Suppressing the Maintenance of Information in Working Memory Alters Long-term Memory Traces

Zachary H. Bretton[1], Hyojeong Kim[1,2], Marie T. Banich[2], and Jarrod A. Lewis-Peacock[1]

Abstract

■ The sensory recruitment hypothesis conceptualizes information in working memory as being activated representations of information in long-term memory. Accordingly, changes made to an item in working memory would be expected to influence its subsequent retention. Here, we tested the hypothesis that suppressing information from working memory, which can reduce short-term access to that information, may also alter its long-term neural representation. We obtained fMRI data (n = 25; 13 female / 12 male participants) while participants completed a working memory removal task with scene images as stimuli, followed by a final surprise recognition test of the examined items. We applied a multivariate pattern analysis to the data to quantify the engagement of suppression on each trial, to track the contents of working memory during suppression, and to assess representational changes afterward. Our analysis confirms previous reports that suppression of information in working memory involves focused attention to target and remove unwanted information. Furthermore, our findings provide new evidence that even a single dose of suppression of an item in working memory can (if engaged with sufficient strength) produce lasting changes in its neural representation, particularly weakening the unique, item-specific features, which leads to forgetting. Our study sheds light on the underlying mechanisms that contribute to the suppression of unwanted thoughts and highlights the dynamic interplay between working memory and long-term memory. ■

INTRODUCTION

Working memory is crucial for managing goal-relevant information but has a finite capacity and benefits from the removal of unwanted information (Lewis-Peacock, Kessler, & Oberauer, 2018; Luck & Vogel, 2013; Cowan, 2001). Information can be removed from working memory in different ways, for example, by replacing it with another thought, suppressing that particular thought, or clearing the mind of all thoughts (Kim, Smolker, Smith, Banich, & Lewis-Peacock, 2020; Banich, Mackiewicz Seghete, Depue, & Burgess, 2015). Each method has a distinct neural signature and behavioral impact. Suppression is most successful at reducing access to information in the short term, but its long-term effects remain unknown. This study evaluates the consequences for long-term memory when items are actively removed from working memory via suppression or replacement. It is vital to highlight the distinction between the transient nature of suppression effects in working memory and the potentially enduring impacts on long-term memory. Although suppression may not always manifest immediate behavioral effects, such as forgetting, its neural correlates in working memory and long-term memory can be highly informative. Studies suggest that fMRI data can capture subtle neural changes that may precede and possibly predict behavioral changes, including those extending into long-term memory (Paller & Wagner, 2002). This fact is particularly relevant in the context of suppression-induced forgetting of information in long-term memory, a phenomenon where existing literature shows that multiple repetitions of suppression cues are often required to observe a behavioral forgetting effect (Benoit & Anderson, 2012; Depue, Curran, & Banich, 2007; Anderson & Green, 2001). However, our inquiry diverges from these prior studies as it probes the active mechanisms of suppression within working memory (a process we term "maintenance suppression") and its potential ripple effects on long-term memory, which might operate under different dynamics than retrieval suppression from long-term memory. Therefore, our study leverages fMRI data to explore how maintenance suppression might alter neural representations both in working memory and long-term memory, even if such changes are not observable at the behavioral level. We employed multi-voxel pattern analysis (MVPA) of fMRI data (Haxby, Connolly, & Guntupalli, 2014; Lewis-Peacock & Norman, 2014b) to assess changes in the neural representation of items that are either maintained, replaced, or suppressed from working memory. We hypothesized that if an item's neural representation is altered during removal from working memory, its subsequent accessibility in long-term memory may also be reduced.

Our hypothesis is grounded in the sensory recruitment theory of working memory (Serences, 2016; D'Esposito &

[1]University of Texas at Austin, [2]University of Colorado

Journal of Cognitive Neuroscience 36:10, pp. 2117–2136
https://doi.org/10.1162/jocn_a_02206

Postle, 2015; Sreenivasan, Curtis, & D'Esposito, 2014; Serences, Ester, Vogel, & Awh, 2009), positing that working memory employs the same cortical regions used for sensory perception and long-term memory storage. Supporting this idea, neuroimaging research has shown the involvement of sensory regions, such as the visual cortex, in working memory tasks with visual stimuli (Albers, Kok, Toni, Dijkerman, & de Lange, 2013; Harrison & Tong, 2009; Serences et al., 2009). Furthermore, both working memory and long-term memory engage a neural activity in the dorsolateral prefrontal cortex (dLPFC) and medial temporal lobe, linking working memory maintenance and long-term memory formation (Melrose et al., 2020; Axmacher, Schmitz, Weinreich, Elger, & Fell, 2008; Blumenfeld & Ranganath, 2006; Ranganath, Cohen, & Brozinsky, 2005). In line with this overlap between neural mechanisms for working and long-term memory, research using the directed forgetting paradigm, in which an individual mentally manipulates in working memory prior associations learned between items in long-term memory, suggests that intentional forgetting can lead to reduced recollection in long-term memory (Levy & Anderson, 2008; Anderson & Green, 2001). These studies indicate that cognitive operations like suppression may not just be a short-term strategy for managing information in working memory but could have lasting impacts on how information is stored and retrieved in long-term memory.

One vantage point from which to consider how manipulating information in working memory might affect its long-term memory representation is to note that memory representations can be characterized by item-level and category-level features. Item-level features are specific details like sensory attributes, whereas category-level features are broader semantic aspects providing context and meaning, with dissociable neural representations (Martin, 2007). Prior evidence suggests these feature levels may be independently modified by suppression. Results from our prior study suggest that suppression in working memory may target item-level representations (Kim, Smolker, et al., 2020). Specifically, after an item is maintained in working memory, the neural representation of a subsequently encoded item from that same category suffers from proactive interference. The representation of the initial item lingers in memory and interferes with the encoding of a similar item. However, if that initial item is instead suppressed, proactive interference is eliminated and subsequent encoding of same-category items is even facilitated relative to different-category items. The inference from this result is that suppression of a working memory item may target its item-specific features for weakening while preserving category-level features of the item in memory, which facilitates the encoding of a new category exemplar. Consistent with this idea of targeting item-specific features for weakening is an item-method, directed-forgetting study that found increased activation of item-unique features for items following a forget cue versus a remember cue (Wang, Placek, & Lewis-Peacock, 2019). As is common in directed-forgetting research, no explicit strategy was given to participants for how to forget. Nonetheless, the to-be-forgotten items were more poorly remembered, suggesting that emphasizing item-specific features during the attempt to forget may have facilitated the targeted weakening of these features in memory, leading to failures in recognition of the items on a subsequent memory test.

Related evidence from studies examining the suppression of retrieval from long-term memory suggests a dissociation in which these two aspects of a memory representation are differentially affected depending on the stage of processing at which suppression occurs. Initial suppression of information from long-term memory weakens memory traces by inhibiting their sensory features (Depue et al., 2007), which is a key component of item-level representations (Anderson & Hanslmayr, 2014). This effect is further evidenced by research showing that retrieval suppression reduces perceptual priming for suppressed items (Taubenfeld, Anderson, & Levy, 2019; Gagnepain, Henson, & Anderson, 2014). This reduction in perceptual priming indicates a loss of item-specific information, suggesting that suppression of retrieval of information from long-term memory can selectively impair item-level features. It is unclear whether the maintenance suppression of information in working memory would similarly produce long-lasting impairment of the item-level features of that information. The present study sought to test this hypothesis.

We designed a neuroimaging study to evaluate whether suppressing an item in working memory—as compared with keeping that item in mind or replacing it with another item (see Figure 1)—would alter its item-specific representation in long-term memory and produce forgetting on a subsequent recognition memory test. Participants (n = 25; 13 female / 12 male participants) were initially exposed once to a set of scene pictures in the fMRI scanner. Then, they performed a working memory task in which these pictures reappeared once and participants were instructed to either maintain, suppress, or replace each item. At the end of the experiment, they performed a recognition memory test for those pictures while still in the scanner. We used MVPA to differentiate and track patterns of brain activity corresponding to (1) the contents of working memory (category-level decoding), (2) the cognitive control operation being engaged in working memory (operation-level decoding), and (3) the long-term memory representation of items both before and after the working memory task (item-level decoding). We tracked the strength of memory representation on each trial and the classifier evidence for the engagement of the cognitive operation. We then analyzed how the variable engagement of the operation on a given trial predicted the strength of the memory representation in working memory and the subsequent memory outcomes in long-term memory. Finally, we evaluated potential changes in long-term memory representations in scene-related occipito-temporal cortex of the items from the working memory task by

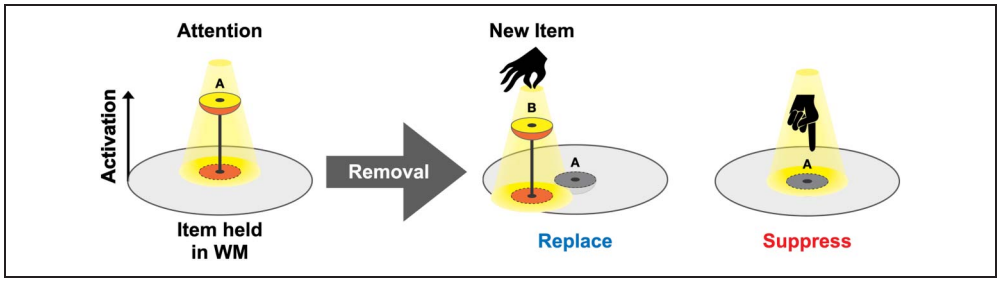

Figure 1. Visualization of removal operations. Depicts the hypothesized attentional states in working memory during the removal operations: replace and suppress used in the study. When the item "A" is goal relevant, it is active and bound to its context in working memory. During removal, the removal target "A" is diminished either by redirecting focal attention to a new item "B" or by suppressing the target in the focal attention. (Adopted and adjusted from Lewis-Peacock et al., 2018).

comparing their before (from the initial exposure) and after (from the memory test) item-specific neural patterns, separately for items that were later remembered and for items that were later forgotten.

METHODS

Participant Information

Twenty-five healthy participants (13 female, 12 male participants; age $M = 20.413$ years, $SD = 2.44$ years; all right-handed) were recruited from the Austin, Texas, area for this fMRI study. Three participants were excluded because of poor fMRI classification performance or confounds such as excessive motion. All participants had normal or corrected-to-normal vision, provided informed consent, and received $60 in compensation.

Stimuli

Our fMRI experiment employed colored images (400 × 400 pixels) from two main categories, each with two subcategories: scenes (manmade and natural) and faces (male and female). Scenes served as the primary stimuli and are the focus of the current study, whereas faces were utilized in a subset of trials, the replace trials, specifically serving as the items to be brought into mind to replace the scene images.

The selection of scenes as the primary stimulus type was methodologically motivated. Scenes offer high decodability and separability, particularly important for our primary measures involving category-level evidence in working memory and representational similarity analyses (RSAs) of item-specific features, making them a reliable choice for a memory experiment (Epstein & Kanwisher, 1998). The scenes used were recognizable locales, such as tropical beaches, or well-known landmarks like the Golden Gate Bridge.

Faces were chosen as replacement stimuli on replace trials because of their high decodability and their ease of recognition and training within a constrained session. The faces used were recognizable celebrities, enabling us to select faces that participants could readily recall, thus minimizing variability across trials (Bruce & Young, 1986). The ability of participants to recognize and differentiate faces ensures that the experimental design remains straightforward and minimizes potential confounds (like being unable to visualize the replacement face).

All images were sourced from various platforms, including the Bank of Standardized Stimuli and Google Images. Two hundred forty images were used in the study: 90 scenes per scene subcategory (manmade, natural) and 30 faces per subcategory (male, female). Half of the face images were used during the initial exposure phase (explained below) and the other half as targets during the replace trials.

In the initial exposure phase, participants were introduced to 60 stimuli, comprising 30 scenes and 30 faces. The faces presented during this phase were specifically chosen to be used later in the replace trials of the working memory manipulation task, where participants were prompted to replace the encoded scene with one of these familiar faces. The premanipulation recognition test assessed memory for the "old" items from the initial exposure, presenting 30 old scenes alongside 90 "new" scenes that were designated for use in the subsequent working memory manipulation task. The faces were balanced, with 30 old faces (from initial exposure) matched against 30 new faces. During the working memory manipulation task, the 90 new scenes from the premanipulation recognition were divided equally among the three operations: maintain, replace, and suppress, with 30 scenes allocated to each operation. In the replace trials, participants were instructed to bring to mind a specific face from the initial exposure phase, effectively replacing the current scene in their working memory with this familiar face. Finally, the postmanipulation recognition test aimed to assess long-term memory for the scenes across all phases, involving 180 scenes in total. This included 30 old scenes from the initial exposure, 90 old scenes (30 from each operation), and an additional 60 new scenes, providing a comprehensive evaluation of memory retention and the impact of each working memory operation.

To control for stimulus-specific effects like familiarity, the study images were fully randomized across trials,

conditions, and participants. A consistent subset of 15 male faces and 15 female faces was used as replacement items across participants. All stimuli were presented on a gray background, with task-related words and fixation crosses displayed in white font.

Experimental Design and fMRI Procedure

The experiment consisted of five phases in the MRI scanner, completed in order: a 12-min resting-state scan (not analyzed here), an initial exposure phase (completed during an magnetization prepared rapid gradient echo anatomical scan), a premanipulation recognition test, the main manipulation phase, and a postmanipulation recognition test (Figure 2). In the initial exposure phase, participants saw the 30 faces that will be used for replacement and 30 scenes that will not be manipulated in the main study phase. The recognition memory tests were identical except that they were conducted on items from the initial exposure phase (for the premanipulation test) or on items from the main manipulation phase (for the postmanipulation test). This procedure allowed us to evaluate any

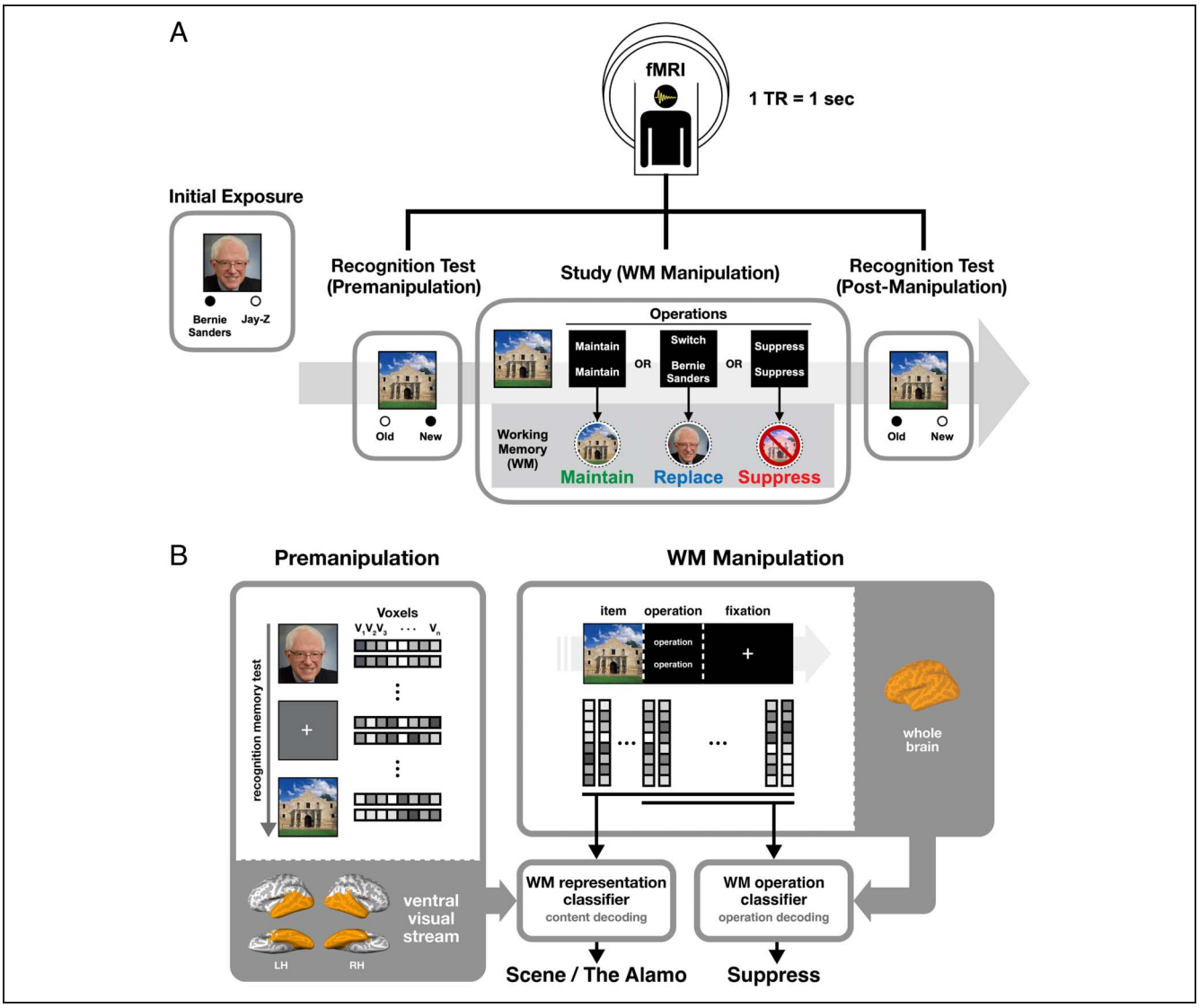

Figure 2. fMRI study design and classifier procedures. (A) Illustrates the experimental procedure and design. During the initial exposure phase (during which the anatomical scan was collected), participants saw a stream of images (scenes/faces) presented one by one with two name options under each and were asked to select the correct name. In the first recognition test (premanipulation) phase, participants performed a surprise recognition memory test for previously seen items, along with novel items. During the manipulation phase, the task involved viewing a scene, then a screen indicating the operation (maintain, replace, suppress) to apply to the scene, followed by a fixation cross. Finally, the second recognition test (poststudy) contained another surprise recognition memory test for all items previously seen, including those that were initially exposed in the premanipulation phase and those that were introduced during the manipulation phase, along with novel items. (B) Brain data from the premanipulation task (left) were used for MVPA for category-level decoding, trained on faces, scenes, and rest, within the VVS ROI. This classifier was then applied to decode information in the focus of attention in working memory during the manipulation phase. Whole brain data from the manipulation phase was used for classification of the cognitive operation being performed on each trial (adopted and adjusted from Kim, Smolker, et al., 2020).

neural and behavioral changes to the items. We also evaluated for possible neural changes to the items during the main manipulation phase. Neural data from the visual cortex in the premanipulation test phase were used as a baseline for category-level classification using MVPAs and item-specific RSA. Lastly, the cognitive operations (maintain vs. replace vs. suppress) were also decoded during the main manipulation phase as in Kim, Smolker, and colleagues (2020).

Participants were asked to perform different tasks during each of the phases of the study. During the initial exposure phase, a stream of images was presented one by one with two potential labels (e.g., "Bernie Sanders" and "Jay-Z") appearing under each, and participants were asked to select the correct label. During the recognition test phases, participants determined whether a presented item had ever been seen previously ("old") or not ("new") and then specified their confidence level ("sure" or "unsure"). During the main manipulation phase, the participant's task on each trial was to view a scene item for 2 sec and then, after it disappeared, they were instructed to manipulate it in working memory for 4 sec, with no response required at the end of the trial. The manipulation was either to: maintain the scene in mind, replace it with a particular face (denoted by its label from the initial exposure phase), or suppress it. The manipulation instruction consisted of two words, one in the top and one in the bottom halves of the screen, presented over a gray background. For the maintain and suppress conditions, the top and bottom words both indicated the name of the instruction (e.g., "maintain / maintain" or "suppress / suppress"). In the replace condition, the top word was "switch" and the bottom word indicated the face's label (e.g., "Bernie Sanders"). Before the experiment began, participants were briefed on the procedures for each type of trial. For "maintain" trials, they were instructed to keep the given item in their mind throughout the trial. In "replace" trials, participants were told to bring the specific cued face to mind, replacing the scene they had just encoded. During "suppress" trials, participants were guided to "suppress as you would suppress a cough," with explicit instructions not to empty their minds entirely or to substitute the item in mind with a thought from long-term memory. Instead, they were to focus on actively pushing the item out of their working memory. Each trial was then followed by a jittered intertrial fixation lasting either 5, 6, or 7 sec, consisting of a white fixation cross centered over a gray background.

Long-term Memory Tests

The two recognition memory tests were designed to match each other, with the only differences being the items that were tested and the total trial count. Participants saw an image on the screen for 2 sec and were told to respond if the item was "old" (meaning they had seen it at some point earlier in the experiment) or "new" (meaning this was a new image to them), along with the confidence of their response ("sure" or "unsure"). Only "sure old" responses were considered hits for an old item and "sure new" for novel items, consistent with prior work (Kim, Schlichting, Preston, & Lewis-Peacock, 2020; Kim, Lewis-Peacock, Norman, & Turk-Browne, 2014; Lewis-Peacock & Norman, 2014a). In the premanipulation memory test, participants began with a test of 60 face images, split across two runs (4 min 6 sec each). Half of the faces were new, and the other half were old faces that subsequently served as replacement items in the main study phase. Then there were four runs of memory tests for scene images (4 min 6 sec each), with 30 old scenes that had been shown in the initial exposure phase and 90 new scenes that subsequently appeared in the main manipulation phase. This yielded a 1:3 ratio of old:new scenes items. The postmanipulation memory test contained a test of 180 scene images, of which 90 were "old and manipulated" (having been seen twice before, first as a "new" image in the prestudy memory test and then again when it was manipulated in the main study phase), 30 were "old but unmanipulated" (having been seen twice before, first in the initial exposure phase and then again in the prestudy memory test, but not in the main study phase), along with 60 novel scenes. There were six runs of memory tests for this post-study phase (4 min 6 sec each). This yielded a 2:1 ratio of old:new scene items.

Data Acquisition

MRI data were acquired on a Siemens Skyra 3.0 Tesla scanner at the Biomedical Imaging Center on the campus of The University of Texas at Austin. fMRI scans were acquired using a sequence with the following parameters: TR (repetition time) = 1000 msec, echo time = 30 msec, field of view = 230 mm—100% phase, multiband acceleration factor = 4, with a $2.4 \times 2.4 \times 2.4$ mm^3 voxel size, acquired across 56 axial slices and aligned along the anterior commissure-posterior commissure line. There were 17 runs: two resting-state runs and 15 task runs. Six premanipulation and postmanipulation recognition memory test runs were acquired, each consisting of 246 EPIs, for 1476 images. The main manipulation phase consisted of three runs with 366 EPIs each, for 1098 images. Total data acquisition time was 91 min 30 sec for each participant, along with an additional 15 min of setup and breakdown time.

fMRI Preprocessing

fMRI data set was formatted in Brain Imaging Data Structure (BIDS) and preprocessed via fMRIPrep 21.0.1 (Esteban et al., 2019), which is based on Nipype 1.6.1 (Esteban et al., 2022; Gorgolewski et al., 2011), FMRIB Software Library (FSL; Version 6.0.5.1:57b01774), and FreeSurfer (Version 6.0.1). For the distortion correction of the magnetic field, each participant's B0 inhomogeneity field map was estimated

from the phase-drift map(s) measure with two consecutive gradient-recalled echo acquisitions. The corresponding phase-map(s) were phase-unwrapped with a prelude (FSL).

A high-resolution, T1-weighted (T1w) image was corrected for intensity non-uniformity with N4BiasFieldCorrection (Tustison et al., 2010) and skull-stripped (target template: OASIS30ANTs) with Nipype in the antsBrainExtraction.sh workflow, both of which were distributed by ANTs (Version 2.3.3; Avants, Epstein, Grossman, & Gee, 2008). The brain-extracted T1w was used as a T1w reference for data alignment and also to define brain tissue segmentation of cerebrospinal fluid (CSF), white matter, and gray matter (GM) with FAST in FSL. Brain surfaces were reconstructed from the T1w with recon-all in FreeSurfer, and the brain mask estimated in the previous step was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical GM of Mindboggle (Klein et al., 2017). To normalize the data to Montreal Neurological Institute (MNI) standard space, volume-based spatial transformation parameters were estimated through nonlinear registration with antsRegistration in ANTs (International Consortium for Brain Mapping 152 Nonlinear Asymmetrical template Version 2009c [RRID: SCR_008796]; Fonov, Evans, McKinstry, Almli, & Collins, 2009; TemplateFlow ID: MNI152NLin2009cAsym), using brain-extracted versions of both T1w reference and the T1w template.

Functional EPI images from the 17 runs per participant (across all task phases) were estimated and resampled with the preprocessing steps, including motion and slice-time corrections, distortion correction when available, co-registrations to anatomical T1w space, and normalization to MNI standard space. Head motion was estimated with MCFLIRT (rigid transformation with six motion parameters) based on the EPI reference volume and its skull-stripped version, generated by aligning and averaging one single-band reference. The slice-time correction was estimated centered by 0.455 sec (0.5 of slice acquisition range = 0–0.91 sec) for each volume using 3dTshift from AFNI (RRID: SCR_005927). The field coefficients from the estimated field map were then aligned with rigid registration to the EPI reference run. The EPI reference was then co-registered with the T1w reference by boundary-based registration (six degrees of freedom) using bbregister in FreeSurfer (Greve & Fischl, 2009). All EPI data were resampled into the final MNI volumetric space with a single interpolation step by composing all the pertinent transformations obtained during the estimations. Gridded resamplings were performed using antsApplyTransforms in ANTs, configured with Lanczos interpolation to minimize the smoothing effects of other kernels (Lanczos, 1964).

Several confounding factors were calculated based on the preprocessed BOLD (EPI) images: framewise displacement (FD), dynamic voxel-wise amplitude response signal (DVARS), and three region-wise global signals. FD and DVARS were calculated for each functional run (Power, 2014) using Nipype. FD was computed using two formulations, the absolute sum of relative motions (Power, 2014) and the relative root mean square displacement between affines (Jenkinson, Bannister, Brady, & Smith, 2002). The three global signals were extracted within the CSF, white matter, and whole-brain masks. In addition, physiological regressors were extracted for component-based noise correction (CompCor; Behzadi, Restom, Liau, & Liu, 2007). Principal components were estimated from the preprocessed BOLD time series after high-pass filtering (a discrete cosine filter with a 128-sec cutoff) for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). The tCompCor components are then calculated from the top 2% variable voxels within the brain mask. For the aCompCor, three probabilistic masks (CSF, white matter, and combined CSF + white matter) are generated in anatomical space. The mask of pixels, which was obtained by dilating a GM mask extracted from the FreeSurfer's age segmentation, was subtracted from the aCompCor masks to remove any voxels containing a minimal fraction of GM (Behzadi et al., 2007). Finally, these aCompCor masks were resampled into EPI space and binarized by thresholding at 0.99. The components were also calculated separately within the white matter and CSF masks. For each CompCor decomposition, the k components with the largest singular values were retained, such that the retained components' time series was sufficient to explain 50% of variance across the nuisance masks (CSF, white matter, combined, or temporal), and the remaining components were dropped from consideration. The confound time series derived from head-motion estimates and global signals were also included with the temporal derivatives and quadratic terms for each (Satterthwaite et al., 2013). For our univariate and multivariate analyses, we used fMRIPrep-preprocessed images. During analysis, confound regressors generated by fMRIPrep were employed to clean the data, effectively addressing physiological and motion-related noise while preserving task-related signals.

ROIs

To characterize the visual representations of the stimuli used in the task (faces and scenes), we focused our analyses on the ventral visual stream (VVS) in the occipitotemporal lobes. This mask consists of anatomically defined regions from the Desikan-Killiany Atlas (packaged with FSL): intracalcarine cortex, lingual gyrus, lateral occipital cortex, occipital fusiform gyrus, occipital pole, parahippocampal gyrus, temporal fusiform cortex, temporal, occipital fusiform cortex, inferior temporal gyrus, middle temporal gyrus, superior temporal gyrus, and temporal pole. Individual masks of these regions were extracted from each participant's parcellated cortical MNI map (from fMRIPrep) and summed together to construct masks for each participant across hemispheres. The ROI masks were then binarized so that voxels within

the mask had a value of 1 and voxels outside the mask had a value of 0 (VVS: M = 15,349 voxels, SD = 461). For the RSA analyses, we included three additional ROIs: one covering various areas in the pFC (M = 12,507 voxels, SD = 406; lateral orbitofrontal, medial orbitofrontal, rostral middle frontal, caudal middle frontal, superior frontal, and frontal pole), one covering the parietal cortex (M = 7742 voxels, SD = 402; inferior parietal, superior parietal, and precuneus), and one covering the hippocampus (M = 854 voxels, SD = 25). These regions were selected based on existing literature suggesting their involvement in representing category-level information and higher-order processing (Zhou, Mohan, & Freedman, 2022; Rademaker, Chunharas, & Serences, 2019; Davachi, 2006).

Statistical Analyses

Univariate Analyses

fMRI univariate analyses were carried out using the Python package Nilearn (Version 0.10.0; general linear model [GLM] module). The preprocessed and smoothed (8-mm FWHM) BOLD data normalized in MNI space were filtered with a group-level GM mask derived from a combination of each participant's GM probability segmentation via fMRIPrep. The GLM modeled three operations across all trials within 4 sec during which the operation of the item is occurring and rest periods between trials were serving as a baseline. The TRs during the presentation time of the stimulus served as an explanatory variable of no interest. Following the methods from our previous studies (Kim, Smolker, et al., 2020; Banich et al., 2015), we explored three main contrasts from the GLM beta estimates to confirm if we could replicate previously established univariate maps of each operation (suppress vs. replace + maintain, replace vs. suppress + maintain, and suppress + replace vs. maintain). For this analysis, contrast maps were initially thresholded voxel-wise by implementing a false discovery rate correction at $p <$ .05. Subsequently, a cluster extent threshold of 20 voxels was applied to retain only those clusters of significant voxels that met or exceeded this size. The voxel-wise threshold was set to the most stringent value obtained from the individual conditions, which was 2.404.

Multivariate Analyses

MVPA was applied to classify task-related signals in working memory with a nonmultinomial (one-versus-others) logistic regression model with L2 regularization. Using the scikit-learn toolbox in Python (Kumar et al., 2020), our models operated on a one-versus-all (also known as one-versus-others) basis within a multiclass framework. This approach allows for the differentiation of each category from the others in the model. A single penalty value was selected for each participant based on the maximum performance on cross-validation analyses. The features from the preprocessed and standardized data (MNI or T1w) were selected by voxel-wise ANOVA on the training data with a threshold of $p < .01$ and then applied to both training and testing in the classifications. The working memory contents and operations were classified within each participant, and additional classifications were conducted between-participant and between-experiment models to test the generalizability of the working memory operation activation patterns.

Working memory content classification. A three-category classifier (face, scene, and rest) was built on the prestudy memory test fMRI data from the VVS (Grill-Spector & Weiner, 2014; Goodale & Milner, 1992), using a one-versus-all strategy, to decode the representations of visual items in working memory (Gayet, Paffen, & Van der Stigchel, 2018). The intertrial interval time points in which a fixation was presented were modeled as the "rest" category. The classifier's performance was initially validated using twofold cross-validation, where the premanipulation memory data were split into training and testing sets for each iteration, ensuring the model's accuracy before generalizing it to the main manipulation phase.

To balance the number of samples in each class, half of the scene trials were randomly selected from the training data, as there were twice as many scene trials as face trials. As a result, an equal number of trials (60 trials, with 2 TR per trial = 120 samples per category) were included in the cross-validation, and the remaining scenes were dropped. This was done to ensure balance and fairness in the model training process. Feature selection was performed for each training set using a voxel-wise ANOVA across classes (threshold: $p = .05$) with the regressors shifted forward 5 TR (5 sec) to account for hemodynamic lag. To find the optimal L2 penalty value to best fit the classifier model for each participant, cross-validation was done with different penalties (range: 0.001–1000). Classification accuracy was reliably above chance, assessed via the area under the receiver operating characteristic (ROC) curve (AUC). ROC/AUC scores of the classifier evidence were significantly above baseline (0.5) at the category level (M = 0.897, SEM = 0.009; one-sample t test against 0.5: $t(21)$ = 46.156, p = 1.34e−22, d = 10.07).

All data from the premanipulation memory test (but still subsampled to maintain balance between the classes) were then used to retrain the classifiers and then applied to the manipulation phase data. The training again included ANOVA-based feature selection for each voxel, and an individualized optimal penalty derived from the cross-validation analysis was used for each participant. These classifiers were used to decode every time point in the manipulation phase data. In addition, to further investigate the role of category-specific regions, we applied the same decoding approach to parahippocampal (M = 616 voxels, SD = 57) and fusiform (M = 3289 voxels, SD = 462) ROIs, known for their preferential responses to

scenes and faces, respectively. A 14-sec time window beginning at trial onset was used to evaluate the trajectory of classification evidence (i.e., class probabilities). Previous research has shown that using neural activity patterns to decode the information in working memory reveals only the information held in the focus of attention (LaRocque, Lewis-Peacock, & Postle, 2014; Lewis-Peacock, Drysdale, Oberauer, & Postle, 2012). Therefore, by tracking the category evidence for an item being manipulated on a given trial, we can assess the operation's impact on the item in the focus of attention.

Within-participant working memory operation classification. The working memory operation classifiers were built with the manipulation phase data from the GM mask across the whole brain. The classifier was trained and validated through a three-fold leave-one-run-out cross-validation approach. In this method, the classifier was trained on two of the three runs and tested on the remaining run. This process was repeated until each run had served as the test set. The classifier regressor included the three operations (maintain, replace, and suppress) in which the operation period (4 sec [4 TRs]) and the subsequent fixation period (jittered from 5 to 7 sec) were modeled on each trial with being shifted forward 5 TR to account for hemodynamic lag. We included the fixation period in the regressor because previous work has shown that relevant information can be captured during this window (Kim, Smolker, et al., 2020). The operation classifier was sensitive across all operations, AUC averaged across operations: $M = 0.807$, $SEM = 0.022$; one-sample t test for maintain: $t(21) = 11.773$, $p = 1.034e{-}10$, $d = 2.569$; replace: $t(21) = 15.368$, $p = 6.740e{-}13$, $d = 3.353$; suppress: $t(21) = 12.379$, $p = 4.097e{-}11$, $d = 2.701$, indicating that the operations were reliably differentiated from each other.

Between-participant working memory operation classification. All data from all participants ($n = 22$) were normalized to MNI standard brain space and concatenated so that all voxels were aligned across participants. The classification was done with a 22-fold leave-one-participant-out cross-validation with the data in the whole-brain GM mask segmented from the standardized MNI brain. The ANOVA-based feature selection was applied to the anatomically aligned data in which the first half of all runs for each participant were concatenated across participants. The other half of the data was used for training and testing the classifier. The operation regressors were shifted forward 5 sec to account for hemodynamic lag within each participant and concatenated. The between-participant classifier was able to decode and differentiate each operation successfully, AUC averaged across operations: $M = 0.658$, $SEM = 0.0103$; one-sample t test for maintain: $t(21) = 7.768$, $p = 1.316e{-}07$, $d = 1.695$; replace: $t(21) = 10.344$, $p = 1.068e{-}09$, $d = 2.257$; suppress: $t(21) = 8.458$, $p = 3.333e{-}08$, $d = 1.846$.

Between-experiment working memory operation classification. We also performed an across-experiment classification analysis for these working memory operations. The goal was to confirm that the patterns of neural activation engaged for these working memory operations were consistent across different experiments and different sites of data collection (University of Colorado Boulder and University of Texas at Austin). We trained across-participant classifiers on fMRI data collected previously in Boulder ($n = 50$; Kim, Smolker, et al., 2020) and tested on data from the present study collected in Austin. Training data was anatomically aligned to MNI space, z scored, feature-selected for the top 10,000 voxels (group-level feature mask), and then reduced to 70 components via PCA. Testing data were anatomically aligned to MNI space, resampled to match the training voxel space (2.4 mm^3 to 2 mm^3, via AFNI's 3dresample function), masked with the group-feature mask and transformed to 70 components using the same transformations from the fitted PCA. One-versus-others L2 logistic regression classification with a penalty of 50 was used, and results indicated that each operation was successfully decodable in each participant.

RSA

To evaluate possible changes to the neural representations in the long-term memory of the items that were manipulated in the main study task, we applied RSA to data from the premanipulation phase and the postmanipulation phase. We used a custom pipeline in Python from our laboratory (Kim, Smolker, et al., 2020) to compute a comparative metric of "fidelity" that quantifies the degree to which neural representations remain consistent, from the premanipulation memory test through to the postmanipulation memory test. By applying feature weighting to the data before RSA (Kaniuth & Hebart, 2022), this fidelity measure can be evaluated at both the item level and category level, offering a nuanced understanding of representational changes over time.

We defined a template pattern of activity for each item (90 items total) from the premanipulation memory test data and used this to evaluate item-specific representations in the manipulation phase and also in the postmanipulation memory test. To evaluate representational changes at different levels of granularity, we used a feature-weighting procedure to create both category-level and item-level templates for each stimulus (Figure 6). We first employed a GLM to select category-specific voxels within the VVS ROI using a scene versus face t-contrast, with a voxel-wise threshold of $p < .05$ and cluster correction with a voxel extent of 10. These identified voxels were then weighted to create these templates at the category and item levels. For category-level templates, voxel patterns were weighted using the beta contrasts from the scene > faces comparison. In the item-level templates, each item was assigned a unique regressor and weighted by contrasting them against the 89 other items in the

study, leveraging least squares - all for this modeling (Abdulrahman & Henson, 2016). These weighted templates were then used for our RSA analyses. To quantify representational fidelity, these templates were compared against neural patterns captured during the postmanipulation memory test. The analysis yielded a similarity score, calculated as normalized correlation coefficients (Pearson r) using Fisher's z transformation for group-level statistical analysis.

To ensure the reliability and sensitivity of these template weights, we evaluated if the weighted templates allowed us to discriminate between the item of interest from other stimuli during the manipulation phase. RSA similarity scores for a given item were higher than the similarity to all other items from the manipulation phase, $t(21) = 6.955, p = 7.19e{-}07$.

To further explore the relationship between representational similarity and memory outcomes, items were sorted based on their memory status (remembered or forgotten), allowing us to scrutinize how representational similarity might differ across memory outcomes. To assess this relationship, we performed a paired t test to subject-specific average fidelity scores, contrasting memory status across the operations. However, $n = 4$ participants were excluded from this test because of missing "forgotten" items in one or more operations, restricting the sample to $n = 18$. To address this limitation, we employed mixed-effects modeling (via *statsmodels*) for each memory operation (maintain, replace, and suppress), analyzing all trials across all $n = 22$ participants (Seabold & Perktold, 2010). In our mixed-effects model, "subjects" were treated as a random effect to accommodate individual differences, and "memory outcome" served as a fixed effect to assess the impact of the operations on representational fidelity. Our model included both intercepts and slopes for "memory effect" as random effects by participant, specified as (1 + Memory | Participant). This adjustment allows the model to account for individual variations in both the baseline fidelity of neural representations and their sensitivity to memory operations. The use of this additional analysis enabled us to discern the influence of different memory operations on representation fidelity using all available data and confirm our initial findings.

Variability in Engagement of Working Memory Operations

There is variability inherent in engaging complex working memory operations (Armbruster-Genç, Ueltzhöffer, & Fiebach, 2016) which reflects the dynamic nature of cognitive resource allocation (Waschke, Kloosterman, Obleser, & Garrett, 2021; Garrett et al., 2013). Attentional fluctuations from trial to trial can impact the maintenance of information in working memory (Hakim, deBettencourt, Awh, & Vogel, 2020; deBettencourt, Keene, Awh, & Vogel, 2019). To assess fluctuations in control engagement, we computed the trial-by-trial variability in operation classifier evidence values across the 30 trials of each operation (from a 4-sec window after the onset of the instruction, z scored across trials). These classifier evidence values represented the predicted data for each trial, reflecting the classifier's confidence in identifying the operation being performed. The variance was calculated using the NumPy var() function, which computes the mean of the squared deviations from the mean (var = mean(abs(a – a.mean()) **2)). The magnitude of the operation classifier's predicted evidence on a given trial reflects the strength of operation engagement on that particular trial, whereas the variability in predicted evidence across trials indicates the consistency of engagement for each operation. To statistically assess the significance of the observed differences in operation variance, we conducted a one-way ANOVA and follow-up t tests for pairwise comparisons between each operation.

To investigate the impact of operation engagement on neural representation and memory outcomes, we conducted regression analyses using the same classifier evidence values that informed the variance calculation. These evidence values, for the operation and category classifiers, were averaged over the same 4-sec window following the onset of the operation cue to ensure consistent time points across analyses. This approach allowed us to determine the extent to which trial-level engagement, as reflected by the operation classifier evidence, influenced the fidelity of working memory representations during the manipulation phase, as well as the accuracy of subsequent memory recognition. To increase our ability to detect trial-specific effects, we pooled data from all participants and then performed bootstrap resampling to evaluate the population reliability of the result (Kim, Schlichting, et al., 2020; Fisher & Hall, 1991; Efron, 1979). On each bootstrap iteration (of 10,000 in total), we randomly sampled (with replacement) a collection of participants' data to match the size of our sample. For the subsequent memory analysis, $n = 4$ participants were excluded because they had at least one operation with no forgotten items and thus could not provide samples for all conditions, leaving a sample of $n = 18$. Statistical significance was calculated with a nonparametric test across bootstrap iterations, evaluating the stability of an effect of interest by calculating the proportion of iterations in which the effect was found. Finally, to verify that the results were not driven by variance across participants, we repeated the primary analyses using standardized (z scored) classifier evidence for each participant to remove participant-specific effects (Kim, Schlichting, et al., 2020; Kim et al., 2014). Results from the primary analyses were qualitatively similar and confirmed.

Statistics and Reproducibility

The fMRI experiment was performed once by each participant. No replication was conducted. We conducted a post hoc power analysis to assess the likelihood of detecting the premanipulation versus postmanipulation change in

neural representations. Using a one-tailed t test for matched samples, we found that our study had a power of 0.663, with an observed effect size of 0.507. Using a one-tailed t test for unmatched samples, we found that our analysis had a power of 0.758, with an observed effect size of 0.218. Although these are both slightly below the commonly accepted threshold for adequate power (0.8), our statistical tests (along with bootstrapping) converge to corroborate our findings.

RESULTS

Removal Operations Are Stable across Trials and Participants

Our first objective was to confirm findings from prior research that the working memory operations of interest (maintain, replace, and suppress) in the current study were neurally distinct. A group-level mass univariate analysis produced results consistent with our previous studies (Kim, Smolker, et al., 2020; Banich et al., 2015; Figure 3). The replace operation showed activation of the fusiform face area (FFA), consistent with replacing scenes with faces, as well as activity in the posterior cingulate cortex and OFC, indicative of integration of working memory content and updating. In contrast, the suppress operation showed bilateral activation in the dlPFC and a decrease in activity in scene-responsive regions (parahippocampal place area), consistent with the engagement of cognitive control mechanisms.

Neural distinctiveness of these working memory operations was confirmed by training fMRI pattern classifiers on whole-brain data separately for each participant (AUC) averaged across operations, M = 0.807, SEM =

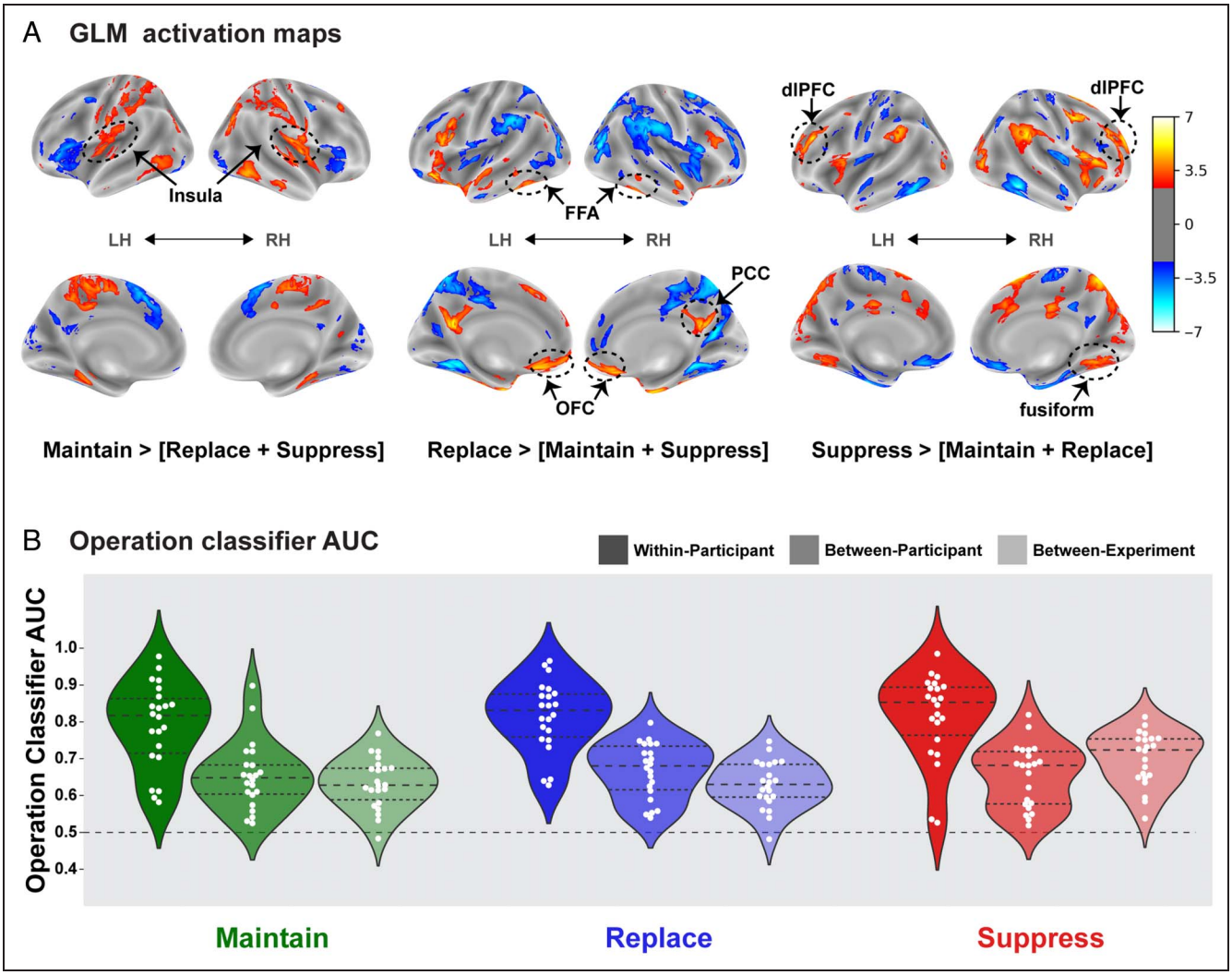

Figure 3. GLM activation maps and MVPA decoding of working memory operations. (A) Univariate contrasts of fMRI data by operation. Key regions previously shown to be activated by each operation are highlighted. In addition, we identified activity in the FFA, which is part of the ventral visual processing stream. (B) Operation classifier performance as the AUC where chance is 0.5. The within-participant and between-participant classification was performed on the data from the current study (n = 22), and the between-experiment classification was performed by training on data (n = 50) from Kim, Smolker, and colleagues (2020) and testing on the data from the current study. PCC = posterior cingulate cortex; LH/RH = left/right hemisphere.

0.022; one-sample t tests for maintain: $t(21) = 11.773$, $p = 1.034e{-}10$, $d = 2.569$; replace: $t(21) = 15.368$, $p = 6.739e{-}13$, $d = 3.353$; and suppress: $t(21) = 12.379$, $p = 4.097e{-}11$, $d = 2.701$ (Figure 3B). Classification was also successful across participants, $M = 0.658$, $SEM = 0.010$; one-sample t test for maintain: $t(21) = 7.768$, $p = 1.316e{-}07$, $d = 1.695$; replace: $t(21) = 10.344$, $p = 1.068e{-}09$, $d = 2.257$; suppress: $t(21) = 8.458$, $p = 3.333e{-}08$, $d = 1.846$, and across experiments, $M = 0.6541$, $SEM = 0.0146$; one-sample t test for maintain: $t(21) = 9.040$, $p = 1.101e{-}08$, $d = 1.973$; replace: $t(21) = 9.428$, $p = 5.377e{-}09$, $d = 2.057$; suppress: $t(21) = 12.929$, $p = 1.818e{-}11$, $d = 2.821$.

Removal Operations Have Divergent Impacts on the Contents of Working Memory

The next set of analyses, which were also confirmatory in nature, was designed to assess how these removal operations influence the representation of information in working memory. Consistent with our prior work (Kim, Smolker, et al., 2020), the decoding of items being replaced from working memory dropped below the decoding for items being maintained; in contrast, the decoding of items being suppressed from working memory did not (Figure 4). The replace operation showed category evidence that dropped below both of the other operations. Suppress, however, did not diverge from maintain, 8–10 sec: $t(21) = 1.540$, $p = .138$; 10–12 sec: $t(21) = 0.307$, $p = .762$, indicating that suppression did not remove category information from working memory while the operation was being engaged.

Further analyses, focusing on the fusiform and parahippocampal ROIs—selected for their propensities toward faces and scenes, respectively—revealed that both regions decoded both face and scene information effectively, underscoring their capacity to represent more than just their specialized content. Specifically, although the fusiform ROI demonstrated an increase in face evidence, aligning with its specialization, face evidence in the parahippocampal ROI also followed an expected trajectory, albeit less prominently. We observed in the parahippocampal ROI that scene evidence during suppress trials numerically fell below that of maintain trials, but there were no statistical differences, 8–10 sec: $t(21) = -2.653$, $p = .223$; 10–12 sec: $t(21) = -2.454$, $p = .344$. These results were therefore consistent with what we observed in the larger VVS that suppression did not remove category information from working memory while the operation was being engaged.

The Variability of Control Engagement Influences the Impact of Removal Operations

We next considered the consistency of operation engagement across trials and its potential influence on memory representations. This analysis is rooted in the literature

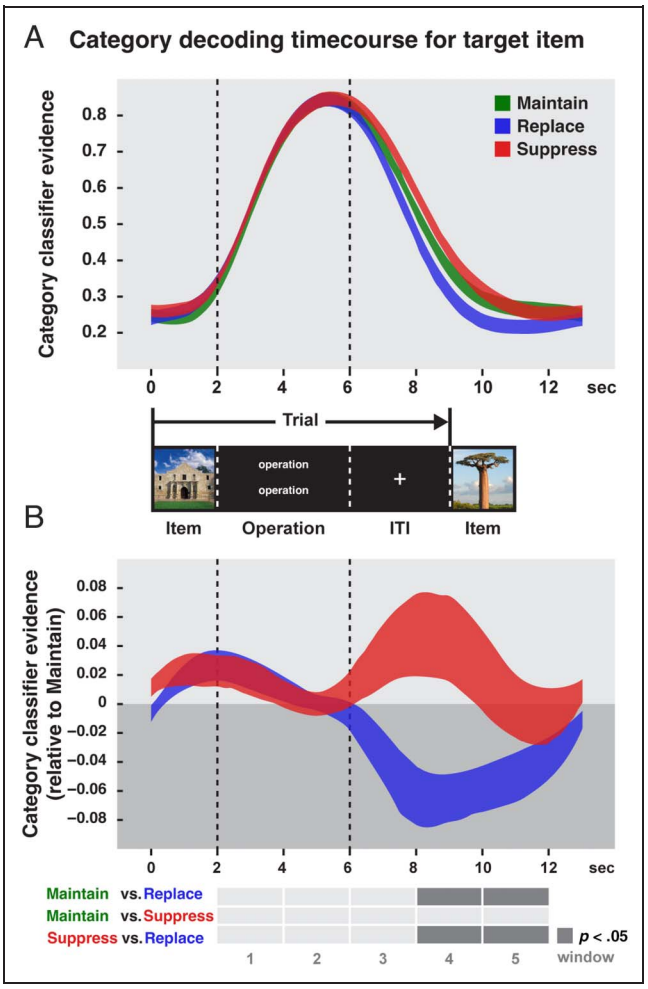

Figure 4. Neural decoding of working memory representations during removal. (A) Group-averaged classifier evidence scores from the category-level fMRI pattern classifier, unshifted for hemodynamic lag. (B) Relative classifier evidence scores for replace and suppress trials relative to maintain trials. Statistical significance between operations is indicated in the boxes below the graph. The dark gray cells indicate significant results ($p < .05$, Bonferroni-corrected) from a one-sample t test (suppress vs. maintain and replace vs. maintain) and paired t test (suppress vs. replace). The width of each line represents the mean ± 1 *SEM*.

suggesting that neural variability is a key dimension for understanding brain–behavior relationships (Waschke et al., 2021). Variabilities in internal physiological states and external environmental constraints may engender differential engagement in control across individual trials (Braver, 2012), and consistency in attention control, a facet of cognitive variability, is an important aspect intertwined with various cognitive abilities (Unsworth, 2015). In a novel analysis in the present article, we assessed if the consistency to which an operation is engaged influences the extent to which the item representation is altered and whether those effects vary across the operations. We first evaluated the consistency of operation engagement across trials using trial-to-trial variability of classifier evidence for each operation. Control engagement varied across operations, $F(2, 21) = 41.390$, $p =$

2.458e-18, and suppression was more variable than maintenance, $t(21) = 2.392, p = .026$, and replacement, $t(21) = 2.438$, $p = .0237$, but replacement did not differ from maintenance, $t(21) = 0.534$, $p = .599$.

Having established the variability in operation engagement, we then examined how the strength of engagement within each operation—suppression and replacement, in particular—affects the neural representation of memory items. The strength of engagement of the maintenance operation on a given trial did not influence the decoding of that representation in working memory. However, greater engagement of the suppression and replacement operations did (Figure 5A). Specifically, stronger engagement in the replacement of an item from working memory was associated with weaker decoding of the removed item (scene; 95% CI [−0.10, −0.03]) and stronger decoding of the new item (face; 95% CI [0.10, 0.19], data not shown). Stronger engagement in the suppression of an item from working memory was associated with higher decoding of that item during suppression (95% CI [0.01, 0.09]). Although these findings for maintenance suppression may appear counterintuitive at first glance, they align well with prior work and will be further elaborated in the Discussion section.

Suppression of an Item from Working Memory Can Weaken Its Subsequent Accessibility

Removing items from working memory did not lead to more forgetting, on average, than did maintaining them (22.9% forgotten for replace and suppress trials, versus 20.3% for maintain trials; all pairwise ps $> .05$). When accounting for variability in operation engagement, however, we found a negative relationship between suppression engagement and subsequent memory such that stronger engagement of suppression was associated with more forgetting (95% CI [−0.40, −0.06]; see Figure 5C). This relationship was not mediated by the strength of the representation in working memory during suppression (95% CI [−0.42, −0.08]). These results indicate that greater attempts at suppression lead to greater category evidence in working memory for the items being suppressed but worse long-term recognition memory for those items. No relationship was found between the engagement of either maintenance (95% CI [−0.25, 0.09]) or replacement (95% CI [−0.17, 0.15]) and memory. Separate from the level of control engagement, stronger representations of items being maintained in working memory were associated with better memory (see Figure 5B). This relationship was not found for either of the two removal operations, whose effects were driven instead by the engagement of those control operations.

Maintenance Suppression Can Lead to the Weakening of Item-level Features in Long-term Memory

To test our hypothesis that maintenance suppression selectively impairs item-level features in long-term memory, we computed the feature-weighted neural pattern similarity of memory representations before and after they were operated on in working memory (see Methods section for details). As shown in Figure 6, the fidelity of category-level representations in the ventral visual cortex did not differ for remembered versus forgotten items in the suppressed items. The item-level representation fidelity was selectively impaired, however, for suppressed items that were subsequently forgotten, $t(17) = 2.151$, $p = .046$, bootstrap 95% CI [0.424, 4.633]. We analyzed these data in two ways: pooled across participants (evaluating the average premanipulation/postmanipulation changes across participants, using a paired t test) and

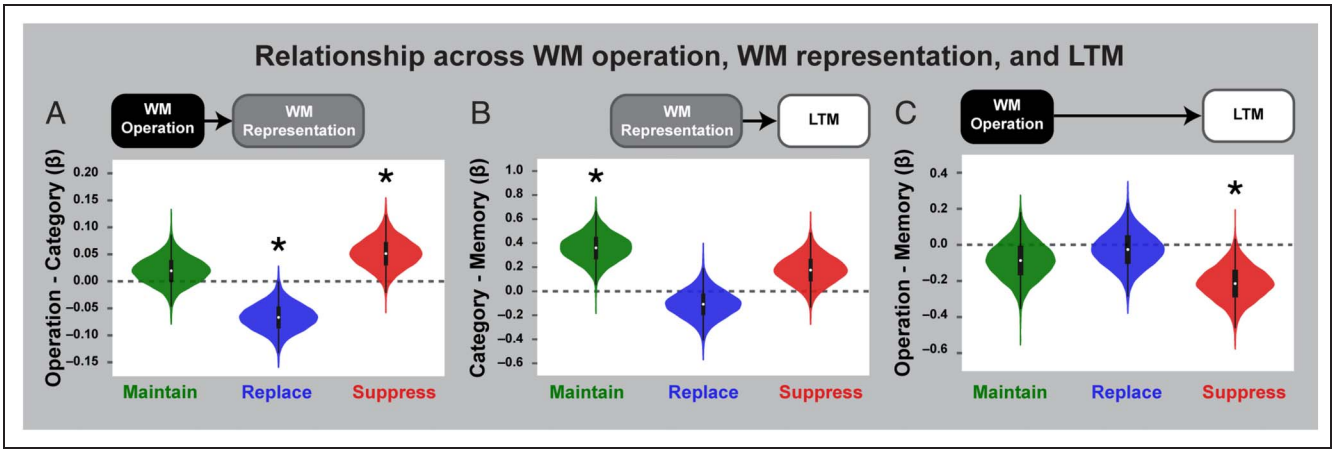

Figure 5. Increased engagement of maintenance suppression predicts stronger representation in working memory and more forgetting. (A) Linear regression linking operation engagement and representation of the item in working memory. Statistics are based on bootstrap analyses with 10,000 iterations (*$p < .05$). (B) Logistic regression linking working memory representations and subsequent memory. (C) Logistic regression linking operation engagement and subsequent memory. Trial-level operation and representation evidence were taken from a 4TR window following the onset of the operation cue. WM = working memory.

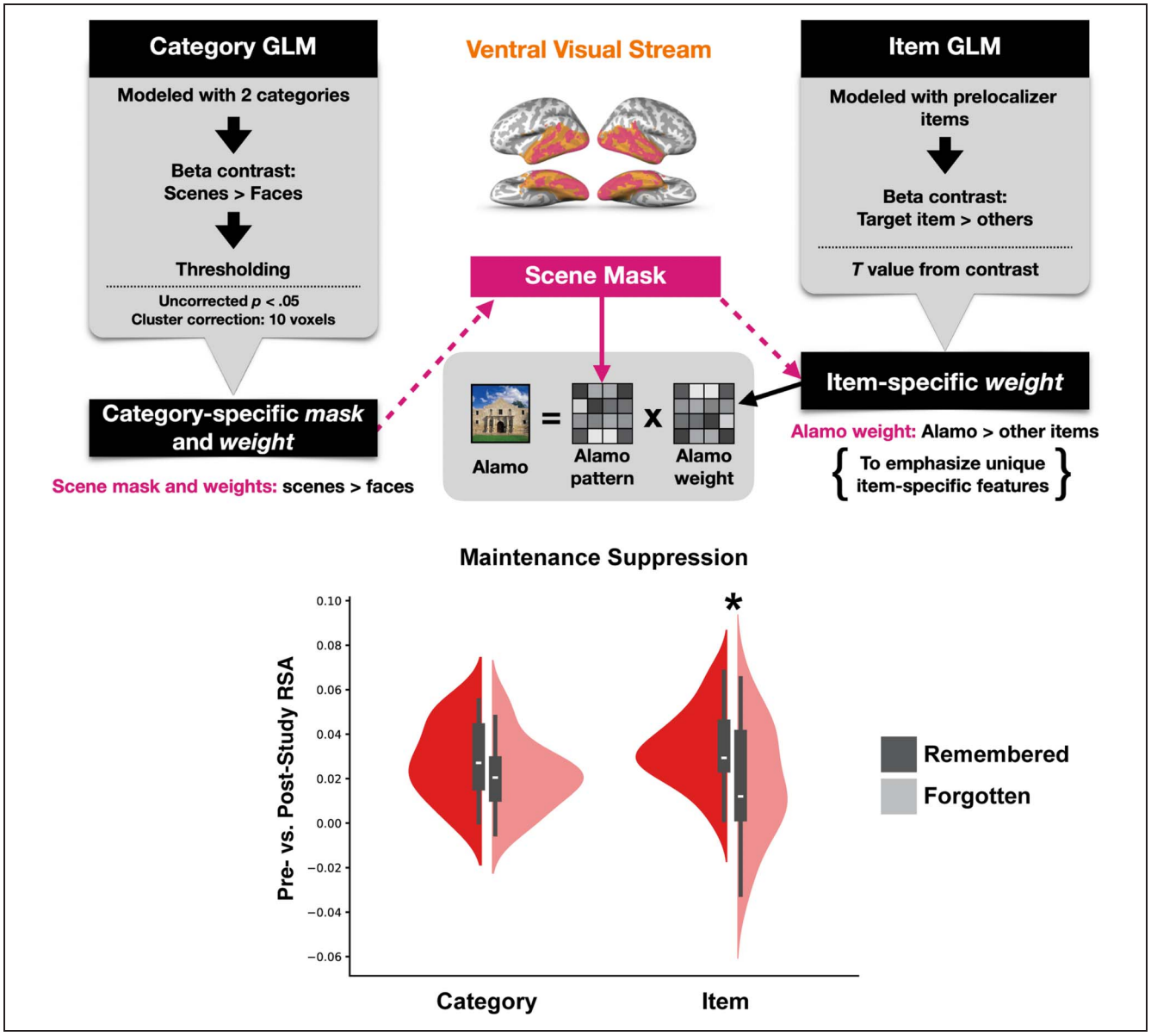

Figure 6. Feature-weighted representational changes in long-term memory. Top: Illustration of the method to extract item- and category-level representational templates to evaluate representational changes following an operation. Bottom: Category-level and item-level RSA results for premanipulation versus postmanipulation changes to item representations. Items are sorted by memory outcome (*$p < .05$). Violins represent the kernel density estimate of the underlying distributions. The violin interior represents a box plot of the data, with the mean indicated by a white dash and inner-quartile ranges indicated by a thick bar.

pooled across items (evaluating the premanipulation/postmanipulation changes in the representation of all items, regardless of participant, using mixed effects modeling). Pooling data across items confirmed the significant effect of maintenance suppression on item-level representation fidelity when the item was forgotten ($\beta = 0.014$, $SE = 0.006$, $z = 2.421$, $p = .015$, 95% CI [0.003, 0.026]). Moreover, we confirmed the lack of significant change in category-level representation fidelity because of maintenance suppression ($\beta = 0.007$, $SE = 0.004$, $z = 1.568$, $p = .117$, 95% CI [−0.002, 0.015]).

Further analyses were conducted to investigate the effects of the maintain and replace operations. These operations did not significantly impact the neural representation fidelity (regardless of memory outcome), either at the item or category level. Specifically, in the maintain condition, item-level ($\beta = 0.004$, $SE = 0.015$, $z = 0.261$, $p = .794$, 95% CI [−0.025, 0.033]) and category-level changes were not significant ($\beta = 0.004$, $SE = 0.004$, $z = 0.929$, $p = 0.353$, 95% CI [−0.005, 0.013]). Similarly, the replace condition showed no significant effects on item-level ($\beta = -0.001$, $SE = 0.007$, $z = -0.070$, $p = 0.944$, 95% CI [−0.015, 0.014]) or category-level ($\beta = -0.002$, $SE = 0.005$, $z = -0.392$, $p = 0.695$, 95% CI [−0.013, 0.008]) fidelity. These null results suggest that the observed impairments in neural representation fidelity

are specific to the maintenance suppression operation and do not extend to maintain or replace trials.

This finding in maintenance suppression was also selective to the ventral visual cortex and did not extend to pFC, parietal cortex, or hippocampus (see Methods section for how these regions were defined). Maintenance suppression did not impact the category-level fidelity of forgotten versus remembered items in pFC (95% CI [−0.824, 2.989]), parietal regions (95% CI [−1.560, 2.710]), or hippocampus (95% CI [−3.172, 0.911]), nor did it impact the item-level fidelity in pFC (95% CI [−0.857, 2.992]), parietal cortex (95% CI [−1.067, 3.284]), or hippocampus (95% CI [−1.708, 2.224]). These analyses support the regional specificity of maintenance suppression effects on representations in sensory regions.

DISCUSSION

Our study set out to examine the influence of two working memory removal operations—maintenance suppression and replacement—on the long-term memory of visual stimuli. First, MVPA of fMRI data indicated that when an item in working memory was replaced by another, its neural representation became deactivated. Importantly, this deactivation did not appear to have a lasting impact on the item's representation or accessibility in long-term memory. Conversely, maintenance suppression did impact long-term memory. When an item was suppressed from working memory, its neural representation was not deactivated, but rather it remained activated to a similar degree as on maintain trials, suggesting that the suppressed item likely remained in the focus of attention (Lewis-Peacock et al., 2012). These findings corroborate prior studies showing that suppressing items from working memory is an active process requiring focused attention on the unwanted information (Kim, Smolker, et al., 2020; Banich et al., 2015; Benoit & Anderson, 2012). In fact, our analyses showed that greater attempts to suppress an item in working memory led to a seemingly paradoxical increase in neural activation in working memory that was also associated with subsequent forgetting in long-term memory (Figure 5). This pattern of results has been reported previously in an item-method, directed-forgetting study with visual stimuli where a forget cue produced greater neural activation of the item in the ventral temporal cortex, relative to a remember cue, and this led to more forgetting on an item-recognition test (Wang et al., 2019). In the present study, a single dose of maintenance suppression selectively weakened the item-level features, but not category-level features, in the ventral temporal cortex of items that were subsequently forgotten. However, there were no representational changes observed in brain areas associated with the control of memories including the hippocampus, pFC, and parietal cortex. In summary, our data illustrate that removing information from working memory can be accomplished using different cognitive strategies (suppression or replacement) with distinct neural signatures with divergent impacts on long-term memory representations.

We conducted several analyses to validate the distinctiveness and consistency of the working memory removal operations being studied here. First, we performed a group-level mass univariate analysis of the fMRI data to identify unique patterns of brain activity associated with maintaining, replacing, and suppressing information in working memory. These results were consistent with previous studies on the same cognitive operations (Kim, Smolker, et al., 2020; Banich et al., 2015; Benoit & Anderson, 2012). Next, we used MVPA to show that the neural patterns of activity underlying each operation were distinct and consistent within individuals. Moreover, these operation activity patterns were sufficiently consistent to be decodable across participants in this study and our prior study (Kim, Smolker, et al., 2020).

Another important aspect of our findings lies in the consideration of variability in cognitive control across trials. It has been shown that variability in attention and control can significantly impact perceptions, decision-making, and maintenance of information in working memory (Hakim et al., 2020; Wolff et al., 2019; Arazi, Censor, & Dinstein, 2017; Armbruster-Genç et al., 2016). In the present study, we evaluated the impact of variability in the application of maintenance suppression in working memory on the long-term memory representations of the suppressed items. To do this, we used the operation classifier evidence on each trial to show that trials with greater evidence that the maintenance suppression operation was effectively engaged were associated with a stronger representation of the to-be-suppressed item in working memory and, subsequently, worse long-term memory recognition for that item. Our findings propose that the variability in patterns of neural activity during attempts to suppress information in working memory could serve as a predictive neural signature for the success of the suppression effort, and by extension, its enduring impact on memory. These findings align with an emerging appreciation that variability in neuroimaging measurements across trials is not mere noise but a meaningful signal that can offer novel insights into brain function (Waschke et al., 2021; Armbruster-Genç et al., 2016; Garrett et al., 2013; Braver, 2012).

To investigate the lasting impact of maintenance suppression on long-term memory, we compared the representation of items from the premanipulation scan to the postmanipulation scan. Our results were consistent with Kim, Smolker, and colleagues (2020) in elucidating the unique effects of maintenance suppression within working memory. However, we extend this line of research by showing that maintenance suppression can also have a specific and durable impact on long-term memory. When an item was suppressed and later forgotten, we observed a significant weakening of its item-level features in the sensory cortex. This effect diverges from the effects noted for items that were either maintained or replaced. In these conditions,

we did not observe any changes in item- or category-level features, suggesting that forgetting in these conditions is likely driven by more passive mechanisms like decay or general interference (Pertzov, Manohar, & Husain, 2017). It is important to note, however, that although these more passive mechanisms may not manifest as measurable changes in our fMRI data (Zhang & Luck, 2009), they do affect the likelihood of successful recall. Therefore, the absence of item- or category-level impacts in maintained or replaced items does not guarantee their successful retention and highlights a possible limitation in our neural measures. Nonetheless, the representational changes we observed for maintenance suppression point to a targeted, active mechanism of forgetting, aligning with prior observations on the active processes underpinning memory suppression (Banich et al., 2015; Benoit & Anderson, 2012).

The results of this study are generally consistent with the sensory recruitment hypothesis (Christophel, Klink, Spitzer, Roelfsema, & Haynes, 2017; Serences, 2016; D'Esposito & Postle, 2015; Harrison & Tong, 2009; Serences et al., 2009; D'Esposito, 2007). We observed suppression-induced representational changes in sensory areas (ventral temporal cortex) responsible for initial encoding and representation of the visual stimuli, but not in regions associated with the control of attention and memory such as the hippocampus, pFC, and parietal cortex (Zhou et al., 2022; Rademaker et al., 2019; Bowman & Zeithamova, 2018). Consistent with the biased-competition model (Polk, Drake, Jonides, Smith, & Smith, 2008) and the attentional theory proposed by Zanto and Gazzaley (2009), we suggest that attentional signals may selectively target item-level features in suppression, leaving the category-level features largely intact. It is noteworthy that our cognitive frameworks for "stimulus categories" are dynamic (Jern & Kemp, 2013), even though we innately recognize and remember patterns within them (Brady & Oliva, 2008). This adaptability might explain why, under maintenance suppression, specific item details fade but the broader categorical structures remain. These preserved category-level features could function as an "organizational scaffold" for long-term memory (Reagh & Ranganath, 2023). This idea is in line with prior work that showed regions in the inferior temporal cortex, such as the FFA and parahippocampal place area, are not just passive recipients of sensory information but are actively modulated during the encoding and maintenance phases of working memory tasks (Ranganath et al., 2005; Ranganath, Cohen, Dam, & D'Esposito, 2004; Ranganath, DeGutis, & D'Esposito, 2004). This aligns with our findings from the fusiform and parahippocampal ROIs, where scene evidence did not significantly fall below maintain during the maintenance suppression trials, indicating these regions' role in retaining broad informational content (Christophel et al., 2017). The retention of category-level features during successful maintenance suppression could thus be conceptualized as an efficient strategy for memory function. The selective weakening of item-level features, but not category features, could minimize the erasure of useful categorical frameworks that could facilitate future encoding and retrieval of related information. This idea is supported by Hemmer and Persaud (2014), who argue that integrating categorical knowledge not only aids in episodic memory reconstruction but also can enhance memory accuracy. Our findings suggest that maintenance suppression is not merely a "blunt tool" for forgetting but rather a fine-tuned process that contributes to the remembering of related information (Rademaker et al., 2019; Gayet et al., 2018; Hemmer & Persaud, 2014; Ranganath et al., 2005).

Much prior work has emphasized memory suppression in the context of retrieval from long-term memory (e.g., Anderson & Hanslmayr, 2014). Such paradigms shed light on mechanisms affecting both episodic memory and what has been termed as conceptual implicit memory. This idea refers to the memory of information's meaning without explicit recall of encountering it. Prior work has revealed that suppressing memories can impact this conceptual implicit memory, suggesting alterations in the semantic representations of suppressed content (Taubenfeld et al., 2019). However, these studies typically necessitate multiple manipulations of an item before discernible effects on memory are observed (Taubenfeld et al., 2019; Depue et al., 2007; Anderson & Green, 2001). Drawing a distinction, our research focuses on the suppression of information within working memory. Rather than requiring numerous attempts, our study reveals that a singular event of maintenance suppression can manifest lasting changes in item-level representations and recognition memory performance. These results not only demonstrate the immediacy of the impact of maintenance suppression on long-term memory but also helps to build on our understanding of dynamics between shared representations in working and long-term memory (Vo et al., 2022).

Directed forgetting research offers another perspective on memory modulation in which forgetting is accomplished not by altering an item's representation, but rather by changes to the contextual representation to which that item is bound in episodic memory (Hubbard & Sahakyan, 2021, 2023; Sahakyan & Kelley, 2002). In a recent study using the item-method, directed-forgetting procedure, participants were shown words interspersed with scene images that acted as "context tags." When participants were later instructed to forget specific items, there was an enhanced reactivation of their associated context (the "scene" activity pattern), whereas item-specific information (the "word" activity pattern) was reduced. Greater differences between context and item activity were predictive of successful forgetting (Chiu, Wang, Beck, Lewis-Peacock, & Sahakyan, 2021). These results suggest an important role of context in memory modification such that intentional forgetting may involve dissociating items from their contexts. It is possible that context unbinding

also plays a role in maintenance suppression. We have no way of assessing this in the present study, but we plan to address this in future research.

There are several key limitations to our study that warrant noting. First, our investigation exclusively focused on the effects of suppression and replacement of visual nonverbal stimuli, specifically scenes (e.g., landscapes and monuments). The importance of this limitation lies in the question of generalizability. It remains unclear whether our findings extend to other types of stimuli, such as other classes of visual information, or verbal and/or auditory information. Future research should explore these different sensory and content domains to determine the breadth of the mechanisms we have identified. Such future studies are crucial for establishing whether the weakening of item-level features during maintenance suppression is a general phenomenon. Second, although we discerned nuanced impacts, a consistent long-term memory effect from the removal operations was not universally observed. The lack of reliable long-term behavioral effects could stem from a ceiling effect, as some participants did not forget any items in a given condition. Addressing this ceiling effect in future studies could involve incorporating a more complex array of stimuli where participants would be asked to manipulate more than one item at a time, thereby increasing the task's difficulty. Third, and finally, we did not include emotionally relevant stimuli, either negative or positive, in our experiment. Adding emotional content is an important consideration for future work, particularly given the different neural and cognitive processing pathways for emotionally charged memories. Understanding how maintenance suppression affects these types of memories could have important clinical implications, especially for conditions that involve the intrusion of unwanted, emotionally charged memories, such as those that occur in individuals with posttraumatic stress disorder (PTSD). This is a topic of ongoing research in our laboratories.

This study may hold relevance for clinical applications. One such application concerns possible treatments for PTSD aimed at reducing the frequency and impact of intrusive thoughts (Foa & McLean, 2016; Hayes, VanElzakker, & Shin, 2012). Our demonstration that maintenance suppression can weaken item-level features in long-term memory offers a potential strategy for suppressing aspects of traumatic memories. However, the application and timing of suppression must be carefully considered. The memory suppression examined in related research using the Think/No-Think paradigm (Anderson, 2004; Anderson & Green, 2001) is focused on suppressing the retrieval of information from long-term memory, not on suppressing the representation of information in working memory. In the retrieval-induced forgetting literature, there is evidence that disruptions in memory control associated with trauma are associated with reduced activation in the right middle frontal gyrus during memory suppression attempts in trauma-exposed individuals (Sullivan et al., 2019). Another recent study indicates a generalized disruption in PTSD of the regulation signal controlling the reactivation of unwanted memories, suggesting a deficit in memory control following trauma (Mary et al., 2020). Although these studies explored the impacts of retrieval suppression, they provide a broader context on memory control and its potential malfunction in PTSD, thus accentuating the importance of understanding suppression mechanisms at different memory stages. However, often in PTSD, the suppression of retrieval of unwanted information is not possible, and such information enters working memory (so-called "intrusions"). Our work fills a crucial gap and suggests that acting on information that intrudes into working memory, for example, might enable a change in the long-term memory representation of that information which could reduce the likelihood of future intrusions.

Conclusion

In conclusion, our study focuses on the long-term memory consequences of suppressing information from working memory. Our findings reveal that maintenance suppression selectively weakens item-level features of forgotten items in the sensory regions of the brain that encoded them. As such, these results are consistent with the sensory recruitment hypothesis of working memory, and they complement prior observations that suppressing the retrieval of items from long-term memory can weaken their item-specific features. It is an important area of future research to evaluate how the neural mechanisms of maintenance suppression differ from those of retrieval suppression.

Acknowledgments

We thank Paige Stetson for her assistance in fMRI data collection.

Corresponding author: Zachary H. Bretton, Institute for Neuroscience, University of Texas at Austin, 1 University Station, Stop C7000, Austin, TX 78712-1139, or via e-mail: zbretton@utexas.edu.

Data Availability Statement

All de-identified neuroimaging data are available from the authors upon request.

Author Contributions

Zachary H. Bretton: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Project administration; Software; Validation; Visualization; Writing—Original draft; Writing—Review & editing. Hyojeong Kim: Conceptualization; Methodology; Software; Visualization; Writing—Original draft; Writing—Review & editing. Marie T. Banich: Conceptualization; Funding acquisition;

Methodology; Writing—Original draft; Writing—Review & editing. Jarrod A. Lewis-Peacock: Conceptualization; Funding acquisition; Methodology; Project administration; Resources; Supervision; Writing—Original draft; Writing—Review & editing.

Funding Information

This work was supported by the National Eye Institute (https://dx.doi.org/10.13039/100000053), grant number: R01EY028746; and National Institute of Mental Health (https://dx.doi.org/10.13039/100000025), grant numbers: R56MH125642, T32MH106454.

Diversity in Citation Practices

Retrospective analysis of the citations in every article published in this journal from 2010 to 2021 reveals a persistent pattern of gender imbalance: Although the proportions of authorship teams (categorized by estimated gender identification of first author/last author) publishing in the *Journal of Cognitive Neuroscience* (*JoCN*) during this period were M(an)/M = .407, W(oman)/M = .32, M/W = .115, and W/W = .159, the comparable proportions for the articles that these authorship teams cited were M/M = .549, W/M = .257, M/W = .109, and W/W = .085 (Postle and Fulvio, *JoCN*, 34:1, pp. 1–3). Consequently, *JoCN* encourages all authors to consider gender balance explicitly when selecting which articles to cite and gives them the opportunity to report their article's gender citation balance. The authors of this article report its proportions of citations by gender category to be: M/M = .623; W/M = .208; M/W = .104; W/W = .065.

REFERENCES

Abdulrahman, H., & Henson, R. N. (2016). Effect of trial-to-trial variability on optimal event-related fMRI design: Implications for beta-series correlation and multi-voxel pattern analysis. *Neuroimage*, *125*, 756–766. https://doi.org/10.1016/j.neuroimage.2015.11.009, PubMed: 26549299

Albers, A. M., Kok, P., Toni, I., Dijkerman, H. C., & de Lange, F. P. (2013). Shared representations for working memory and mental imagery in early visual cortex. *Current Biology*, *23*, 1427–1431. https://doi.org/10.1016/j.cub.2013.05.065, PubMed: 23871239

Anderson, M. C. (2004). Neural systems underlying the suppression of unwanted memories. *Science*, *303*, 232–235. https://doi.org/10.1126/science.1089504, PubMed: 14716015

Anderson, M. C., & Green, C. (2001). Suppressing unwanted memories by executive control. *Nature*, *410*, 366–369. https://doi.org/10.1038/35066572, PubMed: 11268212

Anderson, M. C., & Hanslmayr, S. (2014). Neural mechanisms of motivated forgetting. *Trends in Cognitive Sciences*, *18*, 279–292. https://doi.org/10.1016/j.tics.2014.03.002, PubMed: 24747000

Arazi, A., Censor, N., & Dinstein, I. (2017). Neural variability quenching predicts individual perceptual abilities. *Journal of Neuroscience*, *37*, 97–109. https://doi.org/10.1523/JNEUROSCI.1671-16.2016, PubMed: 28053033

Armbruster-Genç, D. J. N., Ueltzhöffer, K., & Fiebach, C. J. (2016). Brain signal variability differentially affects cognitive flexibility and cognitive stability. *Journal of Neuroscience*, *36*, 3978–3987. https://doi.org/10.1523/JNEUROSCI.2517-14.2016, PubMed: 27053205

Avants, B., Epstein, C., Grossman, M., & Gee, J. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis*, *12*, 26–41. https://doi.org/10.1016/j.media.2007.06.004, PubMed: 17659998

Axmacher, N., Schmitz, D. P., Weinreich, I., Elger, C. E., & Fell, J. (2008). Interaction of working memory and long-term memory in the medial temporal lobe. *Cerebral Cortex*, *18*, 2868–2878. https://doi.org/10.1093/cercor/bhn045, PubMed: 18403397

Banich, M. T., Mackiewicz Seghete, K. L., Depue, B. E., & Burgess, G. C. (2015). Multiple modes of clearing one's mind of current thoughts: Overlapping and distinct neural systems. *Neuropsychologia*, *69*, 105–117. https://doi.org/10.1016/j.neuropsychologia.2015.01.039, PubMed: 25637772

Behzadi, Y., Restom, K., Liau, J., & Liu, T. T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *Neuroimage*, *37*, 90–101. https://doi.org/10.1016/j.neuroimage.2007.04.042, PubMed: 17560126

Benoit, R. G., & Anderson, M. C. (2012). Opposing mechanisms support the voluntary forgetting of unwanted memories. *Neuron*, *76*, 450–460. https://doi.org/10.1016/j.neuron.2012.07.025, PubMed: 23083745

Blumenfeld, R. S., & Ranganath, C. (2006). Dorsolateral prefrontal cortex promotes long-term memory formation through its role in working memory organization. *Journal of Neuroscience*, *26*, 916–925. https://doi.org/10.1523/JNEUROSCI.2353-05.2006, PubMed: 16421311

Bowman, C. R., & Zeithamova, D. (2018). Abstract memory representations in the ventromedial prefrontal cortex and hippocampus support concept generalization. *Journal of Neuroscience*, *38*, 2605–2614. https://doi.org/10.1523/JNEUROSCI.2811-17.2018, PubMed: 29437891

Brady, T. F., & Oliva, A. (2008). Statistical learning using real-world scenes: Extracting categorical regularities without conscious intent. *Psychological Science*, *19*, 678–685. https://doi.org/10.1111/j.1467-9280.2008.02142.x, PubMed: 18727783

Braver, T. S. (2012). The variable nature of cognitive control: A dual mechanisms framework. *Trends in Cognitive Sciences*, *16*, 106–113. https://doi.org/10.1016/j.tics.2011.12.010, PubMed: 22245618

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology*, *77*, 305–327. https://doi.org/10.1111/j.2044-8295.1986.tb02199.x, PubMed: 3756376

Chiu, Y.-C., Wang, T. H., Beck, D. M., Lewis-Peacock, J. A., & Sahakyan, L. (2021). Separation of item and context in item-method directed forgetting. *Neuroimage*, *235*, 117983. https://doi.org/10.1016/j.neuroimage.2021.117983, PubMed: 33762219

Christophel, T. B., Klink, P. C., Spitzer, B., Roelfsema, P. R., & Haynes, J.-D. (2017). The distributed nature of working memory. *Trends in Cognitive Sciences*, *21*, 111–124. https://doi.org/10.1016/j.tics.2016.12.007, PubMed: 28063661

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114. https://doi.org/10.1017/S0140525X01003922, PubMed: 11515286

D'Esposito, M. (2007). From cognitive to neural models of working memory. *Philosophical Transactions of the Royal Society of London, Series B: Biological Sciences*, *362*,

761–772. https://doi.org/10.1098/rstb.2007.2086, PubMed: 17400538
D'Esposito, M., & Postle, B. R. (2015). The cognitive neuroscience of working memory. *Annual Review of Psychology*, *66*, 115–142. https://doi.org/10.1146/annurev-psych-010814-015031, PubMed: 25251486
Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, *16*, 693–700. https://doi.org/10.1016/j.conb.2006.10.012, PubMed: 17097284
deBettencourt, M. T., Keene, P. A., Awh, E., & Vogel, E. K. (2019). Real-time triggering reveals concurrent lapses of attention and working memory. *Nature Human Behaviour*, *3*, 808–816. https://doi.org/10.1038/s41562-019-0606-6, PubMed: 31110335
Depue, B. E., Curran, T., & Banich, M. T. (2007). Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. *Science*, *317*, 215–219. https://doi.org/10.1126/science.1139560, PubMed: 17626877
Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *Annals of Statistics*, *7*, 1–26. https://doi.org/10.1214/aos/1176344552
Epstein, R., & Kanwisher, N. (1998). A cortical representation of the local visual environment. *Nature*, *392*, 598–601. https://doi.org/10.1038/33402, PubMed: 9560155
Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., et al. (2019). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods*, *16*, 111–116. https://doi.org/10.1038/s41592-018-0235-4, PubMed: 30532080
Esteban, O., Markiewicz, C. J., Burns, C., Goncalves, M., Jarecka, D., Ziegler, E., et al. (2022). *nipy/nipype: 1.8.1* [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.6555085
Fisher, N. I., & Hall, P. (1991). Bootstrap algorithms for small samples. *Journal of Statistical Planning and Inference*, *27*, 157–169. https://doi.org/10.1016/0378-3758(91)90013-5
Foa, E. B., & McLean, C. P. (2016). The efficacy of exposure therapy for anxiety-related disorders and its underlying mechanisms: The case of OCD and PTSD. *Annual Review of Clinical Psychology*, *12*, 1–28. https://doi.org/10.1146/annurev-clinpsy-021815-093533, PubMed: 26565122
Fonov, V., Evans, A., McKinstry, R., Almli, C., & Collins, D. (2009). Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *Neuroimage*, *47 (Suppl. 1)*, S102. https://doi.org/10.1016/S1053-8119(09)70884-5
Gagnepain, P., Henson, R. N., & Anderson, M. C. (2014). Suppressing unwanted memories reduces their unconscious influence via targeted cortical inhibition. *Proceedings of the National Academy of Sciences, U.S.A.*, *111*, E1310–E1319. https://doi.org/10.1073/pnas.1311468111, PubMed: 24639546
Garrett, D. D., Samanez-Larkin, G. R., MacDonald, S. W. S., Lindenberger, U., McIntosh, A. R., & Grady, C. L. (2013). Moment-to-moment brain signal variability: A next frontier in human brain mapping? *Neuroscience & Biobehavioral Reviews*, *37*, 610–624. https://doi.org/10.1016/j.neubiorev.2013.02.015, PubMed: 23458776
Gayet, S., Paffen, C. L. E., & Van der Stigchel, S. (2018). Visual working memory storage recruits sensory processing areas. *Trends in Cognitive Sciences*, *22*, 189–190. https://doi.org/10.1016/j.tics.2017.09.011, PubMed: 29050827
Goodale, M. A., & Milner, A. D. (1992). Separate visual pathways for perception and action. *Trends in Neurosciences*, *15*, 20–25. https://doi.org/10.1016/0166-2236(92)90344-8, PubMed: 1374953
Gorgolewski, K., Burns, C. D., Madison, C., Clark, D., Halchenko, Y. O., Waskom, M. L., et al. (2011). Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Frontiers in Neuroinformatics*, *5*, 13. https://doi.org/10.3389/fninf.2011.00013, PubMed: 21897815
Greve, D. N., & Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, *48*, 63–72. https://doi.org/10.1016/j.neuroimage.2009.06.060, PubMed: 19573611
Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, *15*, 536–548. https://doi.org/10.1038/nrn3747, PubMed: 24962370
Hakim, N., deBettencourt, M. T., Awh, E., & Vogel, E. K. (2020). Attention fluctuations impact ongoing maintenance of information in working memory. *Psychonomic Bulletin & Review*, *27*, 1269–1278. https://doi.org/10.3758/s13423-020-01790-z, PubMed: 32808159
Harrison, S. A., & Tong, F. (2009). Decoding reveals the contents of visual working memory in early visual areas. *Nature*, *458*, 632–635. https://doi.org/10.1038/nature07832, PubMed: 19225460
Haxby, J. V., Connolly, A. C., & Guntupalli, J. S. (2014). Decoding neural representational spaces using multivariate pattern analysis. *Annual Review of Neuroscience*, *37*, 435–456. https://doi.org/10.1146/annurev-neuro-062012-170325, PubMed: 25002277
Hayes, J. P., VanElzakker, M. B., & Shin, L. M. (2012). Emotion and cognition interactions in PTSD: A review of neurocognitive and neuroimaging studies. *Frontiers in Integrative Neuroscience*, *6*, 89. https://doi.org/10.3389/fnint.2012.00089, PubMed: 23087624
Hemmer, P., & Persaud, K. (2014). Interaction between categorical knowledge and episodic memory across domains. *Frontiers in Psychology*, *5*, 584. https://doi.org/10.3389/fpsyg.2014.00584, PubMed: 24966848
Hubbard, R. J., & Sahakyan, L. (2021). Separable neural mechanisms support intentional forgetting and thought substitution. *Cortex*, *142*, 317–331. https://doi.org/10.1016/j.cortex.2021.06.013, PubMed: 34343901
Hubbard, R. J., & Sahakyan, L. (2023). Differential recruitment of inhibitory control processes by directed forgetting and thought substitution. *Journal of Neuroscience*, *43*, 1963–1975. https://doi.org/10.1523/JNEUROSCI.0696-22.2023, PubMed: 36810228
Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, *17*, 825–841. https://doi.org/10.1006/nimg.2002.1132, PubMed: 12377157
Jern, A., & Kemp, C. (2013). A probabilistic account of exemplar and category generation. *Cognitive Psychology*, *66*, 85–125. https://doi.org/10.1016/j.cogpsych.2012.09.003, PubMed: 23108001
Kaniuth, P., & Hebart, M. N. (2022). Feature-reweighted representational similarity analysis: A method for improving the fit between computational models, brains, and behavior. *Neuroimage*, *257*, 119294. https://doi.org/10.1016/j.neuroimage.2022.119294, PubMed: 35580810
Kim, G., Lewis-Peacock, J. A., Norman, K. A., & Turk-Browne, N. B. (2014). Pruning of memories by context-based prediction error. *Proceedings of the National Academy of Sciences, U.S.A.*, *111*, 8997–9002. https://doi.org/10.1073/pnas.1319438111, PubMed: 24889631
Kim, H., Schlichting, M. L., Preston, A. R., & Lewis-Peacock, J. A. (2020). Predictability changes what we remember in familiar temporal contexts. *Journal of Cognitive Neuroscience*, *32*, 124–140. https://doi.org/10.1162/jocn_a_01473, PubMed: 31560266
Kim, H., Smolker, H. R., Smith, L. L., Banich, M. T., & Lewis-Peacock, J. A. (2020). Changes to information in working

memory depend on distinct removal operations. *Nature Communications*, *11*, 6239. https://doi.org/10.1038/s41467-020-20085-4, PubMed: 33288756

Klein, A., Ghosh, S. S., Bao, F. S., Giard, J., Häme, Y., Stavsky, E., et al. (2017). Mindboggling morphometry of human brains. *PLoS Computational Biology*, *13*, e1005350. https://doi.org/10.1371/journal.pcbi.1005350, PubMed: 28231282

Kumar, M., Ellis, C. T., Lu, Q., Zhang, H., Capotă, M., Willke, T. L., et al. (2020). BrainIAK tutorials: User-friendly learning materials for advanced fMRI analysis. *PLoS Computational Biology*, *16*, e1007549. https://doi.org/10.1371/journal.pcbi.1007549, PubMed: 31940340

Lanczos, C. (1964). Evaluation of noisy data. *Journal of the Society for Industrial and Applied Mathematics: Series B, Numerical Analysis*, *1*, 76–85. https://doi.org/10.1137/0701007

LaRocque, J. J., Lewis-Peacock, J. A., & Postle, B. R. (2014). Multiple neural states of representation in short-term memory? It's a matter of attention. *Frontiers in Human Neuroscience*, *8*, 5. https://doi.org/10.3389/fnhum.2014.00005, PubMed: 24478671

Levy, B. J., & Anderson, M. C. (2008). Individual differences in the suppression of unwanted memories: The executive deficit hypothesis. *Acta Psychologica*, *127*, 623–635. https://doi.org/10.1016/j.actpsy.2007.12.004, PubMed: 18242571

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*, 61–79. https://doi.org/10.1162/jocn_a_00140, PubMed: 21955164

Lewis-Peacock, J. A., Kessler, Y., & Oberauer, K. (2018). The removal of information from working memory. *Annals of the New York Academy of Sciences*, *1424*, 33–44. https://doi.org/10.1111/nyas.13714, PubMed: 29741212

Lewis-Peacock, J. A., & Norman, K. A. (2014a). Competition between items in working memory leads to forgetting. *Nature Communications*, *5*, 5768. https://doi.org/10.1038/ncomms6768, PubMed: 25519874

Lewis-Peacock, J. A., & Norman, K. A. (2014b). Multi-voxel pattern analysis of fMRI data. In *The cognitive neurosciences* (pp. 911–920). Cambridge, MA: MIT Press. https://doi.org/10.7551/mitpress/11442.001.0001

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: From psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences*, *17*, 391–400. https://doi.org/10.1016/j.tics.2013.06.006, PubMed: 23850263

Martin, A. (2007). The representation of object concepts in the brain. *Annual Review of Psychology*, *58*, 25–45. https://doi.org/10.1146/annurev.psych.57.102904.190143, PubMed: 16968210

Mary, A., Dayan, J., Leone, G., Postel, C., Fraisse, F., Malle, C., et al. (2020). Resilience after trauma: The role of memory suppression. *Science*, *367*, eaay8477. https://doi.org/10.1126/science.aay8477, PubMed: 32054733

Melrose, R. J., Zahniser, E., Wilkins, S. S., Veliz, J., Hasratian, A. S., Sultzer, D. L., et al. (2020). Prefrontal working memory activity predicts episodic memory performance: A neuroimaging study. *Behavioural Brain Research*, *379*, 112307. https://doi.org/10.1016/j.bbr.2019.112307, PubMed: 31678217

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, *6*, 93–102. https://doi.org/10.1016/S1364-6613(00)01845-3, PubMed: 15866193

Pertzov, Y., Manohar, S., & Husain, M. (2017). Rapid forgetting results from competition over time between items in visual working memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 528–536. https://doi.org/10.1037/xlm0000328, PubMed: 27668485

Polk, T. A., Drake, R. M., Jonides, J. J., Smith, M. R., & Smith, E. E. (2008). Attention enhances the neural processing of relevant features and suppresses the processing of irrelevant features in humans: A functional magnetic resonance imaging study of the Stroop task. *Journal of Neuroscience*, *28*, 13786–13792. https://doi.org/10.1523/JNEUROSCI.1026-08.2008, PubMed: 19091969

Power, J. D. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. *Neuroimage*, *84*, 320–341. https://doi.org/10.1016/j.neuroimage.2013.08.048, PubMed: 23994314

Rademaker, R. L., Chunharas, C., & Serences, J. T. (2019). Coexisting representations of sensory and mnemonic information in human visual cortex. *Nature Neuroscience*, *22*, 1336–1344. https://doi.org/10.1038/s41593-019-0428-x, PubMed: 31263205

Ranganath, C., Cohen, M. X., & Brozinsky, C. J. (2005). Working memory maintenance contributes to long-term memory formation: Neural and behavioral evidence. *Journal of Cognitive Neuroscience*, *17*, 994–1010. https://doi.org/10.1162/0898929054475118, PubMed: 16102232

Ranganath, C., Cohen, M. X., Dam, C., & D'Esposito, M. (2004). Inferior temporal, prefrontal, and hippocampal contributions to visual working memory maintenance and associative memory retrieval. *Journal of Neuroscience*, *24*, 3917–3925. https://doi.org/10.1523/JNEUROSCI.5053-03.2004, PubMed: 15102907

Ranganath, C., DeGutis, J., & D'Esposito, M. (2004). Category-specific modulation of inferior temporal activity during working memory encoding and maintenance. *Cognitive Brain Research*, *20*, 37–45. https://doi.org/10.1016/j.cogbrainres.2003.11.017, PubMed: 15130587

Reagh, Z. M., & Ranganath, C. (2023). Flexible reuse of cortico-hippocampal representations during encoding and recall of naturalistic events. *Nature Communications*, *14*, 1279. https://doi.org/10.1038/s41467-023-36805-5, PubMed: 36890146

Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*, 1064–1072. https://doi.org/10.1037/0278-7393.28.6.1064, PubMed: 12450332

Satterthwaite, T. D., Elliott, M. A., Gerraty, R. T., Ruparel, K., Loughead, J., Calkins, M. E., et al. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. *Neuroimage*, *64*, 240–256. https://doi.org/10.1016/j.neuroimage.2012.08.052, PubMed: 22926292

Seabold, S., & Perktold, J. (2010). Statsmodels: Econometric and statistical modeling with Python. In *Proceedings of the 9th Python in Science Conference* (pp. 92–96). https://doi.org/10.25080/Majora-92bf1922-011

Serences, J. T. (2016). Neural mechanisms of information storage in visual short-term memory. *Vision Research*, *128*, 53–67. https://doi.org/10.1016/j.visres.2016.09.010, PubMed: 27668990

Serences, J. T., Ester, E. F., Vogel, E. K., & Awh, E. (2009). Stimulus-specific delay activity in human primary visual cortex. *Psychological Science*, *20*, 207–214. https://doi.org/10.1111/j.1467-9280.2009.02276.x, PubMed: 19170936

Sreenivasan, K. K., Curtis, C. E., & D'Esposito, M. (2014). Revisiting the role of persistent neural activity during working memory. *Trends in Cognitive Sciences*, *18*, 82–89. https://doi.org/10.1016/j.tics.2013.12.001, PubMed: 24439529

Sullivan, D. R., Marx, B., Chen, M. S., Depue, B. E., Hayes, S. M., & Hayes, J. P. (2019). Behavioral and neural correlates of

memory suppression in PTSD. *Journal of Psychiatric Research*, *112*, 30–37. https://doi.org/10.1016/j.jpsychires.2019.02.015, PubMed: 30844595

Taubenfeld, A., Anderson, M. C., & Levy, D. A. (2019). The impact of retrieval suppression on conceptual implicit memory. *Memory*, *27*, 686–697. https://doi.org/10.1080/09658211.2018.1554079, PubMed: 30522403

Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., et al. (2010). N4ITK: Improved N3 bias correction. *IEEE Transactions on Medical Imaging*, *29*, 1310–1320. https://doi.org/10.1109/TMI.2010.2046908, PubMed: 20378467

Unsworth, N. (2015). Consistency of attentional control as an important cognitive trait: A latent variable analysis. *Intelligence*, *49*, 110–128. https://doi.org/10.1016/j.intell.2015.01.005

Vo, V. A., Sutterer, D. W., Foster, J. J., Sprague, T. C., Awh, E., & Serences, J. T. (2022). Shared representational formats for information maintained in working memory and information retrieved from long-term memory. *Cerebral Cortex*, *32*, 1077–1092. https://doi.org/10.1093/cercor/bhab267, PubMed: 34428283

Wang, T. H., Placek, K., & Lewis-Peacock, J. A. (2019). More is less: Increased processing of unwanted memories facilitates forgetting. *Journal of Neuroscience*, *39*, 3551–3560. https://doi.org/10.1523/JNEUROSCI.2033-18.2019, PubMed: 30858162

Waschke, L., Kloosterman, N. A., Obleser, J., & Garrett, D. D. (2021). Behavior needs neural variability. *Neuron*, *109*, 751–766. https://doi.org/10.1016/j.neuron.2021.01.023, PubMed: 33596406

Wolff, A., Yao, L., Gomez-Pilar, J., Shoaran, M., Jiang, N., & Northoff, G. (2019). Neural variability quenching during decision-making: Neural individuality and its prestimulus complexity. *Neuroimage*, *192*, 1–14. https://doi.org/10.1016/j.neuroimage.2019.02.070, PubMed: 30844503

Zanto, T. P., & Gazzaley, A. (2009). Neural suppression of irrelevant information underlies optimal working memory performance. *Journal of Neuroscience*, *29*, 3059–3066. https://doi.org/10.1523/JNEUROSCI.4621-08.2009, PubMed: 19279242

Zhang, W., & Luck, S. J. (2009). Sudden death and gradual decay in visual working memory. *Psychological Science*, *20*, 423–428. https://doi.org/10.1111/j.1467-9280.2009.02322.x, PubMed: 19320861

Zhou, Y., Mohan, K., & Freedman, D. J. (2022). Abstract encoding of categorical decisions in medial superior temporal and lateral intraparietal cortices. *Journal of Neuroscience*, *42*, 9069–9081. https://doi.org/10.1523/JNEUROSCI.0017-22.2022, PubMed: 36261285