Cerebral Cortex, June 2019;29: 2682–2693

doi: 10.1093/cercor/bhy137
Advance Access Publication Date: 17 May 2018
Original Article

ORIGINAL ARTICLE

Neural Overlap in Item Representations Across Episodes Impairs Context Memory

Ghootae Kim[1], Kenneth A. Norman[2,3] and Nicholas B. Turk-Browne[2,3,4]

[1]Department of Psychology, University of Oregon, Eugene, OR 97403, USA, [2]Department of Psychology, Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA, [3]Princeton Neuroscience Institute, Princeton University, Princeton, NJ 08544, USA and [4]Department of Psychology, Yale University, New Haven, CT 06520, USA

Address correspondence to Dr Ghootae Kim, Lewis Integrative Science Building (Room 348), Eugene, OR 97403, USA. Email: ghootaek@uoregon.edu

Abstract

We frequently encounter the same item in different contexts, and when that happens, memories of earlier encounters can get reactivated. We examined how existing memories are changed as a result of such reactivation. We hypothesized that when an item's initial and subsequent neural representations overlap, this allows the initial item to become associated with novel contextual information, interfering with later retrieval of the initial context. Specifically, we predicted a negative relationship between representational similarity across repeated experiences of an item and subsequent source memory for the initial context. We tested this hypothesis in an fMRI study, in which objects were presented multiple times during different tasks. We measured the similarity of the neural patterns in lateral occipital cortex that were elicited by the first and second presentations of objects, and related this neural overlap score to subsequent source memory. Consistent with our hypothesis, greater item-specific pattern similarity was linked to worse source memory for the initial task. In contrast, greater reactivation of the initial context was associated with better source memory. Our findings suggest that the influence of novel experiences on an existing context memory depends on how reliably a shared component (i.e., item) is represented across these episodes.

Key words: item representation, multivariate pattern analysis, neural overlap, source memory

Introduction

Our experience is highly repetitive, with the same objects appearing repeatedly over time and often in different contexts. For example, we might move a piece of furniture to many different apartments, see somebody from work at the grocery store, or look for our car in various parking lots. How does experiencing a familiar item in a novel context affect pre-existing memories of the item and its prior contexts?

It has long been known that memory for the initial context in which an item was experienced can be impaired by a later encounter of the item in a new context. Such retroactive interference has been widely investigated using the AB/AC paradigm (McGovern 1964; Postman and Underwood 1973; Richter et al. 2016). In this paradigm, participants learn an episode with components A and B, then another episode with components A and C. As in a previous example, we might encounter a colleague at the grocery store (item A in context C), with whom we previously chatted in the office (item A in context B). Because of the shared component A, learning AC can trigger retrieval of the previously learned AB memory. How does this memory reinstatement relate to retroactive interference (i.e., forgetting of B and/or AB)? One prominent account is that reactivation of a prior context B during later AC learning builds resistance to interference, leading to better subsequent retrieval of the initial context B when cued with A (Koen and Rugg 2016; Kuhl et al. 2010).

Here we investigate a different, though not mutually exclusive, account of how memory reinstatement relates to retroactive interference. We focus on the fact that mental representations of an item can differ over time even when we putatively experience the "same" item. In the AB/AC paradigm, for example, this would correspond to variance in the extent to which the representation of A during AC learning is the same as the representation of A during the prior AB learning. Although neural overlap across repeated presentations of an item can be associated with better memory for that item (Ward et al. 2013; Xue et al. 2010), it is unknown how this overlap affects memory for previously formed item-context associations.

We hypothesize that retroactive interference occurs when the same item representation is reinstated across episodes with different contexts. Specifically, reinstatement of the item representation engaged by the initial processing of A (from AB learning) during AC learning allows these reinstated item features to become associated with the novel context (C), which interferes with later retrieval of the initial context B (note that we are using "item reinstatement" descriptively, to refer to the degree to which the same item features are activated by the initial and subsequent presentation of the item; this overlap could be due to retrieval of features from memory or bottom-up perception). In contrast, if the representation of A during AC learning differs from that of the prior AB episode, memory of the initial context B might be less affected by retroactive interference. In short, we predict a negative relationship between item-specific representational overlap and subsequent source memory for the initial context.

This hypothesized negative relationship stands in contrast to previous findings of a positive relationship between context reinstatement and subsequent source memory (Kuhl et al. 2010; Koen and Rugg 2016). Importantly, those studies measured neural reinstatement of contextual features, tested subsequent memory for those same contextual features, and found that context memory is strengthened by its reactivation. In contrast, for our study, we set out to measure reinstatement of item features across contexts. In keeping with prior theoretical and empirical work (Hupbach et al. 2007; Gershman et al. 2013; Sederberg et al. 2011; St. Jacques et al. 2013), we hypothesized that activating the representation of a previously seen item in a new context "opens a window" where the existing item-context memory can be modified or over-written; we further hypothesized that this effect would be modulated by degree of overlap in the item representation across contexts (with higher overlap yielding more interference).

To test this hypothesis, we presented objects (A) sequentially during 2 different orienting tasks (B and C). These tasks served as the contexts to which the items could be bound (Johnson et al. 1997). Using fMRI, we measured pattern similarity for a given item across the 2 task contexts in the lateral occipital cortex (LOC), which is thought to represent the visual features of objects (Grill-Spector et al. 2001). We then related these item-wise pattern similarity scores to subsequent source memory for the initial task. In addition to testing for the hypothesized negative effect of item reactivation, we also tested for the positive effect of context reactivation observed in previous studies (Koen and Rugg 2016; Kuhl et al. 2010). Such a

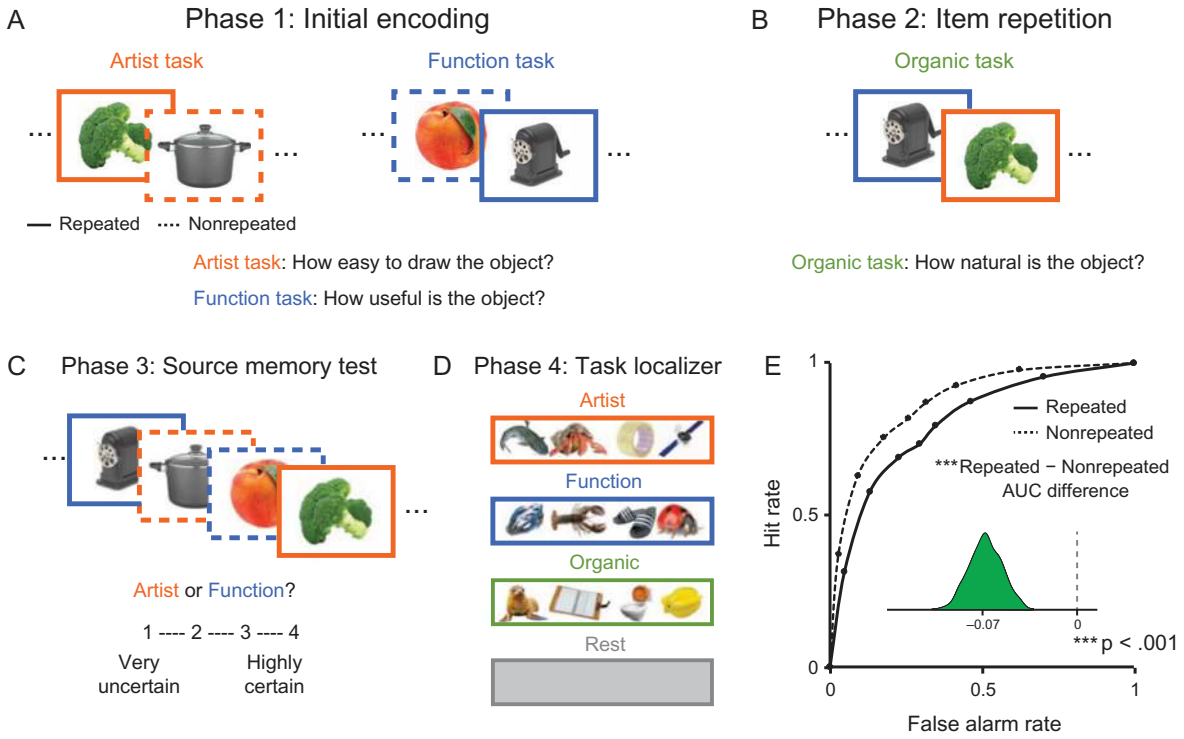

Figure 1. Experimental design and behavioral results. (*A*) During initial encoding, object images were randomly assigned to 1 of the 2 orienting tasks (artist and function tasks). (*B*) In the item repetition phase, half of objects from the first phase were presented again while participants performed a third (organic) task. (*C*) In the subsequent source memory test, judgments were collected about which task had been performed first on each object, both for objects presented twice (repeated condition) and objects presented once (nonrepeated condition). (*D*) In the task localizer, a new set of objects was presented in each of the 3 tasks to define task-specific neural activity patterns. (*E*) The area under the curve (AUC) of source memory judgments was calculated; lower AUC indicates worse memory, and so memory for the first task was worse in the repeated versus nonrepeated condition. The inset plot depicts the sampling distribution of the repeated minus nonrepeated AUC difference from random-effects bootstrap resampling of participants. Almost all resampled AUC differences were below zero (green area), indicating a reliable retroactive interference effect. ***$P < 0.001$. Note that the colored rectangles were presented here for visualization.

dissociation would provide strong evidence that item and context reactivation have differential effects on source memory. Consistent with our main hypothesis, we found a negative relationship between item reactivation and subsequent source memory: greater item-wise pattern similarity was associated with worse source memory for the initial task. In contrast, we observed that context reactivation has an opposite effect on source memory: greater reactivation of the initial task led to better source memory for the task.

Materials and Methods

Overview

This study consisted of 4 phases. In an initial encoding phase (phase 1), participants were exposed to a sequence of object images and performed 1 of 2 orienting tasks (artist or function task). These tasks served as the initial context to which the object items could be bound during the encoding phase. In the item repetition phase (Fig. 1*B*), half of the objects from each task in the initial encoding phase (i.e., 48 objects for each of artist and function) were presented again, and participants determined how organic the object was on a 4-point scale: 1 = very artificial, 2 = artificial, 3 = natural, 4 = very natural. We used this organic judgment as opposed to the more common living/nonliving distinction because not all natural things are living (e.g., a log). We measured how reliably the initial representations of the items were reinstated in this phase by calculating item-specific pattern similarity between the initial and repeated presentations. In the memory test phase (phase 3), source memory for the initial task (i.e., artist or function) was measured and related to the item-specific pattern similarity scores calculated in the second phase. A final task localizer phase (phase 4) was used to train the task classifier, and to generate template neural patterns for each task, which were used to regress out task-related information in measuring item-wise pattern similarity.

Participants

Overall, 31 adults (16 women, all right-handed, mean age 21.65 years) participated for monetary compensation. All participants had normal or corrected-to-normal vision and provided informed consent. The Princeton University IRB approved the study protocol.

Stimuli

Participants were shown color photographs of natural and manmade real-world objects. Stimuli were displayed on a projection screen behind the scanner bore, viewed with a mirror on the head coil (subtending 8.8 × 8.8). Participants fixated a central dot that remained onscreen throughout.

Procedure

Participants completed one scanning session with 4 phases: initial encoding, item repetition, source memory test, and task localizer. During the initial encoding phase (Fig. 1A), participants viewed a series of objects that were randomly assigned to 1 of 2 orienting tasks: How easy would it be to draw the object? (artist task) or How useful is the object? (function task). Participants responded on a 4-point scale (artist/function): 1 = very easy/very useless, 2 = easy/useless, 3 = hard/useful, 4 = very hard/very useful. We used these tasks because previous studies have shown that they are highly decodable with fMRI (Johnson et al. 2009; McDuff et al. 2009; Koen and Rugg 2016). Four runs of encoding were collected, and each run contained 2 blocks of objects from each of the 2 tasks (the order of the 4 blocks was randomized). The task was instructed with a cue at the beginning of the block (e.g., "Artist task"). Each object stimulus was presented for 1 s, followed by a blank interval of 2 s. There were 12 trials per block (36 s duration), followed by 15 s of rest. The total duration of each run was 3 min 42 s.

In the item repetition phase (Fig. 1*B*), half of the objects from each task in the initial encoding phase (i.e., 48 objects for each of artist and function) were presented again, and participants determined how organic the object was on a 4-point scale: 1 = very artificial, 2 = artificial, 3 = natural, 4 = very natural. Each of the 96 objects was presented for 1 s, followed by a blank interval of 3.5 s: We used a longer SOA in this phase with the goal of providing more time for item and task reactivation that could impact source memory. All stimuli were presented in a single run without a rest period, lasting 7 min 33 s.

The source memory test (Fig. 1*C*) came as a surprise to participants. It contained the 96 objects that had been presented in both the encoding and repetition phases (repeated condition) and the 96 objects shown only in the encoding phase (nonrepeated condition). On each trial, one object was presented on the screen and participant's memory was measured in a 2-step procedure: First, a choice option was shown below the object (i.e., "Artist or Function?") and participants were instructed to specify which task had been performed on the object during the initial encoding phase. Second, immediately after the task response, a 4-point confidence scale (1 = very unsure, 2 = unsure, 3 = sure, and 4 = very sure) was presented below the object, and participants reported their confidence level. Each object was presented for 6 s, though participants were encouraged to respond within 5 s. If they failed to respond on a given trial, the object was omitted from later analyses (around 3% of total trials). We did not measure item recognition in the memory test: That is, all objects were old, and we did not ask participants to report whether an object was old or new. In a behavioral pilot, we included novel lure object images and measured item recognition, but found that incorrect recognition rates were negligible (false alarms = 6.4%; misses = 7.3%). We thus excluded the recognition memory test from the full fMRI study.

After the memory test, participants completed 3 runs of a functional localizer (Fig. 1*D*), in which new object images were presented in 1 of the 3 tasks: artist, function, and organic. Each run contained 6 blocks, with 2 blocks from each of the 3 tasks in a random order. Each object was presented for 1 s, followed by a blank interval of 2 s. There were 12 trials per block (36 s duration). Each block was followed by 15 s of fixation, which was treated as a baseline "rest" category. Total run duration was 5 min 24 s.

Behavioral Analysis

We measured memory performance by dividing responses from the source memory test into 8 levels of confidence: 4 = very sure "artist" to −3: very sure "function". These judgments were quantified using receiver operating characteristic (ROC) analyses (Green and Swets 1966; Macmillan and Creelman 2005). For each of the repeated and nonrepeated conditions, we created an ROC curve across the 8 confidence levels and calculated the area under the curve (AUC). Calculating these curves precisely requires a substantial amount of data, and thus we pooled trials across participants beforehand. We assessed the reliability of the AUC difference between conditions across participants using a bootstrapping approach in which entire

participants were resampled with replacement 1000 times (Efron 1979), and AUC was computed (for each resampling) based on the trials pooled across all resampled participants. This provided a population-level confidence interval (CI) for each effect, and also allowed for null hypothesis testing based on the proportion of bootstrapped samples in which the effect was reversed.

Data Acquisition

Experiments were run with the Psychophysics Toolbox (http://psychtoolbox.org). Neuroimaging data were acquired using a 3 T MRI scanner (Siemens Skyra) with a 16-channel head coil. A scout anatomical scan was used to align axial functional slices. Functional images covering the whole brain were acquired with a T2* gradient-echo EPI sequence (TR = 1.5 s; TE = 28 ms; flip = 64; iPAT = 2; matrix = 64 × 64; slices = 26; thickness = 4 mm, resolution = 3 × 3 mm^2). High-resolution (MPRAGE) and coplanar (FLASH) T1 anatomical scans were acquired for registration, along with field maps to correct B0 inhomogeneities.

Preprocessing

fMRI data were preprocessed with FSL (http://fsl.fmrib.ox.ac.uk). Functional scans were corrected for slice-acquisition time and head motion, high-pass filtered (128 s period cut-off), spatially normalized (5 mm FWHM), and aligned to the middle volume.

Selection of ROIs

We defined ROIs for object processing (LOC) and for task processing. LOC was defined anatomically from the Harvard-Oxford cortical atlas in FSL and transformed into each participant's space. We defined the task ROI on an individual-subject basis in 3 steps: 1) We picked voxels selective to each of the 3 tasks (artist, function, and organic) by performing a general linear model (GLM) analysis of the localizer, with regressors for each task and rest. We ran 3 contrasts (artist vs. others, function vs. others, and organic vs. others) and selected voxels whose absolute z-values were above 2.3 ($P < 0.01$). 2) We then took the union of the surviving voxels of each contrast. 3) The resulting image was masked to include gray matter and hippocampus based on the Harvard-Oxford subcortical atlas, and overlapping regions with LOC were excluded to minimize any potential confounding effects of item reactivation. We provide more detailed information about the task ROI in the Supplementary Materials.

Measuring Item Reactivation

We measured how reliably the initial representation of each item was reinstated in the second phase by calculating the Pearson correlation of the patterns of activity elicited in the LOC on the initial and repeated presentations, 4.5 s after stimulus onset. We did not estimate univariate activation for single trials using a trial-wise GLM. To maximize the number of items that could be presented in limited scanning time, we used relatively short SOAs (3 s for phase 1 and the localizer, and 4.5 s for phase 2), knowing that such a design would impede GLM analyses. Indeed, a separate pilot fMRI study with an SOA of 4 s that compared item specificity for scene images (i.e., same-item > different-item pattern similarity) on time-shifted, preprocessed raw data versus trial-wise parameter estimates of univariate activation from a GLM found reliably great item specificity with the former approach. This mirrors the reliable item specificity we observed in the present study based on preprocessed raw data, and is consistent with our previous studies (Kim et al. 2014, 2017; see Results).

A side effect of using short SOAs is that item-specific patterns from phase 1 will contain fading traces of information from preceding trials (Chan et al. 2017), and pattern similarity between phases 1 and 2 might be influenced by retrieval of this preceding-trial information. Crucially, even if this happens, we do not think that preceding-trial retrieval could artifactually give rise to the predicted negative relationship between phases 1 and 2 pattern similarity and subsequent source memory: Because items were blocked by task, successive items were studied using the same task, thus, if anything, retrieval of the preceding item should help source memory, not hurt it.

Measuring Task Information

We measured the amount of task information in the first phase (to index task encoding) and in the second phase (to index task reactivation) using multivariate classification, which was conducted with the Princeton Multi-Voxel Analysis Toolbox (www.pni.princeton.edu/mvpa) using penalized logistic regression with L2-norm regularization (penalty = 1). To classify the initial encoding and item repetition phases, we trained the classifiers on the localizer runs and tested on each of the phases. First, we trained a separate model for each of 4 categories (i.e., artist vs. others, function vs. others, organic vs. others, and rest vs. others) on the localizer runs and tested on each of the initial encoding and item repetition phases. For each fMRI volume in the test set, each classifier estimated the extent to which the activity pattern matched the activity patterns for the 2 categories (e.g., artist vs. others) on which it was trained (from 0 to 1). We refer to these category-level pattern match value as "classifier evidence." We operationalized task (re)activation as difference in the classifier evidence of each item's initial task versus the other one. For example, for an item whose initial task was artist, we subtracted function evidence from artist evidence and vice versa.

To validate our approach, we first performed 2 kinds of classification analyses: 1) within the localizer (using cross-validation) and 2) across the localizer and initial encoding phases. For the within-localizer cross-validation, we trained the 4 classifiers (artist, function, organic, and rest vs. others) using 2 of the localizer runs and tested them on the remaining run (and then swapped training and test runs), and measured classification accuracy (chance level = 0.25). Specifically, we counted a trial as hit when classifier evidence of a target task was greater than the others (e.g., for an item whose orienting task was artist, we coded it as hit when artist evidence is greater than the other categories). We then applied the same procedures for across-phase classification: The 4 classifiers trained on 3 runs of the localizer were tested on the initial encoding phase, and classification accuracy was measured (chance level = 0.25). In principle, we can also measure task (re)activation by training classifiers on phase 1; we provided the results of this analysis in the Supplementary Materials.

Relating Neural Measures to Source Memory

We measured a linear relationship between each of the neural measures (item reactivation, task encoding, and task reactivation) and subsequent source memory. First, we measured memory strength by dividing the source memory responses into 8 levels of confidence: 4 = very sure source correct (e.g., artist response to artist task items) to −3: very sure source

incorrect (e.g., function response to artist task items). We then examined the relationship between each of the neural measures and memory strength using the linear regression. The resulting beta coefficient represents the direction of the relationship between the neural measure and memory: positive sign = positive relationship (i.e., better memory when a neural measure is higher) and negative sign = negative relationship (i.e., worse memory when a neural measure is higher). We performed this analysis by pooling trials across participants to measure a reliable relationship, and we assessed the population-level reliability of the result using a bootstrap procedure where we resampled participants with replacement. When running this analysis, we standardized each neural measure within each participant before pooling across participants, to ensure that any relationship we observe between a neural measure and subsequent memory reflects within-participant variance as opposed to across-participant variance.

To examine the possibility that relationships between neural measures and memory depended on the initial orienting task, we divided trials by the initial tasks and measured the relationships for each of the tasks separately. For both item-wise pattern similarity (based on residuals after regressing out task templates) and task reactivation, there was no significant interaction with initial task (item reactivation: difference in β = −0.07, CI = [−0.21, 0.07], bootstrap P = 0.320; task reactivation: difference in β = −0.05, CI = [−0.22, 0.13], bootstrap P = 0.594). Thus, henceforth we report results collapsed across task.

Simulations of Approaches for Measuring Item-Specific Reactivation

Our main hypothesis concerns the relationship between the reactivation of an item's representation and subsequent source memory. To test this, we needed a way of measuring reactivation of item features. Our basic approach (as described above) was to compute pattern similarity in LOC for the 2 presentations of each stimulus in the initial encoding (phase 1) and item repetition (phase 2) phases, respectively, and then to correlate this measure with source memory for the task from initial encoding (e.g., phase 1). However, this analysis is complicated by the fact that LOC pattern similarity is potentially influenced by 2 factors—reactivation of item features and reactivation of generic artist or function task features—which could influence memory in different (potentially opposing) ways. That is, reactivation of the item representation could lead to retroactive interference (for reasons described in the Introduction) whereas reactivation of task features could boost subsequent source memory (as in Koen and Rugg 2016; Kuhl et al. 2010). For this reason, it was essential to use an analysis procedure that could separate the effects of item versus task reactivation.

To compare different procedures, some that we devised and others used in the literature (Kim et al. 2014; Koen and Rugg 2016), we ran simulations of 2 situations: 1) where subsequent memory was affected by item reactivation but not task reactivation, and 2) where subsequent memory was affected by task reactivation but not item reactivation. Our goal was to find an analysis procedure that would report a positive result only in the first situation (i.e., item reactivation is driving subsequent memory). In other words, through these simulations, we aimed to make sure that our measure of item reactivation truly reflects "item-specific" information: That is, it would then provide a positive result in the case where item reactivation affects memory, but not in the case where task reactivation (but not item reactivation) influences memory. Therefore, any procedure meeting these criteria (a positive result when item reactivation drives memory, and a null result if task reactivation influences memory) would be valid for our purpose. In this sense, we would like to note that our simulations were a form of control analysis, and were not intended to fully model the joint contributions of item and task reactivation or their interaction. The simulation results are summarized here, and the simulation methods are described in fuller detail in the Supplementary Material.

Approach 1: Same-Item Minus Different-Item Pattern Similarity

Similar to our study, Koen and Rugg (2016) investigated the effects of item-specific pattern reactivation on source memory. To measure item-specific reactivation (while excluding task reactivation), they computed pattern similarity for pairs of the same item (e.g., A-first-presentation to A-second-presentation) and subtracted out the average pattern similarity of that item with other items encoded with the task (e.g., A-first-presentation to B-second-presentation, B-first-presentation to A-second-presentation, where A and B were encoded in the same task). They then related this difference measure to source memory.

Intuitively, one might think that comparing A to other items (B, C) encoded with the same task would control for task reactivation. However, in our simulations, we found that this method yielded significant results both when memory was driven by item reactivation and when memory was driven by task reactivation but not item reactivation (Ps < 0.001). This result can be explained by the fact that same minus different pattern similarity controls for the average level of task reactivation, but it does not fully control for trial-by-trial variability in task reactivation (see Supplementary Materials). Given that this analysis does not decisively discriminate item and task reactivation effects on memory in our simulations, we explored other approaches.

Approach 2: Permutation Analysis

We previously used permutation analysis to track reactivation of item features (Kim et al. 2014). This involves scrambling the pairings of items (across phases 1 and 2) 1 000 times and, for each scramble, recalculating pattern similarity and its relationship to memory. A z-score of the original effect (based on the intact pairings) with respect to this null distribution can be calculated.

This analysis produced different results across the 2 simulations (situations A and B). When item reactivation drives memory, the original effect was reliably greater than the permuted effects (P < 0.001). In contrast, when task reactivation drives memory, the original effect was not different from the permuted effects (P = 0.53). This pattern of results confirms that the permutation analysis can identify relationships between item feature reactivation and memory and is not misled by trial-by-trial variance in task reactivation.

Approach 3: Regression + Permutation Analysis

Although the permutation analysis handles the case where task reactivation in the second phase varies across items, it can be misled if task activation at encoding (for a given item) is correlated with task reactivation for that item at retrieval. That is, if a task is highly active during encoding, it might be reactivated more during the item repetition, and permuting the item pairings will eliminate this, giving the appearance of item-specific information. Indeed, when we simulated this situation (i.e., where reactivation of task features but not item features is correlated with subsequent memory, and task activation at

encoding for a given item is correlated with task reactivation), the permutation test yielded a significant result ($P < 0.001$). The fact that the permutation test can show a significant result when (in the simulated data) there is no actual relationship between item feature reactivation and subsequent memory indicates that this test is also unsuitable for our purposes.

To address this issue, we adopted a different approach of regressing out the localizer template activity patterns for each task from every item's representation prior to calculating pattern similarity and its relationship to source memory. After removing task information in this manner on a trial-by-trial basis, our simulation where item-specific reactivation drives memory survives the permutation test ($P < 0.001$), but both forms of the simulation where task reactivation drives memory (i.e., where there is item-wise variance in task reactivation alone, and where there is also correlated item-wise variance in task encoding and task reactivation) both fail for the first time ($Ps > 0.37$). So far in this simulation, we have regressed out the task template determined a priori, whereas in our experiment, the task templates were estimated by averaging activity patterns from the artist and function tasks (respectively) in the localizer. To examine the impact of this difference, we re-ran the same regression + permutation simulation based on estimated task templates, which were acquired by averaging the noisy patterns from phase 1 and 2. This new approach provided the same qualitative results as above: a positive result only for the case where item-specific reactivation drives memory ($P < 0.001$), and null results for both cases where task reactivation influences memory ($Ps > 0.55$).

Having confirmed the validity of regression + permutation analysis, we applied it to our data in following steps: First, based on the localizer, we defined an activation template for each of initial tasks (i.e., artist and function) over LOC voxels by averaging patterns from the corresponding task. Second, we regressed out the template for each task from the corresponding patterns in the first and second phases (e.g., for an item whose initial task was artist, we regressed out the artist template from both patterns for that item). We then measured item-wise pattern similarity in the residual patterns, and related it to source memory by measuring a linear relationship between the two. Third, we performed a permutation analysis by scrambling the original pairings of first and second phase trials within each task and participant (e.g., permuting items whose initial task was artist). For each of 1000 permutations, we recalculated pattern similarity and related it to memory. Finally, a z-score of the original, task-residualized relationship (beta coefficient score) based on the intact pairings was calculated with respect to the null distribution of beta coefficients.

Searchlight Analysis

We measured the relationship between item-wise pattern similarity and memory based on the assumption that the item-wise pattern similarity measure contains item-specific information, and we focused on the LOC given that this region has been shown to represent item-specific information (Cichy et al. 2011; Eger et al. 2008). However, it is possible that other brain regions might show the same relationship. Thus, we performed an exploratory searchlight analysis.

Each participant's EPI volume was aligned to standard space, and we swept a 14-mm-voxel cubic searchlight (radius = 3 voxels) throughout the EPI volume. In each searchlight, we first computed an item-specificity score using a 2-step permutation analysis (see Simulations of Approaches for Measuring Item-Specific Reactivation): We regressed out generic task information from patterns in the first and second phases, and recalculated the same-item pattern similarity. We then permuted the original item pairings across the first and second phases and remeasured the item-wise pattern similarity. Finally, we tested whether the same-item pattern similarity from intact pairings is greater than the item-wise pattern similarity scores of permuted pairings. Second, we measured a relationship between item-wise pattern similarity and source memory based on the 2-step permutation analysis. We assigned the final 2 outcomes (i.e., an item-specificity score and relationship between item-wise pattern similarity and source memory) to the center voxel of each searchlight. The 2 resulting maps were masked to exclude white matter. We then examined the reliability of each of the 2 analyses across participants by applying a bootstrap test for every voxel. Finally, we selected voxels with $Ps < 0.001$ (uncorrected) in both analyses.

Results

Subsequent Source Memory Behavior

Overall, participants successfully discriminated the correct and incorrect initial task (mean $A' = 0.56$, bootstrap $P < 0.001$). Consistent with the idea that item repetition increases retroactive interference, source memory for the initial task was lower for repeated compared with nonrepeated items (AUC difference = –0.07, CI = [–0.09, –0.04], bootstrap $P < 0.001$; Fig. 1E).

Initial Pattern Similarity Results

We first confirmed that LOC activation patterns contained information about items (Cichy et al. 2011; Eger et al. 2008), by showing that pattern similarity between the first and second phases was greater when the item was the same versus when 2 different items from the same task were compared (bootstrap $P < 0.001$). To address the possibility that this pattern might reflect item-wise variance in task activation rather than item-specific information, we performed a 2-step permutation analysis (see Materials and Methods). First, we regressed out generic task information from patterns in the first and second phases, and recalculated the same-item pattern similarity. Second, we permuted the original item pairings across the first and second phases and remeasured the item-wise pattern similarity. Finally, we tested whether the same-item pattern similarity from intact pairings is greater than the item-wise pattern similarity scores of permuted pairings. Indeed, the original same-item pattern similarity was reliably more positive than a null distribution of permuted pattern similarity scores (average z-score = 5.26, CI = [1.65, 9.56], bootstrap $P = 0.002$ Fig. 2A).

Task Classifier Results

Within-localizer cross-validation: We examined the validity of the task classification by checking the within-localizer cross-validation accuracy. Overall classification accuracy was reliably greater than the chance level of 0.25 (accuracy = 0.47, bootstrap $P < 0.001$). We then measured cross-validation accuracy for each of 3 tasks (Table 1), and all of the accuracy measures were well above chance (bootstrap $P < 0.001$). A one-way ANOVA showed a significant main effect of task ($F_{2,60} = 16.89$, $P < 0.001$). Bootstrap tests revealed that accuracy of the organic task was significantly lower than the other tasks (bootstrap $Ps < 0.001$),

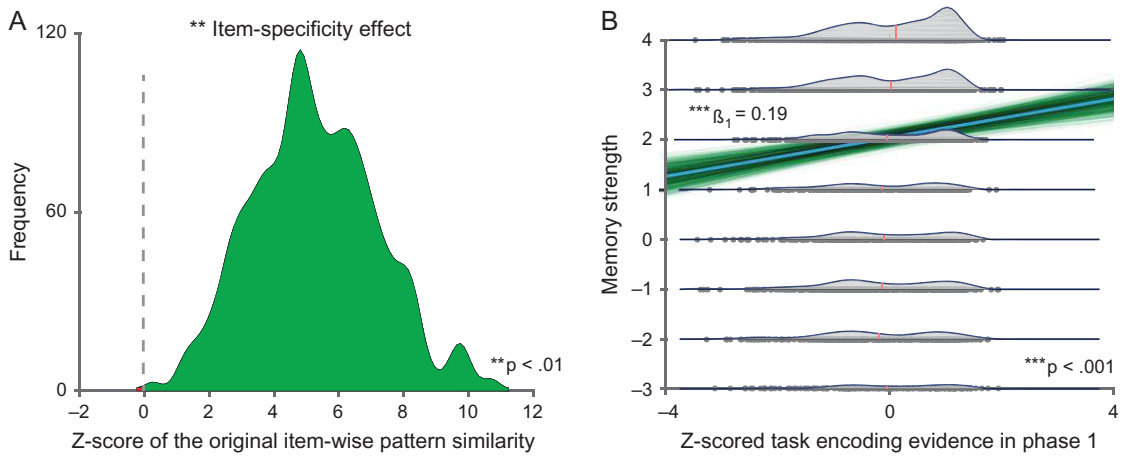

Figure 2. Manipulation checks. (*A*) Item-wise pattern similarity in LOC based on intact pairings was higher than pattern similarity scores based on permuted pairings. (*B*) Greater task encoding strength in phase 1 was associated with better task memory on the final test, as reflected in a positive beta coefficient. The thick blue regression line is from the original data. Thin translucent green regression lines are from the 95% confidence interval of 1000 bootstrap results. Each line is overlaid with others, and darkness depicts the density. Dark gray dots represent memory responses. Light gray histograms represent a distribution of encoding evidence for each level of memory strength, and pink lines mark the mean value of each of encoding evidence. $^{**}P < .01$, $^{***}P < 0.001$.

Table 1 Task classification accuracy

Artist	Function	Organic
	Within-localizer cross-validation	
0.52 (0.49, 0.56)	0.48 (0.44, 0.52)	0.39 (0.36, 0.43)
	Across-phase classification	
0.67 (0.63, 0.70)	0.67 (0.63, 0.70)	

Chance performance was 0.25. 95% confidence intervals are provided in parentheses.

and accuracy of artist task was numerically higher compared with function task (bootstrap $P = 0.164$).

Across-phase classification: We also checked accuracy of across-phase classification by training on the full localizer dataset and testing on the initial encoding phase. Overall classification accuracy was reliably greater than the chance level of 0.25 (accuracy = 0.67, bootstrap $P < 0.001$). There was no significant difference between artist and function tasks (bootstrap $P = 0.762$; Table 1).

Task Encoding Results

Several previous studies (Gordon et al. 2014; Kuhl et al. 2012; Kim et al. 2014; Koen and Rugg 2016) have reported positive relationships between multivariate measures of encoding strength and subsequent memory. Consistent with this, greater classifier encoding strength for the corresponding task (e.g., artist minus function classifier evidence) during phase 1 was associated with better task memory on the final test ($\beta = 0.19$, CI = [0.11, 0.27], bootstrap $P < 0.001$; Fig. 2*B*). We tested whether the positive relationship varied as a function of repetition condition. There was no significant difference in the relationships for repeated versus nonrepeated items (difference in beta coefficient = 0.06, CI = [−0.04, 0.16], bootstrap $P = 0.232$).

Relationship Between Item Reactivation and Subsequent Source Memory

We hypothesized that reactivation of the item representation from the first phase during the second phase would lead to retroactive interference with the source memory for the initial task. Consistent with our hypothesis, higher pattern similarity between the first and second phases was associated with worse source memory ($\beta = -0.08$, CI = [−0.15, −0.01], bootstrap $P = 0.006$; Fig. 3C).

This observed negative relationship might be driven by item-wise variance in task reactivation rather than item reactivation. We controlled for this confound using a 2-step permutation analysis (see Materials and Methods): First, we regressed out generic task information from patterns in the first and second phases (Fig. 4A), and recalculated pattern similarity and its relationship to source memory. Second, we scrambled the original pairings of items across the first and second phases and recalculated the relationship for each permutation (Fig. 4C). If item reactivation modulates subsequent source memory, as hypothesized, then the relationship based on residual intact pairings (excluding task information) should be more negative than the null distribution of relationships from permuted pairings (excluding both task and item information).

After the first step of regressing out task information, the relationship for intact pairings remained negative ($\beta = -0.08$, CI = [−0.16, −0.01], bootstrap $P = 0.028$; Fig. 4*B*). In addition, this relationship tended to be more negative than the null distribution of permuted relationships after the second step (average z-score = −1.07, CI = [−2.28, 0.02], bootstrap $P = 0.054$; Fig. 4*D*). The finding that the negative relationship remained robust even after the conservative process of excluding potential task effects, combined with an opposite effect of task reactivation on source memory (see following sections), strongly supports our argument that the negative relationship was driven by item reinstatement rather than task reactivation.

Searchlight Results

The above analyses were performed within the LOC given that this region has been shown to represent item-specific information (Cichy et al. 2011; Eger et al. 2008). However, item-wise pattern similarity in other brain regions might influence subsequent source memory as well. To examine this possibility, we swept a cubic searchlight through a whole brain. For each searchlight, we performed the 2 main analyses. First, given the assumption

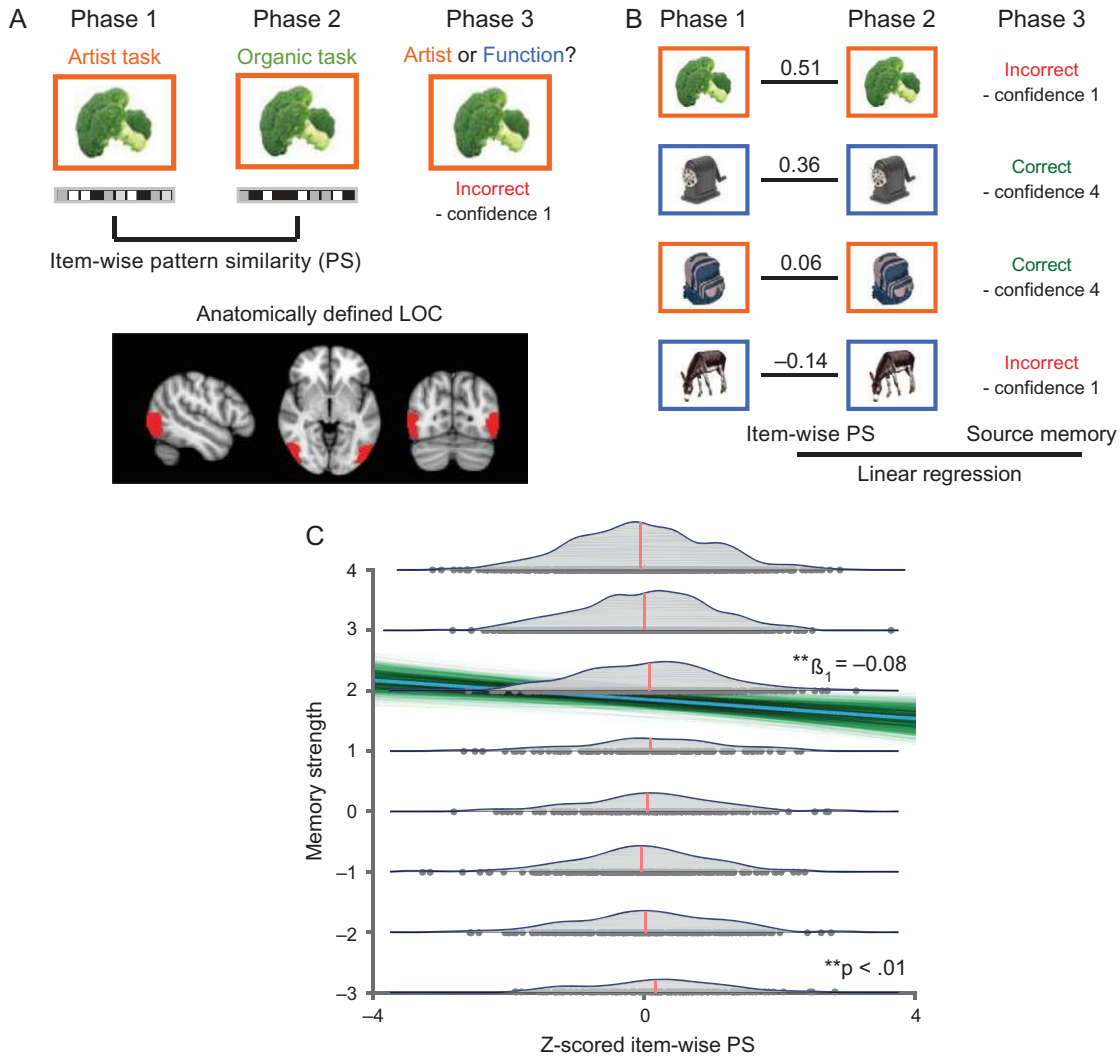

Figure 3. Relating pattern similarity to source memory across items. (A) Pattern similarity was measured for each item between the first and second phases. (B) A trial-by-trial relationship between item-wise pattern similarity and subsequent source memory was measured using the linear regression. (C) There was a reliable negative relationship between item-wise pattern similarity and source memory across items. **P < .01.

that item-wise pattern similarity contains item-specific information, we computed item-specificity score based on the regression + permutation procedure. That is, we examined whether item-wise pattern similarity of the original item pairings across the first and second phases is greater than those of permuted pairings. Second, we measured the trial-by-trial relationship between item-wise pattern similarity and source memory based on the regression + permutation approach. Note that we performed both analyses in a 2-tailed manner: That is, the analyses could identify both positive and negative effects (e.g., for the item-specificity test, an effect is positive when pattern similarity of the original parings is greater than those of permuted parings, or negative if the pattern is opposite). Several clusters in the bilateral LOC (right LOC: 50 voxels and 7 voxels, left LOC: 46 voxels), and right occipital fusiform gyrus (2 voxels) survived statistical tests of both analyses (bootstrap *Ps* < .001 uncorrected). Replicating the main findings above, we found both a positive main effect of item specificity and a negative relationship between item-wise pattern similarity and source memory in all of the clusters (Fig. 5). No clusters showed the other possible combinations of effects (i.e., positive item specificity and positive relationship, negative item specificity and positive relationship, and negative item specificity and negative relationship).

Relationship Between Task Reactivation and Subsequent Source Memory

In addition to examining the effect of item reinstatement on initial source memory, we also considered the effect of task reactivation. We measured the amount of information about the initial task when each item was repeated in the second phase by calculating classifier evidence of the initial task compared with that of the other task (e.g., artist minus function task classifier evidence) using multivariate classification trained on the localizer (Fig. 6A,B). In contrast to the negative effect of item reinstatement, greater task reactivation was associated with better subsequent source memory (β = 0.11, CI = [0.01, 0.20], bootstrap P = 0.028; Fig. 6C). This positive effect of task reactivation on source memory was reliably greater than the negative effect of item reactivation (difference in β = 0.19, CI = [0.07, 0.29], bootstrap P = 0.002).

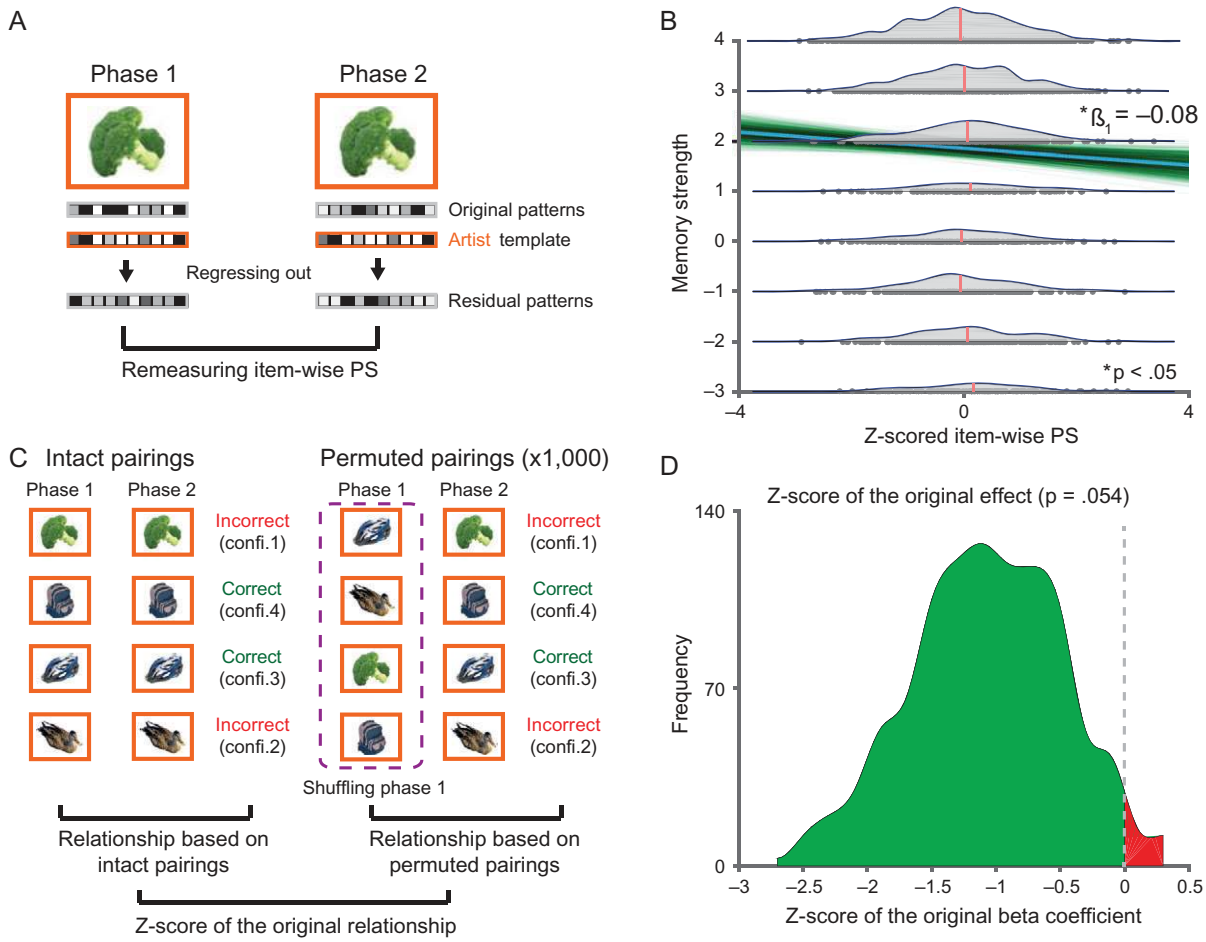

Figure 4. Testing specificity of item reinstatement effects. (*A*) Generic task information was regressed out of the item patterns from the first and second phase using the task templates from the localizer. (*B*) Using the residual patterns, the relationship between item-wise pattern similarity and source memory was recalculated and remained reliably negative. (*C*) We further narrowed in on item-specific variance by submitting the residual item patterns to a permutation test. (*D*) The z-score of the original negative relationship tended to be more negative than the null relationships calculated after permuting the item pairings between the first and second phases. *$P < .05$.

Ruling Out Alternative Accounts

Univariate confounds: We have assumed that item-wise pattern similarity reflects neural overlap of an item representation across the first and second phases, and that this neural overlap affects subsequent retrieval of the initial task. In principle, however, univariate activation in the second phase might have affected both pattern similarity and subsequent source memory. For example, imagine that a participant was in an inattentive state for some of the trials in the second phase. Lower activation for those trials (vs. more attentive trials) could reduce item-wise pattern similarity across the first and second phases (Coutanche 2013; Davis and Poldrack 2013; Davis et al. 2014; Aly and Turk-Browne 2016). Furthermore, subsequent source memory for the initial task on those inattentive trials might be better, because there was less learning of second-phase information, leading to less retroactive interference. In short, if univariate activation in the second phase is a factor affecting both item-wise pattern similarity and source memory, the observed negative relationship between the 2 would be spurious.

To address this issue, we ran an analysis to directly examine the role of univariate activation. We computed the relationship between univariate LOC activation in the first and second phases and subsequent source memory using similar procedures as for pattern similarity. Specifically, for each phase, we measured average activity across LOC voxels on every trial (without regressing out task patterns) and calculated its linear relationship with subsequent source memory for the initial task. Univariate activation was not related to source memory in either phase (phase 1: $\beta = 0.04$, CI = [–0.05, 0.12], bootstrap $P = 0.388$; phase 2: $\beta = -0.04$, CI = [–0.12, 0.04], bootstrap $P = 0.296$). The difference in univariate activation between the 2 phases (phase 1 minus phase 2) also failed to predict source memory ($\beta = 0.06$, CI = [–0.03, 0.15], bootstrap $P = 0.184$). If the observed negative relationship between item-wise pattern similarity and source memory was driven by univariate activation, this relationship should have been more robust than for pattern similarity; to the contrary, the effect was not significant and in fact in the wrong direction numerically. Thus, these findings are consistent with our interpretation that the negative relationship between item-wise pattern similarity and subsequent source memory reflects neural overlap of items across the 2 phases.

Task encoding confounds: Even though the negative relationship between item-wise pattern similarity and source memory remained significant after task-related information was carefully removed from patterns using the regression + permutation

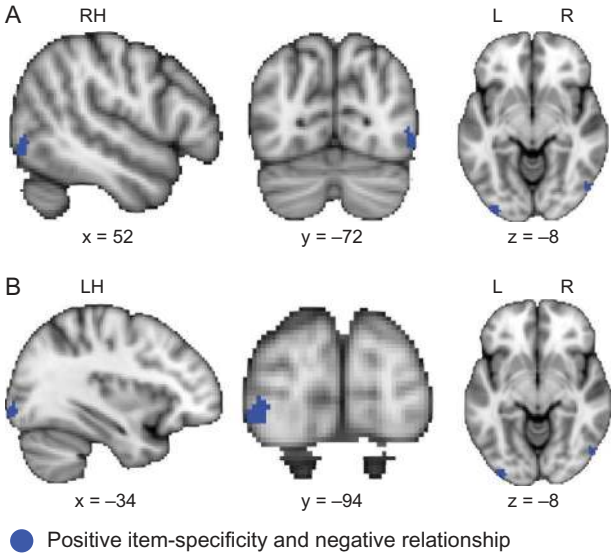

Figure 5. Exploratory searchlight results. (*A*) Right LOC (50 voxels) and (*B*) left LOC (46 voxels) showed a positive main effect of item specificity and a negative relationship between item-wise pattern similarity and source memory (*P*s < 0.001, uncorrected). There were also smaller clusters with the same results (data not shown): right LOC (7 voxels; 48, –82, 18), and right occipital fusiform gyrus (2 voxels; 20, –76, –10).

approach, there is another way that the negative relationship could be confounded with initial task encoding. Imagine that a participant did not pay much attention to the orienting task in the first phase. In this scenario, the item representations across the 2 phases will be more item-centered (i.e., less affected by task information), leading to higher item-wise pattern similarity. Because the initial task was not reliably encoded, higher item-wise pattern similarity would then be related to worse source memory for the initial task.

This account makes the straightforward predictions that greater task encoding in phase 1 (reflecting more attention to the task) and task reactivation in phase 2 (reflecting stronger task encoding) should be associated with lower item-wise pattern similarity. However, although numerically negative, these relationships (based on residuals after regressing out task templates) were not reliable (phase 1 task activation: $\beta = -0.01$, CI = [–0.05, 0.02], bootstrap $P = 0.432$; phase 2 task reactivation: $\beta = -0.02$, CI = [–0.05, 0.01], bootstrap $P = 0.316$), failing to support the alternative account.

Discussion

We investigated how memory for the context in which an object was encountered is influenced by encountering it again in a novel context. Although the basic behavioral finding—impaired subsequent source memory for the initial context—has been demonstrated previously (McGovern 1964; Postman and Underwood 1973; Richter et al. 2016), we tested a novel explanation for this important phenomenon. Specifically, we hypothesized that the extent to which the item is represented the same way across contexts determines how much interference the old context suffers. Consistent with this hypothesis, we found a negative relationship between neural overlap across item repetitions and subsequent source memory for the initial context. In addition to this main finding, we found that greater task reactivation in the second phase leads to better source memory for the initial context, which is consistent with previous studies suggesting that reactivation of a prior context builds resistance to interference from a novel subsequent context (Kuhl et al. 2010; Koen and Rugg 2016).

We used pattern similarity to index how much the initial item representation was reactivated upon repetition. However, taken at face value, this measure does not necessarily reflect item information alone—it can also be influenced by variance in task reactivation. We addressed this issue by showing: 1) that the negative relationship persists after regressing out task information, 2) that this relationship is eliminated after permuting item pairings within task, and 3) that task reactivation per se leads to an opposite (beneficial) effect on memory. Taken together, these results strongly support our main conclusion that reactivation of an item-specific representation leads to retroactive interference.

Using a paradigm similar to ours, Koen and Rugg (2016) also recently investigated the consequence of item repetition for task memory. As in our study, participants performed one task on an item during an initial phase and then performed a different task on that item during a later phase; participants were then required to recall both tasks for each item. Interestingly, Koen and Rugg (2016) found a positive relationship between item-specific reactivation during item repetition and subsequent source memory for the initial task—the opposite of what we found.

However, we do not think that these results are necessarily incompatible with ours. Our hypothesis was that reactivating item-specific perceptual features in a new context might hurt source memory, by linking the item's representation to a new context that later interferes with retrieval of the original context (Hupbach et al. 2007; Gershman et al. 2013; Sederberg et al. 2011; St. Jacques et al. 2013). To test this hypothesis, we used object stimuli with rich perceptual features and focused on a region known to represent item-level features of objects (Cichy et al. 2011; Eger et al. 2008). Crucially, while reinstating item-level object features might be harmful for source memory, reinstating task-specific elaborations might be helpful for source memory. For example, after making a pleasantness judgment (an encoding task in Koen and Rugg's study) about the word "broccoli," later reactivation might contain task information specific to that item, such as that it has an unpleasant taste. This retrieved information related to the initial task is unlikely to be confused with other tasks in the final source memory test. In fact, reactivation of this task-specific elaboration might strengthen its association with the item and ultimately improve source memory.

Speculatively, features of Koen and Rugg's study may have led to task-specific elaborations playing a greater role in their results. For example, their item-specific analysis focused on a task-selective mask, which may have increased the prominence of task-specific elaborations (vs. perceptual features). To assess the importance of this difference, we ran our item-specific pattern similarity analysis in a task-selective mask and did not obtain a significant relationship (in either direction) with source memory (see Supplementary Materials), showing that this difference was not unto itself responsible for the discrepancy in results. Another possibility is that our paradigm was less conducive overall to forming task-specific elaborations. For example, our use of object pictures (vs. words) may have led participants to engage more with the visual features of the item and to form fewer elaborations; also, our task instructions were slightly different (in the function task, we simply asked participants to rate the usefulness of objects as opposed to counting specific uses).

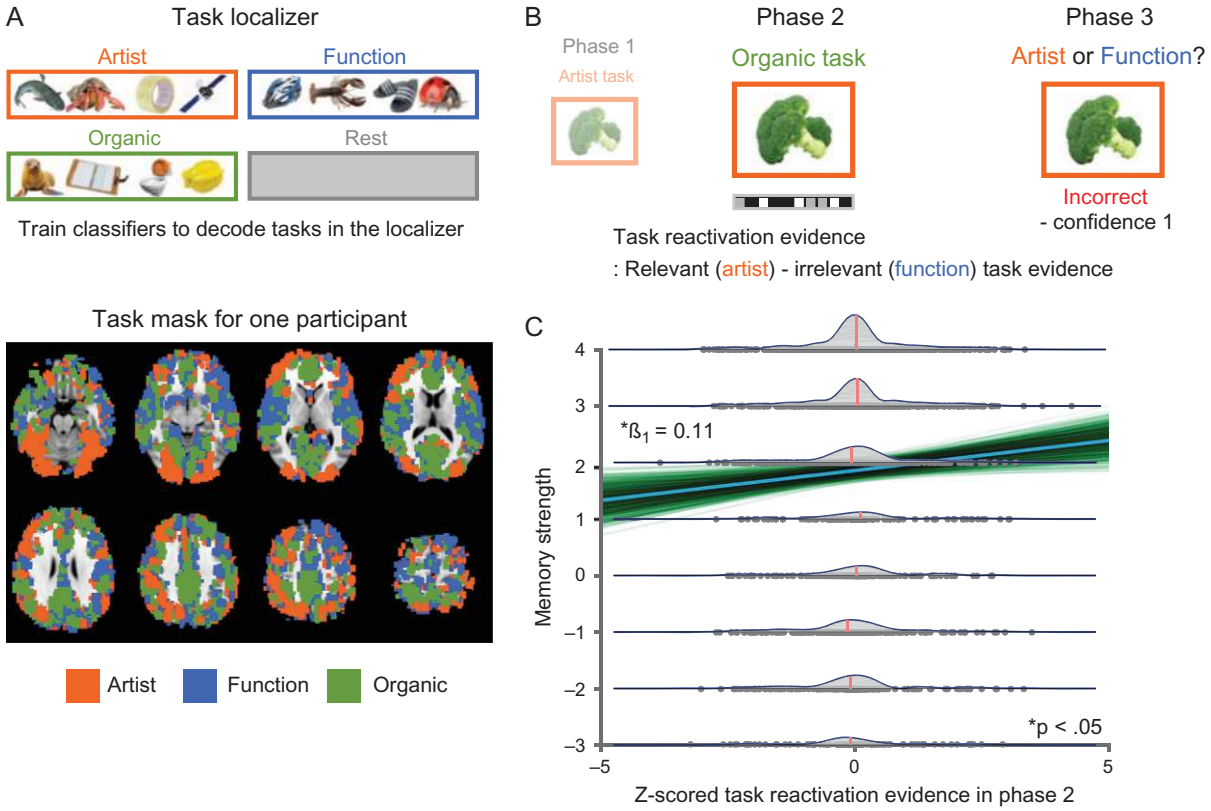

Figure 6. Relating task reactivation to memory. (*A*) We trained the classifiers using the localizer runs based on task-selective voxels. (*B*) We measured reactivation of the initial task (artist or function) in the second phase by testing the classifiers on the second phase data. Task reactivation was operationalized by subtracting relevant task minus irrelevant task classifier evidence (e.g., artist minus function). (*C*) The relationship between this index of task reactivation and subsequent source memory was significantly positive.

There are other accounts that might explain the discrepancy between the previous study and ours. For example, in Koen and Rugg (2016) the initial and second tasks were interleaved, whereas they were blocked in our study. Interleaving might have increased the overall level of task reactivation: For example, switching between an initial and second task could have led to greater residual activation of the initial task. This stronger initial task reactivation might enhance integration of information relating to the initial and second tasks, reducing interference between the 2 (Zeithamova and Preston 2017). In our case, task reactivation might have been weaker because the initial and second tasks were separated, leading to less integration and (consequently) more interference. Another difference between our study and Koen and Rugg (2016) is that—in our study—the artist and function tasks did not recur in phase 2, whereas in Koen and Rugg (2016) all of the tasks occurred in all phases; participants in our study might therefore have been motivated to suppress memories relating to phase 1 tasks during phase 2, leading to even weaker memory reactivation. Prior studies from our lab (Detre et al. 2013; Kim et al. 2014) have shown that weak (but nonzero) activation of a memory can reduce the subsequent accessibility of that memory; these effects may have further contributed to our finding that item reactivation was associated with worse (not better) memory. Although we do not have definitive evidence for one of these accounts, they provide an interesting area for future research.

We argue that reinstatement of initial item features opens a window for novel contextual information to be bound to the initial item features, which subsequently interferes with retrieval of an initial context. Insofar as the organic task is a major feature of the novel (phase 2) context, this implies that greater organic-task activity (for a particular item during phase 2) might have a negative relationship with subsequent source memory for the initial context. However, classifier evidence for the organic task during phase 2 did not predict subsequent source memory (bootstrap $P = 0.414$; see Supplementary Materials), suggesting that other aspects of the novel context (besides the organic task itself) might be responsible for the observed interference in subsequent source memory.

Although the effects of item and task reactivation on source memory went in opposite directions, there was a numerically negative correlation between item reactivation (based on residuals after regressing out task templates) and task reactivation scores (see Results). To test the specificity of each of these effects, we first examined the observed negative effect of item reactivation after controlling for task reactivation with partial correlation, and we found that it remained reliable ($\beta = -0.08$, CI = [−0.15, −0.01], bootstrap $P = 0.022$). When this analysis was reversed, controlling item reactivation, the positive effect of task reactivation was also reliable ($\beta = 0.10$, CI = [0.01, 0.19], bootstrap $P = 0.028$). These results suggest that those opposite effects of item and task reactivation were distinct from each other.

Generally speaking, these results highlight the complex nature of retroactive interference effects. As prior work has shown, there is no simple answer to the question of how new learning modulates the accessibility of existing knowledge:

There is some evidence that reactivation makes old memories susceptible to retroactive interference (Forcato et al. 2007; Gershman et al. 2013; Hupbach et al. 2007; Sederberg et al. 2011), but other studies observed the opposite effect that reactivation alleviates retroactive interference (Kuhl et al. 2010; Koen and Rugg 2016). Still others posit that the degree to which information is activated determines whether it is strengthened or weakened (e.g., Detre et al. 2013). Our findings contribute to this debate by suggesting that—when an item appears in multiple contexts—the specific content of the reactivated memory is a key determinant of retroactive interference: If contextual features uniquely related to the item in the initial experience are strongly reactivated during new learning, this can strengthen memory for these features (e.g., Koen and Rugg 2016; Kuhl et al. 2010; a trend in our study). However, crucially, when shared item features are reactivated, our results show that this can impair memory for the initial context.

Supplementary Material

Supplementary material is available at *Cerebral Cortex* online.

Funding

NIH (R01 MH069456 to K.A.N.) and NIH (R01 EY021755 to N.B.T.B.).

Notes

Conflict of Interest: none.

References

Aly M, Turk-Browne NB. 2016. Attention stabilizes representations in the human hippocampus. Cereb Cortex. 26(2): 783–796.

Chan SCY, Applegate MC, Morton NW, Polyn SM, Norman KA. 2017. Lingering representations of stimuli influence recall organization. Neuropsychologia. 97:72–82.

Cichy RM, Chen Y, Haynes JD. 2011. Encoding the identity and location of objects in human LOC. Neuroimage. 54(3): 2297–2307.

Coutanche MN. 2013. Distinguishing multi-voxel patterns and mean activation: why, how, and what does it tell us? Cogn Affect Behav Neurosci. 13(3):667–673.

Davis T, LaRocque KF, Mumford JA, Norman KA, Wagner AD, Poldrack RA. 2014. What do differences between multi-voxel and univariate analysis mean? How subject-, voxel-, and trial-level variance impact fMRI analysis. Neuroimage. 97: 271–283.

Davis T, Poldrack RA. 2013. Measuring neural representations with fMRI: practices and pitfalls. Ann NY Acad Sci. 1296:108–134.

Detre GJ, Natarajan A, Gershman SJ, Norman KA. 2013. Moderate levels of activation lead to forgetting in the think/no-think paradigm. Neuropsychologia. 51(12):2371–2388.

Efron B. 1979. Bootstrap methods: another look at the jackknife. Ann Stat. 7(1):1–26.

Eger E, Ashburner J, Haynes JD, Dolan RJ, Rees G. 2008. fMRI activity patterns in human LOC carry information about object exemplars within category. J Cogn Neurosci. 20(2):356–370.

Forcato C, Burgos VL, Argibay PF, Molina VA, Pedreira ME, Maldonado H. 2007. Reconsolidation of declarative memory in humans. Learn Mem. 14(4):295–303.

Gershman SJ, Schapiro AC, Hupbach A, Norman KA. 2013. Neural context reinstatement predicts memory misattribution. J Neurosci. 33(20):8590–8595.

Gordon AM, Rissman J, Kiani R, Wagner AD. 2014. Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. Cereb Cortex. 24(12):3350–3364.

Green DM, Swets JA. 1966. Signal detection theory and psychophysics. New York, NY: Wiley.

Grill-Spector K, Kourtzi Z, Kanwisher N. 2001. The lateral occipital complex and its role in object recognition. Vision Res. 41(10–11): 1409–1422.

Hupbach A, Gomez R, Hardt O, Nadel L. 2007. Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. Learn Mem. 14(1–2):47–53.

Johnson M, Kounios J, Nolde S. 1997. Electrophysiological brain activity and memory source monitoring. Neuroreport. 8(5): 1317–1320.

Johnson JD, McDuff SGR, Rugg MD, Norman KA. 2009. Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis. Neuron. 63(5):697–708.

Kim G, Lewis-Peacock JA, Norman KA, Turk-Browne NB. 2014. Pruning of memories by context-based prediction error. Proc Natl Acad Sci USA. 111(24):8997–9002.

Kim G, Norman KA, Turk-Browne NB. 2017. Neural differentiation of incorrectly predicted memories. J Neurosci. 37(8): 2022–2031.

Koen JD, Rugg MD. 2016. Memory reactivation predicts resistance to retroactive interference: evidence from multivariate classification and pattern similarity analyses. J Neurosci. 36(15):4389–4399.

Kuhl BA, Rissman J, Wagner AD. 2012. Multi-voxel patterns of visual category representation during episodic encoding are predictive of subsequent memory. Neuropsychologia. 50(4): 458–469.

Kuhl BA, Shah AT, DuBrow S, Wagner AD. 2010. Resistance to forgetting associated with hippocampus-mediated reactivation during new learning. Nat Neurosci. 13(4):501–506.

Macmillan NA, Creelman CD. 2005. Detection theory: a user's guide. 2nd ed. Mahwah, New Jersey: Lawrence Erlbaum Associates.

McDuff SGR, Frankel HC, Norman KA. 2009. Multivoxel pattern analysis reveals increased memory targeting and reduced use of retrieved details during single-agenda source monitoring. J Neurosci. 29(2):508–516.

McGovern JB. 1964. Extinction of associations in four transfer paradigms. Psychol Monogr Gen A. 78(16):1–21.

Postman L, Underwood BJ. 1973. Critical issues in interference theory. Mem Cognit. 1(1):19–40.

Richter FR, Chanales AJH, Kuhl BA. 2016. Predicting the integration of overlapping memories by decoding mnemonic processing states during learning. Neuroimage. 124(Pt A):323–335.

Sederberg P, Gershman S, Polyn S, Norman K. 2011. Human memory reconsolidation can be explained using the temporal context model. Psychon B Rev. 18(3):455–468.

St. Jacques PLS, Olm C, Schacter DL. 2013. Neural mechanisms of reactivation-induced updating that enhance and distort memory. Proc Natl Acad Sci USA. 110(49):19671–19678.

Ward EJ, Chun MM, Kuhl BA. 2013. Repetition suppression and multi-voxel pattern similarity differentially track implicit and explicit visual memory. J Neurosci. 33(37):14749–14757.

Xue G, Dong Q, Chen C, Lu Z, Mumford JA, Poldrack RA. 2010. Greater neural pattern similarity across repetitions is associated with better memory. Science. 330(6000):97–101.

Zeithamova D, Preston AR. 2017. Temporal proximity promotes integration of overlapping events. J Cogn Neurosci. 29(8): 1311–1323.