JNeurosci
THE JOURNAL OF NEUROSCIENCE

Research Articles | Behavioral/Cognitive

Intervening contexts reshape hippocampal-cortical retrieval states and later source attribution

https://doi.org/10.1523/JNEUROSCI.1413-26.2026

Received: 18 August 2026
Revised: 4 November 2026
Accepted: 11 November 2026

Copyright © 2026 Hermiller and Davachi
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

This Early Release article has been peer reviewed and accepted, but has not been through the composition and copyediting processes.The final version may differ slightly in style or formatting and will contain links to any extended data.

Alerts: Sign up at www.jneurosci.org/alerts to receive customized email alerts when the fully formatted version of this article is published.

TITLE: Intervening contexts reshape hippocampal-cortical retrieval states and later source attribution

Running title: Retrieval contexts reshape memory states

AUTHORS: Molly S Hermiller[1] and Lila Davachi[2,3]
1 Florida State University, Department of Psychology, USA
2 Yale University, Department of Psychology, USA
3 Nathan Kline Psychiatric Institute, USA

Corresponding Author: Molly S. Hermiller
1107 W. Call St
Tallahassee, FL 32303
mhermiller@fsu.edu

Number of Figures: 6
Number of Tables: 1
Number of Supplemental Materials: 7
Abstract word count: 221 (250 max)
Significance word count: 104 (120 max)
Introduction word count: 649 (650 max)
Discussion word count: 1484 (1500 max)

ABSTRACT
Memories are often reactivated in one context and only later judged with respect to where they originally occurred. We tested how an intervening retrieval context shapes neural representations and later source memory. In two fMRI tasks, participants (both male and female) encoded objects within contextual backgrounds, completed an immediate retrieval test in which object cues appeared against either the original background or a different background, and subsequently judged each object's original context on a neutral screen. Shifted cueing contexts reduced source accuracy and biased errors toward the intervening context, despite no detectable difference in immediate recognition or associative-memory performance. Representational similarity analyses showed that encoding patterns reflected original-context structure, whereas immediate-cue-period patterns were broadly organized by the cued (i.e., currently visible) context. On shifted trials, activity patterns showed greater similarity to encoding-derived templates for the cued context than for the original context. Across participants, greater differentiation of original-context similarity between consistent and shifted cueing in posterior hippocampus, parahippocampal cortex, and dorsal anterior cingulate cortex was associated with better source memory after a contextual shift. Within participants, precuneus contextual engagement and dorsal anterior cingulate context alignment predicted later source choices, with similar relationships after adjustment for immediate recognition responses. Later source attribution therefore depended on the contextual state expressed when an item was re-encountered, not simply on reinstatement of the original event.

SIGNIFICANCE STATEMENT
Remembering is often described as bringing up the past, but memories are frequently reactivated in new contexts before we judge where they came from. We show that this intervening context can bias later source attribution even when that context is no longer visible at final test. Using two fMRI tasks, representational similarity analysis and trial-level source-choice models, we found that cue-period activity was organized by the context present during cueing and that specific context-state signals predicted later memory choices, whereas item reinstatement did not. Source-memory errors can therefore reflect contextual information encountered after the original event, not only failures to recover the original context.

JNeurosci Accepted Manuscript

Keywords: episodic memory; source memory; retrieval context; context reinstatement; hippocampus; medial temporal lobe; posterior-medial cortex; representational similarity analysis

JNeurosci Accepted Manuscript

INTRODUCTION

Episodic remembering unfolds in the present, even when the goal is to recover the past. Imagine hearing a song during a night out, then again days later in a café. The song may bring the original night to mind but is now encountered in a different setting. If later asked where you first heard it, your answer may reflect both encounters. This creates a source-memory problem: an item can be reactivated in one context before its original source is judged. What determines which context is ultimately treated as the source?

Classic theories emphasize contextual overlap between encoding and retrieval. Encoding-specificity accounts propose that memory is facilitated when retrieval cues recapitulate features of the original experience (Godden and Baddeley, 1975; Smith and Vela, 2001; Tulving and Thomson, 1973). Retrieved-context models propose that memory search is guided by a dynamic contextual state that evolves over time (Howard and Kahana, 2002; Polyn et al., 2009). At the neural level, successful remembering is accompanied by reinstatement of encoding-related activity patterns across sensory and medial temporal regions (Danker and Anderson, 2010; Johnson et al., 2009; Ritchey et al., 2013), including temporal and event context as well as item content (Folkerts et al., 2018; Kragel and Voss, 2021; Lohnas et al., 2023; Manning et al., 2011). Together, this work identifies reinstatement as a core signature of episodic retrieval.

Yet memories are not always judged when they are reactivated. The original encoding context, the intervening context in which the event is re-encountered, and the later test situation can all contribute to a source judgment. Competing memories and contextual cues can bias later source judgments (Johnson et al., 1993; Mitchell and Johnson, 2009; Postman and Underwood, 1973; Underwood, 1957), while reactivation can promote updating and misattribution (Gershman et al., 2013; Hupbach et al., 2007). Event-boundary research provides a complementary framework: changes in temporal, spatial, or conceptual context segment continuous experience into events and shape how memories are organized or integrated (Clewett et al., 2020; DuBrow and Davachi, 2013; DuBrow and Davachi, 2014; Ezzyat and Davachi, 2011). Neural activity likewise carries information about event and contextual structure, with hippocampal and posterior-medial regions representing spatial, temporal, and situational structure (Baldassano et al., 2017; Cohn-Sheehy et al., 2021). Re-encountering an old item in a new context therefore creates a simple problem: information about the original event is recovered while a second context is being experienced. Either context could later contribute to the source judgment.

Hippocampal–cortical memory networks are well positioned to support these competing contextual representations. The hippocampus and surrounding medial temporal lobe bind items to contextual features and support recovery of those associations (Davachi, 2006; Diana et al., 2007; Eichenbaum et al., 2007; Yonelinas et al., 2019). Posterior-medial regions represent spatial and situational context (Ranganath and Ritchey, 2012; Reagh and Ranganath, 2023; Ritchey et al., 2015), whereas prefrontal and cingulate regions may contribute when mnemonic representations compete (Badre and Wagner, 2007; Mitchell and Johnson, 2009; Schlichting et al., 2015). These systems could therefore carry information about both original and intervening contexts.

Here, we tested this possibility using two fMRI tasks that separated initial encoding, immediate cued retrieval, and later source judgment (Figure 1). During encoding, objects appeared in an original context. During immediate retrieval, studied objects were cued while the background either remained consistent or shifted to a different context. Source memory for the original context was tested later on a neutral screen with neither background visible. Using representational similarity analysis, we asked whether encoding patterns reflected original-context structure, whether cue-period patterns were organized by the current context, and whether cue-period representations predicted later source attribution when original and cued context diverged. We predicted that a context shift would bias cue-period activity toward the intervening context. We then asked whether the way original-context information changed across cueing conditions was related to later source memory. We further tested whether contextual and item-specific signals predicted which source participants ultimately selected.

Neuroscience Accepted Manuscript

MATERIALS and METHODS

Participants

Twenty healthy adults completed the study. Participants reported no history of neurological or psychiatric disorders and no current use of psychoactive medications. Two participants were excluded because more than 10% of functional volumes were censored for excessive head motion, yielding a final sample of 18 participants (12 female, 6 male; age range = 18–35 years). All participants provided written informed consent in accordance with procedures approved by the Columbia University Institutional Review Board and received monetary compensation.

Task Design

Overview

Participants completed two structurally matched episodic encoding–retrieval tasks designed to dissociate the context in which an item was originally encoded from the context present when that item was later re-cued. We use original context to refer to the background present during encoding and cued context to refer to the background present when the item cue appeared during immediate retrieval. On consistent-context trials, the original and cued contexts matched. On shifted-context trials, the item cue appeared in a different contextual background. Critically, source memory for the original context was tested only later, on a neutral gray screen, requiring participants to identify the original context without either the original or cued background visually present.

Participants completed six fMRI blocks, comprising three City and three Soccer blocks. Each block constituted one functional run. Each block contained four encoding–retrieval loops, with six studied objects per loop, yielding 24 studied objects per block and 144 object-encoding events across the experiment. All 24 objects were subsequently presented during immediate cued retrieval. At the end of each block, 12 objects were tested for their original context, while the remaining 12 were assigned to six temporal-order test pairs. Thus, participants completed 72 final source-memory trials, including 36 consistent-context and 36 shifted-context trials.

The tasks differed in stimulus domain and contextual structure while maintaining parallel timing, trial organization, and memory demands. In the City task, participants encoded object–location associations embedded within spatial environments that varied in weather/lighting (i.e., rainy, sunny, night). In the Soccer task, participants encoded object–character associations embedded within team-color environments (i.e., red, green, blue). Across both tasks, the cued context either remained consistent with the original context or shifted before immediate cued retrieval, allowing within-subject dissociation of original-context representations and cued-context organization. Each block contained two encoding-to-cue shifts and two consistent encoding-to-cue episodes. Three contextual states occurred within each block, and once the context changed away from a state, it did not return within that block. On shifted trials, the uninvolved context was the third contextual state that was neither the item's original encoding context nor the context present during immediate retrieval.

To maintain engagement, the tasks were framed using cover stories. In the City task, participants were told they were helping recover objects scattered throughout a virtual city after belongings fell from a moving truck in specific locations. In the Soccer task, participants were told that members of different soccer teams were temporarily storing recovered objects within team-color-associated locker rooms. Participants were instructed to remember both the paired associations and the contextual environment in which each item appeared.

Prior to scanning, participants completed approximately 45 minutes of behavioral training that included one full practice block of each task. Stimuli used in the practice block were unique from those used in the experiment. During fMRI acquisition, participants completed six functional blocks total: three City blocks and three Soccer blocks presented in randomized interleaved order. Each block consisted of four encoding–retrieval episodes (“loops”), followed by two end-of-block memory tests and a low-level number-judgment baseline task. City

J Neurosci Accepted Manuscript

blocks were implemented in Unity (Unity Technologies), whereas Soccer blocks were implemented in PsychoPy. Approximately 24 hours later (range: 22–26 h), participants returned for a delayed memory assessment; those data are not analyzed here.

Encoding Episodes

In each encoding loop, participants encountered six object associations within a single contextual background. Individual encoding events lasted 2 s and were separated by jittered interstimulus intervals averaging 6 s (range = 2–8 s). Each encoding loop lasted approximately 52 s. Each object appeared only once during encoding across the six experimental blocks and was subsequently re-presented during its corresponding immediate-retrieval test; novel lure items used during immediate retrieval did not appear during encoding.

In the City task, participants passively navigated a looped virtual city environment. At six spatially distinct locations, forward movement paused and an object appeared for 2 s above a location marker together with its name. Each object location was separated by at least 60 virtual units to maintain spatial distinctiveness. After object presentation, navigation resumed until the next stop. Each loop began and ended at the same corner of the city block.

In the Soccer task, participants viewed dynamic kaleidoscopic backgrounds that continuously changed in shape and texture while maintaining a stable contextual color state (red, green, or blue). During each loop, six object–character associations were presented sequentially for 2 s each on static snapshots of the kaleidoscope. Objects were accompanied by text labels and displayed alongside one of six characters, with left–right screen positions randomized across trials. Following each encoding event, the kaleidoscope resumed its dynamic motion.

Immediate Cued Retrieval

Each encoding loop was immediately followed by a 5-s visualization prompt and an immediate cued-retrieval phase. The prompt instructed participants to “Visualize or re-imagine the exact item you found, when prompted,” encouraging active recovery of encoding-phase representations before object name cues appeared. Participants then viewed nine object names in randomized order: the six studied items from the preceding encoding loop and three novel lures. Each cue appeared for 2 s and was followed by a 4-s Remember–Familiar–New judgment (Rugg and Yonelinas, 2003). For items endorsed as Remember or Familiar, participants completed an associative memory judgment: in the City task, they navigated a marker to the original object location; in the Soccer task, they selected the character originally paired with the object. The 2-s item-cue period constituted the cue-period activity used in representational similarity analyses. The contextual background present during this cue period is referred to as the cued context.

Original- and Cued-context Manipulation

Contextual environments consisted of weather states in the City task (sunny, rainy, night) and team-color states in the Soccer task (red, green, blue). During immediate cued retrieval, the contextual background either matched the original encoding context or shifted to a different contextual background. Thus, on shifted-context trials, the original and cued contexts were experimentally dissociated.

Three contextual-transition structures were created and randomized across blocks and participants such that each participant experienced all transition structures once per task. Three contextual states occurred within each block. Each block contained two loops in which the context changed between encoding and immediate retrieval and two loops in which it remained consistent. Once a context changed away from one state, it did not return within the same block, creating structured contextual trajectories across encoding and immediate-retrieval episodes. Contextual transitions were dissociated from encoding-loop boundaries: except for the first loop within a block, each new encoding loop began with the context present during the immediately preceding cued-retrieval phase.

Neurosci Accepted Manuscript

End-of-Block Memory Tests

Each block contained four encoding–retrieval loops comprising 24 studied objects. After the four loops, participants completed two end-of-block memory tests: a temporal-order test (not analyzed here) and the source-memory test used for the present behavioral analyses. In the source-memory test, 12 objects were tested individually in randomized order on a neutral gray screen. The item name and three context options were presented for 2 s, during which participants made a three-alternative forced-choice judgment identifying the context in which the object had originally been encoded. Half of the source-tested items had been re-cued during immediate retrieval in a consistent context, and half had been re-cued in a shifted context. Across the three blocks of each task, participants completed 36 source-memory judgments per task.

Stimuli

Objects were drawn from the Bank of Standardized Stimuli (BOSS-2014 v.2) (Brodeur et al., 2014). Food items, animals, and objects implausible within the task narratives were excluded. All objects were presented as images accompanied by text labels. Character images used in the Soccer task were drawn from the video game *Animal Crossing: New Horizons* (Nintendo Co., Ltd., Kyoto, Japan, 2020).

Behavioral Analyses

Immediate recognition responses were classified as Remember, Familiar, or New. Immediate-behavior analyses were restricted to the 1,292 source-test trials that could be linked to the RSA trial table (648 City, 644 Soccer). Four Soccer source-test trials, one from each of four participants, were unavailable because of a task-code error and were excluded from RSA-linked analyses. We examined the probability of a Remember or Familiar response versus New, and Remember versus Familiar among recognized items. R/F/N analyses were restricted to previously studied items subsequently included in the source-memory analysis. Novel-lure responses were not included; therefore, Remember-plus-Familiar responses were interpreted as studied-item hit rates rather than overall recognition accuracy. For recognized City items, location error was quantified as the shortest distance along the virtual track perimeter between the encoded and selected locations. Target accuracy was additionally defined as an error of fewer than 20 track units, consistent with our previous behavioral analysis. For recognized Soccer items, associative accuracy was defined by whether the selected character matched the character paired with the object during encoding. Associative accuracy was not assigned to items receiving a New response. Full immediate-behavior results are reported in Table S1.

Source-memory performance was assessed during the end-of-block source-memory test. For each participant and condition, we quantified the proportion of trials in which participants selected the original context, the cued context, or an uninvolved context. Source accuracy was evaluated with a 2 × 2 repeated-measures ANOVA with within-subject factors of Task (City, Soccer) and Context Condition (consistent, shifted). Shifted-trial response distributions were evaluated with repeated-measures ANOVAs including Task and Choice as within-subject factors. Planned comparisons tested whether, among shifted-trial errors, participants selected the cued context more often than the uninvolved context.

We used mixed-effects models to test whether contextual shift affected immediate recognition or associative performance. We then restricted analyses to shifted-context source-test trials and tested whether immediate recognition and associative-memory measures predicted three later outcomes: final source accuracy (equivalent to original versus not-original choice), original versus cued choice, and, among source errors, cued versus uninvolved choice. Models included participant random intercepts and the relevant task, run, and within-run position covariates.

Context-shift effects on binary immediate-memory outcomes were tested with mixed-effects logistic regression; continuous City track error was analyzed after log(1 + error) transformation. Models included participant random intercepts, run, and within-run trial position; pooled recognition models additionally included task and the Shift × Task interaction. For models relating immediate behavior to later source attribution, Familiar served

Neurosci Accepted Manuscript

as the reference category in the three-level R/F/N model. Task-specific associative-memory models were restricted to recognized trials with a usable response and included Remember-versus-Familiar status as a covariate. City location quality was standardized within participant, with higher values indicating lower track error.

MRI acquisition and preprocessing

MRI data were collected on a Siemens 3T Prisma whole-body scanner equipped with a 64-channel head/neck coil. Functional images were acquired using a multiband gradient-echo echo-planar imaging sequence sensitive to blood oxygenation level-dependent contrast (TR = 1500 ms, TE = 30 ms, flip angle = 65°, voxel size = 2 mm isotropic, 69 ascending axial slices, field of view = 208 × 196 mm, multiband acceleration factor = 3). Each functional run lasted approximately 5.1 min. High-resolution anatomical images included a T1-weighted MPRAGE sequence (TR = 2300 ms, TE = 2.98 ms, TI = 900 ms, flip angle = 9°, 1 mm isotropic) and a T2-weighted structural scan (TR = 3200 ms, TE = 563 ms, 0.7 mm isotropic).

Functional data were preprocessed in AFNI (Cox, 1996) using custom scripts adapted from afni_proc.py. Preprocessing included deobliquing, despiking, slice-time correction, motion correction, anatomical alignment, and intensity scaling. Functional data were retained in native space without spatial smoothing to preserve fine-grained multivoxel pattern information (Dimsdale-Zucker and Ranganath, 2018; Kriegeskorte et al., 2008). Head motion correction was performed using 3dvolreg by aligning all functional volumes to a representative EPI reference volume. Anatomical segmentation of gray matter, white matter, and cerebrospinal fluid (CSF) was performed using 3dSeg, and ventricular and white-matter nuisance signals were derived from participant-specific tissue masks for inclusion in the single-trial models (Jo et al., 2010). Six motion parameters and their temporal derivatives were retained for nuisance modeling. Volumes exceeding a threshold of 0.3 mm framewise displacement were flagged for censoring. Two participants were excluded because greater than 10% of functional volumes were censored for excessive motion. Atlas-based cortical ROI definitions and group-level displays used normalized anatomical information, but all trial-level pattern estimates and ROI similarity analyses were performed in native functional space. Temporal signal-to-noise ratio maps, spatial smoothness estimates, and automated AFNI quality-control outputs were visually inspected for all participants.

Regions of Interest

A priori regions of interest (ROIs) were selected from medial temporal and cortical memory networks implicated in contextual representation, episodic retrieval, and mnemonic control. Cortical ROIs were defined volumetrically using the Brainnetome atlas v1.0 (Fan et al., 2016) and intersected with participant-specific gray matter masks. Cortical regions included retrosplenial cortex (RSC; A23d + A23v), precuneus (A31 + dmPOS), dorsolateral prefrontal cortex (Area 46d), and dorsal anterior cingulate cortex (A24rv + A24cd).

Medial temporal lobe ROIs were generated using Automated Segmentation of Hippocampal Subfields (Yushkevich et al., 2015) applied to each participant's high-resolution anatomical images. Individualized native-space segmentations included anterior hippocampus (aHPC), posterior hippocampus (pHPC), entorhinal cortex (ERC), perirhinal cortex (PRC), and parahippocampal cortex (PHC). ERC and PRC were combined into a single entorhinal/perirhinal cortex mask (EPRC) to improve reliability and ensure sufficient voxel coverage across participants. All segmentations were visually inspected for anatomical accuracy. Medial temporal ROIs were retained separately for the left and right hemispheres to preserve potential hemispheric differences in representational coding, whereas cortical ROIs were bilaterally averaged to improve reliability and reduce model complexity. Bilateral averaging was performed after hemisphere-specific pattern preprocessing and drift correction, so that temporal proximity residualization reflected each hemisphere's actual trial timing structure before the two hemispheric matrices were averaged. This yielded a final set of 12 *a priori* ROIs: left and right anterior hippocampus, posterior hippocampus, entorhinal/perirhinal cortex, and parahippocampal cortex, plus bilateral precuneus, retrosplenial cortex, dACC, and dlPFC. All ROI masks were restricted to participant-specific analyzable voxels, and ROIs containing fewer than 25 voxels were excluded from analysis; no ROIs were removed for this reason.

Neurosci Accepted Manuscript

Single-Trial Activity Estimation

Trial-wise multivoxel activation patterns were estimated using the least-squares-separate (LSS) approach (Abdulrahman and Henson, 2016; Mumford et al., 2012) implemented in AFNI's 3dLSS. Separate LSS models were estimated for encoding and immediate-retrieval cue-period phases within each task and run, for all participants. All representational analyses were conducted in each participant's native EPI space using unsmoothed functional data.

Encoding-phase LSS models included each encoding event as a separate BLOCK(2,1) regressor specified with -stim_times_IM. The remaining regressors modeled task events not of interest: retrieval cue onsets, Remember/Familiar/New recognition responses, source-memory judgment periods, temporal-order judgment periods, contextual-state judgment periods during end-of-block tests, number-judgment baseline events, and loop-level activity. All LSS models additionally included ventricular and white-matter nuisance time series, six motion parameters and their temporal derivatives, and a run-level polynomial baseline; high-motion and high-outlier volumes were censored. These terms and censoring specifications were incorporated into the design matrix used for single-trial beta estimation.

Cue-period retrieval-phase LSS models were structurally parallel to the encoding models, with each old-item retrieval cue modeled as a separate BLOCK(2,1) regressor. Novel lure cues were modeled with a pooled BLOCK(2,1) regressor. Additional regressors modeled encoding events, recognition responses, source-memory and temporal-order judgment periods, number-judgment baseline events, and loop-level activity. The same nuisance regressors, censoring criteria, and polynomial baseline terms described above were included.

Following beta estimation, trial-wise beta vectors were extracted from each ROI mask in native space and stacked into trial × voxel matrices separately for encoding and cue-periods within each run. Object identifiers were used to establish correspondence between encoding and cue-periods, creating the encoding-to-retrieval mapping used in cross-phase analyses.

Pattern-similarity Preprocessing and Temporal-Drift Correction

Before computing similarity matrices, trial-wise beta vectors were preprocessed within each ROI to reduce the influence of outlier voxels, low-quality trials, and run-level drift. First, voxelwise outlier handling was applied separately within each run. For each voxel, beta values across trials were evaluated using a robust median-absolute-deviation z-score, and values exceeding the outlier threshold of > 3 were set to NaN. Second, trials with fewer than 50 finite voxels after voxel-level preprocessing were excluded. Third, run-mean pattern demeaning was performed by subtracting each voxel's mean activation across trials within a run from individual trial values. Pairwise Pearson correlations were then computed only when both patterns shared at least 25 finite voxels; cells failing this minimum-overlap requirement were set to NaN. Correlation coefficients were Fisher r-to-z transformed.

Because multivoxel similarity can be contaminated by temporal proximity(Cai et al., 2019), all similarity matrices were corrected for temporal drift before representational contrasts were computed(Diedrichsen and Kriegeskorte, 2017). For within-phase encoding–encoding and retrieval–retrieval matrices, each finite off-diagonal similarity value was regressed on the absolute elapsed time between the two trials and a same-loop indicator. The fitted temporal trend was subtracted, and the grand mean of finite off-diagonal cells was restored so that corrected matrices retained a meaningful average similarity level. Retrieval–retrieval matrices were drift-corrected in actual retrieval order before being permuted into encoding-trial order for context-organization analyses.

Cross-phase encoding–retrieval matrices were corrected separately. Each encoding–retrieval similarity cell was regressed on the absolute elapsed time between the encoding and retrieval trials and a same-loop indicator. This correction was necessary because same-item encoding–retrieval pairs are necessarily separated by longer temporal lags than many off-diagonal pairs. This correction was applied before computing Item Reinstatement and template-based context measures. For cortical ROIs that were subsequently averaged

Neuroscience Accepted Manuscript

across hemispheres, drift correction was performed separately within each hemisphere before bilateral averaging.

For template-based analyses, voxelwise beta patterns were additionally detrended across within-run trial order before context templates were constructed. This pattern-level detrending and the cell-wise similarity correction addressed different sources of slow temporal variation. We validated the cross-phase correction with an item-identity diagnostic contrasting same-item encoding–retrieval similarity with different-item, across-loop similarity. Before correction, this contrast was systematically negative across all evaluated regions, consistent with a temporal-proximity artifact; after correction, no region showed a significant negative contrast (Figure S1).

Representational Measures

The primary representational analyses quantified complementary forms of contextual structure during encoding and immediate cued retrieval. Original-context organization measured whether encoding-phase activity patterns clustered according to the context in which items were originally encountered. Cued-context organization measured whether immediate-retrieval activity patterns clustered according to the context present with the item cue. Template-based trial-level measures then quantified, on shifted-context trials, the extent to which each cue-period pattern aligned with the item's original context, the cued context, or both. Item Reinstatement was included as a control measure of item-specific encoding–retrieval similarity.

Original-context Organization. Original-context organization was computed from encoding–encoding similarity matrices. For each participant, task, run, and ROI, we contrasted similarity among encoding trials that shared the same original context with similarity among encoding trials associated with different original contexts. Analyses used across-loop trial pairs to minimize residual local temporal dependencies. Positive values indicate that encoding activity patterns were organized by the original contextual environment.

$$Original\ Context\ Organization = \bar{r}_{EE,\ same\ original} - \bar{r}_{EE,\ different\ orginal}$$

Cued-context Organization. Cued-context organization was computed from retrieval–retrieval similarity matrices. For each participant, task, run, and ROI, we contrasted similarity among immediate-retrieval trials that shared the same cued context with similarity among cue-period patterns associated with different cued contexts. Analyses used across-loop trial pairs to minimize residual local temporal dependencies. Positive values indicate that cue-period retrieval activity patterns were organized by the contextual background present with the item cue.

$$Cued\ Context\ Organization = \bar{r}_{RR,\ same\ cued} - \bar{r}_{RR,\ different\ cued}$$

Template-based Context-state Measures. To quantify trial-level context-state alignment, we conducted template-based analyses restricted to shifted-context trials, where the original and cued contexts differed. For each shifted retrieval event, two context templates were constructed from encoding-phase activity patterns. The original-context template was created by averaging encoding patterns from other items experienced in the same original context as the target item. The cued-context template was created by averaging encoding patterns from items experienced in the context later present with the item cue. Although the cued-context template was constructed from encoding patterns, it represented the context that would later be present during immediate retrieval. Template construction excluded the target item itself and included only encoding trials from different loops, minimizing item-specific influences and local temporal dependencies.

Each cue-period pattern was correlated separately with the original-context template and the cued-context template, yielding original-context similarity and cued-context similarity scores.

$$Original\ Context\ Similarity_i = z\left[r(Retrieval\ pattern_i, Template_i^{original})\right]$$

$$Cued\ Context\ Similarity_i = z\left[r(Retrieval\ pattern_i, Template_i^{cued})\right]$$

Neurosci Accepted Manuscript

Present-vs-Past Alignment. Present-vs-Past Alignment was defined as cued-context similarity minus original-context similarity. Positive values indicate stronger alignment with the context present during immediate cueing, whereas negative values indicate stronger alignment with the item's original encoding context:

$$\text{Present-vs-Past Alignment} = \text{cued-context similarity} - \text{original-context similarity}$$

Contextual Engagement. Contextual Engagement was defined as the average of cued-context similarity and original-context similarity. This measure indexes the overall strength of contextual-state expression during immediate retrieval, irrespective of whether the cue-period pattern favors the cued or original context:

$$\text{Contextual Engagement} = (\text{cued-context similarity} + \text{original-context similarity}) / 2$$

Original-context Fidelity. Original-context Fidelity was defined as similarity between a shifted-trial cue-period pattern and the encoding-derived template for the item's original context. This trial-level measure was used in analyses relating cue-period representations to immediate retrieval behavior and in the cross-region trial-wise covariation analyses:

$$\text{Original-context Fidelity} = \text{original-context similarity}$$

Item Reinstatement. Item Reinstatement quantified item-specific encoding–retrieval similarity. For each participant, task, run, and ROI, we computed the mean Fisher-transformed Pearson correlation between encoding and retrieval patterns for matched item pairs and subtracted the mean correlation for non-matching item pairs drawn from the same contextual environment. This contrast isolates item-specific reactivation from contextual similarity shared across items within a context:

$$\text{Item Reinstatement} = \text{mean } r(\text{ER, same item}) - \text{mean } r(\text{ER, different item, same context})$$

Item Reinstatement served as a control measure to test whether contextual organization and context-state effects were separable from item-level encoding–retrieval similarity.

Original-context Differentiation and Source Accuracy

To test whether context-dependent changes in original-context similarity were associated with source memory, we computed a participant-level Original-context Differentiation score. For each participant, ROI, task, and run, we calculated mean original-context template similarity on consistent-context trials minus mean original-context template similarity on shifted-context trials. Run-level contrasts were averaged within task, and City and Soccer estimates were then averaged to yield one neural value per participant and ROI. Higher values therefore indicate a larger consistent-minus-shifted difference in original-context template similarity. Shifted-trial source accuracy was calculated across source-tested shifted trials pooled across City and Soccer. We computed Spearman correlations between Original-context Differentiation and shifted-trial source accuracy across participants, with 10,000-bootstrap 95% confidence intervals and leave-one-subject-out stability estimates. To test whether these relationships were separable from cued-context organization, overall Contextual Engagement, and data quantity, we also computed partial Spearman correlations controlling for cued-context organization, Contextual Engagement, and number of template trials. False-discovery-rate correction was applied across ROIs.

$$\text{Original-context Differentiation} = \text{mean}(\text{original context similarity}_{consistent}) - \text{mean}(\text{original context similarity}_{shifted})$$

Trial-level Source-Choice Models

To test whether neural context states expressed during immediate cued retrieval predicted later source-memory decisions, we modeled source choices from the end-of-block source-memory test on shifted-context trials. Each response was classified relative to the item's original and cued contexts: original if the participant

JNeurosci Accepted Manuscript

selected the encoding context, cued if the participant selected the context present during immediate retrieval, and uninvolved if the participant selected the remaining context.

We fit three planned binary mixed-effects logistic regression contrasts: original versus cued choices, original versus not-original choices, and, among source errors, cued versus uninvolved choices. Models included a subject random intercept, task, within-run trial position, and run as covariates. Neural predictors were z-scored within subject, task, and ROI before model fitting. The primary model was:

$$\text{Choice} \sim z(\text{Contextual Engagement}) + z(\text{Present-vs-Past Alignment}) + z(\text{Item Reinstatement}) + \text{Task} + \text{WithinRun} + \text{Run} + (1 \mid \text{Subject})$$

Across all shifted-context trials, the original-versus-cued model included 501 observations, the original-versus-not-original model included 646 observations, and the cued-versus-uninvolved error model included 381 observations. All models included all 18 participants. Subject random intercepts accounted for repeated observations within participant. Complete ROI-wise model coefficients, trial counts, and predictor × task interaction tests are reported in Table S3.

The primary predictors were Contextual Engagement and Present-vs-Past Alignment, both derived from original- and cued-context template similarity. Item Reinstatement was included as a control predictor. Follow-up template-decomposition models entered original-context similarity and cued-context similarity separately to determine whether significant effects reflected overall contextual engagement or preferential alignment with one context. Fixed-effect coefficients were evaluated with Wald z tests, computed as the coefficient estimate divided by its standard error using the asymptotic normal approximation for logistic mixed-effects models. To test task generality, we additionally tested neural predictor × task interactions for each source-choice contrast, with ROI-wise inference FDR-corrected within predictor and contrast (Table S3).

To test whether immediate retrieval behavior was directly associated with cue-period neural representations, we used linear mixed-effects models to relate immediate retrieval behavior to trial-level Contextual Engagement and original-context template similarity as a function of immediate R/F/N responses or task-specific associative-memory performance separately within each ROI. Remember-versus-Familiar models were restricted to recognized trials; New-versus-Familiar models used Familiar as the reference category. Task-specific associative-memory models included Remember-versus-Familiar status as a covariate. All models included participant random intercepts, run, and within-run trial position; pooled recognition models additionally included task. FDR correction was applied across the 12 ROIs separately within each neural-measure × behavior-test × coefficient family (Table S5). Finally, we repeated the key precuneus and dACC source-choice models with immediate retrieval performance included. Unadjusted and adjusted neural coefficients were compared within identical complete-case samples. These analyses assessed the sensitivity of the neural relationships to measured immediate behavior rather than causal mediation. Each immediate-behavior measure was added separately, and baseline and adjusted neural coefficients were estimated on identical complete-case samples (Table S5).

Trial-wise Covariation of Representational Signals

To test whether representational signals expressed during immediate retrieval covaried across regions, we computed trial-wise covariation between all pairs of the 12 ROIs on shifted-context trials. For each pair of ROIs, we quantified the relationship between trial-level representational measures (Contextual Engagement, Original-context Fidelity, Present-vs-Past Alignment, and Item Reinstatement) using linear regression, controlling for task and within-run trial position. Trial values were centered within subject before estimation, and the resulting slope served as that subject's trial-wise covariation estimate for that edge and measure.

Subject-level trial-wise covariation estimates were averaged across edges to yield full-network composites and tested against zero with one-sample t-tests. Signal specificity was assessed by contrasting each context-state measure's trial-wise covariation against Item Reinstatement trial-wise covariation over the same edges. The

JNeurosci Accepted Manuscript

full 12-ROI network, comprising 66 edges, was the primary unit for trial-wise covariation analyses. P values were corrected within each edge set across measures using the Benjamini–Hochberg false-discovery-rate procedure. Complete task-difference analyses are reported in Table S6.

Exploratory analyses tested whether trial-wise covariation was organized by a priori functional groupings. ROIs were partitioned into posterior-medial, anterior-MTL, and prefrontal/control groups. The posterior-medial group included posterior hippocampus, parahippocampal cortex, precuneus, and retrosplenial cortex; the anterior-MTL group included anterior hippocampus and entorhinal/perirhinal cortex; and the prefrontal/control group included dACC and DLPFC. This grouping was inspired by the posterior-medial/anterior-temporal framework (Ranganath and Ritchey, 2012; Ritchey et al., 2015), but was constrained to the ROI set used in the primary representational analyses.

Experimental Design and Statistical Analysis

MRI data were analyzed using AFNI software(Cox, 1996) and custom MATLAB (The MathWorks, Inc. Natick, MA, USA) scripts. Brain visualizations were generated using AFNI and the BrainNet Viewer MATLAB toolbox(Xia et al., 2013).

For ROI-level representational measures, subject-level estimates were computed separately for each ROI and task and then tested at the group level. Original-context organization, cued-context organization, Present-vs-Past Alignment, Contextual Engagement, Original-context Differentiation, and Item Reinstatement were evaluated using planned ROI-wise tests with false-discovery-rate correction within each analysis family. Task-specific effects were evaluated separately within City and Soccer tasks, and task differences were assessed using paired-sample t-tests.

Bootstrap confidence intervals and leave-one-subject-out stability estimates were used for across-participant brain–behavior correlations because these analyses were based on one neural and one behavioral summary value per participant and are therefore most sensitive to individual-subject leverage. Trial-level source-choice analyses were evaluated with mixed-effects logistic models including subject-level random intercepts, and ROI-wise inference was controlled using false-discovery-rate correction within each planned model family.

RESULTS

Context shifts bias later source memory despite high immediate retrieval performance

We first examined memory performance during the intervening retrieval episode. Across the 1,292 source-tested items included in the RSA-linked dataset, participants responded Remember on 89.0% of trials, Familiar on 6.7%, and New on 4.3%. The proportion of studied items recognized as Remember or Familiar was 96.3% on consistent-context trials and 95.0% on shifted-context trials (Figure 2A). Context shift did not significantly affect recognition ($\beta = -0.31$, $p = .278$) or the likelihood of a Remember rather than Familiar response among recognized items ($\beta = -0.25$, $p = .312$).

Associative memory during immediate retrieval was also high. In City, participants placed objects within 20 track units of their original locations on 88.1% of consistent-context trials and 88.6% of shifted-context trials. Neither continuous location error ($\beta = -0.02$, $p = .860$) nor target accuracy ($\beta = 0.13$, $p = .617$) differed significantly by condition. In Soccer, participants correctly identified the original character on 92.8% of consistent-context trials and 93.1% of shifted-context trials ($\beta = 0.15$, $p = .690$). Thus, we did not detect a context-shift cost in immediate item-associate retrieval.

Context shifts had a much larger effect on the later source test. Overall source accuracy was 52.2%, but performance differed substantially by context condition: participants selected the original context on 63.3% of consistent-context trials compared with 41.0% of shifted-context trials (Figure 2B). A 2 × 2 repeated-measures ANOVA with Task (City, Soccer) and Condition (consistent, shifted) as within-subject factors revealed a significant main effect of Condition, $F(1,17) = 61.20$, $p = 4.94 \times 10^{-7}$, partial $\eta^2 = .78$. Neither the main effect of

Neuroscience Accepted Manuscript

Task nor the Task × Condition interaction was significant (both $p > .65$). Context shifts therefore substantially impaired later source memory despite the high immediate recognition.

The context shift also changed which source participants selected. On shifted trials, responses were distributed across the original (41%), cued (36%), and uninvolved (22%) contexts (Figure 2C). Among source errors, participants selected the cued context more often than the uninvolved context. A 2 × 2 repeated-measures ANOVA with Task and Error Type (cued, uninvolved) confirmed this bias, main effect of Error Type: $F(1,17) = 11.16$, $p = .0039$, partial $\eta^2 = .40$, with no significant task effect or interaction (both $p \geq .885$) (Figure 2D). This cued-context bias was observed in both tasks, City: $t(17) = 2.95$, $p = .0091$, $d = 0.69$; Soccer: $t(17) = 2.37$, $p = .0297$, $d = 0.56$. Context shifts therefore did more than reduce source accuracy: they biased errors toward the context in which the item had been re-encountered.

Immediate retrieval performance also did not reliably predict which source was selected later. On shifted trials, Remember/Familiar/New responses did not significantly predict final source accuracy (both *p* values ≥ .472) or either source-choice contrast (all *p* values ≥ .133; Table S1). Neither City location performance nor Soccer character accuracy significantly predicted final source accuracy or either source-choice contrast (all *p* values ≥ .237; Table S1). These results provide no evidence that the later source-memory bias was explained by the measured immediate retrieval responses. However, immediate recognition and associative accuracy were high, which limits the sensitivity of these analyses to detect differences in retrieval success.

Neural analyses were evaluated across 12 a priori hippocampal, medial temporal, posterior-medial, and frontal/control ROIs, with FDR correction applied within each analysis family. Four measures captured different aspects of cue-period activity. Original-context Fidelity measured similarity to the original encoding context. Present-vs-Past Alignment measured relative similarity to the cued versus original context. Contextual Engagement measured the overall expression of contextual information, regardless of which context was favored. Item Reinstatement measured item-specific encoding-to-cue similarity. Primary neural-behavioral analyses pooled City and Soccer, with task included as a covariate in trial-level models.

Encoding patterns reflect original-context structure, particularly in the City task

We first tested whether encoding activity patterns distinguished items encountered in the same versus different original contexts. Original-context organization was defined as greater encoding-pattern similarity for item pairs encountered in the same original context than for item pairs encountered in different original contexts (Figure 3A). Positive values therefore indicate that encoding patterns clustered according to the environment in which items were originally encountered. Across tasks, original-context organization survived correction in 3 of 12 ROIs: right posterior hippocampus, right anterior hippocampus, and dACC (all $q \leq .021$; Figure 3B). Complete ROI-wise statistics and representational pair coverage are reported in Table S2.

Original-context organization was strongest in the City task. In City, original-context organization survived correction in 10 of 12 ROIs, spanning hippocampal, medial temporal cortical, posterior-medial, and frontal/control regions. In Soccer, effects were uniformly small and no ROI survived correction. Direct City–Soccer comparisons confirmed stronger original-context organization in right posterior hippocampus, right anterior hippocampus, dACC, and right entorhinal/perirhinal cortex (all $q \leq .046$; Figure 3C).

Cue-period activity is broadly organized by the context present at cueing

During immediate cueing, context organization was evident across most ROIs. It was defined as greater cue-period pattern similarity for trials sharing the same cued context than for trials with different cued contexts (Figure 3D). Across tasks, the effect survived correction in 11 of 12 ROIs (Figure 3E). The largest effects were observed in dlPFC, right posterior hippocampus, and precuneus (all $q \leq .001$). Only left parahippocampal cortex did not survive correction. Complete ROI-wise statistics are reported in Table S2.

No ROI showed a corrected City–Soccer difference in cued-context organization (Figure 3F). Because the contextual background was visible during cueing, these effects could reflect perceptual processing, retrieval-

Neurosci Accepted Manuscript

related contextual representations, or both. To ask whether cue-period activity also carried information about the earlier context, we next compared each shifted-trial pattern with encoding-derived templates for the original and cued contexts.

Shifted-trial cue-period patterns favor the intervening context over the original context

On shifted trials, the original and cued contexts provided competing templates. We compared each cue-period pattern with encoding-derived templates for both contexts and defined Present-vs-Past Alignment as cued-context similarity minus original-context similarity. Positive values indicate stronger alignment with the context present during immediate cueing (Figure 4A).

Across the ROI set, shifted-trial cue-period patterns showed greater similarity to the cued-context template than to the original-context template, $dz = 0.67$, $t(17) = 2.85$, $p = .011$ (Figure 4B). This present-over-past alignment was strongest in left posterior hippocampus, $dz = 1.02$, $q = .006$, with convergent effects in left parahippocampal cortex, $dz = 0.82$, $q = .017$, and retrosplenial cortex, $dz = 0.78$, $q = .017$ (Figure 4C). Complete ROI-wise Present-vs-Past Alignment and Item Reinstatement statistics are reported in Table S2.

Cue-period representational measures predict later source attribution

We next related these cue-period measures to later source memory in two ways. Across participants, we asked whether Original-context Differentiation tracked source accuracy after a shift. Within participants, we asked whether trial-to-trial variation in the cue-period measures predicted the source selected later for that item.

Across participants, stronger Original-context Differentiation predicted higher source accuracy on shifted-context trials in right posterior hippocampus, $\rho = .63$, 95% CI [.20, .87], $q = .045$; left parahippocampal cortex, $\rho = .58$, $q = .045$; right parahippocampal cortex, $\rho = .57$, $q = .045$; and dACC, $\rho = .56$, $q = .045$ (Figure 5A). Because these were across-participant correlations, we assessed robustness with bootstrap confidence intervals and leave-one-subject-out resampling. All four correlations remained positive across every leave-one-subject-out sample; for example, the right posterior hippocampal correlation ranged from $\rho = .57$ to .74. In right posterior hippocampus, the relationship held after controlling for cued-context organization, partial $\rho = .67$, and after jointly controlling for cued-context organization, Contextual Engagement, and the number of template trials, partial $\rho = .60$. No Present-vs-Past Alignment brain–behavior correlation survived correction. Complete ROI-wise correlations, bootstrap intervals, leave-one-subject-out ranges, and partial correlations are reported in Table S4.

At the trial level, we modeled which source participants later selected for each item. Because original and cued contexts diverged on shifted trials, each later response could be classified as an original-context choice, a cued-context intrusion, or an uninvolved-context response. We fit mixed-effects logistic models for three source-choice contrasts: original versus cued choices, original versus not-original choices, and, among source errors, cued versus uninvolved choices. These models included all 18 participants and repeated observations within participants: 501 original-vs-cued observations, 646 original-vs-not-original observations, and 381 source-error observations. Each model included Contextual Engagement, Present-vs-Past Alignment, and Item Reinstatement, together with task, run, and within-run trial position. Complete ROI-wise coefficients, trial counts, and predictor × task interaction tests are reported in Table S3.

In precuneus, greater Contextual Engagement predicted a higher likelihood of selecting the original context rather than the cued context ($\beta = 0.92$, $SE = 0.31$, $z = 2.94$, $q = .042$; Figure 5B). Neither Present-vs-Past Alignment nor Item Reinstatement survived correction in this model (Table 1). We then entered original- and cued-template similarity as separate predictors to examine whether the engagement effect was driven by one context. Both coefficients were positive in precuneus and retrosplenial cortex, although only original-template similarity survived correction in the original-versus-not-original model. The Contextual Engagement effect was not driven solely by preferential similarity to the original context, although the decomposition cannot establish

JNeurosci Accepted Manuscript

independent contributions of both contextual representations. Complete template-decomposition models are reported in Table S4.

In dACC, Present-vs-Past Alignment predicted a different type of source error. Among source errors, greater Present-vs-Past Alignment predicted a lower likelihood of choosing the cued context rather than the uninvolved context ($\beta = -0.28$, SE = 0.09, $z = -3.03$, $q = .032$; Figure 5C; Table 1). Stronger cued-over-original alignment was associated with uninvolved-context errors rather than more frequent cued-context intrusions. Item Reinstatement did not survive correction in this or any of the other planned source-choice models.

The focal precuneus and dACC effects were largely unchanged after accounting for immediate Remember/Familiar responses. The precuneus Contextual Engagement coefficient remained similar ($\beta = 0.97$ before adjustment, $p = .0033$; $\beta = 0.98$ after adjustment, $p = .0030$; n = 370 trials). The dACC Present-vs-Past Alignment coefficient was also largely unchanged ($\beta = -0.315$ before adjustment, $p = .0016$; $\beta = -0.311$ after adjustment, $p = .0020$; n = 282 trials). Task-specific associative-memory adjustments produced similar dACC estimates, although the City continuous-error model was no longer nominally significant after adjustment. The precuneus effect remained significant in Soccer but was not significant within City before or after adjustment. Complete matched-sample sensitivity analyses are reported in Table S5.

We also tested whether immediate retrieval responses were directly related to the cue-period measures. A small number of ROI-wise associations survived correction. Remember responses were associated with greater original-context template similarity than Familiar responses in right anterior hippocampus and dACC (both $q = .038$), whereas New responses were associated with greater dACC Contextual Engagement than Familiar responses ($q = .021$). City target accuracy did not show the predicted positive relationship with original-context similarity; the only corrected association was negative in right posterior hippocampus ($q = .045$). No corrected associations were observed for continuous City location quality or Soccer character accuracy (Table S5). These response-stratified analyses were secondary and should be interpreted cautiously. Remember responses predominated, leaving relatively few Familiar and New trials. The primary source-choice models did not depend on these response categories, and the focal precuneus and dACC coefficients changed little when Remember/Familiar response was added to the models.

Context-state signals covary across hippocampal–cortical regions

Context-state measures also covaried across regions from trial to trial. For each shifted-context trial, we estimated trial-wise covariation between every ROI pair separately for Contextual Engagement, Original-context Fidelity, Present-vs-Past Alignment, and Item Reinstatement. Across the full 12-ROI network, all three context-state measures showed reliable positive trial-wise covariation (Figure 6A): Contextual Engagement, $M = 0.32$, $t(17) = 15.6$, $q = 5 \times 10^{-11}$; Original-context Fidelity, $M = 0.31$, $t(17) = 13.8$, $q = 2 \times 10^{-10}$; and Present-vs-Past Alignment, $M = 0.35$, $t(17) = 14.8$, $q = 8 \times 10^{-11}$. Item Reinstatement also showed positive cross-region covariation, $M = 0.16$, $t(17) = 12.1$, $q = 1 \times 10^{-9}$, but its magnitude was smaller than that of each context-state measure. Complete task-difference analyses are reported in Table S6.

No exploratory relationship between cross-region covariation and source behavior survived correction. Covariation also did not moderate the relationship between Original-context Differentiation and source accuracy.

DISCUSSION

Remembering often begins before an explicit memory decision is made. Here, re-encountering an item in a different context reduced later source accuracy and biased errors toward the intervening context, despite high immediate recognition and associative-memory performance. Cue-period neural measures were also associated with later source outcomes. During cueing, activity patterns were broadly organized by the currently visible context, but information related to the original context remained detectable. Across participants, greater Original-context Differentiation between consistent and shifted cueing was associated with better later source

JNeurosci Accepted Manuscript

memory. Trial-level analyses linked precuneus Contextual Engagement and dACC Present-vs-Past Alignment to different later source choices. Source attribution therefore tracked the contextual representations expressed when an item was re-encountered, rather than item reinstatement alone.

Successful remembering is accompanied by reactivation of encoding-related activity patterns across sensory cortex, medial temporal lobe, and distributed cortical networks (Gordon et al., 2014; Rissman and Wagner, 2012; Staresina et al., 2012; Tompary et al., 2016; Wing et al., 2015; Xue, 2018). Here, encoding-related reinstatement coexisted with representations of the current environment. Reactivation can stabilize previously encoded information (Tambini et al., 2010; Tambini and Davachi, 2019), support updating or integration (Bridge and Voss, 2014; Hupbach et al., 2007; Hupbach et al., 2008; Schlichting and Preston, 2016), or contribute to later distortion or misattribution (St Jacques et al., 2013). The temporal structure of the task matters here. We measured cue-period activity while participants re-encountered studied objects in consistent or shifted contexts, but tested original-context memory only later, on a neutral screen. Because source memory was tested later on a neutral screen, the cued-context bias cannot be attributed to a background that remained visible during the source decision. Instead, the intervening cueing episode provided an additional source of contextual information that could influence later source attribution. This fits source-monitoring accounts in which memory judgments depend on attributing recovered information to its source (Johnson et al., 1993), and constructive-memory frameworks in which memories are assembled from multiple sources of information and can be distorted when competing contextual information is available (Bartlett, 1932; Schacter et al., 1998; Schacter et al., 2011; Schacter, 2012). Consistent with this account, neural reinstatement of a prior context has been shown to predict later source misattribution (Gershman et al., 2013).

The behavioral data argue against a simple retrieval-failure explanation. Context shifts did not significantly affect studied-item recognition, City location accuracy, or Soccer character accuracy, and immediate retrieval performance did not reliably predict source choices. The later source bias occurred without measurable deficits in item recognition or task-specific associative memory during the intervening tests. Because immediate performance was high and these measures were obtained after the 2-sec cue period, smaller differences in retrieval quality, attention, or other unmeasured processes could still contribute to the observed neural relationships.

Encoding and cueing showed different forms of context organization. Original-context organization was stronger during City encoding, whereas cue-period patterns were organized by the currently visible context across both City and Soccer. This does not mean that context was more strongly represented during retrieval. The phases differed in stimulus presentation and task demands, and cue-period organization could reflect perceptual processing of the visible background, retrieval-related contextual processing, or both. City's stronger encoding effects may partially reflect its spatially organized contexts, given the roles of hippocampal, parahippocampal, and retrosplenial regions in spatial and situational context (Epstein, 2008; Ranganath and Ritchey, 2012; Ritchey et al., 2015) and of hippocampal and posterior-medial regions in tracking event structure (Baldassano et al., 2017; Cohn-Sheehy et al., 2021; Tarder-Stoll et al., 2024). Both tasks contained dynamic visual input, but that input was structured differently. City views changed as participants moved through a spatial layout, whereas Soccer backgrounds continuously changed in shape and texture within each color context. The tasks also differed in contextual richness and layout, so the present data cannot isolate which feature drove the task difference.

During immediate cueing, activity patterns were more similar to encoding-derived templates for the cued than original context. Because the cued background was visible, this could reflect perceptual processing, retrieval-related contextual processing, or both. The original-context template provides a different comparison. On shifted trials, the original background was absent, yet information related to that context remained detectable. Even when cue-period patterns were more aligned with the current context, information about the prior context remained detectable.

Journal Accepted Manuscript

At the participant level, however, the brain–behavior relationship was not a simple shifted-trial fidelity effect. Better source memory was associated with greater Original-context Differentiation (i.e., greater separation in original-template similarity between consistent and shifted cueing) in posterior hippocampus, parahippocampal cortex, and dACC. These relationships remained positive after controlling for cued-context organization, Contextual Engagement, and template-trial counts. Differentiation is a condition contrast; it cannot tell us whether better memory reflects higher original-context similarity during consistent cueing, lower similarity after a shift, or both.

The continued detectability of original-context information despite a changed background is consistent with transformed-reinstatement accounts, in which retrieval-related activity can be altered by current cues and goals (Favila et al., 2020; Xiao et al., 2017). Competing mnemonic representations can also be differentially expressed or suppressed during retrieval (Kuhl et al., 2011; Wimber et al., 2015). In our paradigm, the competition was contextual rather than item-level. The continued representation of prior contextual information is likewise compatible with contextual-binding accounts, in which the hippocampus binds items to contexts and supports later recovery of those associations (Davachi, 2006; Ranganath, 2010; Yonelinas et al., 2019). Related experiences can also become more overlapping with learning and consolidation (Tompary and Davachi, 2017; Tompary and Davachi, 2024), providing another way for past and present contextual information to coexist.

Precuneus and dACC showed different relationships with later source choice. In precuneus, stronger Contextual Engagement predicted original-context rather than cued-context choices. Decomposing Contextual Engagement into its original- and cued-template components suggested that the precuneus effect was not simply an original-context reinstatement effect, although only original-template similarity survived correction in the original-versus-not-original model. This fits evidence that posterior-medial systems carry event-level and situation-level context (Ranganath and Ritchey, 2012; Ritchey et al., 2015), and track event-level structure across perception and recall (Chen et al., 2017; Tarder-Stoll et al., 2024).

A different pattern emerged in dACC. Greater cued-over-original alignment predicted uninvolved-context rather than cued-context errors. One possibility is that this relationship reflects control demands or uncertainty elicited by competing contextual information, consistent with conflict-monitoring accounts of anterior cingulate function (Botvinick, 2007). Frontal control systems contribute to monitoring and selecting among competing representations (Badre and Wagner, 2007; Mitchell and Johnson, 2009; Moscovitch and Winocur, 2002). Alternatively, alignment may reflect mismatch-related processing triggered by a changed context (Bein et al., 2020; Gershman et al., 2014). Future work measuring neural activity at both cueing and final decision will be important for distinguishing conflict-monitoring, prediction-error, and decision-uncertainty accounts of this effect.

The precuneus and dACC effects changed little after adjustment for Remember/Familiar responses, although task-specific associative-memory adjustments were more variable. A few associations between immediate responses and cue-period measures survived correction, but these response-stratified findings are difficult to interpret because Remember responses predominated and Familiar and New trials were sparse. The primary source-choice models used all eligible shifted trials and did not depend on these response categories.

Context-state measures also showed reliable trial-wise covariation across pairs of ROIs, stronger than that observed for Item Reinstatement. The pattern was broadly distributed rather than confined to posterior-medial, anterior-MTL, or prefrontal/control groupings. This broad distribution is consistent with work showing distributed representations and large-scale network organization during episodic retrieval (Rissman and Wagner, 2012; Westphal et al., 2017). Cross-region covariation did not show corrected relationships with source behavior. These effects show shared trial-to-trial variation across regions. They should not be interpreted as functional connectivity or as a mechanism for the source-memory bias.

The final sample was modest, and the participant-level correlations should be interpreted cautiously despite their bootstrap and leave-one-subject-out stability. The trial-rich, within-subject design provided stronger

J Neurosci. Accepted Manuscript

leverage for the mixed-effects trial-level analyses. City and Soccer also differed in more than contextual domain, limiting attribution of task differences to spatial structure alone. Finally, we measured neural activity during immediate cueing but not the final source decision. The present study examined only one intervening episode. Future work should test what happens after repeated shifts: whether reactivation produces durable updating (Hupbach et al., 2007; Schlichting and Preston, 2016; Tambini and Davachi, 2019), and whether repeated contextual changes integrate memories within new latent or event structures (Gershman et al., 2014; Schapiro et al., 2017). It will also be important to measure whether the intervening context is reinstated again at the final source decision.

In sum, source memory was biased by the context in which an item was re-encountered before the later source decision. Cue-period patterns were strongly organized by the intervening context, Original-context Differentiation was associated with better memory across participants, and distinct contextual signals predicted later source choices. Remembering is therefore not simply a return to the original event; it unfolds in the present and can incorporate contextual information encountered after the event itself.

Data availability

De-identified behavioral data and ROI-level summary data needed to reproduce the reported analyses will be made available upon publication. Raw MRI data are not publicly available because participant consent did not include unrestricted public data sharing; requests for access to minimally processed data should be directed to the corresponding author and will be considered in accordance with institutional and IRB requirements.

Code availability

Analysis code used to generate the reported results and figures will be made available upon publication and is available from the corresponding author upon reasonable request.

Funding Statements

This project was supported by NIMH R01-MH112733 (Davachi), the Columbia University Zuckerman Institute MRI Seed Program (Davachi and Hermiller), and the Florida State University First Year Professor Award (Hermiller).

Author contributions

M.S.H. and L.D. conceived and designed the study. M.S.H. collected and analyzed the data. M.S.H. drafted the manuscript. M.S.H. and L.D. interpreted the results and revised the manuscript.

Competing interests

The authors declare no competing interests.

Preregistration

This study was not preregistered.

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.

Badre, D., Wagner, A.D., 2007. Left ventrolateral prefrontal cortex and the cognitive control of memory. Neuropsychologia. 45**,** 2883-901.

Baldassano, C., et al., 2017. Discovering Event Structure in Continuous Narrative Perception and Memory. Neuron. 95**,** 709-721 e5.

JNeurosci Accepted Manuscript

Bartlett, F.C., 1932. Remembering: A Study in Experimental and Social Psychology., Vol., Cambridge University Press.
Bein, O., Duncan, K., Davachi, L., 2020. Mnemonic prediction errors bias hippocampal states. Nat Commun. 11, 3451.
Botvinick, M.M., 2007. Conflict monitoring and decision making: reconciling two perspectives on anterior cingulate function. Cogn Affect Behav Neurosci. 7, 356-66.
Bridge, D.J., Voss, J.L., 2014. Hippocampal binding of novel information with dominant memory traces can support both memory stability and change. J Neurosci. 34, 2203-13.
Brodeur, M.B., Guerard, K., Bouras, M., 2014. Bank of Standardized Stimuli (BOSS) phase II: 930 new normative photos. PLoS One. 9, e106953.
Cai, M.B., et al., 2019. Representational structure or task structure? Bias in neural representational similarity analysis and a Bayesian method for reducing bias. PLoS Comput Biol. 15, e1006299.
Chen, J., et al., 2017. Shared memories reveal shared structure in neural activity across individuals. Nat Neurosci. 20, 115-125.
Clewett, D., Gasser, C., Davachi, L., 2020. Pupil-linked arousal signals track the temporal organization of events in memory. Nat Commun. 11, 4007.
Cohn-Sheehy, B.I., et al., 2021. The hippocampus constructs narrative memories across distant events. Curr Biol. 31, 4935-4945 e7.
Cox, R.W., 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput Biomed Res. 29, 162-73.
Danker, J.F., Anderson, J.R., 2010. The ghosts of brain states past: remembering reactivates the brain regions engaged during encoding. Psychol Bull. 136, 87-102.
Davachi, L., 2006. Item, context and relational episodic encoding in humans. Curr Opin Neurobiol. 16, 693-700.
Diana, R.A., Yonelinas, A.P., Ranganath, C., 2007. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. Trends Cogn Sci. 11, 379-86.
Diedrichsen, J., Kriegeskorte, N., 2017. Representational models: A common framework for understanding encoding, pattern-component, and representational-similarity analysis. PLoS Comput Biol. 13, e1005508.
Dimsdale-Zucker, H.R., Ranganath, C., 2018. Representational Similarity Analyses. In: Handbook of in Vivo Neural Plasticity Techniques. Handbook of Behavioral Neuroscience, Vol., D. Manahan-Vaughan, ed.^eds. Elsevier, pp. 509-525.
DuBrow, S., Davachi, L., 2013. The influence of context boundaries on memory for the sequential order of events. J Exp Psychol Gen. 142, 1277-86.
DuBrow, S., Davachi, L., 2014. Temporal memory is shaped by encoding stability and intervening item reactivation. J Neurosci. 34, 13998-4005.
Eichenbaum, H., Yonelinas, A.P., Ranganath, C., 2007. The medial temporal lobe and recognition memory. Annu Rev Neurosci. 30, 123-52.
Epstein, R.A., 2008. Parahippocampal and retrosplenial contributions to human spatial navigation. Trends Cogn Sci. 12, 388-96.
Ezzyat, Y., Davachi, L., 2011. What constitutes an episode in episodic memory? Psychol Sci. 22, 243-52.
Fan, L., et al., 2016. The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture. Cereb Cortex. 26, 3508-26.
Favila, S.E., Lee, H., Kuhl, B.A., 2020. Transforming the Concept of Memory Reactivation. Trends Neurosci. 43, 939-950.
Folkerts, S., Rutishauser, U., Howard, M.W., 2018. Human Episodic Memory Retrieval Is Accompanied by a Neural Contiguity Effect. J Neurosci. 38, 4200-4211.
Gershman, S.J., et al., 2013. Neural context reinstatement predicts memory misattribution. J Neurosci. 33, 8590-5.

Neuron. Accepted Manuscript

Gershman, S.J., et al., 2014. Statistical computations underlying the dynamics of memory updating. PLoS Comput Biol. 10, e1003939.
Godden, D.R., Baddeley, A.D., 1975. Context-dependent memory in two natural environments: on land and underwater. British Journal of Psychology. 66, 325-331.
Gordon, A.M., et al., 2014. Cortical reinstatement mediates the relationship between content-specific encoding activity and subsequent recollection decisions. Cereb Cortex. 24, 3350-64.
Howard, M.W., Kahana, M.J., 2002. A distributed representation of temporal context. Journal of Mathematical Psychology. 46, 269-299.
Hupbach, A., et al., 2007. Reconsolidation of episodic memories: a subtle reminder triggers integration of new information. Learn Mem. 14, 47-53.
Hupbach, A., et al., 2008. The dynamics of memory: context-dependent updating. Learn Mem. 15, 574-9.
Jo, H.J., et al., 2010. Mapping sources of correlation in resting state FMRI, with artifact detection and removal. Neuroimage. 52, 571-82.
Johnson, J.D., et al., 2009. Recollection, familiarity, and cortical reinstatement: a multivoxel pattern analysis. Neuron. 63, 697-708.
Johnson, M.K., Hashtroudi, S., Lindsay, D.S., 1993. Source monitoring. Psychol Bull. 114, 3-28.
Kragel, J.E., Voss, J.L., 2021. Temporal context guides visual exploration during scene recognition. J Exp Psychol Gen. 150, 873-889.
Kriegeskorte, N., Mur, M., Bandettini, P., 2008. Representational similarity analysis - connecting the branches of systems neuroscience. Front Syst Neurosci. 2, 4.
Kuhl, B.A., et al., 2011. Fidelity of neural reactivation reveals competition between memories. Proc Natl Acad Sci U S A. 108, 5903-8.
Lohnas, L.J., Healey, M.K., Davachi, L., 2023. Neural temporal context reinstatement of event structure during memory recall. J Exp Psychol Gen. 152, 1840-1872.
Manning, J.R., et al., 2011. Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. Proc Natl Acad Sci U S A. 108, 12893-7.
Mitchell, K.J., Johnson, M.K., 2009. Source monitoring 15 years later: what have we learned from fMRI about the neural mechanisms of source memory? Psychol Bull. 135, 638-77.
Moscovitch, M., Winocur, G., 2002. The frontal cortex and working with memory. In: Principles of frontal lobe function. Vol., S. Knight, ed.^eds. Oxford University Press, pp. 188-209.
Mumford, J.A., et al., 2012. Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses. Neuroimage. 59, 2636-43.
Polyn, S.M., Norman, K.A., Kahana, M.J., 2009. A context maintenance and retrieval model of organizational processes in free recall. Psychol Rev. 116, 129-56.
Postman, L., Underwood, B.J., 1973. Critical issues in interference theory. Mem Cognit. 1, 19-40.
Ranganath, C., 2010. A unified framework for the functional organization of the medial temporal lobes and the phenomenology of episodic memory. Hippocampus. 20, 1263-90.
Ranganath, C., Ritchey, M., 2012. Two cortical systems for memory-guided behaviour. Nat Rev Neurosci. 13, 713-26.
Reagh, Z.M., Ranganath, C., 2023. Flexible reuse of cortico-hippocampal representations during encoding and recall of naturalistic events. Nat Commun. 14, 1279.
Rissman, J., Wagner, A.D., 2012. Distributed representations in memory: insights from functional brain imaging. Annu Rev Psychol. 63, 101-28.
Ritchey, M., et al., 2013. Neural similarity between encoding and retrieval is related to memory via hippocampal interactions. Cereb Cortex. 23, 2818-28.
Ritchey, M., Libby, L.A., Ranganath, C., 2015. Cortico-hippocampal systems involved in memory and cognition: the PMAT framework. Prog Brain Res. 219, 45-64.
Rugg, M.D., Yonelinas, A.P., 2003. Human recognition memory: a cognitive neuroscience perspective. Trends Cogn Sci. 7, 313-319.

Neuroscience Accepted Manuscript

Schacter, D.L., Norman, K.A., Koutstaal, W., 1998. The cognitive neuroscience of constructive memory. Annu Rev Psychol. 49, 289-318.
Schacter, D.L., Guerin, S.A., St Jacques, P.L., 2011. Memory distortion: an adaptive perspective. Trends Cogn Sci. 15, 467-74.
Schacter, D.L., 2012. Adaptive constructive processes and the future of memory. Am Psychol. 67, 603-13.
Schapiro, A.C., et al., 2017. Complementary learning systems within the hippocampus: a neural network modelling approach to reconciling episodic memory with statistical learning. Philos Trans R Soc Lond B Biol Sci. 372, 20160049.
Schlichting, M.L., Mumford, J.A., Preston, A.R., 2015. Learning-related representational changes reveal dissociable integration and separation signatures in the hippocampus and prefrontal cortex. Nat Commun. 6, 8151.
Schlichting, M.L., Preston, A.R., 2016. Hippocampal-medial prefrontal circuit supports memory updating during learning and post-encoding rest. Neurobiol Learn Mem. 134 Pt A, 91-106.
Smith, S.M., Vela, E., 2001. Environmental context-dependent memory: a review and meta-analysis. Psychon Bull Rev. 8, 203-20.
St Jacques, P.L., Olm, C., Schacter, D.L., 2013. Neural mechanisms of reactivation-induced updating that enhance and distort memory. Proc Natl Acad Sci U S A. 110, 19671-8.
Staresina, B.P., et al., 2012. Episodic reinstatement in the medial temporal lobe. J Neurosci. 32, 18150-6.
Tambini, A., Ketz, N., Davachi, L., 2010. Enhanced brain correlations during rest are related to memory for recent experiences. Neuron. 65, 280-90.
Tambini, A., Davachi, L., 2019. Awake Reactivation of Prior Experiences Consolidates Memories and Biases Cognition. Trends Cogn Sci. 23, 876-890.
Tarder-Stoll, H., Baldassano, C., Aly, M., 2024. The brain hierarchically represents the past and future during multistep anticipation. Nat Commun. 15, 9094.
Tompary, A., Duncan, K., Davachi, L., 2016. High-resolution investigation of memory-specific reinstatement in the hippocampus and perirhinal cortex. Hippocampus. 26, 995-1007.
Tompary, A., Davachi, L., 2017. Consolidation Promotes the Emergence of Representational Overlap in the Hippocampus and Medial Prefrontal Cortex. Neuron. 96, 228-241 e5.
Tompary, A., Davachi, L., 2024. Integration of overlapping sequences emerges with consolidation through medial prefrontal cortex neural ensembles and hippocampal-cortical connectivity. Elife. 13.
Tulving, E., Thomson, D.M., 1973. Encoding Specificity and Retrieval Processes in Episodic Memory. Psychological Review. 80, 352-373.
Underwood, B.J., 1957. Interference and forgetting. Psychol Rev. 64, 49-60.
Westphal, A.J., Wang, S., Rissman, J., 2017. Episodic Memory Retrieval Benefits from a Less Modular Brain Network Organization. J Neurosci. 37, 3523-3531.
Wimber, M., et al., 2015. Retrieval induces adaptive forgetting of competing memories via cortical pattern suppression. Nat Neurosci. 18, 582-9.
Wing, E.A., Ritchey, M., Cabeza, R., 2015. Reinstatement of individual past events revealed by the similarity of distributed activation patterns during encoding and retrieval. J Cogn Neurosci. 27, 679-91.
Xia, M., Wang, J., He, Y., 2013. BrainNet Viewer: a network visualization tool for human brain connectomics. PLoS One. 8, e68910.
Xiao, X., et al., 2017. Transformed Neural Pattern Reinstatement during Episodic Memory Retrieval. J Neurosci. 37, 2986-2998.
Xue, G., 2018. The Neural Representations Underlying Human Episodic Memory. Trends Cogn Sci. 22, 544-561.
Yonelinas, A.P., et al., 2019. A contextual binding theory of episodic memory: systems consolidation reconsidered. Nat Rev Neurosci. 20, 364-375.
Yushkevich, P.A., et al., 2015. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. Hum Brain Mapp. 36, 258-87.

JNeurosci Accepted Manuscript

914

915

JNeurosci Accepted Manuscript

Main Figure Legends

Figure 1. Task structure for dissociating original and cued contexts. (A) Participants completed six fMRI task blocks, three City blocks and three Soccer blocks, in randomized interleaved order. Each block contained four encoding loops followed by immediate cued retrieval and end-of-block memory tests. **(B)** During each encoding loop, participants encountered six items (2-sec each) embedded within a contextual background. **(C)** Contexts were defined by weather states in the City task and team-color states in the Soccer task. **(D)** During immediate cued retrieval, studied item cues were presented while the background either remained consistent with the original encoding context or shifted to a different cued context. **(E)** Source memory for the original context was tested later on a neutral gray screen, ensuring that the cued background was not visible at the time of the source decision. **(F)** A priori ROIs included anterior and posterior hippocampus (aHPC; pHPC), entorhinal/perirhinal cortex (EPRC), parahippocampal cortex (PHC), precuneus, retrosplenial cortex (RSC), dorsal anterior cingulate cortex (dACC), and dorsolateral prefrontal cortex (dlPFC). Medial temporal ROIs were analyzed separately by hemisphere; cortical ROIs were analyzed bilaterally. Note, character faces from *Animal Crossing* used in the Soccer task are occluded here for copyright reason.

Figure 2. Context shifts bias later source memory despite high immediate retrieval performance. (A) Remember, Familiar, and New response proportions for studied items during immediate retrieval, shown overall and separately for consistent- and shifted-context trials. **(B)** Source-memory accuracy was higher for consistent-context than shifted-context trials in both tasks. **(C)** On shifted-context trials, responses were distributed across the original, cued, and uninvolved contexts; among source errors, participants selected the cued context more often than the uninvolved context. **(D)** The cued-context intrusion bias was evident in both City and Soccer tasks. Points and lines show individual participants; large markers indicate group means. **Related Supplemental Material: Table S1.**

Figure 3. Encoding patterns reflect original-context structure, whereas immediate-cue-period activity patterns are broadly organized by the cued context. (A) Original-context organization was computed from encoding–encoding pattern similarity as greater similarity among item pairs from the same original context than from different original contexts. Schematic matrices show simulated data for conceptual illustration. **(B)** Across the 12 ROIs, original-context organization during encoding survived correction in right posterior hippocampus, right anterior hippocampus, and dACC. Error bars show ±95% confidence intervals; filled teal points indicate FDR-corrected effects; small dots indicate individual participants. **(C)** Original-context organization was stronger in City than Soccer in four ROIs: right posterior hippocampus, right anterior hippocampus, dACC, and right entorhinal/perirhinal cortex. **(D)** Cued-context organization was computed analogously from retrieval–retrieval cue-period pattern similarity based on the context present with the item cue. **(E)** Cued-context organization during immediate cued retrieval survived correction in 11 of 12 ROIs. Error bars show ±95% confidence intervals; filled teal points indicate FDR-corrected effects; small dots indicate individual participants. **(F)** No ROI showed a significant City–Soccer difference in cued-context organization. **Related Supplemental Material: Figure S1 and Table S2.**

Figure 4. Shifted-trial cue-period patterns favor the intervening context over the original context. **(A)** Present-vs-Past Alignment was computed by comparing each shifted-trial cue-period pattern with encoding-derived templates for the cued context and the original context. Positive values indicate greater similarity to the cued-context template than to the original-context template. **(B)** Across the ROI set, cue-period patterns showed stronger alignment with the cued context than with the original context; 15 of 18 participants showed this effect. **(C)** ROI-level effects, with strongest corrected effects in left posterior hippocampus, left

JNeurosci Accepted Manuscript

parahippocampal cortex, and retrosplenial cortex. Error bars show ±95% confidence intervals unless otherwise noted; filled points indicate FDR-corrected effects. **Related Supplemental Material: Table S2.**

JNeurosci Accepted Manuscript

Figure 5. Cue-period representational measures predict later source attribution across participants and trials. (A) Across participants, greater Original-context Differentiation was associated with higher shifted-trial source accuracy, with the strongest effect in right posterior hippocampus. Bottom: each point represents one participant; the fitted line is shown for visualization, and shading indicates a bootstrap 95% confidence interval. **(B)** On shifted-context trials, greater precuneus Contextual Engagement predicted later original-context choices over cued-context choices. Bottom: the curve shows the predicted probability of an original-context choice from the logistic model; shading indicates the 95% confidence ribbon; rugs show observed trial-level choices. **(C)** Among source errors, greater dACC Present-vs-Past Alignment predicted uninvolved responses rather than specific cued-context intrusions. Bottom: the curve shows the predicted probability of selecting the cued context over the uninvolved context; shading indicates the 95% confidence ribbon; rugs show observed trial-level choices. Filled forest-plot markers indicate FDR-corrected effects. **Related Supplemental Material: Tables S3-S5.**

Figure 6. Context-state signals covary across hippocampal–cortical regions. (A) Mean trial-wise covariation for Contextual Engagement, Original-context Fidelity, Present-vs-Past Alignment, and Item Reinstatement across all ROI pairs. Context-state measures showed stronger cross-ROI trial-wise covariation than Item Reinstatement. Points show individual participants; large markers show group means ±95% confidence intervals. **(B)** Mean context-state trial-wise covariation matrix across the 12 ROIs, averaged across Contextual Engagement, Original-context Fidelity, and Present-vs-Past Alignment. **(C)** Difference matrix showing context-state trial-wise covariation minus Item Reinstatement trial-wise covariation for each ROI pair. Positive values indicate stronger trial-wise covariation for context-state signals than for item-level reinstatement. **(D)** Subject-level context-state minus item-reinstatement trial-wise covariation difference across the full 12-ROI network. Values above zero indicate stronger trial-wise covariation for context-state signals than for Item Reinstatement. **Related Supplemental Material: Table S6.**

Main Table Legends

Table 1. Neural predictor coefficients for focal trial-level source-choice models. Coefficients are from mixed-effects logistic regression models for the precuneus original-versus-cued contrast and the dACC cued-versus-uninvolved error contrast. β values are log-odds coefficients; *z* and *p* are Wald tests; *q* values are FDR-corrected across the 12 ROIs within predictor and contrast. Complete ROI-wise results for all three planned source-choice contrasts are reported in Table S3.

JNeurosci Accepted Manuscript

998 **Table 1. Neural predictor coefficients for focal trial-level source-choice models.**

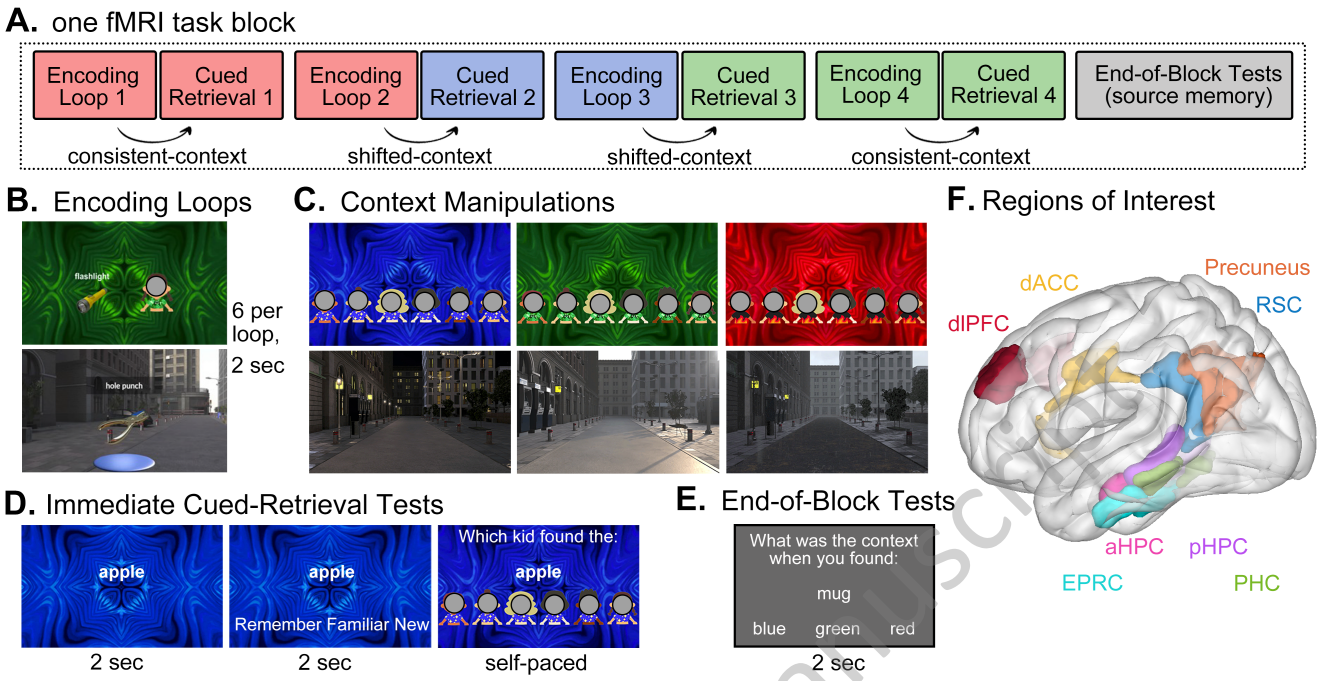
999

ROI / contrast	Predictor	β	SE	z	p	q
Precuneus / original vs cued	Contextual Engagement	0.92	0.31	2.94	.003	.042
	Present-vs-Past Alignment	0.00	0.08	0.06	.949	.965
	Item Reinstatement	0.28	0.14	1.99	.047	.565
dACC / cued vs uninvolved	Contextual Engagement	−0.17	0.41	−0.43	.671	.944
	Present-vs-Past Alignment	−0.28	0.09	−3.03	.003	.032
	Item Reinstatement	−0.06	0.20	−0.29	.769	.903

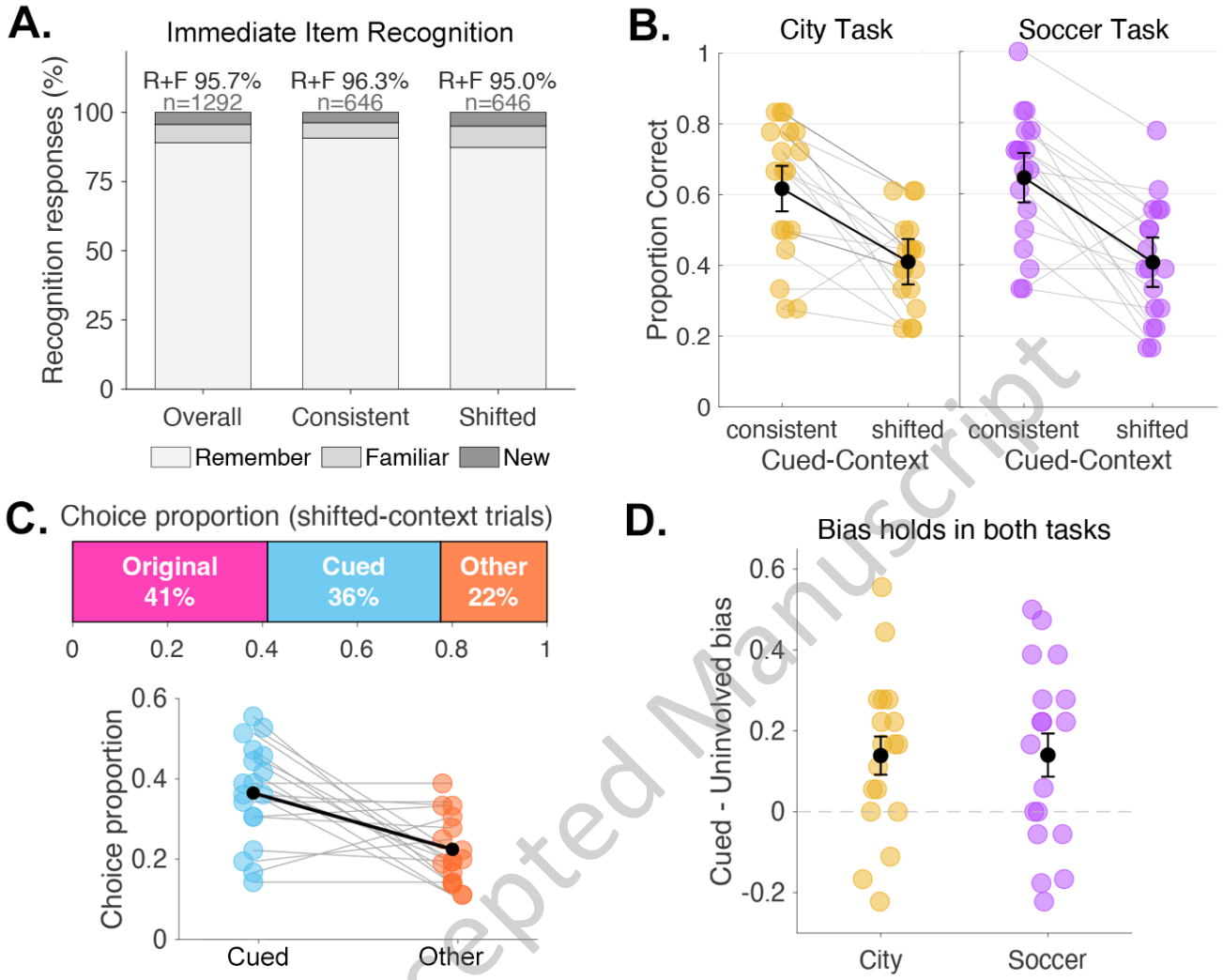
000

001

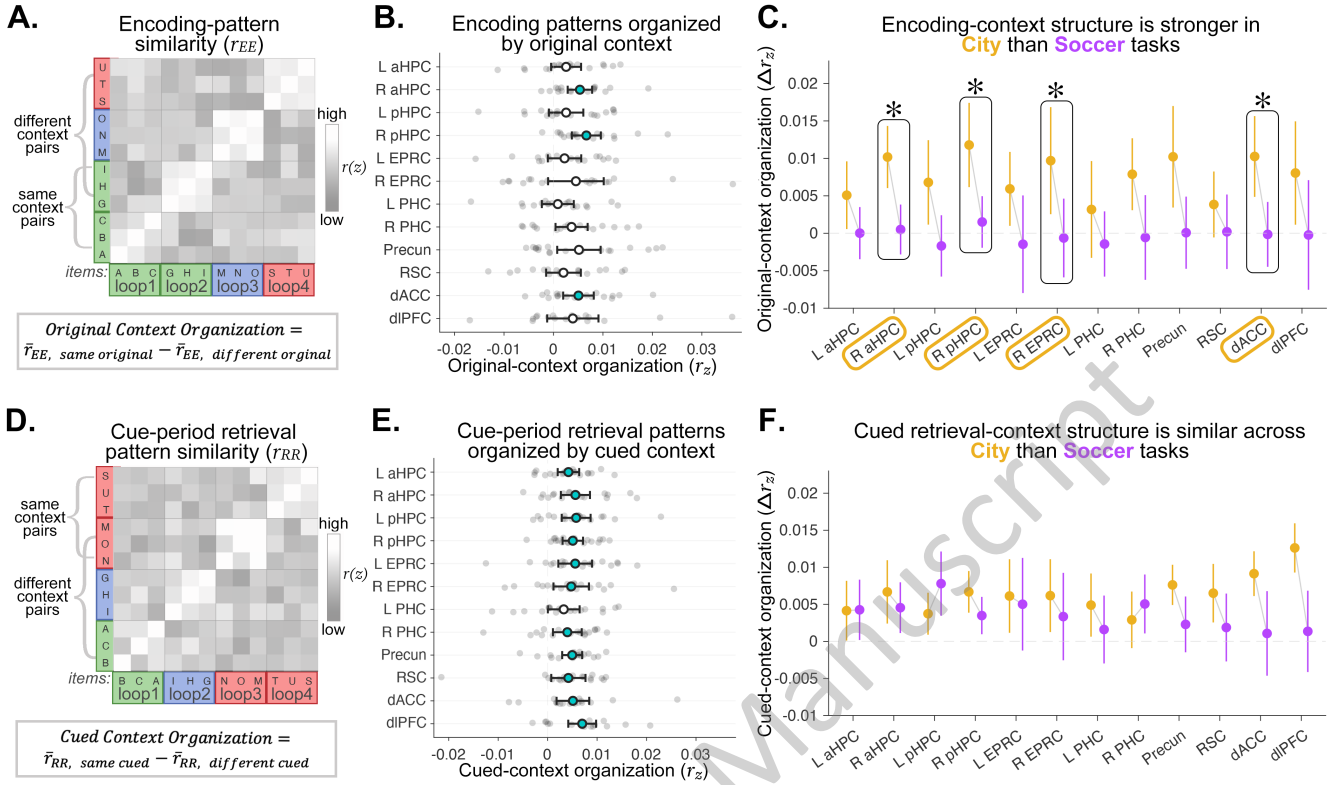
JNeurosci Accepted Manuscript



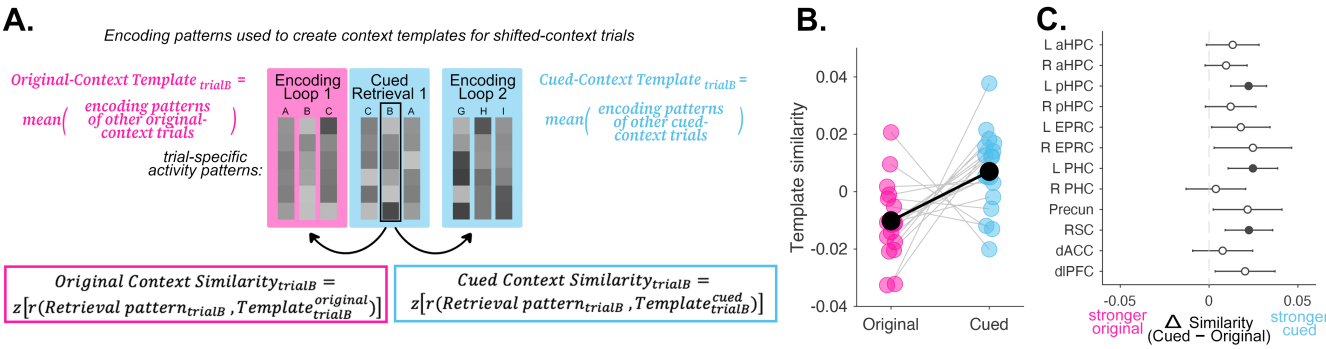
JNeurosci Accepted Manuscript



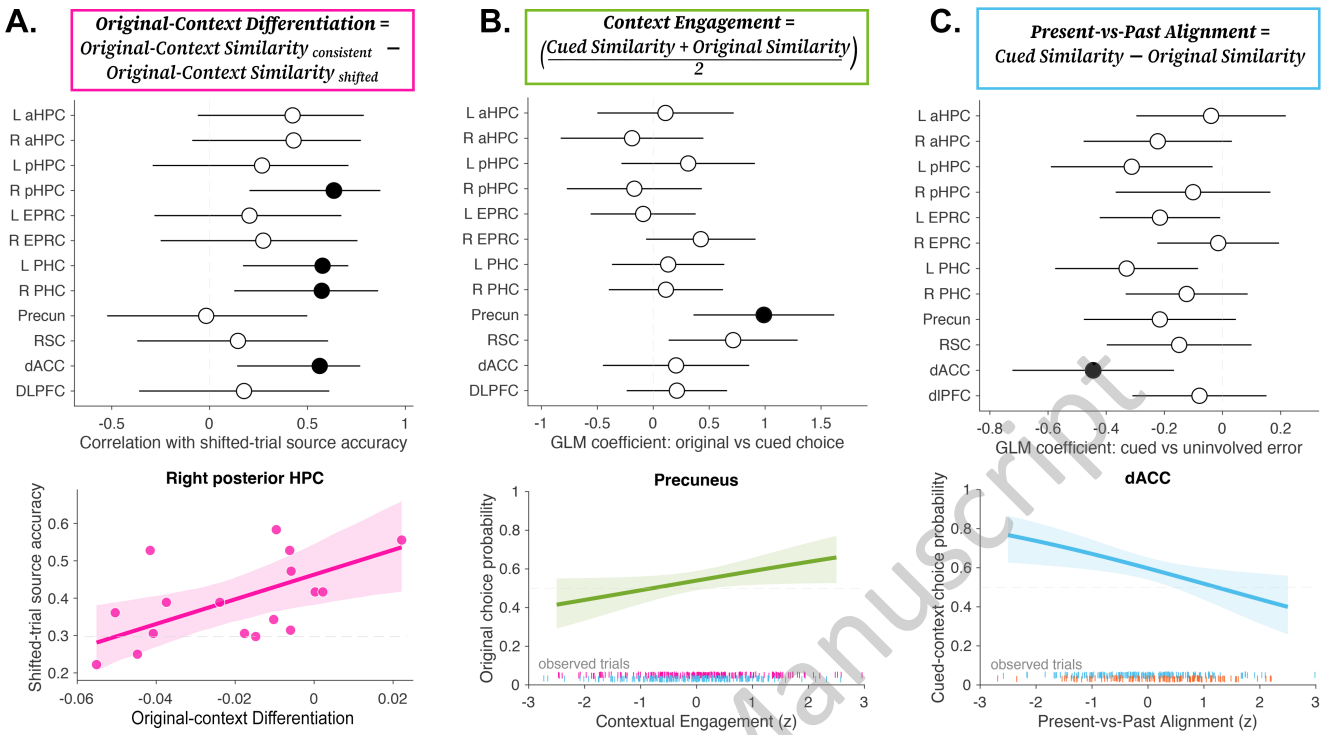
JNeurosci Accepted Manuscript



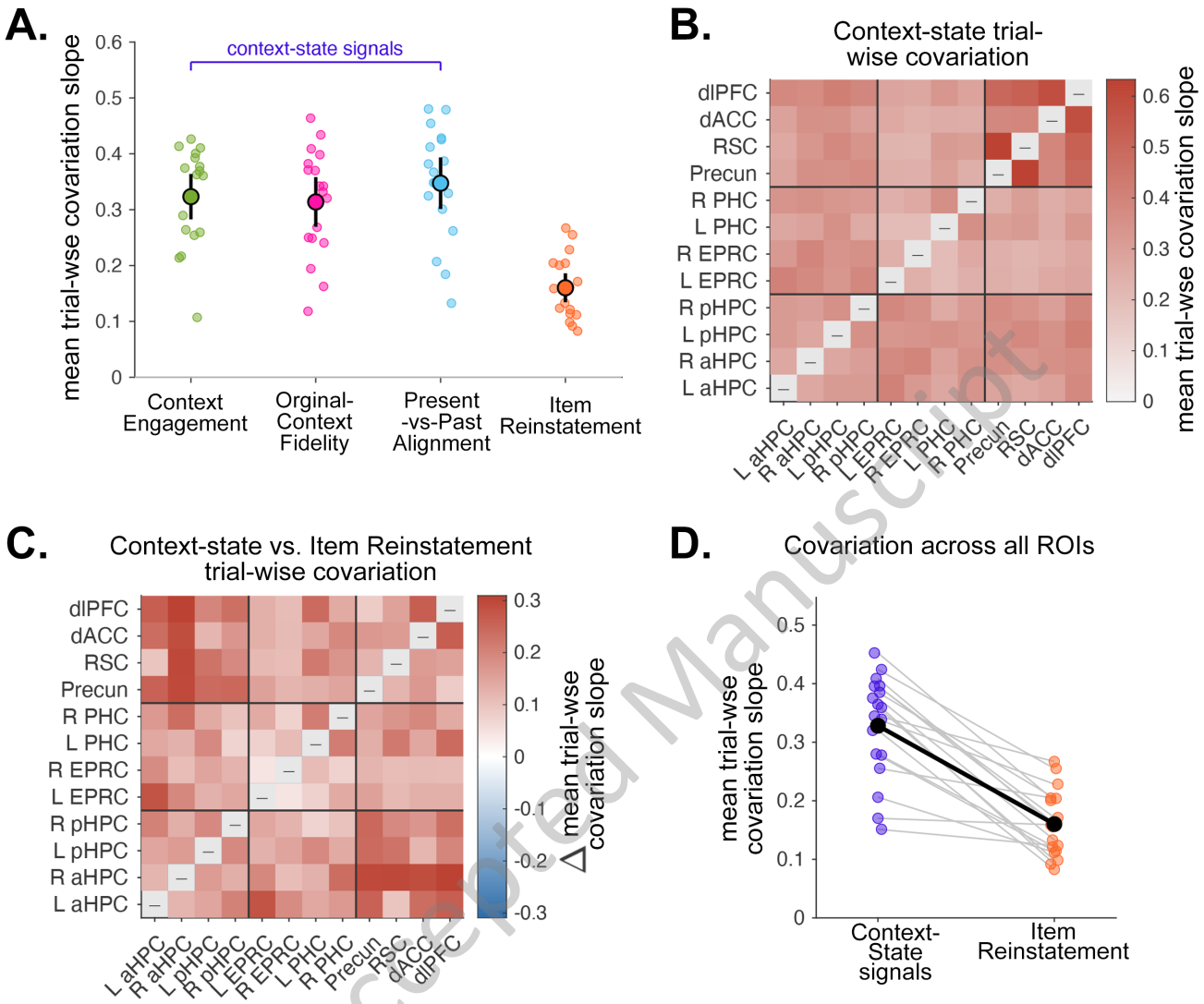
JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript



JNeurosci Accepted Manuscript