty about the estimated parameters. The highest density regions (HDRs) of these posteriors can be used for statistical inference. A 95% HDR represents the range in which the true value of a parameter lies with probability 0.95, given model and data (Morey, Hoekstra, Rouder, Lee, & Wagenmakers, 2016). If zero lies outside the Bayesian HDR there is strong evidence for the existence of the corresponding effect. Although the strength of evidence varies continuously, for simplicity we will describe effects as "credible" if their HDRs exclude zero. We used an MCMC algorithm (implemented in Stan; Carpenter et al., 2017) that estimated the posteriors by sampling parameter values proportional to the product of prior and likelihood. These samples are generated through 4 independent Markov chains, with 1000 warmup samples each, followed by 1000 samples drawn from the posterior distribution which were retained for analysis. Following Gelman and colleagues (2013), we confirmed that the 4 chains converged to the same posterior distribution by verifying that the $\hat{R}$ statistic – reflecting the ratio of between-chain variance to within-chain variance – was < 1.01 for all parameters, and we visually inspected the chains for convergence.

Generation of ROIs

We included voxels within a distributed mask of ROIs – encompassing frontal, fusiform and parietal regions – that were previously reported in fMRI studies investigating either refreshing or elaboration and that had shown subsequent memory effects and/or significant activation differences between repeating and refreshing or elaboration in univariate analyses (see Table 1 for details). The search for those ROIs was performed using the neurosynth.org database

and keyword-based search in pubmed.gov. Anatomical ROIs were generated using an automated parcellation method from *FreeSurfer*. Briefly, a surface mesh model was reconstructed for each subject's brain. Each subject's surface was then auto-parcellated based on the folding pattern of the gyri and sulci. We constructed the combined *frontal-fusiform-parietal* mask using *fslmaths*, encompassing Brodmann areas 4, 6, 8, 9, 10, 44, 45, 46 and 47 for the frontal mask, and Brodmann area 3, 7 and 40 for the parietal mask. The fusiform mask consisted of the fusiform label of the *aparc* atlas.

Multivariate Pattern Analyses of fMRI Data

MVPA provides greater inferential power than classical univariate approaches due to its higher sensitivity at detecting information in neural signals. As a result, MVPA has led to the successful within-category decoding of the contents of WM at an item level (LaRocque, Riggall, Emrich, & Postle, 2017) as well as the characterization of neural representations in different states of WM (Christophel, Iamshchinina, Yan, Allefeld, & Haynes, 2018; Lewis-Peacock, Drysdale, Oberauer, & Postle, 2012; Lewis-Peacock, Drysdale, & Postle, 2015; Rose et al., 2016). The sensitivity of MVPA was further established by a study demonstrating that allegedly category-selective brain regions detected in univariate analyses of the BOLD signal during the delay period of a WM task still carried patterns of activity associated with another category of information that was currently relevant for behavior (Lewis-Peacock & Postle, 2012).

MVPA was performed in MATLAB using the Princeton MVPA toolbox (http://code.google.com/p/princeton-mvpa-toolbox). The classification algorithm used for this analysis was a L2-regularized binary logistic regression (predicting one category vs. the others), that uses Carl Rasmussen's conjugate gradient minimization algorithm, with a penalty term of 50. The classification was performed in the anatomically defined ROI (the frontal-fusiform-parietal

mask) defined above. All neural data were high-pass filtered with a cut-off of 128 seconds and z-scored across trials, within runs, before running MVPA. We performed ANOVA-based feature selection of all active voxels within the ROI and chose the voxels that individually were able to discriminate between the four conditions (repeat, refreshing, elaboration and refreshing with elaboration) significantly ($p < .05$) over the course of the experiment. This univariate feature selection technique has been shown to reliably improve classification accuracy in MVPA of fMRI (Lewis-Peacock & Norman, 2014). To avoid circularity in the data analysis, feature selection was performed separately for each iteration of the cross-validation classifier training algorithm, using independent training and testing sets in each iteration (Kriegeskorte, Simmons, Bellgowan, & Baker, 2009). For subjects who did not show successful classification in the larger ROI, we investigated whether any of the three subregions could differentiate the processes (i.e. frontal, fusiform and parietal). If this were the case, it would indicate that at least one of the subregions contributed more noise than signal to the classification problem. The number of participants for which subregion selection was used is reported in the Results. The average number of feature-selected voxels across young participants was 785.98 (SEM = 90.96) and 1095.3 (SEM = 172.11) across older participants, and the pattern of activity across these voxels was used as the input to each participant's pattern classifiers.

MVPA — multiple process discrimination

The classification procedure used k-fold cross-validation on the data from the WM task. Preprocessed fMRI data from each 6-s processing period (three volumes) from the WM trials, after accounting for a 6-s hemodynamic lag, were used for the analysis. Our analysis scheme incorporated each functional volume (acquired over a 2-s TR) as a separate training event, so that

every trial resulted in three events. Each event was associated with an array of features corresponding to BOLD signals in voxels in the ROI being used.

The k-fold cross-validation scheme (k = 4, for each of the runs) trained a classifier, separately for each participant, on the data of the four conditions (repeat, refreshing, elaboration and refreshing with elaboration) from three runs and then used this classifier to test the data from the withheld run. This process was repeated until every run had been held out for testing. The statistical significance of classifier accuracy was evaluated by performing permutation tests on relabeled training data, in each cross-validation fold, and comparing the resulting distribution of classifier accuracies to the true (unshuffled labels) classifier accuracy with a one-sample, one-tailed t-test. This analysis scheme was performed for the whole ROI and for the single masks of frontal, fusiform and parietal regions. Finally, classification performance was also assessed using receiver operating characteristic (ROC) curves, which rank the classification outputs according to their probability estimates (from strongly favoring the target class to favoring one of the three non-target classes) and chart the relationship between the classifier's *true positive rate* (probability of correctly labeling examples of the target class) and *false positive rate* (probability of incorrectly labeling examples of the non-target class as the target class) across a range of decision boundaries. The area under the curve (AUC) indexes the mean accuracy with which a randomly chosen trial of the target class could be assigned to their correct classes (0.5 = random performance; 1.0 = perfect performance).

MVPA — discrimination from repeat

In order to assess how the neural classification of the refreshing process as well as the neural classification of the elaboration process relates to an individual's task performance, we

used the repeat condition as a reference. First, we extracted classification scores from repeat and refresh trials only, using the classifiers that were trained on all four processes. The same was done for the perceptually identical conditions of repeat and elaboration. Once again, we assessed classifier performance for each binary classification problem using AUC. To evaluate whether the degree of neural separability between the conditions relates to the individuals' memory performance, we performed a logistic regression relating the evidence values from the trained classifier for the respective condition on each trial to the WM outcome of that trial. To increase statistical power for this analysis, we performed a non-parametric bootstrap analysis using data sampled from all participants (see Lewis-Peacock, Cohen, & Norman, 2016).

Researcher Degrees of Freedom

Analyses of neural data involve many decisions, and when these decisions are informed by the data to be analyzed, there is a risk that they are biased in favor of a desired outcome (Simmons, Nelson, & Simonsohn, 2011). Some aspects of our analysis plan (in particular, the decision to use an anatomically defined ROI for the MVPA analyses, rather than whole-brain classification or searchlight analysis) were informed by the data of the young adults. Our analysis of the older adults' data, however, used the exact same analysis pipeline as that for the young adults without any adjustment informed by the older adults' data. Therefore, any convergent finding in both age groups can be thought of as having been directly replicated in a different population. For any finding that differs between age groups, there remains an ambiguity as to whether the divergence reflects a failure to replicate the finding in the young-adult sample, or a genuine age difference. Resolving this ambiguity requires a replication of the entire study with the present analysis plan.

Results

Behavioral Results

We replicated all effects of the young adults reported in a previous study (Bartsch et al. 2018). Figure 2 shows the estimated proportion of correct responses and their corresponding 95% highest posterior density regions for the immediate and delayed memory data. The posterior effect estimates are presented in Table 2 and Table 3. A first question was whether our manipulation of processing half of a memory list had an effect on memory. The credible main effect of processing on immediate and delayed memory supported an effect of our manipulation: Participants had better memory for items that were processed again after initial encoding than for items from the unprocessed triplets (see Table 2 & Table 3 and Figure 2). There was also a main effect of age, such that older adults showed worse memory performance on tests of both WM and LTM.

Working memory performance. We first tested how the effect of refreshing a subset of words in WM compares to the effect of repeated reading of these words. This is the comparison through which Johnson and colleagues evaluated the effect of refreshing on delayed memory (Johnson et al., 2002; Raye et al., 2007). There was a main effect of repeat/refresh (Table 2), but with an advantage of repeating over refreshing. This main effect was further qualified by the two-way interaction of processing and repeat/refresh, indicating that repeated words benefited more from being processed again than refreshed words did. Nevertheless, a beneficial effect of processing was found for both repeated words ($\Delta = 0.34$, 95% HDR = [0.31, 0.37]) and refreshed words ($\Delta = 0.12$, 95% HDR = [0.10, 0.15]). Furthermore, the factor of repeat/refresh interacted with age, indicating that older adults had a greater advantage of repeat over refreshed trials than

young adults. Nevertheless, the repeat-refresh difference appeared for both, young (Δ = 0.16, 95% HDR = [0.13, 0.18]) and older adults (Δ = 0.09, 95% HDR = [0.06, 0.12]).

The BGLMM revealed no credible evidence for a main effect of elaboration on WM performance, or for any of the interactions involving elaboration (see Table 2).

Long-term memory performance. The BGLMM revealed evidence for a main effect of repeat/refresh on LTM performance, but as with WM, there was an advantage for repeating over refreshing (see Table *3*). There was no evidence for any further interaction including the repeat/refresh factor. Hence, contrary to the findings of Johnson and colleagues, refreshing did not lead to better LTM than repeated reading. Note that the above pattern of results also holds for a lenient score of performance in the LTM task, counting all responses showing correct item memory (i.e. the target, same-list items, and other-list items) as correct responses.

Furthermore*,* the analysis of the LTM data revealed evidence for an interaction of elaboration with age (see Table *3*). Follow-up analyses of the interaction revealed that a beneficial effect of elaboration appeared only for young (Δ = 0.05, 95% HDR = [0.02, 0.06]), but not older adults (Δ = -0.01, 95% HDR = [-0.04, 0.3]). In sum, memory was better for trials with instructed elaboration than for those without, but only for the young and not the older adults. The above evidence speaks for an age-dependent beneficial effect of elaboration on LTM that is lost in older age.

To summarize, our results provide no evidence for an effect of refreshing on LTM for either age group; instead we replicated the benefit of elaboration on LTM but only for young adults.

MVPA Results

Young adults

Repeat vs. Refresh vs. Elaborate vs. Refreshing with Elaboration.

The classification scores for each individual were converted to a sensitivity score, accounting for both hits and false alarms, by computing the area under the ROC curve (AUC) for the four-way classification. For 24 of the 30 subjects classification of repeat, refresh, elaborate and refreshing-with-elaboration processes was successful (i.e., significantly better than chance with $p < .05$) in the predefined anatomical ROI (M_{AUC} = 55.92 %, SD_{AUC} = 3.48%, see Figure 3a). Data from the six remaining subjects were nevertheless included in the subsequent analyses, because excluding them would artificially restrict the range of classifier accuracy values that we used as predictors of behavior. The majority of the subjects (N = 19) showed classification in the whole mask of frontal–fusiform–parietal regions, and the remaining subjects (N = 5) showed good classification in the individual regions of the whole mask (Frontal: N = 3, Fusiform: N = 1, Parietal: N = 1). Notably, whole-brain classification was less successful than this ROI-based approached (M_{AUC} = 54.79%, SD_{AUC} = 2.69%, i.e., significantly better than chance with $p < .05$).

Linking neural classification to memory performance

Refresh vs. Repeat

The classifier evidence values for refresh and repeat events were extracted from the four-way classifier and the resulting classification scores for each of the 30 individuals were converted to a sensitivity score, accounting for both hits and false alarms, by computing the area under the ROC curve (AUC) for the *refresh* vs. *repeat* classification. The mean classifier AUC for separating re-reading from refreshing was 0.557 (*SD* = .060, chance = 0.500). We performed a logistic regression relating the evidence values for the two processes on each trial to the WM

outcome of that trial, and applied non-parametric bootstrapping using data sampled from all participants, in order to increase statistical power for this analysis (Efron, 1992; Lewis-Peacock et al., 2016) On each bootstrap iteration (n = 10,000), 30 participants were selected at random, with replacement, and their data were combined into a single *supersubject* for fixed-effects analysis. The stability of these effects across all iterations was analyzed to assess population-level reliability. For refreshing trials, WM outcome was operationalized in two ways: (1) as a binary outcome variable of whether the *refreshing processing benefit* (defined as the contrast between processed and not-processed triplet) of this trial was larger or smaller than the mean *repeat processing benefit* of this individual and (2) as a binary outcome variable of whether there was a *refreshing processing benefit*. None of the analyses revealed a significant relationship between WM outcome variables and neural separability (all ps > 0.31).

In summary, these results indicate that although repeating items benefited WM performance more than refreshing did, the neural pattern separability of these processes was not predictive of the size of this benefit.

Elaborate vs. Repeat

The classifier evidence values for elaborate and repeat events were extracted from the four-way classifier and the resulting classification scores for each of the 30 individuals were converted to an AUC sensitivity score for the elaborate vs. repeat classification. The mean classifier AUC for separating repeating from elaboration was 0.591 (SD = .090, chance = 0.500). To assess how the neural classification of the elaboration process relates to an individual's task performance, we again used the repeat condition as a reference. We performed a logistic regression relating the evidence values for elaboration (relative to repeat) on each trial to the

memory performance of that trial. Elaboration had no behavioral effect on WM, but instead showed a benefit for LTM. Therefore, our analysis focused on the behavioral contrasts in the LTM accuracy data: LTM outcomes were operationalized as a binary outcome variable of whether the *processing benefit of this elaboration* trial was larger or smaller than the mean *repeat processing benefit* of this individual. As shown in Figure 4a, stronger neural evidence for elaboration was related to a larger elaboration benefit in LTM ($\beta = 1.12$, $p=.005$). There was no such relationship for WM outcomes ($\beta = 0.2$, $p=.449$). The more elaborating on stimuli was neurally distinct from merely repeating those stimuli, the larger its beneficial effect on LTM beyond simply re-reading the words.

Older Adults

Repeat vs. Refresh vs. Elaborate vs. Refreshing with Elaboration.

The same analysis pipeline for the four-way problem in the young adults was subsequently applied to the independent sample of 27 older adults. For 17 subjects the classification of *repeat vs. refresh vs. elaborate vs. refreshing with elaboration* was significantly above chance in the in the predefined anatomical ROIs, with a mean classifier *AUC* of 0.5414 (*SD* =0.033, see Figure 3c). This result indicates that the four processes were neurally separable also in older adults, yet a smaller proportion of older subjects showed successful classification of the processes (17/27; 63%), compared to the young adult group (24/30; 80%). A t-test comparing the AUCs of young adults to those of older adults showed no significant difference between the age-groups ($t(35.7) = 1.58$, $p = 0.062$). The majority of these subjects (N = 13) showed good classification in the whole mask of frontal–fusiform–parietal regions, and the remaining subjects (N = 4) showed good classification in the individual regions of the whole mask (Frontal: N = 2,

Fusiform: N = 2). Data from the ten subjects for which the cross-validation classification accuracy was not significantly above chance were nevertheless included in the subsequent analyses.

Linking neural classification to memory performance

Repeat vs. Refreshing

Equivalently to the analysis of the young adults' data, the evidence values for refresh and repeat events were extracted from the four-way classifier, and the resulting classification scores for each of the 30 individuals were converted to a sensitivity score, accounting for both hits and false alarms, by computing the area under the ROC curve (AUC) for the refresh vs. repeat classification. The mean classifier AUC for separating re-reading from refreshing was 0.578 (SD = 0.062, chance = 50%). We performed a logistic regression relating the evidence values on each trial to the WM outcome of that trial and applied non-parametric bootstrapping using data sampled from all participants.

As in the young adults, there was no significant relationship between either of the two WM outcomes (*refresh benefit relative to repeat benefit,* or *refresh benefit* alone) and the neural separability of the refreshing and repeating processes (both ps > 0.38).

In summary, these results replicate the findings in the young adults showing that although repeating items benefited WM performance more than refreshing did, the neural pattern separability of these processes was not predictive of the size of this benefit for the older adults.

Repeat vs. Elaborate

The mean classifier AUC for separating re-reading from elaborating was 0.57 (SD = 0.058, chance = 50%). We performed a logistic regression relating the evidence values on each trial to the WM and LTM outcome of that trial and applied non-parametric bootstrapping using data sampled from all participants. Just as in the young adults, elaboration had no behavioral effect on WM, but in contrast to the young adults, the older adults also showed no benefit of elaboration on LTM. Furthermore, an individual's classifier evidence score for elaboration (relative to repeat) was unrelated to WM performance, measured by the *elaboration-minus-repeat benefit* ($\beta = -0.04$, $p=.847$). As shown in Figure 4b and in contrast to young adults, there was also no relationship to LTM performance, indicated by the same contrast ($\beta = -0.09$, $p=.71$).

Discussion

The goal of the present study was to investigate to what extent elaboration and refreshing are separable processes, given prior reports of their neural overlap as well as their similar proposed roles for WM and LTM. We aimed at investigating whether refreshing and elaboration are distinct in their contribution to WM and LTM formation, whether they elicit separable neural activation patterns in fMRI, and how they relate to age-related memory deficits. We compared the neural and behavioral results of these processes to a control condition of re-reading (repeating) the words during the delay-period of a WM task. In the following, we discuss the effects of refreshing and elaboration on WM and LTM and we argue that these processes are distinct and have distinct consequences on memory performance in young and old adults.

Are refreshing and elaboration distinct processes?

If refreshing and elaboration are two labels for the same process, then the pattern of behavioral effects should be the same for WM and on LTM, and the patterns of brain activity supporting these processes should be indistinguishable. In the present study, in a combined mask of a priori brain regions from frontal, temporal, and parietal lobes, we found successful differentiation of brain activity associated with repeated reading, refreshing and elaboration processes. This neural evidence supports the assumption that refreshing and elaboration are implemented with distinct neural processes. In line with this differentiation, we showed that also the combined condition of refreshing with elaboration was successfully discriminable from brain activation patterns from the other three conditions. Combining two distinct processes resulted thereby in a differentiable third mental process.

Further, as discussed in detail below, refreshing and elaboration resulted in distinct behavioral effects on tests of WM and LTM.

How does elaboration affect WM and LTM?

The elaboration process can be distinguished from mere re-reading by the accompanying distributed patterns of fMRI activity in task-relevant regions of the brain. Whereas elaboration showed no benefit for WM, it did facilitate LTM performance for young but not older adults. Accordingly, in young but not older adults, the degree of neural separability of re-reading vs. elaboration was positively related to the individuals' elaboration benefit in LTM: Greater separation between the neural processes of reading and elaboration was associated with larger LTM benefits of elaboration within subjects (Figure 4). The present results confirm prior studies showing evidence against a WM benefit of elaboration). Furthermore, our findings thereby fail to

provide experimental support for the conclusion from previous studies which found that higher WM performance on complex-span tasks was correlated with individuals' use of elaboration strategies such as imagery and sentence generation (Bailey, Dunlosky, & Hertzog, 2009; Bailey et al., 2008, 2011; Dunlosky & Kane, 2007). This discrepancy could be due to the present study using a simple-span paradigm and previous research relying on complex-span tasks. Alternatively, the correlation reported in previous studies might not reflect a causal effect of elaboration on memory – rather, participants who have good memory have more information in memory to elaborate on.

How does refreshing affect WM and LTM?

We replicated the behavioral findings from Bartsch et al. (2018) that repeating items benefited WM performance more than refreshing did, but that there was an advantage of refreshed items (the processed triplets) over the ones that were not refreshed within a list. We further replicated the lack of a LTM benefit for refreshed compared to repeated items, which stands in contrast to previous fMRI studies (e.g. Raye, Johnson, Mitchell, Greene, & Johnson, 2005). One explanation for this could be the use of larger set sizes in the current study and in Bartsch et al (2018): Raye and colleagues asked their participants to refresh one item at a time, whereas in our studies, the subjects had to refresh three items. Still, the processing benefit in WM indicates that the subjects were able to engage in a process that was beneficial to their memory performance. If this refreshing benefit arises through the build-up of LTM traces, we should see this also in a LTM benefit of refreshing for the processed vs. non-processed words.

How do refreshing and elaboration contribute to age-related memory deficits?

A second goal of the present study was to investigate whether refreshing and elaboration and their impacts on memory are preserved in older adults. As in our young adult sample, the three processes of repeating, refreshing, and elaboration were neurally distinguishable in the predefined mask of frontal, parietal and fusiform regions for a majority of the older adults (Figure 3c). The comparison of these processes provided confirmatory evidence that, like young adults, older adults engaged these processes differently, but as in young adults, their degree of neural separability did not relate to subsequent WM performance. The proportion of non-classifiable subjects was larger in the sample of older adults (10 of 27, 37%) than young adults (6 of 30, 20%). This is not surprising, as age has been associated with a decline in segregation of brain networks (e.g. Damoiseaux, 2017; Lorist, Geerligs, Renken, Saliasi, & Maurits, 2014; Morcom & Johnson, 2015), which further has been related to worse cognitive performance (Chan, Park, Savalia, Petersen, & Wig, 2014; Wang et al., 2010).

The behavioral analyses revealed that, like young adults, older adults benefited from processing the items again in the refreshing condition, compared to the not-processed items, as shown in evidence against an interaction of age with the repeat/refresh by processing interaction (see Table 2). Using a similar paradigm to the present study, with cues directing refreshing to a subset of the memoranda after encoding, a recent study (Loaiza & Souza, 2019) confirmed that older adults are able to focus attention on no-longer perceptually available representations in WM, a critical component of refreshing, in conditions without distraction.

Similar to the young adults, refreshing had no benefit on LTM in older adults. This replicates the age-group specific findings of Johnson (2004), who also found no LTM benefit in older adults when comparing refreshing to re-reading. Refreshing was identified as an

independent process, however, as it was neurally separable from both re-reading and elaboration. As refreshing was not related to LTM performance, even in the young adults, we conclude that deficits in refreshing are not responsible for the LTM deficit in older adults either.

The results on elaboration show that the fMRI classifiers were able to differentiate mere re-reading from elaborating in the older adults (Figure 3c&d). However, there was no LTM benefit of elaboration in older adults, whereas this effect was robust in the young group (Figure 2). We argue therefore that most of the older adults did perform some mental manipulation in the elaboration condition that was different from mere re-reading, but whatever it was did not affect their LTM performance. These results are in line with the *elaboration deficit hypothesis* (Smith, 1980), showing that when having to generate their own elaborations (here mental images), older adults do not benefit in the same way as young adults do. Taken together, our results provide evidence that the LTM deficit of older adults might arise at least in part from a deficit in the process of elaboration. Future research might investigate whether age-related LTM deficits can be compensated by providing more precise elaboration instructions.

Conclusion

Our study revealed that the processes of repeated reading, refreshing, elaboration and refreshing with elaboration are differentiable in brain activation patterns in both young and older adults on the basis of neural activation patterns. Elaboration as a process can be neurally distinguished from mere reading. While it had no impact on WM, elaboration did improve episodic LTM for young adults, and the size of the benefit was related to the neural separability of elaboration: The more differentiated elaboration was from re-reading, the more elaboration benefited LTM. In contrast to the young adults, older adults' episodic LTM did not benefit from

elaboration, even though this process was neurally separable from reading. This suggests that older adults implemented a sub-optimal form of elaboration, and this may be a contributing factor to age-related deficits in LTM.

References

Addis, D. R., & McAndrews, M. P. (2006). Prefrontal and hippocampal contributions to the generation and binding of semantic associations during successful encoding. *NeuroImage*, *33*(4), 1194–1206. https://doi.org/10.1016/j.neuroimage.2006.07.039

Baddeley, A. D. (1986). *Working Memory*. Oxford: Oxford University Press.

Bailey, H., Dunlosky, J., & Hertzog, C. (2009). Does differential strategy use account for age-related deficits in working-memory performance? *Psychology and Aging*, *24*(1), 82–92. https://doi.org/10.1037/a0014078

Bailey, H., Dunlosky, J., & Kane, M. J. (2008). Why does working memory span predict complex cognition? Testing the strategy affordance hypothesis. *Memory & Cognition*, *36*(8), 1383–1390. https://doi.org/10.3758/MC.36.8.1383

Bailey, H., Dunlosky, J., & Kane, M. J. (2011). Contribution of strategy use to performance on complex and simple span tasks. *Memory & Cognition*, *39*(3), 447–461. https://doi.org/10.3758/s13421-010-0034-3

Bartsch, L. M., Singmann, H., & Oberauer, K. (2018). The effects of refreshing and elaboration on working memory performance , and their contributions to long-term memory formation. *Memory & Cognition*. https://doi.org/10.3758/s13421-018-0805-9

Blumenfeld, R. S. (2006). Dorsolateral Prefrontal Cortex Promotes Long-Term Memory Formation through Its Role in Working Memory Organization. *Journal of Neuroscience*, *26*(3), 916–925. https://doi.org/10.1523/JNEUROSCI.2353-05.2006

Blumenfeld, R. S., Parks, C. M., & Yonelinas, A. P. (2010). Putting the Pieces Together : The Role of Dorsolateral Prefrontal Cortex in Relational Memory Encoding, 257–265.

Blumenfeld, R. S., & Ranganath, C. (2007). Prefrontal cortex and long-term memory encoding:

an integrative review of findings from neuropsychology and neuroimaging. *Neuroscientist*, *13*(3), 280–291. https://doi.org/10.1177/1073858407299290

Camos, V., Johnson, M. R., Loaiza, V. M., Portrat, S., Souza, A. S., & Vergauwe, E. (2018). What is attentional refreshing in working memory? *Annals of the New York Academy of Sciences*, 1–14. https://doi.org/10.1111/nyas.13616

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., … Riddell, A. (2017). *Stan* : A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01

Chan, M. Y., Park, D. C., Savalia, N. K., Petersen, S. E., & Wig, G. S. (2014). Decreased segregation of brain systems across the healthy adult lifespan. *Proceedings of the National Academy of Sciences*, *111*(46), E4997 LP-E5006. https://doi.org/10.1073/pnas.1415122111

Christophel, T. B., Iamshchinina, P., Yan, C., Allefeld, C., & Haynes, J. D. (2018). Cortical specialization for attended versus unattended working memory. *Nature Neuroscience*, *21*(4), 494–496. https://doi.org/10.1038/s41593-018-0094-4

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, *11*(6), 671–684. https://doi.org/10.1016/S0022-5371(72)80001-X

D'Esposito, M., & Postle, B. R. (2015). The Cognitive Neuroscience of Working Memory, (66), 115–142. https://doi.org/10.1146/annurev-psych-010814-015031.THE

Damoiseaux, J. S. (2017). Effects of aging on functional and structural brain connectivity. *NeuroImage*, *160*(January), 32–40. https://doi.org/10.1016/j.neuroimage.2017.01.077

Davachi, L., Maril, a, & Wagner, a D. (2001). When keeping in mind supports later bringing to mind: neural markers of phonological rehearsal predict subsequent remembering. *Journal of*

Cognitive Neuroscience, *13*(8), 1059–1070. https://doi.org/10.1162/089892901753294356

Dunlosky, J., & Kane, M. J. (2007). The contributions of strategy use to working memory span: A comparison of strategy assessment methods. *The Quarterly Journal of Experimental Psychology*, *60*(9), 1227–1245. https://doi.org/10.1080/17470210600926075

Efron, B. (1992). Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics* (pp. 569–593). Springer.

Eriksson, J., Vogel, E. K., Lansner, A., Bergström, F., & Nyberg, L. (2015). Neurocognitive Architecture of Working Memory. *Neuron*, *88*(1), 33–46. https://doi.org/10.1016/j.neuron.2015.09.020

Fletcher, P. C., Shallice, T., & Dolan, R. J. (2000). "Sculpting the response space" - An account of left prefrontal activation at encoding. *NeuroImage*, *12*(4), 404–417. https://doi.org/10.1006/nimg.2000.0633

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician. *Journal of Psychiatric Research*, *12*(3), 189–198. https://doi.org/10.1016/0022-3956(75)90026-6

Gallo, D. A., Meadow, N. G., Johnson, E. L., & Foster, K. T. (2008). Deep levels of processing elicit a distinctiveness heuristic: Evidence from the criterial recollection task. *Journal of Memory and Language*, *58*(4), 1095–1111. https://doi.org/10.1016/j.jml.2007.12.001

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). Bayesian data analysis, 3rd edition. Chapman & Hall/CRC.

Greene, R. L. (1987). Effects of maintenance rehearsal on human memory. *Psychological Bulletin*, *102*(3), 403–413. https://doi.org/10.1037/0033-2909.102.3.403

Haxby, J. V, Connolly, A. C., & Guntupalli, J. S. (2014). Decoding Neural Representational

Spaces Using Multivariate Pattern Analysis. *Annual Review of Neuroscience*, (June), 435–456. https://doi.org/10.1146/annurev-neuro-062012-170325

Haxby, J. V, Gobbini, M. I., Furey, M. L., Ishai, A., Schouten, J. L., & Pietrini, P. (2001). Distrubuted and Overlapping Representations of Face and Objects in Ventral Temporal Cortex. *Science*, *293*(5539), 2425–2430. https://doi.org/10.1126/science.1063736

Haynes, J., & Rees, G. (2006). Decoding mental states from brain activity in humans. *Nature Reviews Neuroscience*, *7*(7), 523–534. https://doi.org/10.1038/nrn1931

Hoareau, V., Lemaire, B., Portrat, S., & Plancher, G. (2016). Reconciling Two Computational Models of Working Memory in Aging. *Topics in Cognitive Science*, *8*(1), 264–278. https://doi.org/10.1111/tops.12184

Hoyer, W. J., & Verhaeghen, P. (2006). *Memory aging*. *Handbook of the psychology of aging* (Vol. 6).

Johnson, M. K., Mitchell, K. J., Raye, C. L., & Greene, E. J. (2004). An age-related deficit in prefrontal cortical function associated with refreshing information. *Psychological Science*, *15*(2), 127–132. https://doi.org/10.1016/j.neuroscience.2008.09.014

Johnson, M. K., Raye, C. L., Mitchell, K. J., Greene, E. J., & Adam, W. (2003). fMRI Evidence for an Organization of Prefrontal Cortex by Both Type of Proces and Type of Information, 265–273.

Johnson, M. K., Raye, C. L., Mitchell, K. J., Greene, E. J., Cunningham, W. a, & Sanislow, C. a. (2005). Using fMRI to investigate a component process of reflection: prefrontal correlates of refreshing a just-activated representation. *Cognitive, Affective & Behavioral Neuroscience*, *5*(3), 339–361. https://doi.org/10.3758/CABN.5.3.339

Kaakinen, J. K., & Hyönä, J. (2007). Strategy use in the reading span test: An analysis of eye

movements and reported encoding strategies. *Memory*, *15*(6), 634–646. https://doi.org/10.1080/09658210701457096

Kapur, S., Craik, F. I., Tulving, E., Wilson, A. A., Houle, S., & Brown, G. M. (1994). Neuroanatomical correlates of encoding in episodic memory: levels of processing effect. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(6), 2008–2011. https://doi.org/10.1073/pnas.91.6.2008

Kim, S. Y., & Giovanello, K. S. (2011). The Effects of Attention on Age-Related Relational Memory Deficits: Evidence From a Novel Attentional Manipulation. *Psychology and Aging*, *26*(3), 678–688. https://doi.org/10.1037/a0022326

Klatsky, R. (1988). Theories of information processing and theories of aging. In L. Light & D. M. Burke (Eds.), *Language, memory, and aging* (pp. 1–16). New York: Cambridge University Press.

Kriegeskorte, N., & Bandettini, P. A. (2007). Analyzing for information, not activation, to exploit high-resolution fMRI. *NeuroImage*, *38*(4), 649–662. https://doi.org/10.1016/j.neuroimage.2007.02.022

Kriegeskorte, N., Simmons, W., Bellgowan, P., & Baker, C. (2009). Circular analysis in systems neuroscience – the dangers of double dipping. *Nat Neurosci*, *12*(5), 535–540. https://doi.org/10.1038/nn.2303.Circular

LaRocque, J. J., Riggall, A. C., Emrich, S. M., & Postle, B. R. (2017). Within-Category Decoding of Information in Different Attentional States in Short-Term Memory. *Cerebral Cortex*, *27*(10), 4881–4890. https://doi.org/10.1093/cercor/bhw283

Lehrl, S. (2005). *Mehrfachwahl-Wortschatz-Intelligenztest MWT-B* (5th ed.). Balingen: Spitta Verlag.

Lewis-Peacock, J. A., Cohen, J. D., & Norman, K. A. (2016). Neural evidence of the strategic choice between working memory and episodic memory in prospective remembering. *Neuropsychologia*, *93*(October), 280–288. https://doi.org/10.1016/j.neuropsychologia.2016.11.006

Lewis-Peacock, J. A., Drysdale, A. T., Oberauer, K., & Postle, B. R. (2012). Neural evidence for a distinction between short-term memory and the focus of attention. *Journal of Cognitive Neuroscience*, *24*(1), 61–79. https://doi.org/10.1162/jocn_a_00140

Lewis-Peacock, J. A., Drysdale, A. T., & Postle, B. R. (2015). Neural evidence for the flexible control of mental representations. *Cerebral Cortex*, *25*(10), 3303–3313. https://doi.org/10.1093/cercor/bhu130

Lewis-Peacock, J. A., & Norman, K. A. (2013). Multi-voxel pattern analysis of fMRI data. *The Cognitive Neurosciences V*, (512).

Lewis-Peacock, J. A., & Norman, K. A. (2014). Multi-Voxel Pattern Analysis of fMRI Data. In M. S. Gazzaniga & G. R. Mangun (Eds.), *The Cognitive Neurosciences* (5th ed., pp. 911–920). Cambridge, Massachusetts: MIT Press.

Lewis-Peacock, J. A., & Postle, B. R. (2008). Temporary Activation of Long-Term Memory Supports Working Memory, *28*(35), 8765–8771. https://doi.org/10.1523/JNEUROSCI.1953-08.2008.Temporary

Lewis-Peacock, J. A., & Postle, B. R. (2012). Decoding the internal focus of attention. *Neuropsychologia*, *50*(4), 470–478. https://doi.org/10.1016/j.neuropsychologia.2011.11.006

Loaiza, V. M., & Camos, V. (2016). Does Controlling for Temporal Parameters Change the Levels-of-Processing Effect in Working Memory? *Advances in Cognitive Psychology*, *12*(1), 2–9. https://doi.org/10.5709/acp-0182-3

Loaiza, V. M., McCabe, D. P., Youngblood, J. L., Rose, N. S., & Myerson, J. (2011). The influence of levels of processing on recall from working memory and delayed recall tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1258.

Loaiza, V. M., & Souza, A. S. (2019). An age-related deficit in preserving the benefits of attention in working memory. *Psychology and Aging*. https://doi.org/10.1037/pag0000326

Lorist, M. M., Geerligs, L., Renken, R. J., Saliasi, E., & Maurits, N. M. (2014). A Brain-Wide Study of Age-Related Changes in Functional Connectivity. *Cerebral Cortex*, *25*(7), 1987–1999. https://doi.org/10.1093/cercor/bhu012

Maillet, D., & Rajah, M. N. (2014). Age-related differences in brain activity in the subsequent memory paradigm: A meta-analysis. *Neuroscience and Biobehavioral Reviews*, *45*, 246–257. https://doi.org/10.1016/j.neubiorev.2014.06.006

Morcom, A. M., & Johnson, W. (2015). Neural Reorganization and Compensation in Aging. *Journal of Cognitive Neuroscience*, *27*(7), 1275–1285. https://doi.org/10.1162/jocn

Morey, R. D., Hoekstra, R., Rouder, J. N., Lee, M. D., & Wagenmakers, E.-J. (2016). The fallacy of placing confidence in confidence intervals. *Psychonomic Bulletin & Review*, *23*(1), 103–123. https://doi.org/10.3758/s13423-015-0947-8

Murray, L. J., & Ranganath, C. (2007). The dorsolateral prefrontal cortex contributes to successful relational memory encoding. *J Neurosci*, *27*(20), 5515–5522. https://doi.org/10.1523/JNEUROSCI.0406-07.2007

Naveh-Benjamin, M., & Old, S. R. (2008). 2.41 Aging and Memory. *Learning and Memory: A Comprehensive Reference*, 787–808. https://doi.org/http://dx.doi.org/10.1016/B978-012370509-9.00133-9

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-

voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, *10*(9), 424–430. https://doi.org/10.1016/j.tics.2006.07.005

Penny, W. D., Friston, K. J., Ashburner, J. T., Kiebel, S. J., & Nichols, T. E. (2011). *Statistical parametric mapping: the analysis of functional brain images*. Elsevier.

Plancher, G., Boyer, H., Lemaire, B., & Portrat, S. (2017). Under Which Conditions Can Older Participants Maintain Information In Working Memory? *Experimental Aging Research*, *43*(5), 409–429. https://doi.org/10.1080/0361073X.2017.1369730

Ragland, J. D., Blumenfeld, R. S., Ramsay, I. S., Yonelinas, A. P., Yoon, J., Solomon, M., … Ranganath, C. (2012). Neural correlates of relational and item-specific encoding during working and long-term memory in schizophrenia. *NeuroImage*, *59*(2), 1719–1726. https://doi.org/10.1016/j.neuroimage.2011.08.055

Ranganath, C. (2006). Working memory for visual objects: Complementary roles of inferior temporal, medial temporal, and prefrontal cortex. *Neuroscience*, *139*(1), 277–289. https://doi.org/10.1016/j.neuroscience.2005.06.092

Ranganath, C., & Blumenfeld, R. S. (2005). Doubts about double dissociations between short- and long-term memory. *Trends in Cognitive Sciences*, *9*(8), 374–380. https://doi.org/10.1016/j.tics.2005.06.009

Ranganath, C., Cohen, M. M. X., & Brozinsky, C. C. J. (2005). Working memory maintenance contributes to long-term memory formation: neural and behavioral evidence. *Journal of Cognitive Neuroscience*, *17*(7), 994–1010. https://doi.org/10.1162/0898929054475118

Rankin, J. L., & Collins, M. (1985). Adult age differences in memory elaboration. *Journal of Gerontology*, *40*(4), 451–458.

Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2005). Refreshing:

A Minimal Executive Function. *Control*, *43*(1), 135–145. https://doi.org/10.1016/S0010-9452(08)70451-9

Raye, C. L., Johnson, M. K., Mitchell, K. J., Greene, E. J., & Johnson, M. R. (2007). Refreshing : a Minimal Executive Function. *Cortex*, *43*(1), 135–145. https://doi.org/10.1016/S0010-9452(08)70451-9

Raye, C. L., Johnson, M. K., Mitchell, K. J., Reeder, J. a, & Greene, E. J. (2002). Neuroimaging a single thought: dorsolateral PFC activity associated with refreshing just-activated information. *NeuroImage*, *15*(2), 447–453. https://doi.org/10.1006/nimg.2001.0983

Raye, C. L., Mitchell, K. J., Reeder, J. a, Greene, E. J., & Johnson, M. K. (2008). Refreshing one of several active representations: behavioral and functional magnetic resonance imaging differences between young and older adults. *Journal of Cognitive Neuroscience*, *20*(5), 852–862. https://doi.org/10.1162/jocn.2008.20508

Rose, N. S., Buchsbaum, B. R., & Craik, F. I. M. (2014). Short-term retention of a single word relies on retrieval from long-term memory when both rehearsal and refreshing are disrupted. *Memory & Cognition*, *42*(5), 689–700. https://doi.org/10.3758/s13421-014-0398-x

Rose, N. S., & Craik, F. I. M. (2012). A processing approach to the working memory/long-term memory distinction: Evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*(4), 1019–1029. https://doi.org/10.1037/a0026976

Rose, N. S., Craik, F. I. M., & Buchsbaum, B. R. (2014). Levels of Processing in Working Memory: Differential Involvement of Fronto-Temporal Networks. *Journal of Cognitive Neuroscience*, *27*(3), 522–532. https://doi.org/DOI: 10.1162/jocn_a_00738

Rose, N. S., LaRocque, J. J., Riggall, A. C., Gosseries, O., Starrett, M. J., Meyering, E. E., &

Postle, B. R. (2016). Reactivation of latent working memories with transcranial magnetic stimulation. *Science*, *354*(6316), 1136 LP-1139. Retrieved from http://science.sciencemag.org/content/354/6316/1136.abstract

Rose, N. S., Myerson, J., Roediger III, H. L., Hale, S., Roediger, H. L., & Hale, S. (2010). Similarities and differences between working memory and long-term memory: Evidence from the levels-of-processing span task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(2), 471. https://doi.org/10.1037/a0018405

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*(11), 1359–1366.

Smith, A. D. (1980). Age differences in encoding, storage, and retrieval. *New Directions in Memory and Aging*, 23–46.

Souza, A. S., & Oberauer, K. (2017). The contributions of visual and central attention to visual working memory. *Attention, Perception, & Psychophysics*. https://doi.org/10.3758/s13414-017-1357-y

Souza, A. S., Vergauwe, E., & Oberauer, K. (2018). Where to attend next: guiding refreshing of visual, spatial, and verbal representations in working memory. *Annals of the New York Academy of Sciences*, 20–22. https://doi.org/10.1111/nyas.13621

Stan Development Team. (2018). rstanarm: {Bayesian} applied regression modeling via {Stan}. Retrieved from http://mc-stan.org/

Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of Memory* (pp. 381–403). New York: New York: Academic Press.

Wagner, A. D., Poldrack, R. A., Eldridge, L. L., Desmond, J. E., Glover, G. H., & Gabrieli, J. D.

E. (1998). Lateralization of Prefrontal Activation During Episodic. *NeuroReport*, *9*(16), 3711–3717.

Wang, L., LaViolette, P., O'keefe, K., Putcha, D., Bakkour, A., Van Dijk, K. R. A., … Sperling, R. A. (2010). Intrinsic connectivity between the hippocampus and posteromedial cortex predicts memory performance in cognitively intact older individuals. *Neuroimage*, *51*(2), 910–917.

Wechsler, W. (1982). Manual for the Hamburg-Wechsler intelligence test for adults (HAWIE). *Bern, Switzerland: Huber*.

Zacks, R. T., Hasher, L., & Li, K. Z. H. (2000). Human Memory. *The Handbook of Aging and Cognition*, 293–357. https://doi.org/10.3758/s13423-011-0086-9

Tables

Table 1 ROIs with their corresponding BAs and references of previous reporting in univariate analyses in the literature.

Label	sub region	BA	Labeled region reported in
frontal	inferior frontal	44,45,47	Johnson et al., 2005; Johnson, Mitchell, Raye, & Greene, 2004; Raye, Johnson, Mitchell, Greene, & Johnson, 2007; Raye, Mitchell, Reeder, Greene, & Johnson, 2008
	middle & superior frontal	4,6, 8,9,10,46	Blumenfeld, 2006; Blumenfeld, Parks, & Yonelinas, 2010; Kim & Giovanello, 2011; Murray & Ranganath, 2007
parietal		3,7,40	Johnson et al., 2004; Kim & Giovanello, 2011; Murray & Ranganath, 2007; Raye, Johnson, Mitchell, Reeder, & Greene, 2002; Raye et al., 2007, 2008
fusiform		19, 37	Murray & Ranganath, 2007; Raye et al., 2008

Table 2 The posterior effect estimates and their 95% HDRs of the generalized linear mixed model for binomial response variables for the immediate serial memory data.

	mean parameter on probit scale	**95% HDR**
(Intercept)	0.09	[-0.01, 0.19]
processing	**0.44**	**[0.38, 0.49]**
repeat/refresh	**-0.23**	**[-0.26, -0.19]**
elaboration	0.01	[-0.03, 0.05]
age	**-0.29**	**[-0.44, -0.15]**
processing * repeat/refresh	**-0.30**	**[-0.35, -0.26]**
processing * elaboration	-0.04	[-0.08, 0.01]
repeat/refresh * elaboration	-0.02	[-0.07, 0.02]
age * processing	0.01	[-0.07, 0.08]
age * repeat/refresh	**0.09**	**[0.04, 0.14]**
age * elaboration	0.05	[0.00, 0.10]
processing * repeat/refresh * elaboration	0.02	[-0.02, 0.08]
processing * repeat/refresh * age	-0.05	[-0.11, 0.02]
processing * elaboration * age	0.02	[-0.05, 0.08]
repeat/refresh * elaboration * age	-0.06	[-0.12, 0.00]
processing * repeat/refresh * elaboration * age	0.00	[-0.09, 0.08]

Note. Credible effects, defined as HDRs excluding zero, are printed in bold.

Table 3 The posterior effect estimates and their 95 % HDRs of the generalized linear mixed model for binomial response variables for the delayed memory data.

	mean parameter on probit scale	**95% HDR**
(Intercept)	**-0.29**	**[-0.38 -0.25]**
processing	**0.06**	**[0.01, 0.1]**
repeat/refresh	**-0.05**	**[-0.09, -0.01]**
elaboration	-0.04	[-0.09, 0]
age	**-0.06**	**[-0.11, -0.01]**
processing * repeat/refresh	-0.04	[-0.09, 0.02]
processing * elaboration	-0.01	[-0.07, 0.05]
repeat/refresh * elaboration	0.05	[-0.01, 0.11]
age * processing	0.03	[-0.03, 0.09]
age * repeat/refresh	-0.05	[-0.11, 0.01]
age * elaboration	**0.08**	**[0.01, 0.14]**
processing * repeat/refresh * elaboration	-0.03	[-0.12, 0.05]
processing * repeat/refresh * age	0.05	[-0.03, 0.14]
processing * elaboration * age	0.05	[-0.04, 0.13]
repeat/refresh * elaboration * age	-0.04	[-0.13, 0.05]
processing * repeat/refresh * elaboration * age	-0.04	[-0.15, 0.08]

Note. Credible effects, defined as HDRs excluding zero, are printed in bold.

Figures

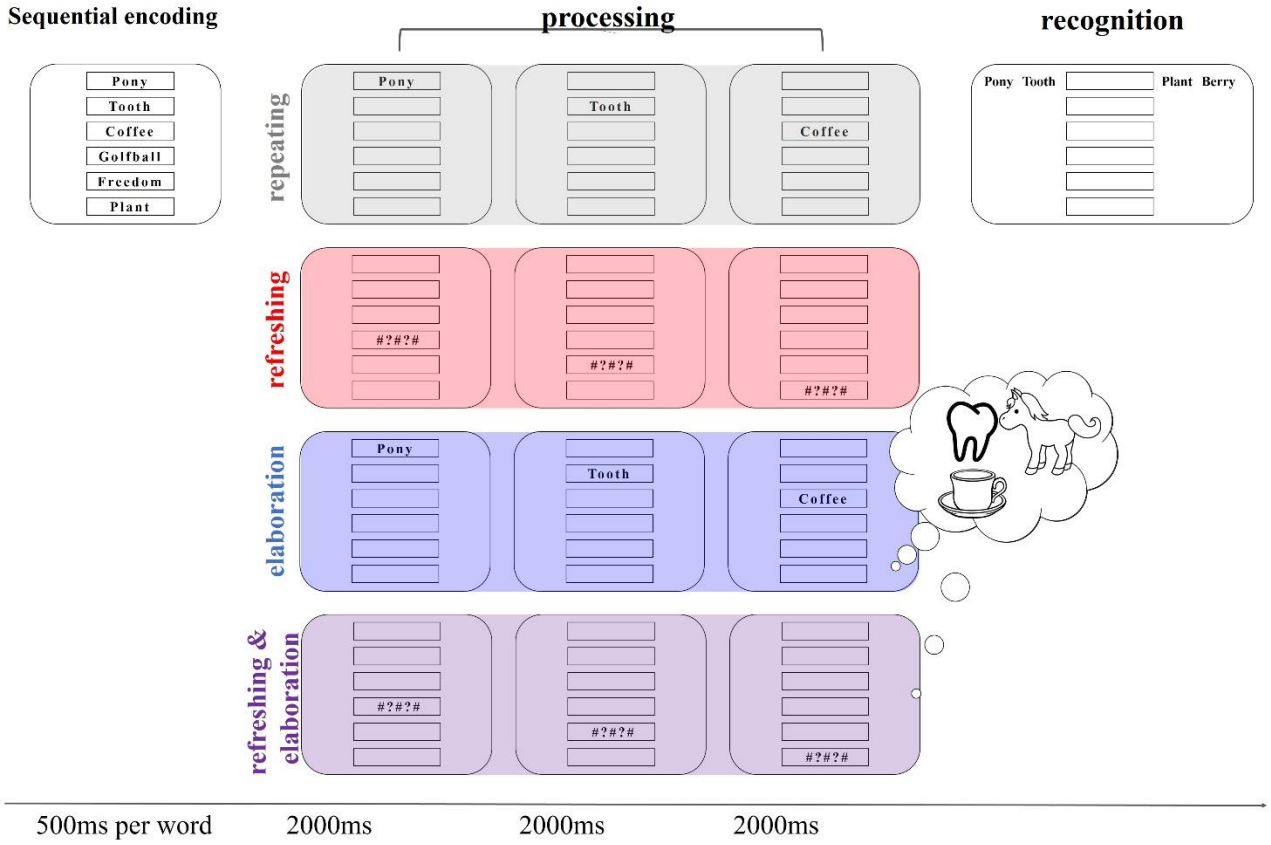

Figure 1 Illustration of the immediate memory paradigm. Subjects were shown a list of six words sequentially, followed by either the first or second triplet being processed according to the four experimental conditions. The trial ended with a recognition test in which each list item was tested in their order of presentation using a 4-alternative forced-choice procedure. The grey panel shows the repeating condition with an example of the first triplet being processed. The red panel shows the refreshing condition, here for the second triplet to be processed. The blue and purple panels show the elaboration and refreshing with elaboration condition, respectively, which were preceded by the instruction to form a mental image, but which were equivalent in the visual input to the respective conditions without the elaboration instruction.

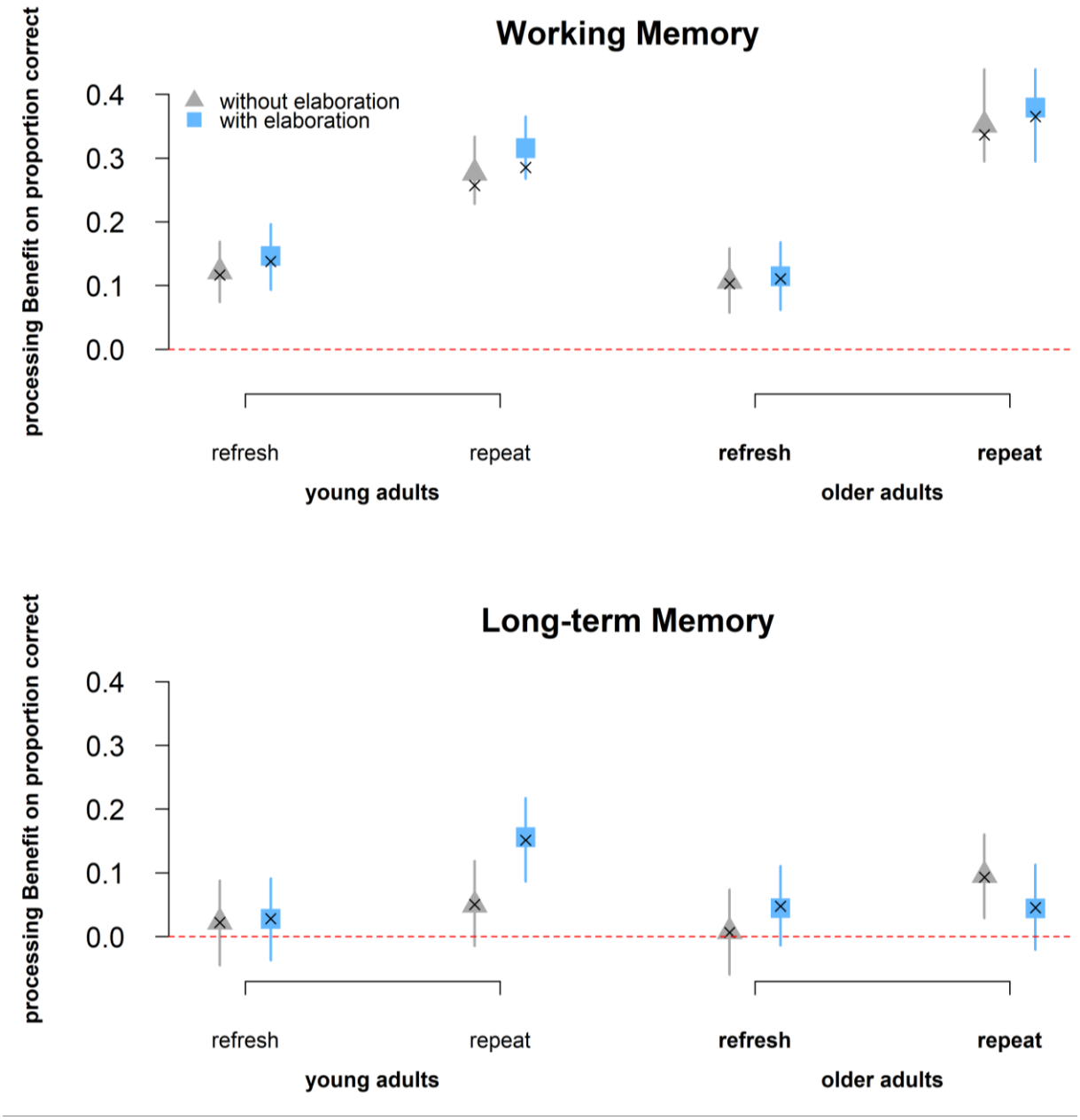

Figure 2 processing benefit in the WM (upper graph) and LTM (lower graph) task. The blue symbols and error bars represent estimated processing benefits and their 95% HDRs from the BGLMM for the conditions with elaboration, the grey symbols represent the same for the ones without elaboration. The crosses represent the observed means. Their overlap indicates that the model adequately describes the data. The red line represents the point of no difference in performance between the processed and the unprocessed triplet.

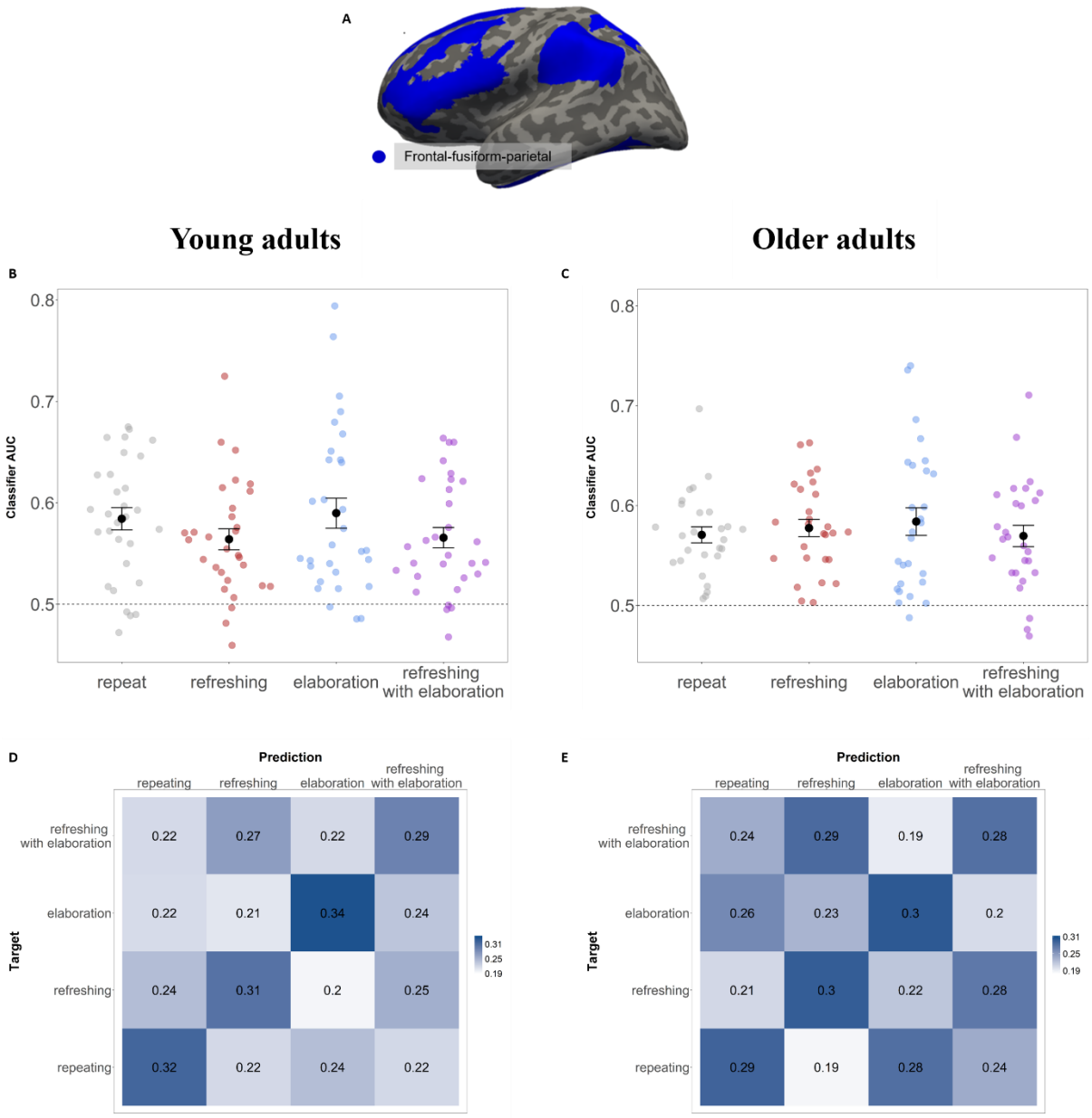

Figure 3 **A** Regions of interest for analysis include voxels in frontal cortex, fusiform gyrus, and parietal cortex (see Methods). **B and C** Classifier for Repeat vs. Refreshing vs. Elaboration vs. Refreshing-with-Elaboration in young and older adults, respectively. Decoding as indicated by the classifier AUC for the subjects. The error bars indicate the standard error. **D and E** Confusion matrix of the classifier accuracy for classifiers trained on Repeat vs. Refreshing vs. Elaboration vs. Refreshing-with-Elaboration in young and older adults. Respectively. Chance performance is at 0.25.

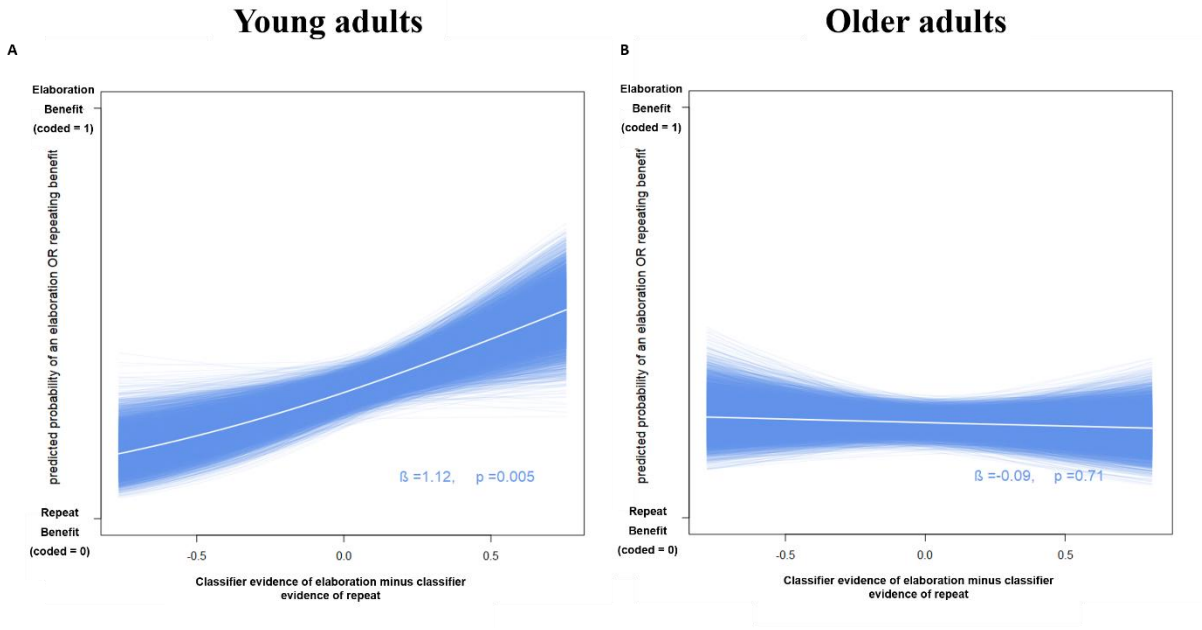

Figure 4. Neural evidence for elaboration predicts elaboration benefit in younger adults. On each of 10,000 bootstrap iterations, a logistic regression evaluated the classifier evidence for elaboration (relative to repeat) to the LTM elaboration benefit (relative to repeat) across each trial. The result from each bootstrap is visualized with a single blue line, and the mean is depicted in white.