

Neural temporal context reinstatement of event structure during memory recall

Lynn J. Lohnas¹, M. Karl Healey², and Lila Davachi^{3,4}

¹Department of Psychology, Syracuse University

²Department of Psychology, Michigan State University

³Department of Psychology, Columbia University

⁴Nathan Kline Institute

Author Note

Data and materials available online are provided primarily from memory.psych.upenn.edu, with more specific links provided as applicable throughout the text. Although analyses have been presented from this dataset previously, all of the analyses presented here are novel. We thank Patrick Crutchley, Jonathan Miller, and Isaac Predisich for assistance with programming the experiments and Adam Broitman, Elizabeth Crutchley, Kylie Hower Alm, Joel Kuhn, and Logan O'Sullivan for help with data collection. We thank Michael Kahana for helpful discussions and sharing this dataset. This research was supported by NIH grant MH106266 to Lynn Lohnas, NIH grant MH055687 to Michael Kahana, and NSF grant 1848972 to Karl Healey.

Correspondence concerning this article should be addressed to: Lynn Lohnas, Department of Psychology, 430 Huntington Hall, Syracuse NY 13244, ljlohnas@syr.edu.

An earlier version of the manuscript was published as a preprint on *bioRxiv* (<https://doi.org/10.1101/2021.07.30.454370>) but the first author retains copyright. Versions of these results were presented at the 2019 Syracuse University Neuroscience Research Day in Syracuse, NY; 2018 Context and Episodic Memory Symposium in Philadelphia, PA; 2018 Annual Meeting of the Cognitive Neuroscience Society in Boston, MA.

Abstract

The transformation of experiences into meaningful events and memories is intertwined with the notion of time. Temporal perception can influence, and be influenced by, segmenting continuous experience into meaningful events. Episodic memories formed from these events become associated with temporal information as well. However, it is less clear how temporal perception contributes to structuring events and organizing memory: whether it plays a more active or passive role, and whether this temporal information is encoded initially during perception or influenced by retrieval processes. To address these questions, we examined how event segmentation influences temporal representations during initial perception and memory retrieval, without testing temporal information explicitly. Using a neural measure of temporal context extracted from scalp electroencephalography in human participants (N=170), we found reduced temporal context similarity between studied items separated by an event boundary when compared to items from the same event. Further, while participants free recalled list items, neural activity reflected reinstatement of temporal context representations from study, including temporal disruption. A computational model of episodic memory, the Context Maintenance and Retrieval model (CMR; Polyn, Norman & Kahana, 2009), predicted these results, and made novel predictions regarding the influence of temporal disruption on recall order. These findings implicate the impact of event structure on memory organization via temporal representations, underscoring the role of temporal information in event segmentation and episodic memory.

Keywords: electroencephalography; episodic memory; event segmentation; free-recall

Neural temporal context reinstatement of event structure during memory recall**Revision of XGE-2021-3771R2 as invited by the action editor, Tim Curran, PhD**

Word count: 19,835

Introduction

An overarching question in cognitive psychology and neuroscience is how the external environment is transformed into internal representations. This question is key to understanding the transition from sensory processing to subjective perception, as well as how perceptual processes interact with memory. Temporal context — the representation of item features surrounding but not comprising the item itself — is a defining feature of episodic memory (Tulving, 1972), which helps account for many episodic memory phenomena (e.g. Healey & Kahana, 2016; Howard et al., 2009; Logan, 2021; Lohnas & Healey, 2021; Lohnas & Kahana, 2014b). Temporal context also plays a role in perception, as a change in nontemporal features can lead to longer prospective temporal judgments (e.g. Block, 1982; Ezzyat & Davachi, 2014; Faber & Gennari, 2017; Lositsky et al., 2016). Further evidence for a common role of temporal context in both memory and perception comes from studies examining how continuous ongoing experience is structured into events (Kurby & Zacks, 2008; Radvansky & Zacks, 2014; Zacks et al., 2007): items in different events not only tend to be linked more weakly in memory (e.g. Ezzyat & Davachi, 2014; Heusser et al., 2018; Speer & Zacks, 2005; Zwaan, 1996), but also tend to be perceived as occurring further apart in time (Clewett et al., 2020; DuBrow & Davachi, 2013; Ezzyat & Davachi, 2014; Faber & Gennari, 2017; Lositsky et al., 2016). However, prospective temporal judgments may involve different cognitive mechanisms than retrospective temporal judgments (Grondin, 2010; Pöppel, 1997), leaving unclear how transient changes in temporal information may inform persistent changes in memory representations. Further, there is no consensus on whether temporal information is secondary or primary in structuring events and organizing memories.

Despite increasing interest and open questions regarding the intersection between these phenomena (Clewett et al., 2019; Frank et al., 2020; Radvansky & Zacks, 2017), few studies have directly examined the three-way interaction between temporal perception, memory, and event

segmentation. Here we consider these interactions, examining how endogenous temporal information and event segmentation interact to organize memory. We present a computational model which formalizes how event boundaries influence temporal information and memory representations. We verify novel predictions of this model using human behavior and neural activity, confirming the impact of event structure on temporal representations during memory encoding and retrieval.

Event segmentation and episodic memory

On a behavioral level, there is a wealth of data suggesting that stimuli presented in the same event share stronger associations in long-term memory than stimuli presented in different events (DuBrow & Davachi, 2013; DuBrow & Davachi, 2014, 2016; Ezzyat & Davachi, 2011; Ezzyat & Davachi, 2014; Heusser et al., 2018; Speer & Zacks, 2005; Zwaan, 1996). For instance, recognition of recently presented information is worse with a change in event, or event boundary, between presentation and test (Swallow et al., 2011; Swallow et al., 2009). Neural data corroborate these findings, as neural activity for pairs of stimuli from the same event is more similar than for stimulus pairs from different events (Baldassano et al., 2017; DuBrow & Davachi, 2013; DuBrow & Davachi, 2014; Ezzyat & Davachi, 2014; Hsieh et al., 2014; Lositsky et al., 2016; Schapiro et al., 2013). Further, brain activity in mnemonic brain regions (e.g. hippocampus) is greater during retrieval of items from another event rather than the current event (Swallow et al., 2011), suggesting that retrieval might be more effortful for information outside of the current event. Taken together, these results suggest that associations are weaker between memories separated by an event boundary, and that overcoming such weakened associations may require more effortful retrieval.

Why might an event boundary weaken associations in memory? One possibility is that stimuli separated by an event boundary may simply share fewer common perceptual or categorical features (Clewett et al., 2019; Zacks et al., 2001). For instance, event boundaries may be caused by physical changes to the environment, such as a change in background scene (Zacks et al., 2007). As another example, DuBrow and Davachi (2014) found evidence that items within the same event form strengthened associations, which can then support reinstatement of one

another and their shared event information. In particular, participants made a recency judgment between two items previously studied with the same category and task. On a neural level, the category of the intervening items was decoded using whole brain multivariate pattern analysis (Norman et al., 2006), and classifier performance predicted the category of these intervening items. On a behavioral level, DuBrow and Davachi (2014) posited that, if testing items from the same event evokes event-level reinstatement, then this reinstatement should facilitate memory recognition of other items from that event. Consistent with this hypothesis, participants recognized an item more quickly when it was preceded by a recency judgment of other same-event items. Taken together, these results reflect the strong associations between items within an event, and how such associations promote memory reinstatement of event-related information. Studies using television episodes, rather than discrete stimuli, have also found such neural evidence of event-level reinstatement (Baldassano et al., 2017; Chen et al., 2017; Zadbood et al., 2017). Although these studies provide support for the stronger associations between items within an event due to their shared features, this explanation is not mutually exclusive with an alternate account, which we next explore: event boundaries weaken temporal associations.

An emerging role of temporal context in event segmentation and episodic memory

Supporting the notion that event boundaries may weaken temporal associations across items, stimuli separated by event boundaries are perceived as occurring farther apart in time than stimuli occurring in the same event (Ezzyat & Davachi, 2014; Faber & Gennari, 2017; Lositsky et al., 2016). Also suggestive of the the importance of temporal information to event structure, if participants are informed that a long amount of time has passed in a narrative, irrespective of other stimulus changes, they are more likely to perceive this as an event boundary in the narrative (Ezzyat & Davachi, 2011; Speer & Zacks, 2005; Zwaan, 1996). Of course, it is possible that changes in temporal information are a by-product of an event boundary, rather than a necessary component. Further, most studies define event boundaries with changes in stimulus or context features, and thus do not separate temporal context from nontemporal context. This motivates our current study of the interaction between temporal associations, memory and event segmentation. We next review studies which provide evidence that event boundaries weaken

temporal associations when changes in stimulus features are minimized.

In a series of studies with more controlled changes to stimuli between events, DuBrow and Davachi (2013) and DuBrow and Davachi (2014) found that event boundaries influenced memory performance and memory representations. They defined an event as a sequence of presented stimuli from the same semantic category and with the same encoding task. In each list, they presented participants with sequences of items, switching back and forth between the two categories and tasks. Critically, they tested participants with pairs of items, where each pair contained items from the same category and task, but only a subset of pairs were from the same event. With these test stimuli, participants exhibited less accurate recency judgments for item pairs across events than within event. This suggests that weakened associations across events are not completely a by-product of fewer shared stimulus features, and points to an important role of temporal information. However, these studies still leave unresolved how and when temporal information influences, or is influenced by, event structure in memory.

Polyn et al. (2009a, 2009b) examined the contributions of temporal and nontemporal features to event structure using model simulations. Although they did not frame their results in terms of event boundaries, like the DuBrow and Davachi studies, participants studied items with one of two encoding tasks, and thus a sequence of items with the same task can be operationalized as an event. Critically, to distinguish between event-level and temporal information, Polyn et al. (2009a) examined predictions of a computational model of episodic memory, the Context Maintenance and Retrieval (CMR) model. CMR assumes that two types of context are updated whenever an item is studied or retrieved: (a) temporal context, reflecting the surrounding temporal information of a given item; (b) task or source context, implemented experimentally as an encoding task. In this way, all items within the same event share similar source context and similar temporal context. By contrast, two neighboring items separated by an event boundary have similar temporal contexts yet distinct source contexts. Two temporally distant items may share the same source context, even though they were presented in different events. Polyn et al. (2009a) compared two variants of the CMR model: (1) one variant assumed that an event boundary evokes a change to source context only; (2) another variant assumed that an event boundary evokes a change to source context as well as a disruption to temporal context.

The second CMR model variant made more accurate predictions of participants' memory performance, and hereafter we refer to this model variant as CMR.

The success of this model variant suggests that an event boundary imposes a perceived shift or disruption in temporal information, even when accounting for differences between stimuli occurring in different events. These results underscore the critical role of temporal information in event representation, both in the moment and in mnemonic representations. These results suggest that temporal information is not just a secondary by-product of event segmentation, but rather may play a critical role in structuring events. Nonetheless, CMR only predicts behavior based on its assumptions of memory representations, and these assumptions may be incorrect. Thus, we sought to examine CMR's predictions using brain activity as well as behavior. Further, we compared predictions of CMR to the less successful model variant which does not assume that an event boundary evokes a change to temporal context.

The current study

Thus far, we have reviewed how event segmentation influences memory, and studies dissociating the contributions of temporal and nontemporal information to event boundaries and memory. An understanding of the interactions between event segmentation and memory remains incomplete without appreciating the role of temporal information. Specifically, it is critical to distinguish between the possibility that temporal representations are a defining feature of stimuli, and thus influenced by event boundaries, from the possibility that temporal perception effects are a by-product of changes to other stimulus features. Distinguishing between these possibilities is not only important in event segmentation, but more broadly may inform the role of temporal information to other perceptual and memory paradigms.

Critically, to our knowledge no research directly links the impact of event segmentation at study, including its impact on temporal disruption, to neural and behavioral measures of memory retrieval. Here we examined these relationships among memory behavior and a neural measure of temporal context (Folkerts et al., 2018; Howard et al., 2012; Manning et al., 2011; Manns et al., 2007). This neural measure allowed us to assess how temporal context states from study were reinstated during memory retrieval to influence behavior. To minimize nontemporal contributions

to event boundaries, stimuli comprising the events and the event boundaries were kept as simple as possible. In particular, participants studied lists of words in which each word was associated with an encoding task or no task, with the task for a given word was indicated by a unique font, color and case. Events were operationalized within a list as a sequence of items with the same encoding task, and a change in event was signified both by the change in task and the visual change in studied words. Previous studies have also used color to operationalize events, by simply changing a color frame surrounding a grayscale image (Heusser et al., 2016; Heusser et al., 2018). Other studies have used encoding tasks, in conjunction with other stimulus feature changes, to promote event segmentation (DuBrow & Davachi, 2013; DuBrow & Davachi, 2014, 2016; Ezzyat & Davachi, 2014; Polyn et al., 2009a, 2009b). Taken together, the current study induces event structure while minimizing changes to stimulus features, thus allowing a more direct test of temporal information on event segmentation and memory.

CMR provides a very good testbed to examine the links between memory, temporal information and event cognition. CMR is a model of episodic memory sharing many assumptions with theories of event cognition (e.g. DuBrow & Davachi, 2014; Ezzyat & Davachi, 2014; Faber & Gennari, 2017; Frank et al., 2020; Lositsky et al., 2016; Swallow et al., 2009). CMR assumes that each studied item is associated with a slowly changing temporal context, as well as a source context reflecting the task features of the items within a shared event. Thus, CMR simulations allowed us to disentangle the interactions between temporal representations and event segmentation. Comparing participants' data to CMR predictions also allows for a more specific characterization of the temporal representations—whether they might rely on local positional information of items within an event or list, or whether they might rely on a more global temporal code.

We compared CMR predictions to data averaged across participants and examined individual variability across participants. If temporal disruption underlies event segmentation and memory representations, then we expect (a) accurate predictions from the CMR model; (b) a disruption to temporal information at study should manifest in neural activity and behavior during recall. To test these hypotheses, we present novel analyses of a neural correlate of temporal context, as posited by the CMR framework (Manning et al., 2011), as well as analyses of

memory behavior which have been used to assess variants of the CMR model (Kahana, 1996; Lohnas & Kahana, 2014a; Polyn et al., 2009a; Sederberg et al., 2008). We generated CMR simulations and predictions from another dataset and its associated best-fit model parameters. We found that CMR predictions were upheld in averaged data from the current study, and participant variability was consistent across predicted measures. Further, CMR predictions were more accurate than a model variant which does not assume that event boundaries evoke temporal context disruptions. Our results clarify how event segmentation impacts temporal representations during memory encoding and retrieval, influencing perception and memory.

Method

Dataset

The data reported here are from the Penn Electrophysiology of Encoding and Retrieval Study (PEERS), which involved three subsequently administered multi-session experiments from 2010–2016. PEERS is a large database on the electrophysiological correlates of memory encoding and retrieval (Kahana et al., 2022).

Participant Characteristics

The present study considered the 172 younger adults (age 18–30) who completed Experiment 1 of PEERS. Participants were right-handed native English speakers.

Sampling Procedures

Participants were recruited through a two-stage process. First, right-handed native English speakers were recruited for a single session to introduce participants to EEG recordings and the free recall task. Participants who did not make an excess of eye movements during item presentation epochs of the introductory session and had a recall probability of less than 0.8 were invited to participate in the full study. Approximately half of the participants recruited for the preliminary session qualified for, and agreed to participate in, the full study. Participants were consented according the University of Pennsylvania’s IRB protocol and were compensated for their participation.

Data Diagnosis

One participant was excluded for not having a neural measure of temporal context in any session (see definitions below and Figure 1), and another was excluded from all behavioral and neural analyses for making too few (< 10) critical recalls (see Figure 4 and surrounding text). This participant had 7 such observations in total, whereas the next fewest participants had 15. For this latter participant, because most analyses include recall behavior, for consistency we excluded this participant from all analyses, rather than from just the recall analyses.

Data Collection

Participants completed sessions each with 16 free recall lists. For each list, 16 words were presented one at a time on a computer screen followed by an immediate free recall test. Generally participants completed 6 sessions, but data collection was incomplete for 1 session each for 5 participants. Based on our criteria of only including sessions with autocorrelated feature vectors (see Neural Feature Selection), 5 participants had 2 included sessions, 8 participants had 3 included sessions, 24 participants had 4 included sessions, 50 had 5 included sessions, and 83 had 6 included sessions. Additional memory tests were administered in each session after immediate free recall of the final list. However, we do not report results from those tests so omit further detail about them.

Each word was accompanied by a cue to perform one of two judgment tasks, either a size judgment task (“Will this item fit into a shoebox?”) or an animacy judgment task (“Does this word refer to something living or not living?”) or no encoding task. The current task was indicated by the color, font and case of the presented item. There were three conditions: no-task lists (participants did not have to perform judgments with the presented items), single-task lists (all items were presented with the same task), and two-task lists (items were presented with either task). In the two-task lists, items were presented successively with the same task in trains of 2–6 items, with train length chosen randomly. The first two lists were two-task lists, and each list started with a different task. The next fourteen lists contained four no-task lists, six one-task lists (three with each task), and four two-task lists. List and task order were counterbalanced across sessions and participants.

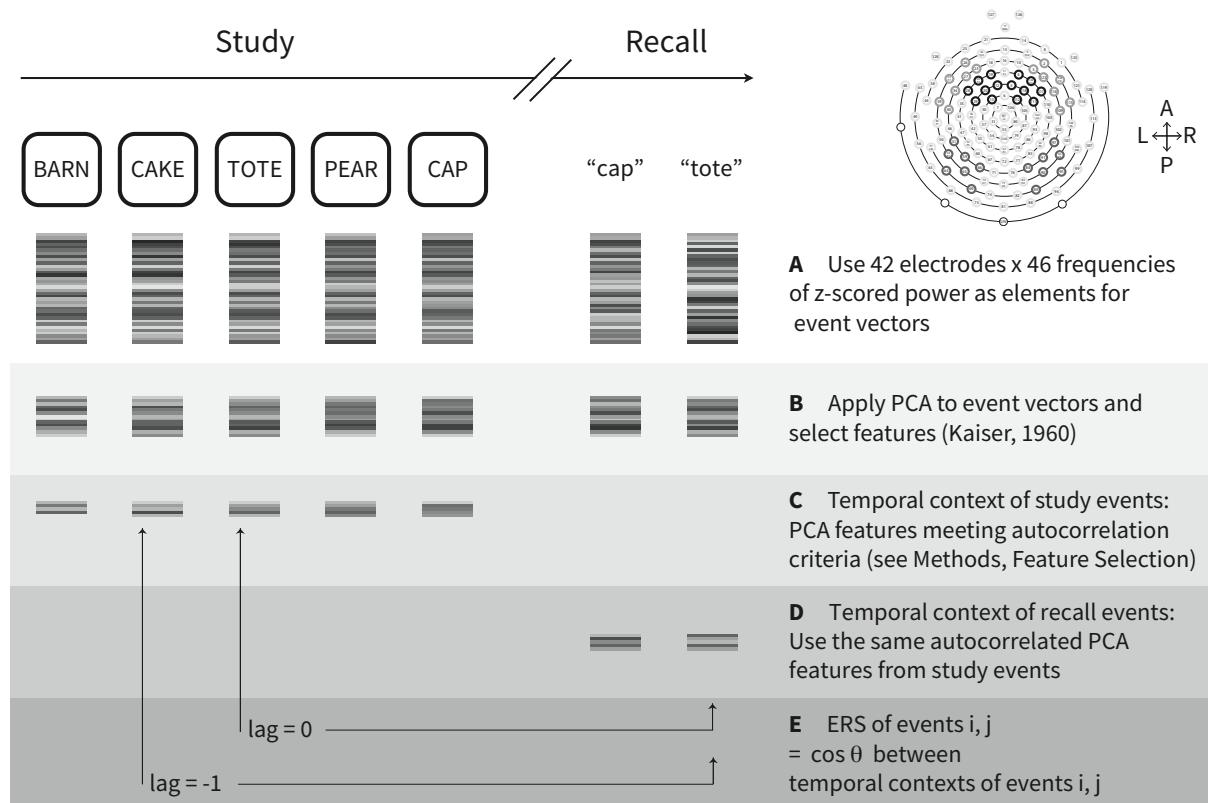


Figure 1

Calculating a neural measure of temporal context. Based on the core assumptions of retrieved context models, the temporal context state of a studied item should (1) be a slowly changing representation of temporal context from earlier studied items; (2) be reinstated if the item is recalled. **A.** We first calculated oscillatory power from electroencephalography (EEG) activity recorded for each studied item or recalled item in control lists. In the upper right panel, the 42 electrodes included in the event vectors are circled in dark gray on the electrode map. L = left, P = posterior, R = right, A = anterior. **B.** By applying PCA, we selected features accounting for a significant amount of variance in the EEG recordings. **C.** To meet the first criterion of a slowly changing representation, we next determined which of the PCA features were autocorrelated across studied items. **D.** To verify the second criterion of a neural measure of temporal context, we next needed to examine this neural signature at recall. Thus, having established a slowly changing neural signature from study of selected PCA features, we then applied those same feature vectors from study events to the recall events. **E.** We assessed whether a studied item’s feature vectors were reinstated when the item was recalled, by calculating the encoding-retrieval similarity (ERS) between each recalled item’s temporal context and temporal context states from study. Retrieved context models predict that the similarity between a recalled item’s retrieved temporal context and temporal contexts at study should be greater for items studied nearby in time, or smaller absolute lag, to the study position of the recalled item (see also Figure 2C,D).

Each word was drawn from a pool of 1638 words (available at memory.psych.upenn.edu/files/wordpools/PEERS_wordpool.zip). Lists were constructed such that varying degrees of semantic relatedness occurred at both adjacent and distant serial positions. Semantic relatedness was determined using the Word Association Space (WAS) model described by Steyvers et al. (2004). WAS similarity values were used to group words into four similarity bins based on the similarity between word pairs (high similarity, $\cos \theta > 0.7$; medium–high similarity, $0.4 < \cos \theta < 0.7$; medium–low similarity, $0.14 < \cos \theta < 0.4$; low similarity, $\cos \theta < 0.14$). Two pairs of items from each of the four groups were arranged such that one pair occurred at adjacent serial positions and the other pair was separated by at least two other items.

For each list, there was a 1500 ms delay before the first word appeared on the screen. Each item was on the screen for 3000 ms, followed by jittered (i.e., variable) inter-stimulus interval of 800–1200 ms (uniform distribution). If the word was associated with a task, participants indicated their response via a keypress. After the last item in the list, there was a jittered delay of 1200–1400 ms, after which a tone sounded, a row of asterisks appeared, and the participant was given 75 seconds to attempt to recall aloud any of the items from the most recent list.

Electrophysiological Recordings

Netstation was used to record EEG from Geodesic Sensor Nets (Electrical Geodesics, Inc.) with 129 electrodes. The signal from all electrodes was digitized at 500 Hz by either the Net Amps 200 or 300 amplifier and referenced to Cz. Prior to any data processing, recordings were rereferenced to the average of all electrodes except those with high impedance or poor contact with the scalp. To eliminate electrical line noise, a first order 2 Hz stopband Butterworth notch filter was applied at 60 Hz.

We excluded any recalls that occurred within 1000 ms of the next recall to prevent overlap of the neural activity between these recalls. In both the neural data and the behavioral data, we excluded recalls from output positions 1–3, as such recalls may reflect recall from short-term memory in immediate free recall (Kahana, 1996), and such earlier immediate recalls may have shorter latencies (Kahana, 2012; Murdock & Okada, 1970). We calculated spectral power from 42 of the 129 electrodes (Figure 1), including electrodes in regions established in successful memory

encoding (Long & Kahana, 2017; Long et al., 2014; Weidemann et al., 2009): bilateral anterior superior (corresponding to dorsolateral prefrontal cortex), bilateral anterior inferior (corresponding to inferior frontal cortex), and bilateral posterior inferior (corresponding to inferior temporal cortex). From these electrodes, we calculated spectral power for each event (defined in the next paragraph) by convolving its EEG time series with Morlet wavelets (wave number = 6) at each of 46 frequencies logarithmically spaced between 2 Hz and 100 Hz. For each frequency and electrode, power was averaged across the entire encoding or recall interval. Then, the power values were z-scored across encoding and recall events separately for each session to remove the effects of these variables. Thus, each study or recall event had a corresponding vector of z-scored power values, concatenated across 42 electrodes at each of the 46 frequencies.

We computed spectral power for defined events of interest: We defined encoding events as the time window from 200 ms to 3000 ms relative to the onset of each item's presentation, and recall events as the time period -1000 ms to -600 ms relative to the verbalization of an item. The time window for presentation events was motivated by the choice of Manning et al. (2011), where the 200 ms delay was meant to account for the time delay between when the word appears on the screen and the participant begins to process the word, but otherwise activity is considered for the entire duration the word is on the screen. For the time window of recall events, we evaluated context reinstatement while varying the onset and duration of the time window. We evaluated time windows beginning from -1000 ms to -500 ms relative to the participant's recall vocalization, ranging in duration from 300 ms to 800 ms (both ranges were assessed in increments of 100 ms). This evaluation of time windows indicated that context reinstatement was strongest for the recall time window of -1000 to -600 ms relative to recall vocalization (see Appendix C).

Neural Feature Selection

We followed the approach of Manning et al. (2011) to determine patterns of neural activity that change gradually with each studied word. First, we applied principal components analysis (PCA) to the set power values across electrodes and frequency bands contributing to each study or recall event, as described above, using control lists only (no-task or single-task lists). We excluded from subsequent analyses those principal components that failed to explain a

substantial proportion of the variance according to the Kaiser criterion (Kaiser, 1960). Next, we quantified the extent to which each principal component changed slowly with each studied item, based on its autocorrelation (Equation 1, Figure 1C). If a principal component was not sufficiently autocorrelated across studied items, it was excluded because it did not meet the critical criterion of temporal context, to change slowly with each studied item. In this way, we calculated a set of autocorrelated feature vectors consistent with the notion of temporal context.

We follow the terminology of Manning et al. (2011) and refer to such autocorrelated principal components as *feature vectors*. To determine which of the components were feature vectors, for each feature x within each list i , we computed the Pearson's lag 1 autocorrelation coefficient (r_i) and associated P value. We then combined the autocorrelation coefficients across lists into a summary autocorrelation measure \bar{r} :

$$\bar{r} = F^{-1}\left(\sum_i^L F(r_i)\right) \quad (1)$$

where F and F^{-1} are the Fisher z-prime transformation and Fisher inverse transformation, respectively. We also computed a summary measure for P across lists, \bar{p} , by applying the inverse Normal transformation to the P values then summing across the transformed P values. We defined \bar{p} as where the sum of the transformed P values fell on the cumulative normal distribution function. Finally, we used \bar{r} and \bar{p} as inclusion criteria, and only included x as an autocorrelated feature vector if it satisfied $\bar{r} > 0$ and $\bar{p} < 0.1$.

The neural measure of temporal context was considered separately for each session. If there were not at least five feature vectors, the session was excluded from further neural and behavioral analysis. If the session produced at least five feature vectors, we applied a PCA transformation matrix, determined from the control lists, to calculate temporal context vectors from two-task lists. Of the 1027 possible sessions, 1023 sessions produced feature vectors. Of these sessions, 879 sessions had at least 5 feature vectors. The threshold of at least 5 feature vectors aims to ensure that the feature vectors are of high enough dimensionality to observe the potential properties of interest including context reinstatement (Manning et al., 2011). Further, 3 of the 4 sessions with no feature vectors had lower than average dimensionality from PCA (8, 12, 58; $M \pm STD = 86.8 \pm 15.7$), and all participants with such sessions did contribute at least 1

other session to the analyses presented here. Thus, the lower dimensionality from these sessions may reflect noisier EEG data, or at least they suggest that they do not reflect solely participants who fail to exhibit neural features of temporal context.

Similarity values

We defined the neural similarity between two feature vectors (in the participants' data) or two temporal context vectors (in CMR) as the cosine of the angle between those two vectors (Manning et al., 2011). When comparing similarity values between neighboring items within events versus across events, we only calculated an item's similarity to its lag = +1 and lag = -1 neighbors if both the preceding item and the following item were valid list positions. Thus, the first list item was never included in similarity value calculations as the item immediately following an event boundary (which we term a boundary item), and the last list item was never included in similarity values as the item preceding an event boundary (which we term a preboundary item).

Neural similarity between studied items. As confirmation of the approach for defining neural similarity during study, we calculated neural similarity as a function of study lag in control lists and two-task lists (see Figure A1). When comparing neural similarity during study of two-task lists for within versus across events, we excluded similarity values of preboundary items in events of 2 items, as the across-event similarity was included for the subsequent boundary item, and the within-event similarity was included for the preceding boundary item (aside from the first event). We only calculated within-event similarity for preboundary and boundary items; otherwise, the within-event similarity measure would include many more items and may not be as comparable to across-event similarity. Finally, for preboundary items for which the across-event similarity was already included for the boundary item, we only included this value once as an across-event similarity value (i.e. we did not double-count these values).

For calculations neural similarity across lists between item pairs of the same or different tasks in control lists, we excluded the no-task lists to make the comparisons between control lists and two-task lists more comparable, as two-task lists did not include items with no task.

Encoding-retrieval similarity (ERS). In addition to the general exclusion noted above, we did not calculate similarity between a recalled item and any of its study neighbors

which were already recalled. Including ERS at study for a previously recalled item may be problematic, as such similarity values may reflect shared features from retrieval, not from study (cf. Folkerts et al., 2018). However, this exclusion did not take into account the items excluded in output positions 1-3 or items recalled less than 1 s earlier than the previous recall, as those items did not contribute to the initial PCA analysis, and thus presumably did not contribute significant variability to the neural measure of temporal context.

Analytic strategy

To compare within-participant conditions across participants with a large sample size, we used standard paired *t*-tests. We calculated effect size using a variant of Cohen's *d* based on the pooled standard deviation, using the second, fourth and fifth equations given on p. 7 of Fritz et al. (2012). The formula for 'very similar' standard deviations between groups was used when standard deviations were within 5% of one another; otherwise, standard deviations were at least 7% apart from one another, and the formula for standard deviations which 'differ' was used.

All correlation analyses used robust regression, a regression measure less sensitive to potential outliers. Unlike standard Pearson's regression, this analysis does not yield the same correlation and significance values if the dependent and independent measures are switched (i.e., the correlation of *x* and *y* is not the same as the correlation of *y* and *x*). In our analyses we defined the independent measure, plotted on the *x* axis, as the measure occurring earlier in time. Our regression analyses were motivated by CMR predictions, whereby a participant exhibiting a stronger impact of temporal disruption in neural reinstatement should also exhibit stronger temporal disruption during study, and stronger temporal disruption in recall behavior. Thus, for each of our regression analyses we had a hypothesized direction of the correlation, and we report one-tailed *p*-values.

For comparable comparison between ERS and behavioral lag-CRPs during recall, we excluded the items recalled at the first three output positions. Further, for the lag-CRPs in two-task lists, we considered transitions from a preboundary or boundary item to any possible item, not just those within the same or neighboring event. Thus, the colors and legends in Figure 5 is meant to reflect the most likely transition. However, only neighboring items are always

consistent with the legend. For instance, if a boundary item is in an event of length 3, then a transition of lag = +4 would not be to an item in the same event.

Transparency and Openness

The raw behavioral data is available at

<http://memory.psych.upenn.edu/files/PEERS.data.tgz> and the raw electrophysiology data is available at http://memory.psych.upenn.edu/mediawiki/index.php?title=Data_Request&paper=WeidKaha16.

The code used for the behavioral simulations and analyses is available at

<http://memory.psych.upenn.edu/CMR>, and the analysis scripts used for calculating the behavioral lag-CRP analyses is available at <https://github.com/vucml/EMBAM>. Remaining materials are available upon request.

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. Because we present analyses of an existing dataset, the sample size was not determined specifically for our current set of analyses, and this study was not preregistered. However, as described above, the large sample size and large number of observations per participant gave us confidence that we would have sufficient statistical power for our analyses. Although analyses of the PEERS dataset have been reported previously (e.g. Lohnas et al., 2015; Long & Kahana, 2017; Long et al., 2014; Miller et al., 2012; Weidemann & Kahana, 2016), all of the analyses presented here are novel.

Results

Temporal context in control lists

We first assessed behavioral and neural measures of temporal context in control lists (see Figure 1). These lists did not impose a strong event structure because participants performed the same (or no) encoding task for every studied item in each list (e.g., compare with Figure 3A). Thus, we used the control lists to assess the contribution temporal information to episodic memory encoding (study) and retrieval. We then build upon these analyses to address how event segmentation influences temporal context and memory organization.

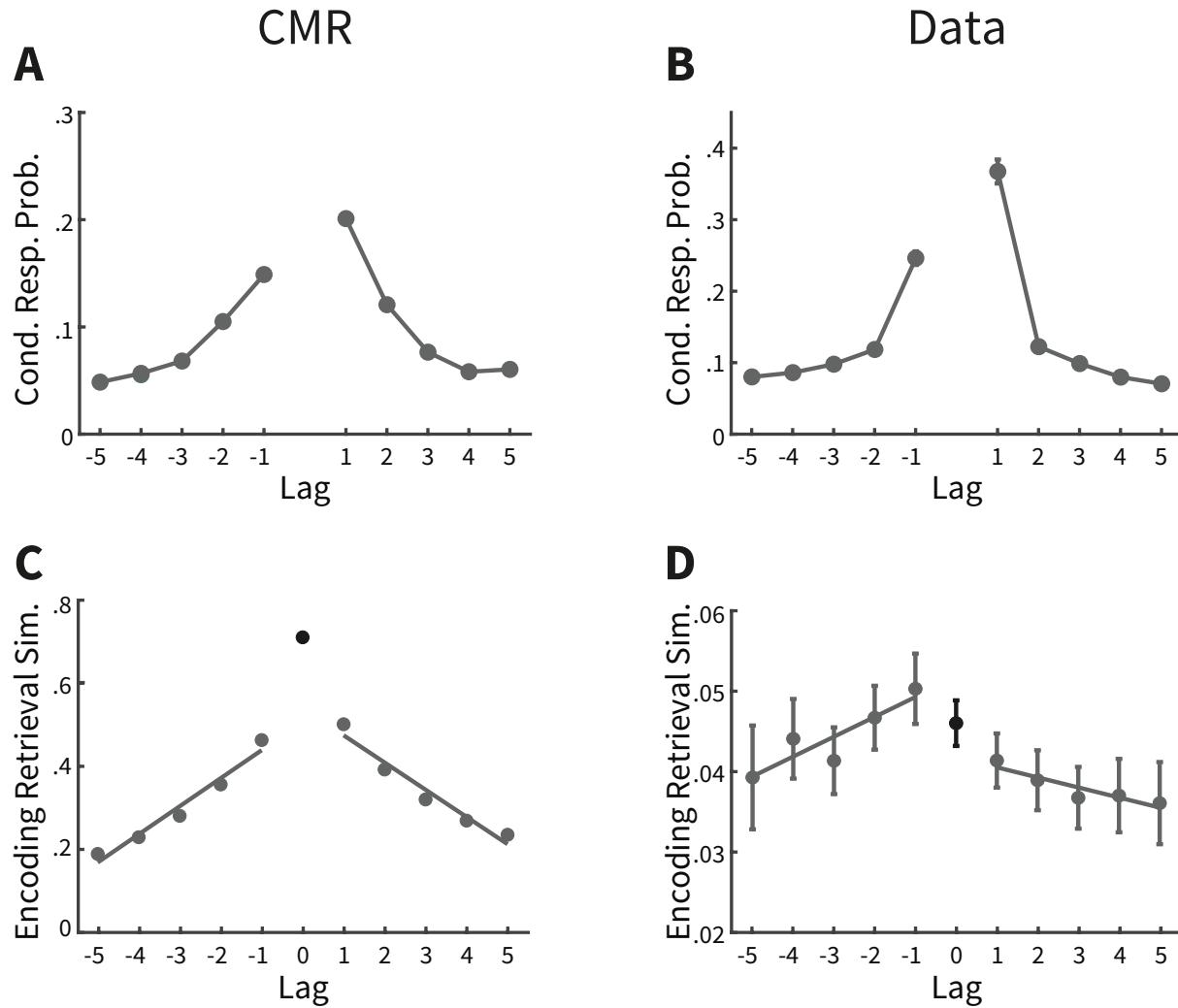


Figure 2

Behavioral and neural correlates of temporal context reinstatement in control lists.

A. Predictions of the context maintenance and retrieval (CMR) model of recall transitions in control lists. The probability of making transitions between successive recalls is plotted as a function of lag, or difference in the serial positions of the successively recalled items. These response probabilities are determined conditional on which lags are available for recall. **B.** Conditional response probability as a function of lag in the behavioral data. **C.** Encoding-retrieval similarity (ERS) between the temporal context state of a recalled item and the temporal contexts of its neighbors during study. Lag refers to the distance in serial position between two items from study (see Figure 1E). CMR predicts that temporal context states will be more similar between the recalled item and neighboring items from study. **D.** CMR's predictions are upheld when measuring the neural measure of temporal context in participants' data. Cond. Resp. Prob. = Conditional response probability. Error bars represent Loftus and Masson (1994) 95% confidence intervals.

Evidence of temporal context in recall behavior

After studying each list, participants performed free recall, recalling as many items as possible from the just-studied list in any order. Despite the open-ended instructions, recall order tends to reflect the temporal order in which items were presented (Healey & Kahana, 2014; Healey et al., 2019; Kahana, 1996; Kahana et al., 2008; Unsworth et al., 2012; Ward et al., 2010). Contributions of temporal organization can be measured by calculating the probability of a recall transition between two items, based on their difference in serial positions at study and conditional on their availability (lag-CRP; Kahana, 1996). Figure 2B shows the lag-CRP from the control lists, demonstrating two ubiquitous and critical features of this function (Kahana et al., 2008). First, the lag-CRP tends to be greatest at smaller absolute lags, indicating the increased transition probability between items studied nearby each other. Second, the lag-CRP is asymmetric, with greater transition probability in the forward direction (positive lag) than the backward direction (negative lag).

Figure 2A presents CMR's prediction of the lag-CRP and these two critical features. Rather than find a set of model parameters which best capture the phenomena of this data set, here we took a stricter approach by simulating a dataset and best-fit parameters generated from another study with a similar experimental design (Polyn et al., 2009a). This approach shares similarities to the generalization criterion method, whereby estimates from a dataset with one design (here, no or single-task lists) are validated using data with a different design (two-task lists; Ahn et al., 2008; Busemeyer & Wang, 2000). Here, we consider the model to be 'trained' on one data set, and 'test' model predictions on a different data set, for both similar effects to the original data set as well as novel effects. Before we consider novel predictions of CMR, we first verify that CMR predicts the same core effects in this study as in Polyn et al. (2009a), despite some minor differences in experimental procedures between the two studies. CMR predicts the temporal contiguity effect, or tendency to recall items from smaller absolute lags. CMR predicts this effect due to its core assumptions that temporal context changes slowly with each studied item, and recall of an item leads to retrieval of its associated context states from study. Thus, when the current context cues recall of the next item, the just-recalled item's context forms a part of this retrieval cue. As a result, CMR is more likely to recall items with shared temporal context

states to the just-recalled item, including that item's neighbors from study. CMR predicts the forward asymmetry in the lag-CRP because the context of a particular item i is incorporated into the context state of item $i + 1$, and thus a recalled item generally has a temporal context more similar to the items presented after it. Taken together, we interpret these behavioral recall dynamics as evidence for a role of temporal context in control lists.

A neural signature of temporal context during study

According to CMR, the behavioral effects in Figure 2 rely on a temporal context representation which changes slowly with each studied item (see Figure A1A). We assessed this prediction by defining an electrophysiological measure of temporal context consistent with the definitions of Manning et al. (2011). Whereas Manning et al. (2011) examined temporal context with intracranial EEG, here we examined temporal context with scalp EEG, which unlike intracranial EEG is noninvasive. To our knowledge, our results provide the first evidence of a temporal context measure using scalp EEG. Further, this lays the foundation to then assess how this neural measure of temporal context interacts with event structure and memory representations.

The measure of temporal context was designed to meet several criteria consistent with CMR's assumptions (see Method for full details). Separately for each participant and session, we first defined a vector of power values for each study event and recall event, where values were concatenated across a range of frequencies and included activity from electrodes implicated in mnemonic processing (Figure 1A; Long & Kahana, 2017; Long et al., 2014; Weidemann et al., 2009). We then applied PCA to the matrix of power vectors across study and recall events, and excluded principal components that contributed a low level of variance in principal components space (Kaiser, 1960). Next, we quantified the extent to which each of the included principal components was autocorrelated. We only used those principal components meeting the threshold for being substantially autocorrelated across each word presentation, consistent with CMR's assumption that temporal context changes slowly with each studied item. Our success in finding such autocorrelated feature vectors for 171/172 (99%) of participants attests to the validity of this approach (see also Figure A1B).

As described thus far, although feature vectors have properties consistent with temporal context, they also have features consistent with a *positional code* account, whereby each feature vector would code the serial position in the list associated with each item (Anderson & Matessa, 1997; Brown et al., 2000; Burgess & Hitch, 1999; Farrell, 2012; Henson, 1998). Like feature vectors of temporal context, feature vectors of positional codes would also be autocorrelated across items within a list. However, unlike temporal context, a purely positional code would be shared across lists. Thus, the feature vector of the item in Serial Position i in List j should exhibit greater similarity to the item in Serial Position $i + 1$ not only from the same list j , but also to the item in Serial Position $i + 1$ in List $j + 1$ or List $j + 2$ (Burgess & Hitch, 1999; Conrad, 1960; Henson, 1996; Osth & Dennis, 2015). By contrast, a temporal context account would predict that the similarity between item i and items $i + 1$ should decrease as the number of lists between i and $i + 1$ increases (Howard et al., 2008; Lohnas et al., 2015; Unsworth, 2008). This within-list vs. across-list contrast should be most pronounced in neighboring serial positions. Thus, to test this possibility, we calculated the neural similarity between items with differing serial positions of $lag = 1$ and with differing list numbers, i.e. list-lags $\in \{0, 1, 2, 3\}$. Further, to ensure a reduction across lists did not reflect differences in source context across lists, we calculated similarity only between pairs of items presented with the same encoding task and in control lists. Consistent with a temporal context account, neural similarity was greater for between item pairs with smaller list-lags (see Figure B1 and surrounding text).

Another way in which positional information might contribute to the current effects is that early list positions may contribute more to the positional code than later positions, such that these feature vectors really just reflect drift from primacy positions (Henson, 1998). However, the autocorrelated property of feature vectors is present across serial positions, suggesting that this effect is not driven by positional information from specific serial positions (see Figure B2 and surrounding text).

It is also important to rule out the possibility that feature vectors incorporate task information, that is, that they might change slowly with studied items but not for all task types. We found that neural similarity in feature vectors changed slowly over time for each task type and no task (Figure A1C). First, we verified that neural similarity decreased by lag, being

significantly greater for item pairs with lag = 1 than lags 3 to 5, in lists with no task ($M = 0.099, SD = 0.040, t(169) = 32.3, CI = [0.0928, 0.1049], p < .00001, d = 0.64$), lists with the size task ($M = 0.086, SD = 0.037, t(169) = 30.6, CI = [0.0809, 0.0921], p < .00001, d = 0.56$), and lists with the animacy task ($M = 0.084, SD = 0.037, t(169) = 29.9, CI = [0.0787, 0.0898], p < .00001, d = 0.55$).

Further, we calculated mean neural similarity between feature vectors between items of the same or different tasks. If task information contributed to the feature vectors, then we would expect neural similarity to be greater for item pairs studied with the same task when compared to items with different tasks. We conducted this analysis between item pairs studied in neighboring lists (*list – lag* = 1) because control lists are only comprised of items of the same task type or no task type. We thus excluded control lists of items studied without a task. To avoid potential positional differences, we calculated neural similarity between items with a difference in serial position or *lag* = 1. In control lists, neural similarity between items of the same task did not differ from neural similarity of items studied with different tasks ($p > .3$). Further, the mean magnitude of feature values did not differ between lists studied with the size task versus the animacy task ($p > .5$). Taken together, these analyses suggest that the autocorrelated feature vectors reflect temporal context information. To further discern the role of these posited temporal context features in memory processes, we next examined their properties during memory retrieval.

Reinstatement of temporal context in control lists

We next verified CMR's core prediction that temporal context is reinstated during recall. This prediction suggests that temporal information is not just a by-product of study, but rather contributes to memory representations and retrieval. To test this prediction, for each recalled item we calculated the similarity between the temporal context state of that item when it was originally studied to the current temporal context state as the item was being recalled. In addition, we calculated the similarity between the current temporal context and the temporal contexts of the item's neighbors from study (see Figure 1E). According to CMR, because an item's recall leads to reinstatement of temporal context state from study, then the ERS between the current context and an item's context from study should reflect the temporal history of studied

items, such that items with smaller absolute lags should have greater similarity (see Figure 2C).

We assessed CMR's prediction with the autocorrelated feature vectors, our alleged neural measure of temporal context. Specifically, for each recalled item, we calculated the ERS between the recalled item's feature vector to both the feature vectors from study of itself (lag = 0) and to its neighbors of lag $\in \{-5, -4, \dots, 4, 5\}$, for those items not yet recalled (see Figure 2D; also see Method). For negative lags (i.e., the similarity between an item and the items studied before it), CMR predicts that ERS should increase with study-recall lag. This is because a recalled item's retrieved temporal context should have greater overlap in temporal context, i.e. greater ERS, with other items studied nearby in time to that item. This prediction is critical to distinguish the retrieval of *context* information, as predicted in CMR, from the retrieval of *content*, or item, information (Manning et al., 2011). To test this prediction, we compared the ERS between at lag = -1 to the ERS at more distant lags -3 to -5. Consistent with CMR's prediction, ERS was significantly greater at lag = -1 than the average ERS at lags -3 to -5

($M = 0.009, SD = 0.039, t(169) = 2.95, CI = [0.0029, 0.0146], p = 0.004, d = 0.083$). We also evaluated ERS at positive lags, predicting that ERS should decrease with lag in the forward direction. Following the logic with negative lags, CMR predicts that context states of items studied nearby in time should share more temporal context, thus leading to greater ERS.

Paralleling the test of ERS with negative lag, we compared the ERS at lag = +1 to the ERS at lags 3 to 5, and found that ERS was greater at lag = +1

($M = 0.005, SD = 0.031, t(169) = 1.99, CI = [0.0000, 0.0095], p = 0.048, d = 0.043$). This result is not only consistent with CMR's prediction, but also helps to rule out the possibility that our neural measure of temporal context at study only reflects autocorrelated noise (Manning et al., 2011).

The slight negative asymmetry in the neural CRP shown in Figure 2D appears at odds with the striking forward asymmetry in behavioral CRP shown in Figure 2B. Furthermore, ERS at retrieval does not peak for the lag = 0 matching item. Both of these results should appear less surprising when considering that we designed our neural features to separate representations of context from item content. Thus, the neural CRP should not match the behavioral lag-CRP, but rather should represent the underlying neural representation of context. CMR predicts that such

a context representation should decay symmetrically as the absolute value of lag increases (Figure 2C). CMR predicts the forward asymmetry in the lag-CRP because both temporal context and content information contribute to the recall of an item (see Appendix A). Item information promotes recall of items following the just-recalled item, and combines with the symmetric support from context information to favor recall of items from forward lags over backwards lags (cf. Howard & Kahana, 2002; Manning et al., 2011).

Nonetheless, the apparent leftward shift of the neural CRP may reflect a more nuanced understanding of the process of contextual updating. This negative shift could arise because of the persistence of item representations that penetrate the representation of neural context. Although we attempted to rule this out by removing lag = -1 items when participants recalled items in forward pairs, this may not have completely eliminated contributions from the representation of items in feature vectors, especially for items recalled early in the list or within 1 s of each other (see Method). Alternatively, this may reflect our choice of presentation time window as in Manning et al. (2011). Importantly, this asymmetry most likely does not reflect our selection criterion to have ERS at lag=-1 exceed the mean ERS across lags -3, -4, -5. Although such a criterion could, at least in principle, bias us to find a time window with a larger lag=-1 value, all of the considered 36 time windows had a value of lag=-1 numerically greater than lag = 1. Future work remains to characterize the contributions of these factors and relate neural symmetry to behavioral asymmetry. However, regardless of the asymmetry, this does not detract from the critical result that temporal context is reinstated during free recall.

The influence of event boundaries on temporal context representations

Having established neural and behavioral measures of temporal context in control lists, we next turned to the critical analyses of the influence of event boundaries on temporal context states. CMR assumes that each item is associated with a temporal context and a source context, yet these two contexts interact. Specifically, a change in events (and thus a change in source context) leads to a disruption in temporal context. As a result, CMR predicts that the similarity in temporal context between neighboring studied items should be less if those items are separated by an event boundary. Further, CMR predicts that the state of temporal context, incorporating

the disruption to temporal context during study, should be reinstated during recall. We examined these predictions in two-task lists, where participants performed 1 of 2 semantic encoding tasks with each studied item, switching back and forth between the 2 tasks throughout the list (see Figure 3A). In this way, an event is operationalized as a sequence of items presented with the same task, where task was indicated to the participant by the color, font and case of the word. We define an event boundary as a change in encoding tasks. Thus, we term a *boundary* item as the first item presented with the changed encoding task, and a *preboundary* item as the final item in an event before the task switch.

In the experimental data, we calculated temporal context for each item in each two-task list using the feature vectors from the control lists (see Figure 1). We excluded the two-task lists when calculating the features vectors, so that the autocorrelation in our posited context measure could not be driven by a change in encoding task. Because each control list is comprised of only one encoding task type (no task, animacy task, or size task), two items within the same control list could not have reduced similarity with lag due to a task change, because there were no task changes in these lists. Further, our strict criterion of autocorrelation across control lists of three task types (see Method) would exclude feature vectors only autocorrelated for lists of a single task type. Thus, feature vectors reflect information slowly changing with each studied item irrespective of task type, a defining feature of temporal context but not source context. As a result, a reduction in neural similarity between feature vectors also reflects reduced temporal context similarity due to the change in task, but not the perceptual features of the task change. This approach shares similarities to the generalization criterion method, where we ‘train’ feature vectors on control lists, yet ‘test’ the validity of these features in two-task lists (Ahn et al., 2008; Busemeyer & Wang, 2000). By only using control lists to calculate the temporal context feature vectors, this makes it more likely to extract feature vectors which are less sensitive to changes in features specific to a single encoding task.

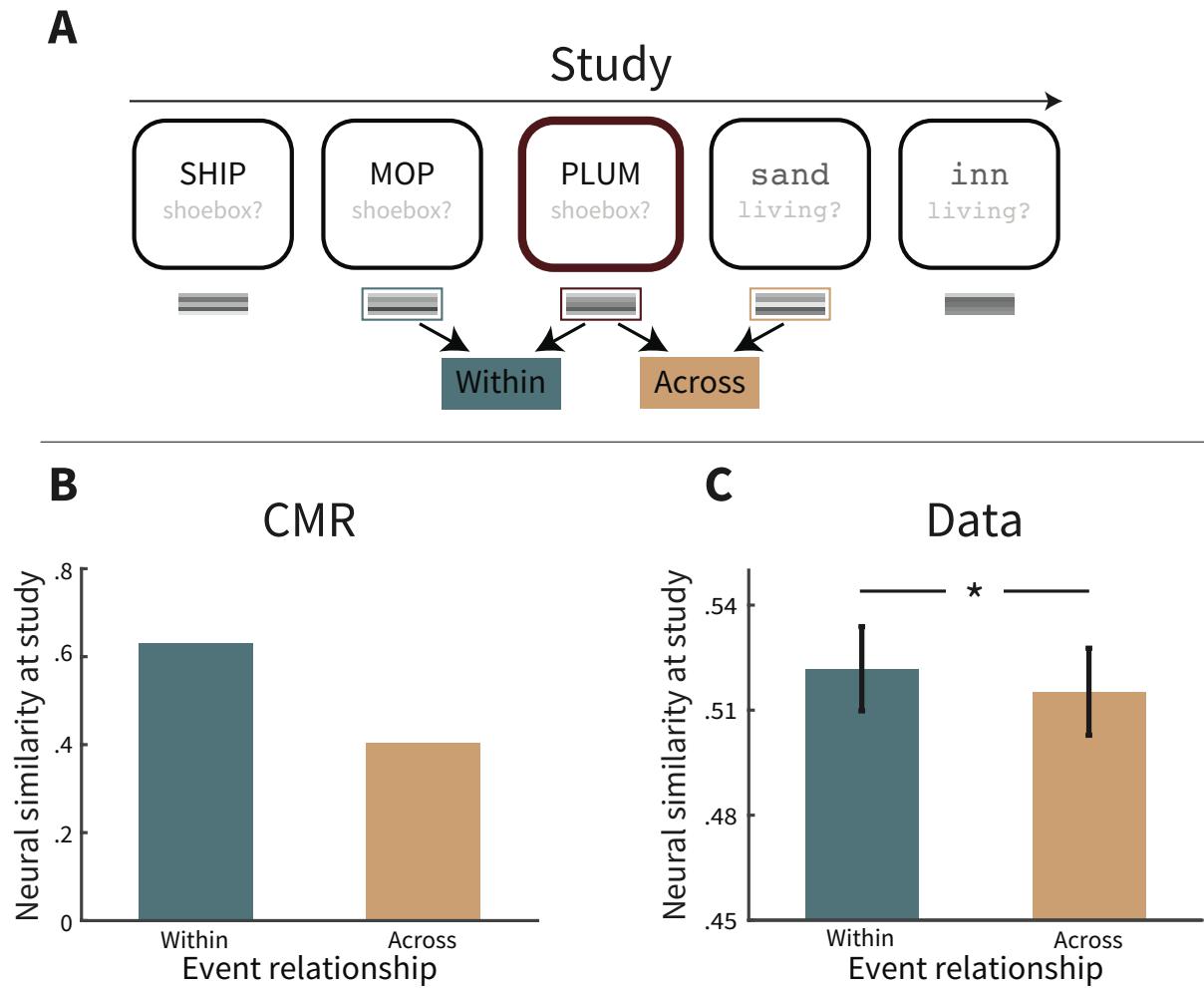


Figure 3

Disruption of temporal context by event boundaries during study. **A.** In two-task lists, participants perform one of two encoding tasks with each presented word (size task or animacy task); a sequence of words with the same task is assumed to comprise an event, and the change in task is assumed to form an event boundary. Here the sample items are shown to calculate neural similarity for the neighbors of a preboundary item, as one of its neighbors (the preceding item) was presented within the same event (Within), and its other neighbor (the following item) was presented across a different event (Across). Task text is for illustrative purposes only; to participants this was implicit from the color, font and case of the word. **B.** Collapsed across preboundary and boundary items, CMR predicts that neural similarity is greater between two neighboring items within the same event than two items across different events. **C.** Collapsed across preboundary and boundary items, neural similarity is greater between two neighboring items within the same event than two items across different events. Error bars represent ± 1 standard error of the mean. * $p < .05$.

We first verified that the temporal context features, defined in control lists, still maintained the critical property of autocorrelation in the two-task lists. We calculated neural similarity during study between items from the same event, and found that neural similarity was significantly greater for item pairs with $lag = 1$ than lags 3 to 5

$(M = 0.070, SD = 0.064, t(169) = 14.41, CI = [0.0606, 0.0799], p < 0.0001, d = 0.45$; Figure A1B).

We also verified that these changes were not driven by a single task type (Figure A1C). Similarly to control lists, in two-task lists neural similarity was significantly greater for item pairs with $lag = 1$ than lags 3 to 5 for the size task

$(M = 0.070, SD = 0.095, t(169) = 9.7, CI = [0.0559, 0.0846], p < .00001, d = 0.43)$ and for the animacy task $(M = 0.069, SD = 0.087, t(169) = 10.3, CI = [0.0555, 0.0817], p < .00001, d = 0.42)$.

For CMR simulations, temporal context states are determined with model equations (see Appendix A), and the similarity between item pairs in a control list is identical to the similarity between item pairs from the same event in a two-task list (Figure A1A). In contrast to CMR predictions, neural similarity between item pairs is reduced in two-task lists. Most likely, this reflects the exclusion of two-task lists when first applying PCA to calculate temporal context vectors. Participant data is more variable than model predictions, and the exclusion of the variance provided by two-task lists could cause overall reduced similarity between neighboring items. Because these similarity values are only calculated for item pairs within the same event, and thus with the same encoding task features, the reduced similarity cannot be explained by fewer shared task features across events.

To further ensure that the reduced neural similarity in two-task lists does not reflect differences in task information, we compared mean neural similarity values at study between item pairs in control lists to similarity values of item pairs in two-task lists, again with $list - lag = 1$ and $lag = 1$. In two-task lists, we excluded boundary items to avoid concerns that these items may exhibit other control processes independent of task information. Neural similarity values did not differ for item pairs of the same task for one task vs. two task lists ($p > .5$), nor for item pairs of different tasks ($p > .5$). This reinforces that task features do not influence differences of neural similarity across list types. To confirm that task features do not influence differences in neural similarity across task types, we also compared neural similarity for item pairs of the same type or

different task types, $list - lag = 1$ and $lag = 1$, paralleling the analysis in control lists. For this comparison as well, now in two-task lists, neural similarity did not differ between item pairs of the same task type vs. different task types ($p > .5$). These analyses suggest that, despite the reduced neural similarity in two-task lists, such a reduction is not explained by assuming that feature vectors incorporate task information. Having defined event structure and feature vectors of temporal context in the two-task lists, changing slowly with each studied item, we now assess CMR predictions in the experimental data.

Event boundaries modulate temporal context during study

CMR assumes that an event boundary leads to a disruption in temporal context, making temporal context after the event boundary less similar to the prior temporal context state. Thus, holding lag constant, the neural similarity in temporal context between two items should be less when those items are separated by an event boundary. CMR predicts that neural similarity should be less across boundaries at any lag, yet because context similarity also decreases with lag, at larger lags this difference becomes more subtle. Thus, CMR predicts the most salient influence of event boundaries for neighboring pairs of items, and here we examine this stricter test of CMR's predictions at study lag = 1. Figure 3B shows CMR's prediction of the neural similarity between pairs of successive items that border an event boundary, as a function of being presented in the same event or different events. Although we show CMR's prediction from a single set of parameters (see Table A1), CMR always predicts greater similarity between neighboring pairs in the same event than neighboring pairs in different events, arising from the core model assumption that a change in source context, or event boundary, leads to a disruption in temporal context (see Appendix A).

We tested CMR's prediction by calculating the neural similarity between temporal context feature vectors of successive items bordering an event boundary during study (see Figure 3C). We found that neural similarity was greater between item pairs studied within the same event than item pairs studied across different events

$(M = 0.007, SD = 0.040, t(169) = 2.12, CI = [0.0005, 0.0126], p = 0.035, d = 0.041)$. This result is consistent with CMR's underlying assumption that there is a disruption to temporal context at

the event boundary, thus leading to reduced temporal similarity of items separated by an event boundary. This also suggests that, for items separated by an event boundary, their weakened neural similarity may reflect their weakened temporal associations.

It is important to consider the alternate explanation that our posited temporal context vectors actually reflect task features that change at the event boundary, including the encoding task and the visual properties of the studied word. If this were the case, then a reduction in neural similarity would reflect a reduction in task, not temporal context, similarity. We took several steps to rule out this alternate explanation. First, temporal context feature vectors are only calculated from control lists, where each list only has a single encoding task. Thus, if a feature vector is autocorrelated across items within a list, it cannot be driven by changes to task features alone. Also attesting to minimal contributions of task across task types, the magnitudes of feature vectors did not differ between the two types of control list types with a single task, and the control list feature vectors inform those of the two-task lists.

Second, because we defined each feature vector by having a sufficiently high item-to-item autocorrelation when summed across lists (see Method), it is less likely that a feature vector's autocorrelation was driven by changes in a single task type. Instead, this metric supported features which were autocorrelated across items irrespective of within-list task type. Third, neural similarity between feature vectors decreased as a function of study lag for item pairs in two-task lists from the same event (and thus with the same task), as well as item pairs in control lists for each task type (also always presented with the same task; see Figure A1). Finally, we ensured that neural similarity differences between items of the same task type or different task types were at equal levels between control lists and two-task lists (see previous section).

Taken together, these results suggest the changes in neural similarity across events reflect a disruption to temporal context. Further, these results suggest that the purported influence of event structure on retrospective temporal judgments may reflect changes to temporal representations during initial perceptual processing. However, to fully appreciate the role of event structure on temporal information on subsequent memory, we will also need to examine these properties during memory retrieval.

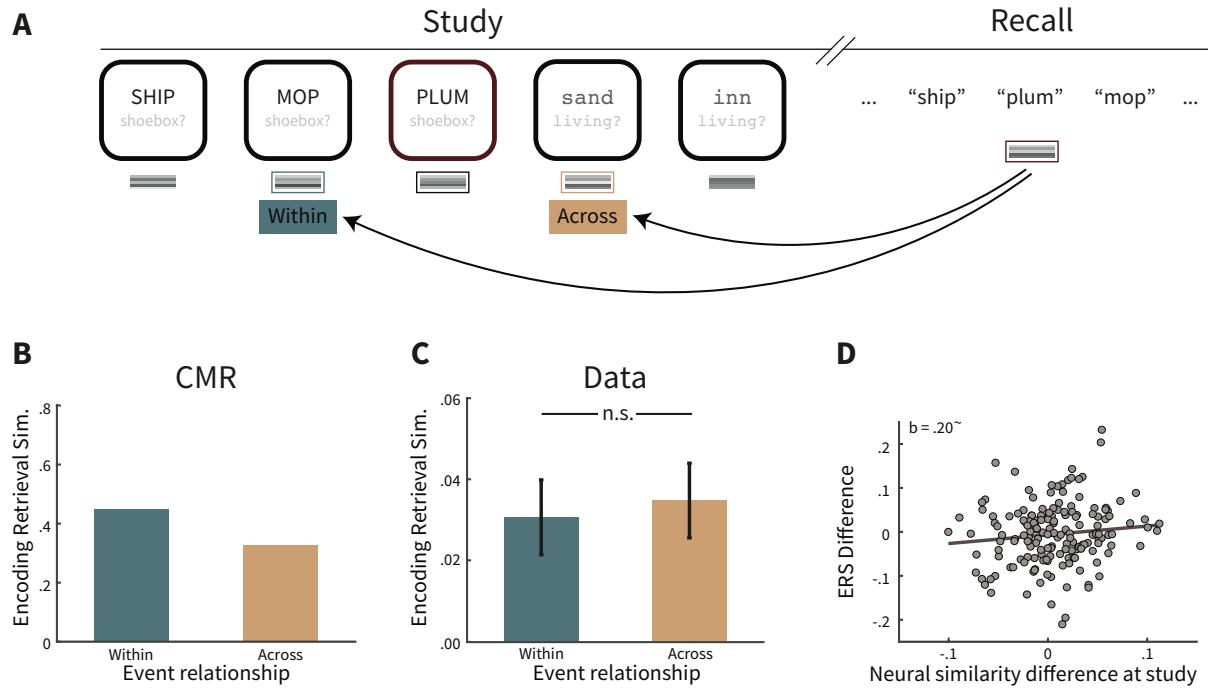


Figure 4

Encoding-retrieval similarity (ERS) in two-task lists. **A.** Calculation of ERS of the current state of context after recall of a preboundary item with its associated neighbors at study: both the preceding neighbor presented with the same task, and thus within the same event (Within), and with its subsequent neighbor across a different event (Across). Task text is for illustrative purposes only; to participants this was implicit from the color, font and case of the word. **B.** CMR predicts that the retrieval of an item bordering an event boundary (e.g. ‘plum’ in A) leads to retrieval of that item’s temporal context states from study, including the disruption to temporal context caused by the event boundary. Thus, the current state of temporal context—which incorporates the item’s retrieved temporal context—should be more similar to the context of adjacent studied item within the same event (e.g. ‘mop’) than the adjacent studied item from a different event (e.g. ‘sand’). However, the difference by event relationship is more subtle than during study (compare with 3B). **C.** Mean ERS in the behavioral data was not significant by event relationship. Error bars represent ± 1 standard error of the mean. **D.** Participants exhibiting greater disruption to temporal context during study also exhibit a greater reinstatement of disruption in temporal context during recall. $\sim p = .06$ (one-tailed).

Reinstatement of event disruptions to temporal context

We next queried temporal context representations during recall, motivated by CMR's assumption that recall of an item evokes retrieval of its context states from study. In the two-task lists, the retrieved context includes the source (task) context, and the temporal context modulated by event boundaries. Thus, CMR predicts that if a preboundary or boundary item is recalled, then the retrieved temporal context of the recalled item should show greater similarity with its within-event neighbor from study in comparison its across-event neighbor from study (see Figure 4A,B). However, the predicted difference in ERS is more subtle during recall than during study (i.e., compare to Figure 3B). In retrieved context models such as CMR, the extent of context reinstatement for each item is defined by a parameter ranging from 0 to 1, with 0 indicating no context reinstatement and 1 indicating perfect reinstatement. If context reinstatement was perfect, then the predictions of study and recall would be the same. To best account for human behavior, the context reinstatement parameter is less than perfect (here, set to .510; see Table A1). As a result, the difference for within-event similarity vs. across-event similarity during recall is less than (a perfect reflection of) the difference in context from study.

We next assessed CMR's prediction in participant EEG data for recall of boundary items and preboundary items (see Figure 4C). ERS was not significantly different for a recalled item if the similarity was calculated with its studied neighbor from the same versus a different event ($M = -0.004, SD = 0.070, t(169) = -0.78, CI = [-0.0147, 0.0064], p = 0.439$). Yet this is not entirely inconsistent with CMR's prediction that the difference in similarity reinstated during recall is less than the ERS difference at study. To further probe whether this nonsignificant difference in ERS might reflect a meaningful signal in temporal context, we examined the across-participant variability in temporal context reinstatement. In particular, we hypothesized that those participants exhibiting greater disruptions to temporal context during study should also exhibit greater reinstatement of those disruptions during recall, even if the mean difference in temporal context was not significant. To test this hypothesis, we calculated each participant's difference in neural similarity at study (within vs. across event in Figure 3C), and correlated this with each participant's ERS difference at recall (see Figure 4C). We used robust regression, a regression method designed to be less sensitive to potential outliers by assigning a weight to each

data point, and inherently downweighting potential outliers. We found that these two difference measures trended towards a positive correlation ($N = 170, b = 0.20$, one-tailed $p = 0.063$; Figure 4D). This suggests that, if a participant experiences the task changes as more salient disruptions to temporal information associated with items, then that participant also reinstates such temporal disruptions when recalling those items. Thus, this result is consistent with our hypothesis that the disruptions to temporal context by event boundaries from study were reinstated during recall. More broadly, this suggests that the influence of event segmentation on retrospective temporal judgments reflects reinstatement of temporal context encoded from study.

Event disruptions to temporal context from study influence recall behavior

We next examined a novel prediction of CMR concerning the impact of event boundaries on free recall behavior, in particular recall transitions. This prediction builds on previous findings establishing that much of the variability of recall transitions in free recall can be explained by the temporal relationships between studied items, as participants are more likely to transition between items presented nearby on the study list (Figure 2B; Healey et al., 2019; Kahana, 1996, 2012). CMR assumes that temporal context drives this temporal organization (see Figure 2A), and so CMR also assumes that recall transitions should be modulated by temporal disruptions imposed by event boundaries. As shown in Figure 5A, CMR predicts that recall transitions *from a preboundary item* should be less likely to the item at lag = +1 in the two-task lists, as compared to transitions of lag = +1 in the control lists (see Figure 2A). This striking prediction stands in contrast to the forward asymmetry usually seen in free recall (Healey et al., 2019; Kahana, 1996, 2012). Yet, according to CMR, an event boundary disrupts temporal context between the preboundary item and the next item (at lag = +1), and so these items do not overlap as much in their temporal context states. As a result, when the preboundary item is recalled and its temporal context is reinstated, the retrieval cue incorporating this context will overlap less with the context of the lag = +1 item. Thus, this state of context does not promote recall of the lag = +1 item as strongly as in a control list. Further, more temporal context is shared between the lag = -1 item and the preboundary item, because these items are from the same event. Thus, CMR predicts that transitions from a preboundary item to the item studied before it, the

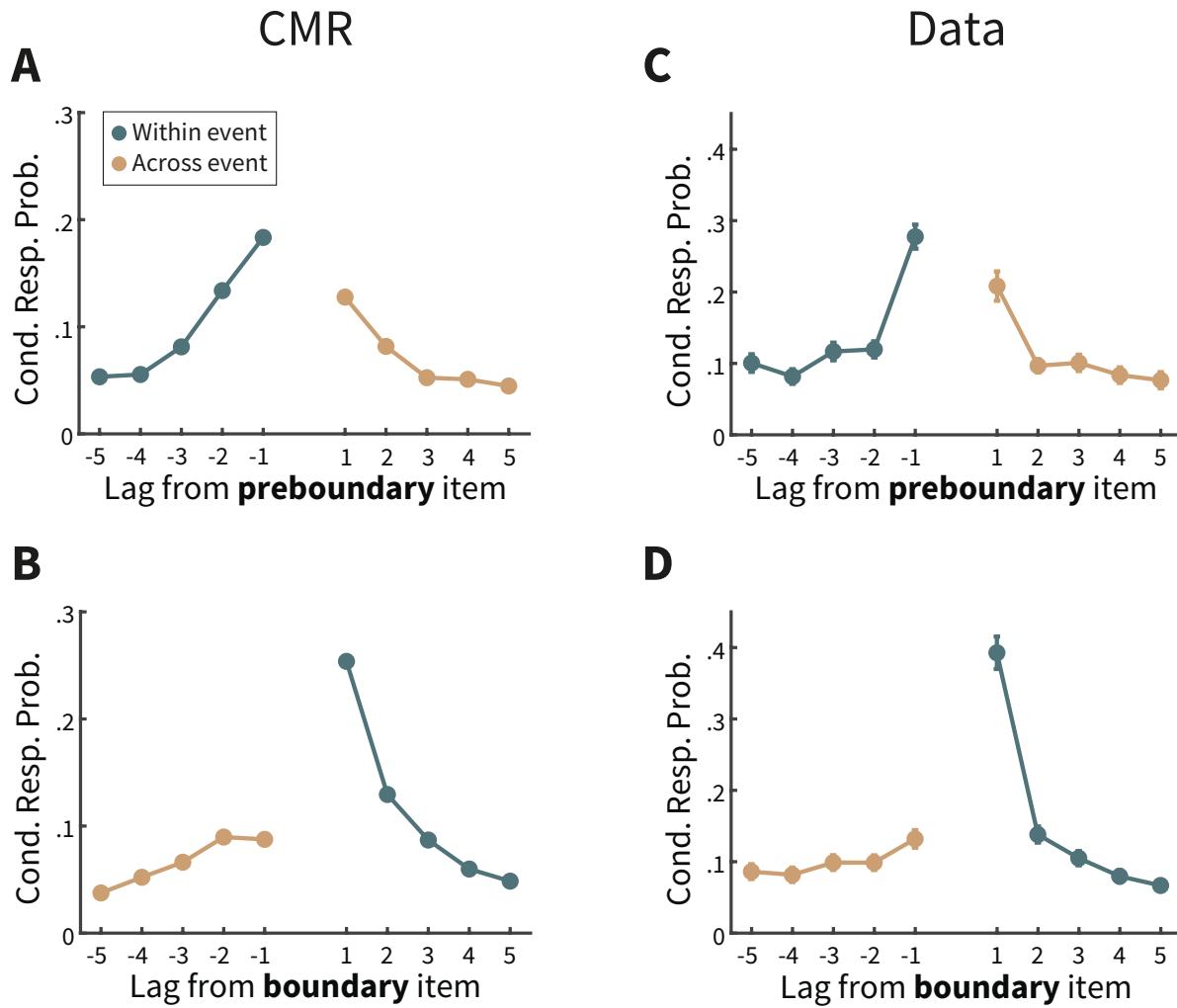
lag = -1 item, is more likely when compared to control lists or even to the lag = +1 item.

In a complementary way, CMR predicts that recall transitions *from boundary items* are modulated as well (see Figure 5B). A transition from a boundary item to its neighbor at lag = +1 should be more likely than in control lists, because these items share both event information and temporal information. Following similar logic, CMR also predicts that a transition from a boundary item should be less likely to the item at lag = -1, because such items were presented in a different event and thus share less temporal context with the just-recalled boundary item.

Next, we examined whether CMR's predictions were upheld in participants' data. To assess these effects statistically, we defined the *temporal modulation score* as the difference in lag-CRP values at $| \text{lag} | = 1$ for transitions made within event minus transitions made across event. (Thus, for preboundary items this score is defined as CRP values at lag = -1 minus those at lag = +1; for boundary item this score is defined as CRP values at lag = +1 minus those at lag = -1). As a baseline, we compared this value to the lag-CRP values at the same lags from the control lists. We found that the distribution of temporal modulation scores from preboundary items was significantly greater in two-task lists than matched lags in control lists

$(M = 0.190, SD = 0.154, t(169) = 16.09, CI = [0.1670, 0.2137], p < .0001, d = 1.41)$. Qualitatively, the lag-CRP in the experimental data (see Figure 5C) exhibits a similar pattern to CMR's prediction, with larger values for transition probabilities for negative lags over positive lags. In addition, the temporal modulation scores from boundary items were also significantly greater in two-task lists than matched lags in control lists (see Figure 5D;

$(M = 0.139, SD = 0.166, t(169) = 10.97, CI = [0.1144, 0.1646], p < .0001, d = 0.98)$. Thus, the recall transitions in two-task lists of the participants' data are consistent with CMR's assumption. During study, event boundaries disrupted temporal representations, and at recall these temporal representations, incorporating the disruption, are reinstated. As a result, this promoted transitions between items with more similar temporal context states, typically within the same event.

**Figure 5**

Recall transitions in two-task lists. **A.** The context maintenance and retrieval (CMR) model predicts that transitions from a preboundary item are more likely in the forward direction, which are more likely to be other items within the same event (darker teal lines) than to items in the following event (lighter orange lines), in contrast to the established bias to make forward transitions (see Figure 2B). **B.** CMR predicts that transitions from a boundary item are more likely in the forward direction, which are more likely to be items in the same event (darker teal lines) than in the backward direction, which are more likely to be items from the preceding event (lighter orange lines), leading to an exaggerated tendency to make forward transitions. **C,D.** Consistent with CMR predictions, participants are more likely to recall items not separated by an event boundary, and more likely to be within the same event (darker teal lines) than items separated by an event boundary and from a different event (lighter orange lines). For more distant items with darker teal lines, items may be from the following event. Cond. Resp. Prob. = Conditional response probability. Error bars represent Loftus and Masson (1994) 95% confidence intervals.

Relating neural temporal context to recall behavior

Having established measures suggestive of the impact of event boundaries on temporal context — neurally and behaviorally — we next assessed whether there was a causal role of the reinstated temporal disruptions, reflecting event structure from study. In particular, we asked whether the neural measures predicted the behavioral measures. According to CMR, both the neural modulation and the behavioral modulation should be greater when there is a greater disruption to temporal context by event information. Although CMR predicts average data, the extent of disruption may vary by participant. If this were the case, then those participants exhibiting a greater modulation by event boundary in their ERS difference should exhibit a greater modulation by event boundaries in their recall transitions. In other words, if each participant's neural measures (see Figure 4) and behavioral measures (see Figure 5) both reflected modulations of temporal context by event boundaries, and if such information influenced memory, then the neural and behavioral measures should be correlated across participants.

We defined a participant's *neural* modulation by temporal context as the difference in ERS for within-event transitions versus across-event transitions (i.e., the difference in the two bars plotted in Figure 4). We defined a participant's *behavioral* modulation by temporal context by summing the temporal modulation scores of preboundary and boundary items. Here we leveraged the variability across participants in the extent to which event boundaries modulate their temporal context states, and we predicted a positive correlation between neural modulation and temporal modulation. Again using robust regression to account for potential outliers, we found that across participants these two measures were weakly correlated (see Figure 6, $N = 170, b = .41$, one-tailed $p = .045$). This suggests that variance in the temporal modulation scores can be explained by the neural temporal context measure. Although there is variability across participants in the extent to which event boundaries modulate their behavioral and neural activity, the correlation across participants suggests that a disruption to temporal context may underlie both effects.

This correlation also argues against the possibility that the significant differences in the temporal modulation scores simply reflect recall organization based on shared event information and shared encoding task, rather than shared temporal context. If temporal context did not

contribute to recall transitions, then we would not expect recall behavior to correlate with the neural measure of temporal context. As further attestation to this point, if the behavioral modulation was driven by shared source or task context, then we would expect the ERS difference to correlate with the degree to which participants transition to items of the same task, irrespective of temporal lag. We calculated this task, or source, clustering score for transitions made from boundary or preboundary items. Across participants, source clustering scores were not correlated with the neural modulation scores ($N = 170, b = 0.003, p > 0.4$), further suggesting that shared task features alone were not driving neural activity. Rather, these correlations are most consistent with CMR's critical assumption that a change in source context, or event boundary, disrupts temporal context.

As another approach to assess the importance of temporal context for memory organization, we examined predictions of a model variant which makes the same assumptions as CMR except that event boundaries do not evoke disruptions in temporal context (best-fit parameters of the pure association model from Polyn et al., 2009a, Figure A3). This model variant was unable to capture the critical findings in behavioral and neural data. Taken together, the results reveal the influence of event structure on temporal context during initial perception, and on memory representations to influence retrieval and recall dynamics.

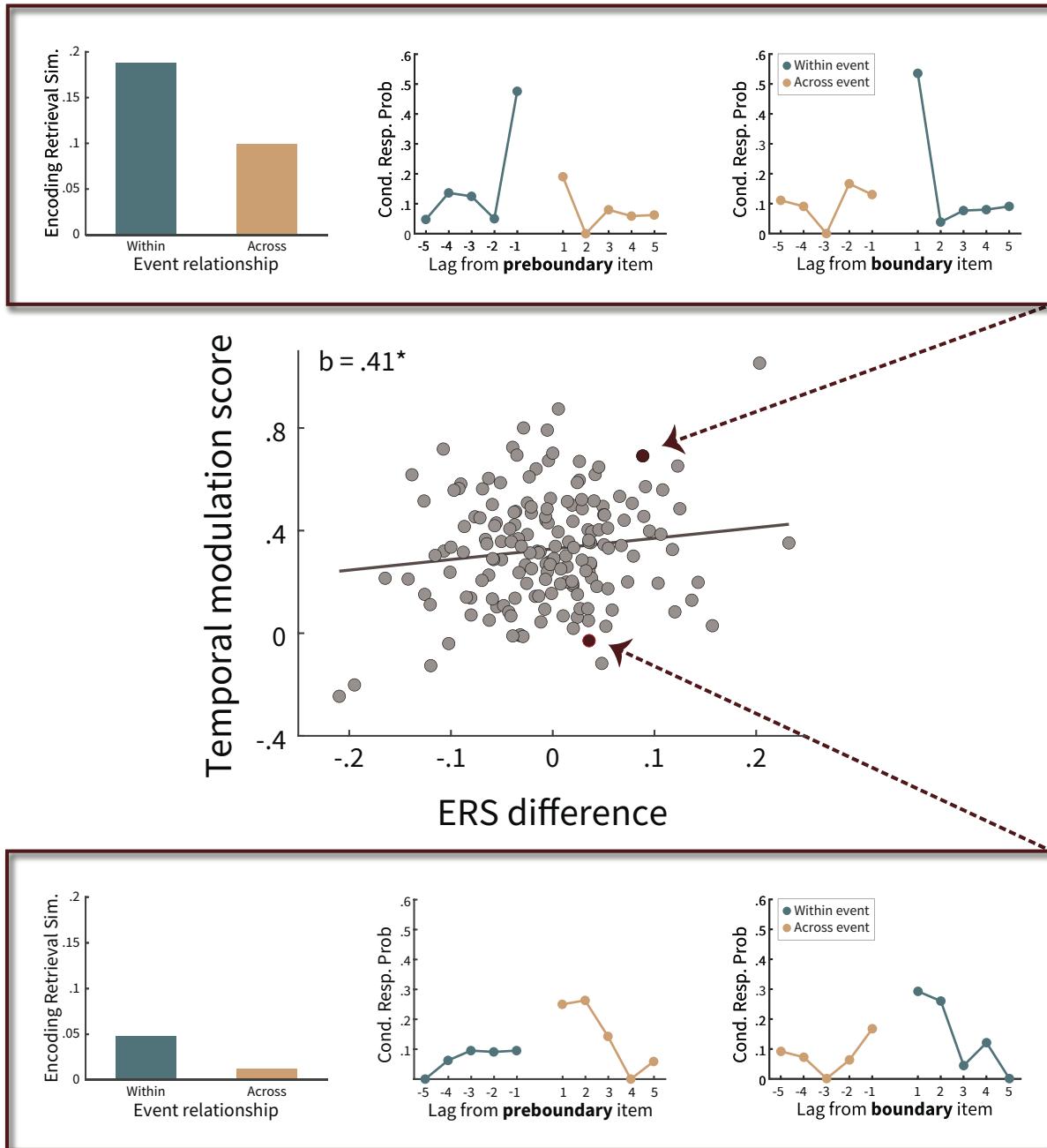


Figure 6

Influence of event boundaries on neural and behavioral measures of temporal context. In this correlation plot, each dot corresponds to a participant. The x-axis reflects the neural measure of event boundary modulation on temporal context; the y-axis reflects a behavioral measure of event boundary modulation on temporal context (see text for details). The top and bottom panels show for two participants the encoding-retrieval similarity (ERS) at recall, used for calculating the x-axis, and conditional response probability (Cond. Resp. Prob.) as a function of lag, used for calculating the y-axis. Top panel: This participant has a high ERS difference at recall and a high temporal modulation score. Bottom panel: This participant has a low ERS difference at recall and a low temporal modulation score, as recall transitions are similar irrespective of whether the transition is from a preboundary item (bottom middle panel) or a boundary item (bottom right panel). * $p < .05$ (one-tailed).

Discussion

How differences emerge between the objective environment and internal experience remains a broad yet fundamental question in cognitive psychology. Such differences can impact perception of information in the moment, as well as how the information becomes represented in memory. Appreciating the interactions between ongoing perception and subsequent memory provides insight into both processes (e.g. Clewett et al., 2019; Zacks et al., 2007). Here we examined the interaction between event segmentation and episodic memory through the lens of temporal context: how online event segmentation influences temporal context, and the consequence of context changes to mnemonic representations. We examined these effects in memory behavior and in neural activity by recording EEG as participants studied and recalled lists of words. To discern the unique contribution of temporal information to event segmentation and memory organization, we compared these results to neural and behavioral predictions of a computational cognitive model. Consistent with model predictions, these results reveal that temporal information plays a primary role in event segmentation and memory: event boundaries disrupt temporal context, impacting memory retrieval even when temporal information is not queried directly.

Our measure of perception with temporal information differs from most other studies relating event segmentation and temporal representations, which ask participants explicitly to make temporal judgments (Ezzyat & Davachi, 2014; Faber & Gennari, 2017; Lositsky et al., 2016). Instead, we measured a neural correlate of temporal context with EEG. This enabled us to assess temporal perception both prospectively during initial perception and encoding, as well as retrospectively during memory retrieval. Further, by not asking participants directly about time, this neural measure allowed us to query how temporal information influences memory dynamics even when such information is not as critical to task performance. The calculation of this temporal measure was motivated by retrieved context models such as CMR. These models assume that context changes slowly with each studied item, and that an item's context state is retrieved when the item is recalled (Figure 1; Howard & Kahana, 2002; Manning et al., 2011). We first established this measure of temporal context in control lists, which only had one task type per list and thus presumably did not impose a strong event structure during study. We demonstrated for

the first time a neural correlate of temporal context in scalp EEG (see Figure 2).

Here we operationalized event boundaries with controlled changes to presented information to help minimize the impact of these changes on memory performance between events: a change in the encoding task performed with each presented item, where the task was indicated by the color, font and case of the item. Event boundaries are often operationalized by more salient changes, such as a change in the semantic category of presented information (e.g. DuBrow & Davachi, 2013; DuBrow & Davachi, 2014; Ezzyat & Davachi, 2014), or features of the presented information such as size or location (e.g. Faber & Gennari, 2015, 2017; Heusser et al., 2016; Heusser et al., 2018; Lositsky et al., 2016; Radvansky & Copeland, 2006). By contrast, we minimized stimulus changes between events to better isolate the contribution of temporal information to event structure and memory. Nonetheless, our operationalization of event boundary — somewhat unpredictable and reflected as a change in encoding task and change in visual stimulus features such as color — is consistent with previous studies which use these features to impose event structure (e.g. DuBrow & Davachi, 2013; DuBrow & Davachi, 2014; Frank et al., 2020; Heusser et al., 2018; Radvansky & Zacks, 2017; Zacks et al., 2011; Zacks et al., 2007).

With our conservative definition of event boundaries, we first verified that EEG activity can provide a neural measure of temporal context by extracting EEG activity from electrodes across mnemonic regions (Long & Kahana, 2017; Long et al., 2014; Weidemann et al., 2009). We next verified several CMR predictions regarding the influence of event boundaries on temporal information. CMR represents event, task and temporal information as context states, and makes testable predictions regarding the interactions between temporal context and event structure. Rather than fit parameters to simulate the current study, we set a stricter threshold of CMR predictions: we simulated data from a pre-existing set of parameters and data, and then assessed CMR predictions in the current study. As a first step to test CMR predictions, we found that neighboring items at study have reduced neural similarity in temporal context when separated by an event boundary (see Figure 3).

Critically, we examined how disruption to temporal context during study subsequently influenced neural and behavioral activity during memory retrieval. While freely recalling list

items, participants exhibited neural activity consistent with reinstatement of temporal context from study. In control lists, the neural temporal context of a recalled item was most similar to the temporal context of its neighbors from study (see Figure 2D), consistent with predictions of the retrieved context model framework (see Figure 2C). In addition, in the lists with two tasks and event structure, we found evidence that participants reinstated temporal context states from study, including temporal context disruptions between events. Specifically, participants exhibiting a larger decrease in temporal context similarity across events during study trended towards exhibiting a greater decrease in this measure during recall (see Figure 4D). This provides support for temporal context reinstatement during recall, as those participants more influenced by the disruption of temporal information during study also reinstates such information during memory retrieval. Although previous studies have demonstrated neural evidence of event-related reinstatement during memory retrieval (Baldassano et al., 2017; Chen et al., 2017; DuBrow & Davachi, 2014; Zadbood et al., 2017), here we show that temporal information related to the event is reinstated as well. Importantly, the free recall test does not explicitly ask participants to remember this temporal information or the event structure. Thus, our results suggest that recall of an item automatically evokes retrieval of temporal context states from study, consistent with CMR's assumption.

Patterns of recall behavior, and their relation to neural activity, also attested to the influence of these neural temporal context states on memory organization. Although in a free recall task, participants may recall items in any order, recall order in two-task lists reflected the influence of event segmentation. Specifically, recall transitions were less likely between items studied in different events than items studied in the same event (see Figure 5C,D). CMR predicts this effect with the presented parameter set (see Figure 5A,B), and would also predict this effect with most parameter sets, due to its core assumption that an event boundary disrupts temporal context and thus weakens memory associations. This pattern of behavior was most striking for successive recalls between neighboring items (i.e. lag = ± 1). Because event boundaries disrupt associative transitions, these results are consistent with findings that items across events have weaker memory associations (Baldassano et al., 2017; DuBrow & Davachi, 2013; DuBrow & Davachi, 2014; Swallow et al., 2011; Swallow et al., 2009). Further, because transitions between

temporal neighbors are impacted, and these relate to the neural measure of temporal context, we also interpret our results as consistent with longer temporal duration judgments across event boundaries (Clewett et al., 2020; DuBrow & Davachi, 2013; Ezzyat & Davachi, 2014; Faber & Gennari, 2017; Lositsky et al., 2016).

We found evidence of these relationships in mean participant data, as well as in across-participant variability. From the viewpoint of CMR, a participant exhibiting a greater difference in neural similarity during study experienced larger disruptions to temporal context, which should manifest during the recall test in both neural activity and memory performance. Future work remains to characterize how and why such changes vary by participant and by individual event, as well as whether more subtle changes to context might be better inferred as a shift in, rather than a disruption to, context (DuBrow et al., 2017). Nonetheless, to our knowledge, this is one of the first studies to link directly, through neural activity and behavior, how event structure can influence temporal representations to impact memory performance. Critically, these results suggest that the influence of event boundaries on memory associations and temporal judgments reflect the direct impact of temporal context, even though event boundaries are defined by nontemporal features. Further, whereas retrospective temporal judgments have been posited to reflect properties of memory distinct from those operating during online perceptual processing (Grondin, 2010; Pöppel, 1997), our results suggest that the foundation is laid for this critical temporal information during initial perception and encoding.

We assessed the success of CMR's predictions based on qualitative patterns in mean data, as well as expected correlations across participants. It is not surprising that the neural measure of temporal context, extracted from brain activity across regions and timepoints, is more variable than the mean activity predicted by CMR. Variability in neural activity would produce weaker neural similarity and ERS values when compared to model predictions. Further, because neural analyses of two-task lists used feature vectors calculated from control lists only, any variability in neural activity across lists would lead to weaker neural similarity in two-task lists. By contrast, CMR does not assume any variability across lists. Further, we assert that such noise may also account for the variability in the neural similarity difference (at study) or ERS difference (at recall) in two-task lists between neighboring items from the same or different events. Although a

surprisingly high number of participants do not exhibit a negative difference as predicted by CMR, this could again be due to a noisier measure in participant data when compared to model predictions. Alternatively, greater similarity across boundaries may reflect rapid reinstatement of the prior event at the event boundary (Ben-Yakov et al., 2014; Sols et al., 2017). Yet if the rapid reinstatement were the primary cognitive mechanism, we would not expect disruptions in neural similarity during study, nor decreased transitions across event boundaries during retrieval. Thus, although there could be tension at boundaries between temporal disruption after the prior event vs. reinstatement of the prior event, we would argue that disruption accounts for a greater amount of variance in the current data, consistent with prior studies examining the influence of event segmentation on temporal perception (e.g. DuBrow & Davachi, 2013; DuBrow & Davachi, 2014; Ezzyat & Davachi, 2014; Faber & Gennari, 2017; Lositsky et al., 2016). With respect to model development, future work could incorporate other posited mechanisms of event segmentation and other cognitive processes which may contribute to the variability in the neural representations. With respect to EEG, future work should extend and replicate these results while incorporating other contributions from purely electrophysiological signals unrelated to mnemonic activity measures. Here we sought to find meaning in the qualitative pattern of results purely due to temporal representations.

We took several steps to ensure that we were assessing a measure of temporal context and how it was impacted by source context and event segmentation, rather than simply measuring source context itself. That is, we would expect a neural measure of source context — like a neural measure of temporal context — to exhibit lower neural similarity between item pairs separated by an event boundary, and so it is important to consider the possibility that a change in temporal context truly reflects temporal information. Our approach to measuring temporal context was motivated by CMR, which assumes each item is associated with a temporal context and a source context. Source context is defined by the encoding task of the current item and a change in source (task) context necessarily causes a disruption to temporal context. A CMR model variant, the “pure association model variant”, which assumes that a change in source context does not cause a disruption to temporal context fails to predict recall behavior correctly, as it overpredicts recall transitions between same-source items (Polyn et al., 2009a). Thus, on a theoretical level,

CMR motivated the expected results because it did not consider temporal context alone, but rather interactions between temporal context and source context. We also verified in the present study that the “full” CMR model provides a better fit to the current set of findings than the pure association model variant.

Experimentally, we calculated slowly changing feature vectors of temporal context in control lists which do not have task changes. Thus, if a feature vector changes slowly in a list without task changes, it is unlikely that this feature vector primarily incorporates source features. It is also not a concern that an individual feature vector only reflects a single task type, both because we defined feature vectors with strong autocorrelation across all control lists with three possible task types, and because these feature vectors were calculated from electrodes previously implicated in memory behavior even when collapsing across the two single-task list types (Long et al., 2014). In addition, when examining the impact of event boundaries on feature vectors during recall, we found this measure to relate to temporal, but not task, recall organization. Taken together, we interpret our neural measure of temporal context to reflect the impact of, but not the representations of, source context.

We also assert that the current results cannot be explained by positional information. According to retrieved context models such as CMR, temporal context changes slowly with each studied item. However, this also means that temporal context shares properties with some positional accounts of recall, whereby each item is associated with a positional code in the list. If the positional codes change slowly with each studied item, then neural feature vectors may reflect positional information, not purely temporal information. Yet neural similarity of feature vectors decreased across lists, suggesting that the neural representations changed slowly over time, not just as a function of within-list position. Although it is possible that within-list positional code information is extracted from temporal context representations (Logan & Cox, 2021), our results rule out the possibility that feature vectors are purely positional codes maintained across lists. Thus, these results may generalize to studies of event segmentation beyond shorter lists of discrete items, where discrete positional codes may play less of a role (e.g. Baldassano et al., 2017; Ezzyat & Davachi, 2014; Zacks et al., 2001).

We also examined contributions of positional information by considering variability across

serial positions in the primary analyses. Recall probability is greater in some serial positions than others (Figure A2), and thus items in these positions may contribute more to ERS analyses. We conducted several analyses to ensure that our results did not reflect properties unique to items from these serial positions (see Appendix B). Further, CMR also incorporates variability in recall across serial positions. When recall begins, the current context cues recall, and context is a recency-weighted sum of studied items. Thus, this context state promotes recall of recently presented items, leading CMR to predict greater recall of recency items. In addition, CMR assumes that context-to-item associations are greater for early list positions. Given an early list and a mid-list item with the same strength in context, it is more likely that context will cue, and that CMR will recall, an early-list item with greater context-item weight (see Appendix A).

Several possible mechanisms may induce the increase weighting of the context-to-item associations in CMR, and an investigation of these mechanisms has been actively investigated for decades. Primacy items may benefit from greater attention or novelty (Davelaar et al., 2005; Farrell, 2012; Lewandowsky & Farrell, 2008). As another explanation, primacy items may also benefit from greater activation energy, or less fatigue, at the beginning of the list (Brown et al., 2000; Lohnas et al., 2020; Page & Norris, 1998; Tulving & Rosenbaum, 2006). If participants are silently rehearsing items during study, then primacy items benefit from more rehearsals (Rundus, 1971; Tan & Ward, 2000). In addition, primacy items may benefit from having less interference, as fewer items precede their study (Brown et al., 2007; Murdock, 1960; Neath, 1993).

Although CMR's primacy mechanism does not distinguish between these possibilities, its primacy mechanism does have a strong implication for context states. Critically, context representations are not influenced by the primacy mechanism, only the context-to-item associations are. However, neural activity can differ between items studied earlier and later in the list (Reddy et al., 2021; Rushby et al., 2002; Sederberg et al., 2006; Serruya et al., 2014; Umbach et al., 2020; Wiswede et al., 2007), which may raise concerns that feature vectors—the posited measure of temporal context—may incorporate neural activity from primacy items. Yet several follow-up analyses which incorporate serial position information suggest that our results are not driven by activity unique to early-list items (see Appendix B). Taken together, these results support our claim that feature vectors are consistent with assumptions of temporal context in

retrieved context models, at the exclusion of source or positional information.

Our neural definition of temporal context was motivated by prior work incorporating assumptions from retrieved context models such as CMR (Folkerts et al., 2018; Howard et al., 2012; Manning et al., 2011). However, other definitions of temporal context have been queried with neural data. Notably, Kragel et al. (2021) queried brain regions which represent temporal context by exhibiting discriminable activity between the serial positions of items within a list, between lists within a session, and across sessions. In the current study, we were interested in a temporal context measure which changed slowly with each studied item, in order to address whether these representations would change more across events than within events. If we employed the temporal context measure defined by Kragel et al. (2021), this measure might confound the start of an event with the start of a list, and might not preserve pairwise similarities between items over time. Future work remains to determine how these brain regions provide complementary or shared information to inform representations of temporal context in event structure and episodic memory.

At the same time, our results are broadly consistent with established properties and neural regions implicated in temporal context. Although EEG recordings generally have poor spatial resolution, our results add confidence that a neural measure of temporal context can be measured at the scalp. Here we recorded activity from electrodes over mnemonic brain regions in the temporal lobe and frontal lobe (Long & Kahana, 2017; Long et al., 2014; Weidemann et al., 2009). Previously, studies recording intracranial EEG have reported activity consistent with temporal context from regions in the temporal lobe (Folkerts et al., 2018; Howard et al., 2012; Long et al., 2017; Long et al., 2014; Manning et al., 2011; Manns et al., 2007; Yaffe et al., 2014). Studies using functional magnetic resonance imaging have also found regions of the temporal lobe exhibit properties consistent with temporal context (Hsieh et al., 2014; Jenkins & Ranganath, 2010; Kimura et al., 2010; Konishi et al., 2002; Kragel et al., 2015; Nielson et al., 2015; Turk-Browne et al., 2012). In addition, regions of the frontal lobe have also shown activity consistent with temporal context representations (Cabeza et al., 1997; Eyler Zorilla et al., 1996; Jenkins & Ranganath, 2010; Konishi et al., 2002).

Although overlap between intracranial and scalp recordings of the temporal lobe might be

more intuitive, scalp recordings from in the included frontal lobe regions has been posited to relate to intracranial medial temporal lobe activity (Long et al., 2014). If electrodes from either the frontal lobe or the temporal lobe might reflect medial temporal lobe activity, it is worth noting that the medial temporal lobe, and the hippocampus in particular, is critical for episodic memory (Aggleton & Brown, 1999; Davachi, 2006; Eichenbaum, 2004; Fernandez et al., 1999; Goyal et al., 2018; Sugar & Moser, 2019) and for temporal representations (Eichenbaum, 2014; MacDonald et al., 2011; Tsao et al., 2018; Umbach et al., 2020). Because temporal context is defined by incorporating features of previously studied information, our neural measure of temporal context also is consistent with prior studies implicating the medial temporal lobe, and the hippocampus in particular, in binding of episodic features within and across events (Davachi, 2004; Heusser et al., 2016; Pacheco Estefan et al., 2019; Richmond & Zacks, 2017; Staresina & Davachi, 2006, 2009). Through these mnemonic representations, the medial temporal lobe has also been shown to represent event-related information, including an influence on memory performance (Baldassano et al., 2017; DuBrow & Davachi, 2014; Ezzyat & Davachi, 2014; Lositsky et al., 2016). In addition, regions of the frontal lobe — including lateral regions more easily measurable from the scalp — are also critical for episodic memory (Blumenfeld et al., 2011; Hanslmayr & Staudigl, 2014; Long et al., 2014; McAndrews & Milner, 1991; Paller & Wagner, 2002) including free recall (Long et al., 2010; Sederberg et al., 2007; Staresina & Davachi, 2006), and support event segmentation (Baldassano et al., 2017; Chen et al., 2017; DuBrow & Davachi, 2016; Ezzyat & Davachi, 2014; Kurby & Zacks, 2008; Sols et al., 2017; Zacks et al., 2001). Thus, our results are broadly consistent with prior studies examining the neural correlates of temporal context, episodic memory and event segmentation. Further linking how these regions are important for these three seemingly disparate cognitive functions, our results suggest that temporal context plays a critical role in episodic memory, and that event segmentation influences memory representations to incorporate temporal context.

We presented CMR simulations to provide an intuition for the impact of temporal context on event structure and memory. We do not wish to suggest that CMR is the *only* model which can predict our results, but the current results point to challenges or areas of further development for other models. For instance, CMR's predictions are consistent with other current cognitive

model frameworks which segment sequences of items and make predictions of memory. Whereas other models can infer event structure based on stimulus features and predictability (e.g. Radvansky, 2012; Zacks et al., 2007), CMR needs to be provided the event structure explicitly to update source context and temporal context. Despite the different objectives of these types of models, CMR assumes event boundaries cause a disruption to memory associations, similar to established accounts of event processing such as Event Segmentation Theory and the Event Horizon Model (Kurby & Zacks, 2008; Radvansky, 2012; Radvansky & Zacks, 2017; Zacks et al., 2001). Thus, our findings may be explained by the Event Horizon Model framework, which embodies Event Segmentation Theory. According to this framework, a current event model is held in working memory, and each event boundary updates the model. As a result, it is more difficult to retrieve information outside of the current event or the currently retrieved event (Radvansky & Zacks, 2017; Swallow et al., 2011). Thus, the Event Horizon Model should predict the decrease in recall transitions between neighboring items from different events. However, development of the Event Horizon Model has focused primarily on information repeated across events, whereas in the current study each event was comprised of unique novel items. Currently this model does not make quantitative predictions and is more agnostic with respect to the role of temporal representations, so it remains to be fully developed to make predictions of memory and temporal information.

In a complementary way, existing models of episodic memory may be amenable to incorporating event segmentation findings. For instance, the model of Farrell (2012) accounts for major findings in free recall by assuming that participants naturally segment list items into groups, where items within a group share a common group context. If the Farrell model assumes that successive items with the same task are represented in a group, then it would predict that recall transitions are more likely between items within the same event. However, the Farrell model assumes that the items within a group are recalled in order, and thus the model may have difficulty predicting the increase in lag = -1 transitions for preboundary items (Figure 5C). Another intriguing contrast between CMR and the Farrell model concerns control lists. Unlike CMR, which treats a control list as a single long event, the Farrell model would assume that the list is subdivided into chunks of variable length (see also Romani et al., 2016). If event

segmentation took place even in control lists, this should lead to a greater drift in the feature vectors for some items more so than others. By contrast, as a simplifying assumption CMR usually assumes that context changes at a constant rate for each item. We kept context drift rates constant to be consistent with prior work of successful predictions using the retrieved context framework. The success of CMR in the current simulations suggests that varying this parameter was not necessary to account for the qualitative pattern of the results, and varying the drift rate leads to interactions with other behavioral effects of less interest here (Polyn et al., 2012). Future work will need to distinguish whether the variance in temporal autocorrelation across items in control lists reflects inevitable noise in neural data, or relates meaningfully to an individual's endogenous segmentation.

Recently, Frank et al. (2020) presented Structured Event Memory (SEM), a computational model of event cognition. Like CMR, SEM can incorporate event structure during study to predict memory performance. Unlike CMR, SEM can infer event structure across a range of naturalistic stimuli, and can predict memory performance based on the inferred structure. Although CMR needs to be provided the event structure explicitly, in a list-learning paradigm like the current study, SEM assumes that event structure is inferred correctly based on encoding tasks. Yet even if both models were provided with the event structure of the current study, SEM does not include explicit representations of temporal information nor has it been applied to free recall data. However, it would be an intriguing direction to examine how SEM might account for the effects presented here, in the absence of temporal information. In a complementary way, CMR may be a suitable framework to extend by incorporating more complex stimulus features and event structures. Such an extension would build upon CMR simulations accounting for the neural correlates of task representations (Morton et al., 2013) and temporal representations (Kragel et al., 2015; Manning et al., 2011). Nonetheless, the current study, which kept changes between events as minimal as possible, allowed us to disentangle the role of temporal information in episodic memory and event structure. Thus, we have begun this process by presenting CMR predictions which integrate these two types of representations on a neural level and on a behavioral level. Such predictions capture how temporal representations, shaped by event segmentation, are formed during encoding to influence memory retrieval and recall

performance, even though neither event nor temporal information are tested directly.

Conclusions

An underlying objective of cognitive neuroscience and psychology is to characterize the transformation of external environment into internal experience. This transformation begins during initial perception, and then influences how information is encoded into memory. Context is posited to be important for both perception and memory, and here we link the critical role of temporal context to both perceptual processes and episodic memory representations. Previous studies have suggested that event segmentation influences temporal perception and memory, but the role of temporal context in these processes remained unclear. In particular, it was less clear how and when temporal context influenced memory representations, as well as whether this was an automatic process, or only manifested when tested explicitly. Here we characterized the influence of event structure on temporal representations while participants studied and recalled words. This task imposed event structure without requiring explicit retrieval of temporal or event information.

Temporal perception and episodic memory share a complex relationship that is not fully understood. Our approach of simultaneously considering behavior and neural activity through the lens of a computational model provides novel insight into their interactions. Leveraging these methodological tools, we showed that event segmentation, even when defined with nontemporal features, impacts temporal representations during initial perception and memory encoding. In turn, temporal representations influenced brain activity and behavior during memory retrieval. These results suggest that temporal context plays a primary, not secondary, role in incorporating event structure into episodic memory. Our results underscore the impact of event segmentation on temporal representations, and the role of temporal context in linking initial perceptual processes with memory representations.

Constraints on Generality

We anticipate that these results will generalize beyond the current subject pool of young adults, aged 18–30 and primarily from Philadelphia universities, to young adults in general. On a

behavioral level, online populations exhibit similar behavioral results of temporal contiguity and free recall (e.g. Mulligan et al., 2022; Mundorf et al., 2021). On a neural level, we also anticipate the findings to generalize across young adult populations and to other recording techniques. We chose brain regions from a study which found mnemonic activity consistent across scalp EEG and intracranial EEG, where the latter was collected with a broader range of ages and geographic locations than PEERS (Burke et al., 2014; Long et al., 2014). We next consider the generalization of the stimuli, which in the current study were common nouns such that each word can be considered its own episodic memory. Although these simple stimuli were tested within minutes after being studied, we expect the behavioral results to generalize to other types of stimuli and longer timescales. First, our work builds on findings of the behavioral temporal contiguity effect, which generalizes to autobiographical experiences and timescales of months or years (Cortis Mack et al., 2017; Moreton & Ward, 2010; Uitvlugt & Healey, 2019). Second, behavioral results of event segmentation with more dynamic stimuli and longer timescales are consistent with studies using simple stimuli (for recent reviews see Clewett et al., 2019; Frank et al., 2020; Radvansky & Zacks, 2017).

Public Significance Statement

Our internal, subjective representation of an experience is not always an accurate depiction of what actually occurred. This article examines two types of representations which are susceptible to inaccuracy: memory for events and the perception of time. When there is a salient change in the environment, this can cause memory for information prior to the change to be more weakly associated to information occurring after the change. The salient change can also lead to the perception that more time has passed. However, it is less clear if temporal information directly influences memory, or plays a more passive role. We examined a direct link between the role of salient changes, temporal perception and memory. Subjects studied words and then had to recall the words from memory. To measure temporal perception, we examined brain activity while subjects performed these tasks. When there was a change study task, color and font of the words, this led to a larger update in subjects' representation of time. Brain activity of temporal representations when subjects studied a word were reinstated when subjects recalled the word

from memory, including the larger updates after changes. These changes also influenced the order in which subjects recalled words from studied lists. The results suggest that temporal information plays a primary role in updating and organizing memories.

Context

The first author (LL) came up with the main ideas for this work during her postdoctoral fellowship with the senior author (LD), who has several recent articles examining interactions between event segmentation and memory performance. The perspective of the first author was inspired by her prior published articles using the retrieved context model framework to account for free recall phenomena. The middle author (KH) had recently analyzed EEG correlates of context reinstatement, and also had published articles with the retrieved context model framework. The first authors' main ideas were developed, refined and framed conceptually thanks to the other two authors.

CRediT Taxonomy

Lynn Lohnas: Conceptualization, Methodology, Software, Formal analysis, Writing - Original Draft, Visualization, Funding acquisition **Karl Healey:** Conceptualization, Methodology, Software, Validation, Writing - Review & Editing, Funding acquisition **Lila Davachi:** Conceptualization, Resources, Writing - Review & Editing, Supervision

References

Aggleton, J. P., & Brown, M. W. (1999). Episodic memory, amnesia, and the hippocampal-anterior thalamic axis. *Behavioral and Brain Sciences*, 22, 425–489.

Ahn, W.-Y., Busemeyer, J. R., Wagenmakers, E.-J., & Stout, J. C. (2008). Comparison of decision learning models using the generalization criterion method. *Cognitive Science*, 32, 1376–1402. <https://doi.org/10.1080/03640210802352992>

Anderson, J. R., & Matessa, M. (1997). A production system theory of serial memory. *Psychological Review*, 104, 728–748.

Baldassano, C., Chen, J., Zadbood, A., Pillow, J. W., Hasson, U., & Norman, K. A. (2017). Discovering event structure in continuous narrative perception and memory. *Neuron*, 95, 709–721. <https://doi.org/10.1016/j.neuron.2017.06.041>

Ben-Yakov, A., Rubinson, M., & Dudai, Y. (2014). Shifting gears in hippocampus: Temporal dissociation between familiarity and novelty signatures in a single event. *Journal of Neuroscience*, 34(39), 12973–12981. <https://doi.org/10.1523/JNEUROSCI.1892-14.2014>

Block, R. A. (1982). Temporal judgments and contextual change. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(6), 530–544. <https://doi.org/10.1037/0278-7393.8.6.530>

Blumenfeld, R. S., Parks, C. M., Yonelinas, A. P., & Ranganath, C. P. (2011). Putting the pieces together: The role of dorsolateral prefrontal cortex in relational memory encoding. *Journal of cognitive neuroscience*, 23, 257–265. <https://doi.org/10.1162/jocn.2010.21459>

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, 114(3), 539–576.

Brown, G. D. A., Preece, T., & Hulme, C. (2000). Oscillator-based memory for serial order. *Psychological Review*, 107(1), 127–181. <https://doi.org/10.1037/0033-295X.107.1.127>

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, 106(3), 551–581.

Burke, J. F., Long, N. M., Zaghloul, K. A., Sharan, A. D., Sperling, M. R., & Kahana, M. J. (2014). Human intracranial high-frequency activity maps episodic memory formation in

space and time. *NeuroImage*, 85, 834–843.
<https://doi.org/10.1016/j.neuroimage.2013.06.067>

Busemeyer, J. R., & Wang, Y.-M. (2000). Model comparisons and model selections based on generalization criterion methodology. *Journal of Mathematical Psychology*, 44, 171–189.
<https://doi.org/10.1006jmps.1999.128>

Cabeza, R., Mangels, J., Nyberg, L., Habib, R., Houle, S., McIntosh, A. R., & Tulving, E. (1997). Brain regions differentially involved in remembering what and when: A PET study. *Neuron*, 19(4), 863–870.

Chen, J., Leong, Y. C., Honey, C. J., Yong, C. H., Norman, K. A., & Hasson, U. (2017). Shared memories reveal shared structure in neural activity across individuals. *Nature Neuroscience*, 20, 115–125. <https://doi.org/10.1038/nn.4450>

Clewett, D., Gasser, C., & Davachi, L. (2020). Pupil-linked arousal signals track the temporal organization of events in memory. *Nature Communications*, 11(4007), 1–14.
<https://doi.org/10.1038/s41467-020-17851-9>

Clewett, D., DuBrow, S., & Davachi, L. (2019). Transcending time in the brain: How event memories are constructed from experience. *Hippocampus*.
<https://doi.org/10.1002/hipo.23074>

Conrad, R. (1960). Serial order intrusions in immediate memory. *British Journal of Psychology*, 51, 45–48. <https://doi.org/10.1111/j.2044-8295.1960.tb00723.x>

Cortis Mack, C., Cinel, C., Davies, N., Harding, M., & Ward, G. (2017). Serial position, output order, and list length effects for words presented on smartphones over very long intervals. *Journal Of Memory And Language*, 97, 61–80. <https://doi.org/10.1016/j.jml.2017.07.009>

Davachi, L. (2004). The ensemble that plays together, stays together. *Hippocampus*, 14, 1–3.
<https://doi.org/10.1002/hipo.20004>

Davachi, L. (2006). Item, context and relational episodic encoding in humans. *Current Opinion in Neurobiology*, 16(6), 693–700. <https://doi.org/10.1016/j.conb.2006.10.012>

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of

recency effects. *Psychological Review*, 112, 3–42. <https://doi.org/10.1037/0033-295X.112.1.3>

DuBrow, S., & Davachi, L. (2013). The influence of contextual boundaries on memory for the sequential order of events. *Journal of Experimental Psychology: General*, 142(4), 1277–1286. <https://doi.org/10.1037/a0034024>

DuBrow, S., Rouhani, N., Niv, Y., & Norman, K. A. (2017). Does mental context drift or shift? *Current Opinion in Behavioral Sciences*, 17, 141–146. <https://doi.org/10.1016/j.cobeha.2017.08.003>

DuBrow, S., & Davachi, L. (2014). Temporal memory is shaped by encoding stability and intervening item reactivation. *The Journal of Neuroscience*, 34(42), 13998–14005. <https://doi.org/10.1523/JNEUROSCI.2535-14.2014>

DuBrow, S., & Davachi, L. (2016). Temporal binding within and across events. *Neurobiology of Learning and Memory*, 134, 107–114. <https://doi.org/10.1016/j.nlm.2016.07.011>

Eichenbaum, H. (2004). Hippocampus: Cognitive processes and neural representations that underlie declarative memory. *Neuron*, 44(1), 109–120. <https://doi.org/10.1016/j.neuron.2004.08.028>

Eichenbaum, H. (2014). Time cells in the hippocampus: A new dimension for mapping memories. *Nature Reviews Neuroscience*, 15(11), 732–744. <https://doi.org/10.1038/nrn3827>

Eyler Zorilla, L. T., Aguirre, G. K., Zarahn, E., Cannon, T. D., & D'Esposito, M. (1996). Activation of the prefrontal cortex during judgments of recency: A functional MRI study. *NeuroReport*, 7, 2803–2806. <https://doi.org/10.1097/00001756-199611040-00079>

Ezzyat, Y., & Davachi, L. (2011). What constitutes an episode in episodic memory? *Psychological Science*, 22(2), 243–252. <https://doi.org/10.1177/0956797610393742>

Ezzyat, Y., & Davachi, L. (2014). Similarity breeds proximity: Pattern similarity within and across contexts is related to later mnemonic judgments of temporal proximity. *Neuron*, 81(5), 1179–1189. <https://doi.org/10.1016/j.neuron.2014.01.042>

Faber, M., & Gennari, S. P. (2015). In search of lost time: Reconstructing the unfolding of events from memory. *Cognition*, 143, 193–202. <https://doi.org/10.1016/j.cognition.2015.06.014>

Faber, M., & Gennari, S. P. (2017). Effects of event structure on prospective duration judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(8), 1203–1214.
<https://doi.org/10.3758/BF03205466>

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review, 119*(2), 223–71. <https://doi.org/10.1037/a0027371>

Fernandez, G., Effern, A., Grunwald, T., Pezer, N., Lehnertz, K., Dumpelmann, M., Van Roost, D., & Elger, C. E. (1999). Real-time tracking of memory formation in the human rhinal cortex and hippocampus. *Science, 285*, 1582–1585.
<https://doi.org/10.1126/science.285.5433.1582>

Folkerts, S., Rutishauser, U., & Howard, M. W. (2018). Estimating scale-invariant future in continuous time. *Journal of Neuroscience*.
<https://doi.org/10.1523/JNEUROSCI.2312-17.2018>

Frank, N. T., Norman, K. A., Ranganath, C., Zacks, J. M., & Gershman, S. J. (2020). Structured event memory: A neuro-symbolic model of event cognition. *Psychological Review, 127*(3), 327–361. <https://doi.org/10.1037/rev0000177>

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations and interpretation. *Journal of Experimental Psychology: General, 141*(1), 2–18.
<https://doi.org/10.1037/a0024338>

Goyal, A., Miller, J. F., Watrous, A. J., Lee, S. A., Coffey, T., Sperling, M. R., Sharan, A., Worrell, G., Berry, B., Lega, B., Jobst, B. C., Davis, K. A., Inman, C., Sheth, S. A., Wanda, P. A., Ezzyat, Y., Das, S. R., Stein, J., Gorniak, R., & Jacobs, J. (2018). Electrical stimulation in hippocampus and entorhinal cortex impairs spatial and temporal memory. *Journal of Neuroscience, 19*, 4471–4481.
<https://doi.org/10.1523/JNEUROSCI.3049-17.2018>

Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics, 72*(3).
<https://doi.org/10.3758/APP.72.3.561>

Hanslmayr, S., & Staudigl, T. (2014). How brain oscillations form memories—a processing based perspective on oscillatory subsequent memory effects. *Neuroimage*, 85, 648–655.
<https://doi.org/10.1016/j.neuroimage.2013.05.121>

Healey, M. K., & Kahana, M. J. (2014). Is memory search governed by universal principles or idiosyncratic strategies? *Journal of Experimental Psychology: General*, 143(2), 575–596.
<https://doi.org/10.1037/a0033715>

Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review*, 123, 23–69. <https://doi.org/10.1037/rev0000015>

Healey, M. K., Long, N. M., & Kahana, M. J. (2019). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, 26(3), 699–720.
<https://doi.org/10.3758/s13423-018-1537-3>

Henson, R. N. A. (1996). *Short-term memory for serial order* (Doctoral dissertation). University of Cambridge, England.

Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology*, 36, 73–137. <https://doi.org/10.1006/cogp.1998.0685>

Heusser, A. C., Poppel, D., Ezzyat, Y., & Davachi, L. (2016). Episodic sequence memory is supported by a theta-gamma phase code. *Nature Neuroscience*.
<https://doi.org/10.1038/nn.4374>

Heusser, A. C., Ezzyat, Y., Shiff, I., & Davachi, L. (2018). Perceptual Boundaries Cause Mnemonic Trade-Offs Between Local Boundary Processing and Across-Trial Associative Binding. 44(7), 1075–1090. <https://doi.org/10.1037/xlm0000503>

Howard, M. W., Jing, B., Rao, V. A., Provyn, J. P., & Datey, A. V. (2009). Bridging the gap: Transitive associations between items presented in similar temporal contexts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35, 391–407.
<https://doi.org/10.1037/a0015002>

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 25(4), 923–941. <https://doi.org/10.1037/0278-7393.25.4.923>

Howard, M. W., & Kahana, M. J. (2002). A distributed representation of temporal context. *Journal of Mathematical Psychology*, 46, 269–99. <https://doi.org/10.1006/jmps.2001.1388>

Howard, M. W., Viskontas, I. V., Shankar, K. H., & Fried, I. (2012). Ensembles of human MTL neurons “jump back in time” in response to a repeated stimulus. *Hippocampus*, 22, 1833–1847. <https://doi.org/10.1002/hipo.22018>

Howard, M. W., Youker, T. E., & Venkatadass, V. (2008). The persistence of memory: Contiguity effects across hundreds of seconds. *Psychonomic Bulletin & Review*, 15, 58–63. <https://doi.org/10.3758/PBR.15.1.58>

Hsieh, L.-T., Gruber, M., Jenkins, L., & Ranganath, C. (2014). Hippocampal activity patterns carry information about objects in temporal cortex. *Neuron*, 81(5), 1165–1178. <https://doi.org/10.1016/j.neuron.2014.01.015>

Jenkins, L. J., & Ranganath, C. (2010). Prefrontal and medial temporal lobe activity at encoding predicts temporal context memory. *Journal of Neuroscience*, 30(46), 15558–15565. <https://doi.org/10.1523/JNEUROSCI.1337-10.2010>

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, 24(1), 103–109. <https://doi.org/10.3758/BF03197276>

Kahana, M. J. (2012). *Foundations of human memory* (1st). Oxford University Press.

Kahana, M. J., Howard, M. W., & Polyn, S. M. (2008). Associative retrieval processes in episodic memory. In H. L. Roediger III (Ed.), *Cognitive psychology of memory. Vol. 2 of Learning and memory: A comprehensive reference, 4 vols. (J. Byrne, Editor)*. Elsevier.

Kahana, M. J., Lohnas, L. J., Healey, M. K., Aka, A., Broitman, A. W., Crutchley, E., Crutchley, P., Alm, K. H., Katerman, B. S., Miller, N. E., Kuhn, J. R., Li, Y., Long, N. M., Miller, J. R., Paron, M. D., Pazdera, J. K., Pedisich, I., & Weidemann, C. T. (2022). The Penn electrophysiology of encoding and retrieval study. *PsyArXiv*, 1–33. <https://doi.org/10.31234/osf.io/bu5x8>

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, 20(1), 141–151. <https://doi.org/10.1177/001316446002000116>

Kimura, H. M., Hirose, S., Kunimatsu, A., Chikazoe, J., Jimura, K., Watanabe, T., Abe, O., Ohtomo, K., Miyashita, Y., & Konishi, S. (2010). Differential temporo-parietal cortical

networks that support relational and item-based recency judgments. *NeuroImage*, 49(4), 3474–3480. <https://doi.org/10.1016/j.neuroimage.2009.11.003>

Konishi, S., Uchida, I., Machida, T., Shirouzu, I., & Miyashita, Y. (2002). Neural correlates of recency judgment. *Journal of Neuroscience*, 22(21), 9549–9555.

Kragel, J. E., Ezzyat, Y., Lega, B. C., Sperling, M. R., Worrell, G. A., Gross, R. E., Jobst, B. C., Sheth, S. A., Zaghloul, K. A., Stein, J. M., & Kahana, M. J. (2021). Distinct cortical systems reinstate the content and context of episodic memories. *Nature Communications*, 12, 1–10. <https://doi.org/10.1038/s41467-021-24393-1>

Kragel, J. E., Morton, N. W., & Polyn, S. M. (2015). Neural activity in the medial temporal lobe reveals the fidelity of mental time travel. *Journal of Neuroscience*, 35(7), 2914–26. <https://doi.org/10.1523/JNEUROSCI.3378-14.2015>

Kurby, C. A., & Zacks, J. M. (2008). Segmentation in the perception and memory of events. *Trends in Cognitive Sciences*, 12(2). <https://doi.org/10.1016/j.tics.2007.11.004>

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. *The Psychology of Learning and Motivation*, 49, 1–48. [https://doi.org/10.1016/S0079-7421\(08\)00001-7](https://doi.org/10.1016/S0079-7421(08)00001-7)

Loftus, G. R., & Masson, M. E. J. (1994). Using confidence intervals in within-subject designs. *Psychonomic Bulletin & Review*, 1, 476–490. <https://doi.org/10.3758/BF03210951>

Logan, G. D. (2021). Serial order in perception, memory, and action. *Psychological Review*, 128(1), 1–44. <https://doi.org/10.1037/rev000253>

Logan, G. D., & Cox, G. E. (2021). Serial memory: Putting chains and position codes in context. *Psychological Review*. <https://doi.org/10.1037/rev0000327>

Lohnas, L. J., Davachi, L., & Kahana, M. J. (2020). Neural fatigue influences memory encoding in the human hippocampus. *Neuropsychologia*, 143, 1–7. <https://doi.org/10.1016/j.neuropsychologia.2020.107471>

Lohnas, L. J., & Healey, M. K. (2021). The role of context in episodic memory: Behavior and neurophysiology. *Psychology of Learning and Motivation*, 75, 157–199. <https://doi.org/10.1016/bs.plm.2021.06.003>

Lohnas, L. J., & Kahana, M. J. (2014a). Compound cuing in free recall. *Journal of Experimental Psychology: Learning, Memory and Cognition, 40*(1), 12–24.
<https://doi.org/10.1037/a0033698>

Lohnas, L. J., & Kahana, M. J. (2014b). A retrieved context account of spacing and repetition effects in free recall. *Journal of Experimental Psychology: Learning Memory and Cognition, 40*(3), 755–764. <https://doi.org/10.1037/a0035585>

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Intralist and interlist effects in free recall. *Psychological Review, 122*(2), 337–363.
<https://doi.org/10.1037/a0039036>

Long, N. M., & Kahana, M. J. (2017). Modulation of task demands suggests that semantic processing interferes with the formation of episodic associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 43*(2), 167–176.
<https://doi.org/10.1037/xlm0000300>

Long, N. M., Öztekin, I., & Badre, D. (2010). Separable prefrontal cortex contributions to free recall. *Journal of Neuroscience, 30*(33), 10967–10976.
<https://doi.org/10.1523/JNEUROSCI.2611-10.2010>

Long, N. M., Sperling, M. R., Worrell, G. A., Davis, K. A., Gross, R. E., Lega, B. C., Jobst, B. C., Sheth, S. A., Zaghloul, K., Stein, J. M., & Kahana, M. J. (2017). Contextually mediated spontaneous retrieval is specific to the hippocampus. *Current Biology, 27*, 1–6. <https://doi.org/10.1016/j.cub.2017.02.054>

Long, N. M., Burke, J. F., & Kahana, M. J. (2014). Subsequent memory effect in intracranial and scalp EEG. *NeuroImage, 84*, 488–494. <https://doi.org/10.1016/j.neuroimage.2013.08.052>

Lositsky, O., Chen, J., Toker, D., Honey, C. J., Poppenk, J. L., Hasson, U., & Norman, K. A. (2016). Neural Pattern Change During Encoding of a Narrative Predicts Retrospective Duration Estimates. *eLife, 043075*. <https://doi.org/10.1101/043075>

MacDonald, C., Lepage, K., Eden, U., & Eichenbaum, H. (2011). Hippocampal “time cells” bridge the gap in memory for discontiguous events. *Neuron, 71*(4), 737–749.
<https://doi.org/10.1016/j.neuron.2011.07.012>

Manning, J. R., Polyn, S. M., Baltuch, G., Litt, B., & Kahana, M. J. (2011). Oscillatory patterns in temporal lobe reveal context reinstatement during memory search. *Proceedings of the National Academy of Sciences, USA*, 108(31), 12893–12897.

<https://doi.org/10.1073/pnas.1015174108>

Manns, J. R., Howard, M. W., & Eichenbaum, H. (2007). Gradual changes in hippocampal activity support remembering the order of events. *Neuron*, 56(3), 530–40.

<https://doi.org/10.1016/j.neuron.2007.08.017>

McAndrews, M. P., & Milner, B. (1991). The frontal cortex and memory for temporal order. *Neuropsychologia*, 29(9), 849–859.

Miller, J. F., Weidemann, C. T., & Kahana, M. J. (2012). Recall termination in free recall. *Memory & Cognition*, 40, 540–550. <https://doi.org/10.3758/s13421-011-0178-9>

Moreton, B. J., & Ward, G. (2010). Time scale similarity and long-term memory for autobiographical events. *Psychonomic Bulletin & Review*, 17(4), 510–515.

<https://doi.org/10.3758/PBR.17.4.510>

Morton, N. W., Kahana, M. J., Rosenberg, E. A., Sperling, M. R., Sharan, A. D., & Polyn, S. M. (2013). Category-specific neural oscillations predict recall organization during memory search. *Cerebral Cortex*, 23(10), 2407–2402. <https://doi.org/10.1093/cercor/bhs229>

Mulligan, N. W., Buchin, Z. L., & Zhang, A. L. (2022). The testing effect with free recall: Organization, attention, and order effects. *Journal Of Memory And Language*, 125, 104333. <https://doi.org/10.1016/j.jml.2022.104333>

Mundorf, A. M. D., Lazarus, L. T. T., Mitchell, M. G., & Healey, M. K. (2021). A test of retrieved context theory: Dynamics of recall after incidental encoding. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 47, 1264–1287.

<https://doi.org/10.1037/xlm0001001>

Murdock, B. B. (1960). The distinctiveness of stimuli. *Psychological Review*, 67, 16–31.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, 64, 482–488. <https://doi.org/10.1037/h0045106>

Murdock, B. B., & Okada, R. (1970). Interresponse times in single- trial free recall. *Journal of Verbal Learning and Verbal Behavior*, 86, 263–267. <https://doi.org/10.1037/h0029993>

Neath, I. (1993). Contextual and distinctive processes and the serial position function. *Journal of Memory and Language*, 32, 820–840. <https://doi.org/10.1006/jmla.1993.1041>

Nielson, D. M., Smith, T. A., Sreekumar, V., Dennis, S., & Sederberg, P. B. (2015). Human hippocampus represents space and time during retrieval of real-world memories. *Proceedings of the National Academy of Sciences*, 112(35), 11078–11083. <https://doi.org/10.1073/pnas.1507104112>

Norman, K. A., Polyn, S. M., Detre, G. J., & Haxby, J. V. (2006). Beyond mind-reading: multi-voxel pattern analysis of fMRI data. *Trends in Cognitive Sciences*, 10(9), 424–430. <https://doi.org/10.1016/j.tics.2006.07.005>

Osth, A. F., & Dennis, S. (2015). Prior-list intrusions in serial recall are positional. *Journal of Experimental Psychology : Learning, Memory, and Cognition*, 41(6), 1893–1901. <https://doi.org/10.1037/xlm0000110>

Pacheco Estefan, D., Sánchez-Fibla, M., Duff, A., Principe, A., Rocamora, R., Zhang, H., Axmacher, N., & Verschure, P. F. M. J. (2019). Coordinated representational reinstatement in the human hippocampus and lateral temporal cortex during episodic memory retrieval. *Nature Communications*, 10(2255). <https://doi.org/10.1038/s41467-019-09569-0>

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, 105, 761–781. <https://doi.org/10.1037/0033-295X.105.4.761-781>

Paller, K. A., & Wagner, A. D. (2002). Observing the transformation of experience into memory. *Trends in Cognitive Sciences*, 6(2), 93–102. [https://doi.org/10.1016/S1364-6613\(00\)01845-3](https://doi.org/10.1016/S1364-6613(00)01845-3)

Polyn, S. M., Kragel, J. E., Morton, N. W., McCluey, J. D., & Cohen, Z. D. (2012). The neural dynamics of task context in free recall. *Neuropsychologia*, 50, 447–457. <https://doi.org/10.1016/j.neuropsychologia.2011.08.025>

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009a). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, 116, 129–56. <https://doi.org/10.1037/a0014420>

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009b). Task context and organization in free recall. *Neuropsychologia*, 47, 2158–2163.

<https://doi.org/10.1016/j.neuropsychologia.2009.02.013>

Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences*, 1(2), 56–61. [https://doi.org/10.1016/S1364-6613\(97\)01008-5](https://doi.org/10.1016/S1364-6613(97)01008-5)

Radvansky, G. A., & Copeland, D. E. (2006). Walking through doorways causes forgetting: Situation models and experienced space. *Memory & Cognition*, 34(5), 1150–1156.

<https://doi.org/10.3758/BF03193261>

Radvansky, G. A., & Zacks, J. M. (2014). *Event cognition* (1st). Oxford University Press.

Radvansky, G. A. (2012). Across the event horizon. *Current Directions in Psychological Science*, 21(4), 269–272. <https://doi.org/10.1177/0963721412451274>

Radvansky, G. A., & Zacks, J. M. (2017). Event boundaries in memory and cognition. *Current Opinion in Behavioral Sciences*, 17, 133–140. <https://doi.org/10.1016/j.cobeha.2017.08.006>

Reddy, L., Zoefel, B., Possel, J. K., Peters, J., Dijksterhuis, D. E., Poncet, M., van Straaten, C. W., Baayen, J. C., Idema, S., & Self, M. W. (2021). Human hippocampal neurons track moments in a sequence of events human hippocampal neurons track moments in a sequence of events. *Journal of Neuroscience*, 41(31), 6714–6725.

<https://doi.org/10.1523/JNEUROSCI.3157-20.2021>

Richmond, L. L., & Zacks, J. M. (2017). Constructing experience: Event models from perception to action. *Trends in Cognitive Sciences*, 21(12), 962–980.

<https://doi.org/10.1016/j.tics.2017.08.005>

Romani, S., Katkov, M., & Tsodyks, M. (2016). Practice makes perfect in memory recall. *Learning and Memory*, 23(4), 169–173. <https://doi.org/10.1101/lm.041178.115>

Rundus, D. (1971). An analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 89, 63–77. <https://doi.org/10.1037/h0031185>

Rushby, J. A., Barry, R. J., & Johnstone, S. J. (2002). Event-related potential correlates of serial-position effects during an elaborative memory test. *International Journal of Psychophysiology*, 46, 13–27.

Schapiro, A. C., Roger, T. T., Cordova, N. I., Turk-Browne, N. B., & Botvinick, M. M. (2013). Neural representations of events arise from temporal community structure. *Nature Neuroscience*, 16, 486–492. <https://doi.org/10.1038/nn.3331>

Sederberg, P. B., Gauthier, L. V., Terushkin, V., Miller, J. F., Barnathan, J. A., & Kahana, M. J. (2006). Oscillatory correlates of the primacy effect in episodic memory. *NeuroImage*, 32(3), 1422–1431. <https://doi.org/10.1016/j.neuroimage.2006.04.223>

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, 115(4), 893–12. <https://doi.org/10.1037/a0013396>

Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., Litt, B., Brandt, A., & Kahana, M. J. (2007). Gamma oscillations distinguish true from false memories. *Psychological Science*, 18(11), 927–932. <https://doi.org/10.1111/j.1467-9280.2007.02003.x>

Serruya, M. D., Sederberg, P. B., & Kahana, M. J. (2014). Power shifts track serial position and modulate encoding in human episodic memory. *Cerebral Cortex*, 24, 403–413. <https://doi.org/10.1093/cercor/bhs318>

Sols, I., DuBrow, S., Davachi, L., & Fuentemilla, L. (2017). Event boundaries trigger rapid memory reinstatement of the prior events to promote their representation in long-term memory. *Current Biology*, 27, 3499–3504. <https://doi.org/10.1016/j.cub.2017.09.057>

Speer, N. K., & Zacks, J. M. (2005). Temporal changes as event boundaries: Processing and memory consequences of narrative time shifts. *Journal of Memory and Language*, 53(1), 125–140. <https://doi.org/10.1016/j.jml.2005.02.009>

Staresina, B. P., & Davachi, L. (2006). Differential encoding mechanisms for subsequent associative recognition and free recall. *Journal of Neuroscience*, 26(36), 9162. <https://doi.org/10.1523/JNEUROSCI.2877-06.2006>

Staresina, B. P., & Davachi, L. (2009). Mind the gap: Binding experiences across space and time in the human hippocampus. *Neuron*, 63(2), 267–276. <https://doi.org/10.1016/j.neuron.2009.06.024>

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.), *Cognitive psychology*

and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer. American Psychological Association. <https://doi.org/10.1037/10895-018>

Sugar, J., & Moser, M. (2019). Episodic memory: Neuronal codes for what, where, and when. *Hippocampus*, 29, 1190–1205. <https://doi.org/10.1002/hipo.23132>

Swallow, K. M., Barch, D. M., Head, D., Maley, C. J., Holder, D., & Zacks, J. M. (2011). Changes in events alter how people remember recent information. *Journal of Cognitive Neuroscience*, 23(5), 1052–1064. <https://doi.org/doi.org/10.1162/jocn.2010.21524>

Swallow, K. M., Zacks, J. M., & Abrams, R. A. (2009). Event boundaries in perception affect memory encoding and updating. *Journal of Experimental Psychology: General*, 138(2), 236–257. <https://doi.org/10.1037/a0015631>

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1589–1626. <https://doi.org/10.1037/0278-7393.26.6.1589>

Tsao, A., Sugar, J., Wang, C., Knierim, J. J., Moser, E. I., & Moser, M. (2018). Integrating time from experience in the lateral entorhinal cortex. *Nature*, 561, 57–62. <https://doi.org/10.1038/s41586-018-0459-6>

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

Tulving, E., & Rosenbaum, R. S. (2006). Distinctiveness and memory. In R. R. Hunt & J. B. Worthen (Eds.). Oxford University Press.

Turk-Browne, N. B., Simon, M. G., & Sederberg, P. B. (2012). Scene representations in parahippocampal cortex depend on temporal context. *Journal of Neuroscience*, 32(21), 7202–7207. <https://doi.org/10.1523/JNEUROSCI.0942-12.2012>

Uitvlugt, M. G., & Healey, M. K. (2019). Temporal Proximity Links Unrelated News Events in Memory. *Psychological Science*, 30(1), 92–104. <https://doi.org/10.1177/0956797618808474>

Umbach, G., Kantak, P., Jacobs, J., Kahana, M. J., Pfeiffer, B. E., Sperling, M., & Lega, B. C. (2020). Time cells in the human hippocampus and entorhinal cortex support episodic memory. *Proceedings of the National Academy of Sciences*, 117(45), 28463–28474. <https://doi.org/10.1073/pnas.2013250117>

Unsworth, N. (2008). Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal-contextual search. *Journal of Memory and Language*, 59, 223–236.
<https://doi.org/10.1016/j.jml.2008.04.002>

Unsworth, N., Spillers, G. J., & Brewer, G. A. (2012). Dynamics of context-dependent recall: An examination of internal and external context change. *Journal Of Memory And Language*, 66, 1–16. <https://doi.org/10.1016/j.jml.2011.05.001>

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, 108(3), 550–592.
<https://doi.org/10.1037/0033-295X.108.3.550>

Ward, G., Tan, L., & Grenfell-Essam, R. (2010). Examining the relationship between free recall and immediate serial recall: The effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 36(5), 1207–1241.
<https://doi.org/10.1037/a0020122>

Weidemann, C. T., & Kahana, M. J. (2016). Assessing recognition memory using confidence ratings and response times. *Royal Society Open Science*, 3, 150670.
<https://doi.org/10.1098/rsos.150670>

Weidemann, C. T., Mollison, M. V., & Kahana, M. J. (2009). Electrophysiological correlates of high-level perception during spatial navigation. *Psychonomic Bulletin & Review*, 16(2), 313–319. <https://doi.org/10.3758/PBR.16.2.313>

Wiswede, D., Rüsseler, J., & Münte, T. F. (2007). Serial position effects in free memory recall—an erp-study. *Biological Psychology*, 75, 185–193.
<https://doi.org/10.1016/j.biopspsycho.2007.02.002>

Yaffe, R. B., Kerr, M. S., Damera, S., Sarma, S. V., Inati, S. K., & Zaghloul, K. A. (2014). Reinstatement of distributed cortical oscillations occurs with precise spatiotemporal dynamics during successful memory retrieval. *Proceedings of the National Academy of Sciences*, 111(52), 18727–8732.

Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, 4(6), 651–655. <https://doi.org/10.1038/88486>

Zacks, J. M., Kurby, C. A., Eisenberg, M. L., & Haroutunian, N. (2011). Prediction error associated with the perceptual segmentation of naturalistic events. *Journal of Cognitive Neuroscience*, 23(12), 4057–4066. https://doi.org/10.1162/jocn_a_00078

Zacks, J. M., Speer, N. K., Swallow, K. M., Braver, T. S., & Reynolds, J. R. (2007). Event perception: A mind-brain perspective. *Psychological Bulletin*, 133, 273–293. <https://doi.org/10.1037/0033-2909.133.2.273>

Zadbood, A., Chen, J., Leong, Y. C., Norman, K. A., & Hasson, U. (2017). How we transmit memories to other brains: Constructing shared neural representations via communication. *Cerebral cortex*, 27, 4988–5000. <https://doi.org/10.1093/cercor/bhw202>

Zwaan, R. A. (1996). Processing narrative time shifts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22(5), 1196. <https://doi.org/10.1037/0278-7393.22.5.1196>

Appendix A

Model simulations

Context Maintenance and Retrieval Model (CMR)

Here we provide an overview of the CMR model, highlighting the components most relevant to interactions between temporal context and event segmentation. CMR stores representations of item features, their corresponding contexts, and the associations between items and context. When an item in serial position i is studied, this activates the item's associated feature representation, \mathbf{f}_i . CMR assumes a localist representation, such that item i is represented by a vector with 1 for those features corresponding to the item's serial position and associated encoding task, and 0's everywhere else. This feature representation is used to generate an input to update context, \mathbf{c}_i^{IN} , using an association matrix that links item features (F) to context states (C), M^{FC} , with the simple product $\mathbf{c}_i^{\text{IN}} = M^{FC}\mathbf{f}_i$. This input to context is then used to update context:

$$\mathbf{c}_i = \rho_i \mathbf{c}_{i-1} + \beta \mathbf{c}_i^{\text{IN}}, \quad (\text{A.1})$$

where β is a model parameter, and ρ is set so that $|\mathbf{c}| = 1$ (a mathematical convenience). Larger values of β mean that the input to context will update context to a greater amount. When β is larger, ρ is smaller; as a result, the prior context is downweighted more. Note that in these equations, the index of each context and item feature is from the item i . Context is updated with each studied item, and thus changes slowly over time.

Critical to CMR's ability to capture event segmentation, an event boundary updates temporal context beyond the updating from the studied item alone. Whenever there is a change in the task associated with an item, this causes CMR to present an additional 'item' to the model and update temporal context. However, these boundary items are not stored in memory and cannot be retrieved during the recall period. Nonetheless, they function to update context in a similar way to studied items, in that they update context according to Equation A.1. Whereas temporal context for a studied item is updated by setting $\beta = \beta_{enc}^{\text{temp}}$, temporal context for an event boundary item is updated with value d . This additional item thus disrupts the temporal context state, causing temporal context to drift even further from the current temporal context

state. As a result, the temporal context between two items should be less similar when they are separated by an event boundary (see Figure 3B). However, in two-task lists, item pairs in the same event should have approximately the same levels of similarity and decrease with lag, as if they were presented in a control list; both types of items are presented with the same task and not separated by event boundary (Figure A1A). This latter prediction is upheld in the experimental data, as neural similarity decreases by lag for both types of item pairs (Figure A1B).

An update to context also updates the association matrices between items and contexts (M^{FC} , M^{CF}) as the Hebbian outer product (e.g., $\Delta M^{FC} = \mathbf{c}_i \mathbf{f}_i^\top$). CMR incorporates a primacy gradient for the weight given to the updated context states in M^{CF} , such that context is updated more strongly from early list items:

$$\Delta M_{exp}^{CF} \propto (\phi_s e^{-\phi_d(i-1)} + 1) \mathbf{f}_i \mathbf{c}_i^\top. \quad (\text{A.2})$$

In this way, the ϕ_s gives extra weight to items with smaller values of i in earlier list positions, and ϕ_d scales the rate at which this advantage decays with i . These early-list items may benefit from extra weight strength due to greater novelty, attention or energy (Brown et al., 2000; Farrell, 2012; Lohnas et al., 2020; Page & Norris, 1998; Sederberg et al., 2008; Tulving & Rosenbaum, 2006), and also shares similarities with positional code models attributing greater weight to early list items (for a detailed discussion see Logan & Cox, 2021). In addition to being updated by experimental associations, CMR also stores of pre-experimental semantic associations. However, we omit details of this process as it is less important in the current study where we average across serial positions and the lists contain mostly unrelated words.

Once CMR is presented with a list of items, the model next attempts to ‘recall’ items as a participant would. The model’s current state of context, reflecting the temporal history of studied items, is used to cue recall. Specifically, a feature strength is determined for each item, based on their relative weight in context: $\mathbf{f}_r^{\text{IN}} = M^{CF} \mathbf{c}_r$, where elements of \mathbf{f}_r^{IN} correspond to studied items. These feature strengths are then used as input to a noisy decision process that outputs a single recalled item, where items with larger strengths have a greater probability of being recalled (Usher & McClelland, 2001). As described in the next paragraph, because these feature strengths are determined from the current state of context, items with similar context states to the current

context (i.e. items with shared temporal context or source context) are more likely to be recalled. At the beginning of an immediate free recall period like the current study, this cues items presented at the end of the list, causing CMR to predict the recency effect. Further, the heavier weighting of early list items supports recall of early list items, leading CMR to predict the primacy effect (Figure A2A).

Once CMR recalls an item, this item is then presented to the model again, and updates context according to Equation A.1. Thus, a recalled item generates an input to context, and now this input includes the context from when the just-recalled item was originally studied. In addition, the rate at which context is updated, β , can vary between study and recall. Thus, the temporal context drift rates during study (*encoding*) and *recall* are termed and β_{enc}^{temp} and β_{rec}^{temp} , respectively. Once context is updated from the new item, this new state of context is used to recall another item. In this way, recall of an item i leads to reinstatement of the context of item i , and thus promotes recall of items with similar context states to i , including items with similar temporal contexts (i.e., items presented nearby on the list), as well as items with similar source contexts (i.e., items presented with the same task). This critical assumption of context updating during retrieval leads to CMR's predictions of neural reinstatement (see Figure 2C) and temporal contiguity in the behavioral lag-CRPs (see Figure 2A).

Instead of determining the parameter values that would best capture the present data, here we examined whether CMR could account for the data qualitatively based on best-fit parameters from a data set used previously (Polyn et al., 2009a). In this way, we were not fitting CMR to the data presented here, but rather using pre-existing parameter values and simulated data to predict the pattern of results for this data. To generate the CMR predictions, we presented the model with the same lists as participants viewed in that original study. Thus, the data used to generate CMR predictions included 45 participants with 631 control lists and 631 two-task lists each with list-length = 24.

Pure association model variant

To assess the necessity of the temporal disruption mechanism in CMR, we also examined predictions from a model variant which shares identical properties to CMR except that event

boundaries do not disruption temporal context (i.e. $d = 0$). Thus, this model variant shares core assumptions with CMR: temporal context changes slowly over time, and temporal context states are reinstated during recall to influence neural activity and behavior. Using the best-fit parameters for this model variant from Polyn et al. (2009a) as shown in Table 1, we confirmed these core assumptions based on predictions in control lists (Figure A3A).

Critically, the predictions of the pure association model variant were not consistent with the experimental data in two-task lists. First, similarity in neural context during study was the same between neighboring items regardless of whether they were studied within the same event or different events (Figure A3B), inconsistent with the reduced neural similarity of neighboring items studied in different events (compare with Figure 3C). As a result, ERS did not differ between neighboring same-event and different-event item pairs (Figure A3C).

On a behavioral level, like CMR the pure association model variant captures the enhanced recall probability of same-event items for $|lag| = 1$ (Figure A3D,E,F; for preboundary items $M = 0.032, SD = 0.112, t(169) = 3.72, CI = [0.0150, 0.0489], p = 0.0003, d = 0.33$; for boundary items $M = 0.026, SD = 0.110, t(169) = 3.06, CI = [0.0091, 0.0426], p = 0.003, d = 0.19$). Thus, this increased probability may reflect shared event or task information between the neighboring items. However, the pure association model fails to capture the reduced recall of items from different events (in the experimental data: for preboundary items, $M = 0.158, SD = 0.103, t(169) = 20.11, CI = [0.1740, 0.1429], p < .00001, d = 1.20$; for boundary items, $M = 0.114, SD = 0.101, t(169) = 14.69, CI = [0.1289, 0.0984], p < .00001, d = 1.43$). Instead the model predicted approximately equal recall probability to items in control lists. Taken together, these results underscore that CMR requires the assumption that each event boundary induces disruption to temporal context in order to make qualitatively accurate predictions of neural activity and memory behavior.

| Parameter | Full | Pure Association |
|----------------------|-------|------------------|
| β_{enc}^{temp} | 0.776 | 0.767 |
| β_{rec}^{temp} | 0.510 | 0.468 |
| β^{source} | 0.588 | 0.681 |
| L_{sw}^{CF} | 0.129 | 0.171 |
| d | 0.767 | 0 |
| γ^{FC} | 0.898 | 0.799 |
| s | 2.78 | 2.71 |
| κ | 0.111 | 0.053 |
| λ | 0.338 | 0.272 |
| η | 0.159 | 0.126 |
| τ | 0.174 | 0.145 |
| ϕ_s | 1.07 | 0.881 |
| ϕ_d | 0.981 | 0.641 |

Table A1

Note: Best-fit parameters from the full model of Polyn et al. 2009, determined using a genetic algorithm fitting technique.

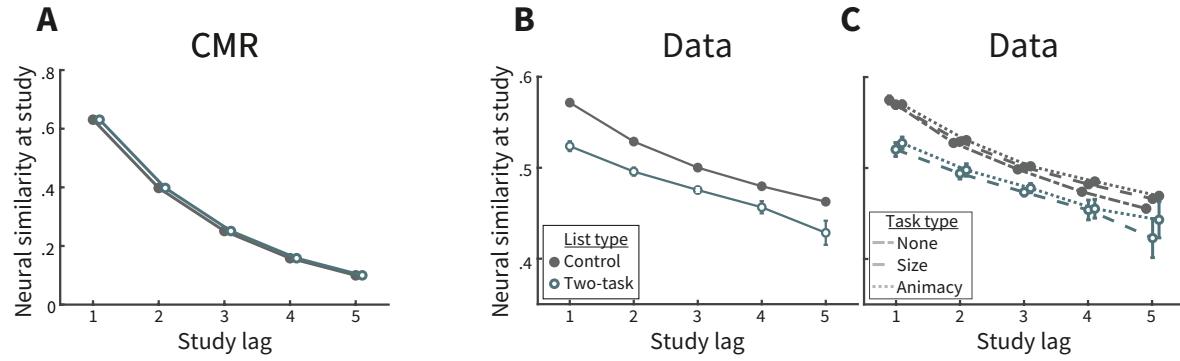


Figure A1

5

Neural similarity during study by study lag. **A.** CMR predicts that neural similarity in temporal context between two items should decrease as a function of lag. In two-task lists, items presented within the same event should have identical neural similarity values as in one-task lists, regardless of task type. **B.** Participant data. As predicted by CMR (and as confirmation of our approach to calculate a neural measure of temporal context), neural similarity decreases as a function of lag for items in control lists, and for items from the same event in two-task lists. **C.** Further subdividing by task type, neural similarity also decreases with lag in participant data. Error bars represent Loftus and Masson (1994) 95% confidence intervals.

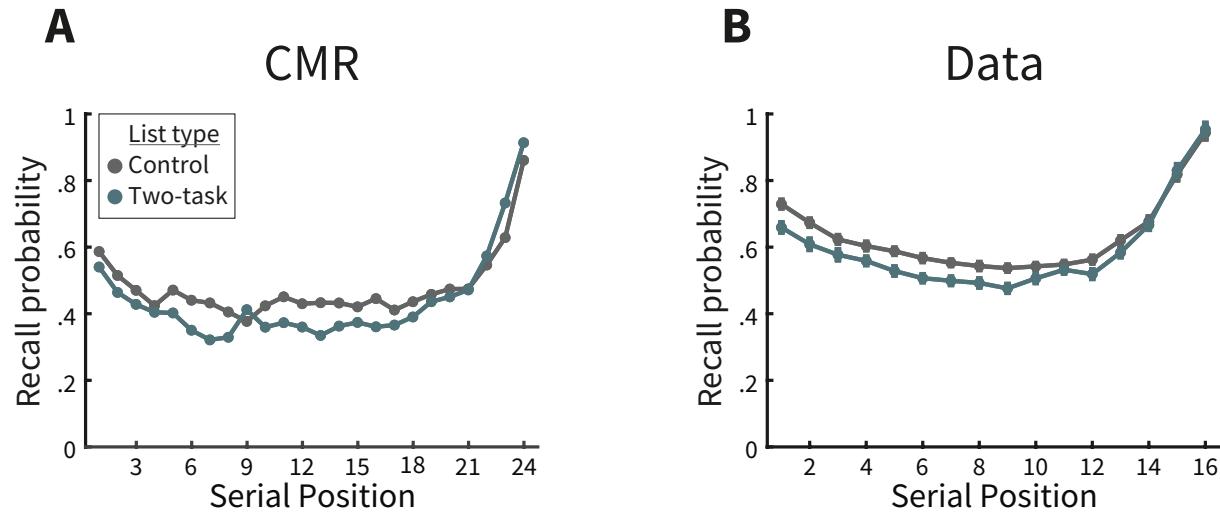


Figure A2

Serial position curves. **A.** CMR predictions of the serial position curve for the control lists (lighter gray) and two-task lists (darker teal). These data were simulated using the best-fit parameters and experimental lists of Polyn et al. 2009, and thus have a longer list-length than the experimental data. Yet critically, CMR predicts greater recall probability for early serial positions (primacy effect) and for late serial positions (recency effect). **B.** Serial position curves in the experimental data for the control lists (lighter gray) and two-task lists (darker teal), also exhibiting a primacy effect and recency effect. Error bars represent Loftus and Masson (1994) 95% confidence intervals.

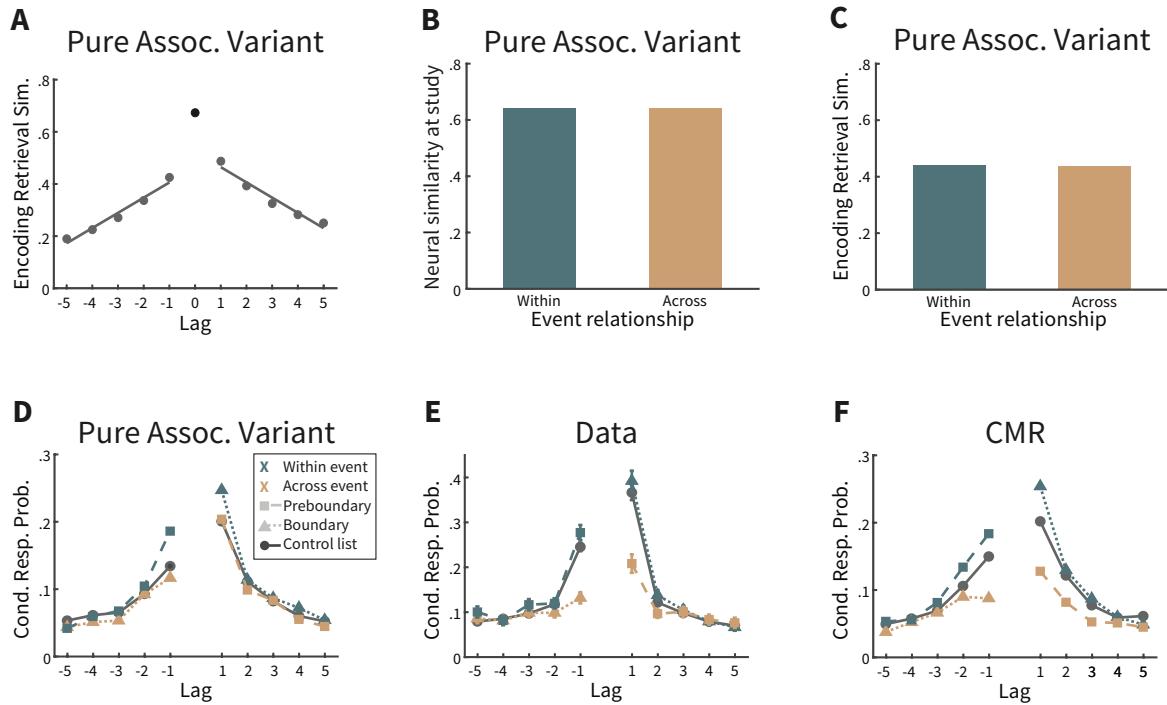


Figure A3

Predictions of the pure association variant of the Context Maintenance and Retrieval (CMR) model, which assumes that a change in task information does not disrupt temporal context or cause an event boundary. **A.** In control lists, encoding-retrieval similarity (ERS) between the temporal context state of a recalled item and the temporal contexts of its neighbors during study. Lag refers to the distance in serial position between two items from study (see Figure 1E). Like CMR, the pure association variant predicts that temporal context states will be more similar between the recalled item and neighboring items from study. **B.** Unlike CMR, the pure association variant predicts that neural similarity is identical for two neighboring items within the same event or two items across different events. **C.** The pure association model variant predicts that the recall of an item bordering an event boundary leads to retrieval of that item's temporal context states from study. Because these temporal context states do not incorporate disruptions, ERS values are nearly identical between items studied with the same task versus items studied with different tasks. **D.** The pure association variant predicts that, when compared to control lists (gray circles), transitions in two-task lists are enhanced for items recalled within the same event (darker teal lines). By contrast, recall transitions across events (lighter orange lines) are equivalent to control lists. Both of these predictions are present for preboundary items (squares) and boundary items (triangles). **E.** In the experimental data (replotted from Figures 2B, 5C,D), participants exhibit reduced recall transitions in two-task lists to items from different events (lighter orange lines), whether transitioning from preboundary items (squares) or boundary items (triangles). By contrast, participants exhibit similar or greater transitions for items recalled within the same event (darker teal lines). **F.** The full CMR model makes both critical predictions of reduced transitions for items from different events and enhanced transitions between items of the same event. For more distant items with darker teal lines, items may be from the following event. Error bars represent Loftus and Masson (1994) 95% confidence intervals. Assoc. = Association. Sim. = Similarity.

Appendix B

Controlling for positional effects

We conducted supplementary analyses to ensure that the feature vectors, changing slowly with each studied item, reflected temporal context rather than serial position information. Neural activity, including activity recorded from the regions of interest used for feature vectors, can change slowly as a function of an item's serial position, as measured oscillatory power changes in EEG (Sederberg et al., 2006; Serruya et al., 2014), event-related potentials (Rushby et al., 2002; Wiswede et al., 2007), or at the level of individual MTL units (Reddy et al., 2021; Umbach et al., 2020).

Whether the feature vectors represented a positional code or temporal context, they should change slowly with each studied item. However, if feature vectors coded for serial position, then the positional codes should reset with each list (Burgess & Hitch, 1999; Conrad, 1960; Henson, 1996; Osth & Dennis, 2015). As a result, the neural similarity between a feature vector for an item studied at serial position i should be most similar to the feature vector for item studied at serial position $i + 1$, irrespective of whether these items were presented in the same list or not. By contrast, if feature vectors change slowly over time, similarity across feature vectors should decrease across lists (Howard et al., 2008; Lohnas et al., 2015; Unsworth, 2008). We found that, for feature vectors of items from successive serial positions in control lists (i.e. lag = 1), similarity decreased with list distance, and thus the feature vectors have properties of temporal context not positional codes (see Figure B1 and also related text in Results; $list - lag = 0 > list - lag = 1 : M = 0.265, SD = 0.131, t(169) = 26.36, CI = [0.2449, 0.2846], p < .00001, d = 1.702, list - lag = 1 > list - lag = 2 : M = 0.065, SD = 0.179, t(169) = 4.76, CI = [0.0383, 0.0926], p < .00001, d = 0.393, list - lag = 2 > list - lag = 3 : M = 0.065, SD = 0.181, t(169) = 4.65, CI = [0.0373, 0.0922], p < .00001, d = 0.369$).

We also conducted several analyses to rule out the possibility that the autocorrelation property of feature vectors, averaged across serial positions, was driven by a subset of serial positions. In particular, we examined the feature values contributing to each feature vector in each list. For each feature, we took the difference in values for each successive pair of serial positions (1-2, 2-3, 3-4, etc.) within each list. We next calculated the differences of these pairwise

differences (so now we have the difference of 1-2 vs. 2-3, 2-3 vs. 3-4, etc.). These difference of difference values help to convey the autocorrelated component of the feature values. For instance, if the 1-2 vs. 2-3 differences are small, this should reflect a greater autocorrelation value in early serial positions, because the 1-2 difference is a good predictor of the 2-3 difference. To equate variability across features, we scaled the absolute value of the difference of difference scores on a range from 0 to 1 (where 0 and 1 are the smallest and largest, respectively, of the absolute value of difference of difference scores for a feature in a list). Across all features, we divided the values at each set of serial positions into deciles (0-0.1, 0.1-0.2, etc.), and plotted a histogram of the deciles for each set of serial positions (Figure B2). If items in primacy positions have the greatest autocorrelation, then those items should have values in lower deciles. However, going against this account, a larger proportion of the values are contained in the largest possible decile for items in earliest serial positions when compared to the largest decile of other serial positions.

Here we did not want to average across lists, features or participants, as mid-decile average activity may be a result of generally mid-decile activity across features, or may average some high-devile features with low-decile features. However, as a result, some participants contribute more features than others. To consider whether the reduced autocorrelation values are upheld across participants, the dashed black line in each plot shows the mean decile value per participant, averaged across all features and lists. Consistent with the results across features, the results across participants also suggest that items from primacy positions do not exhibit greater autocorrelation, rather the earliest serial positions have the greatest mean decile value.

In addition to examining properties of feature vectors themselves, we also examined how neural similarity by study lag (i.e. Figure A1B) may be influenced by differences in serial positions of the feature vectors. Figure B3A shows neural similarity in control lists divided by serial position, and reveals that neural similarity decreases with absolute lag at all serial positions. Figure B3B shows a similar pattern for items within the same event in two-task lists. For these items, approximately 12% of serial position/lag pairs do not follow the expected pattern of decrease in neural similarity as a function of absolute lag. This is not surprising given that not as many serial positions contribute at each lag, both because there are fewer two-task lists than control lists and because we only include two items at a lag if they are studied in the same event.

Further, because feature vectors were calculated from the control lists, this may also lead to noisier neural similarity in two-task lists. Nonetheless, if there were any strong effects of serial position on lag, going in opposition to the expected result of similarity decreasing with lag, we would expect these differences to be more pronounced in the control lists, where it is more likely that items from nonprimacy positions would contribute less noisy data to the similarity values. Thus, we take this set of results to be qualitatively consistent with CMR's prediction that neural similarity decreases with lag, irrespective of serial position.

In addition, in two-task lists we recalculated neural similarity at study excluding items from the first event, thus greatly reducing the influence of early list items with greater effects of differential neural activity. Even with this exclusion, neural similarity was significantly greater for neighboring items with the same event than different events (Figure B5A; $M = 0.007, SD = 0.048, t(169) = 1.98, CI = [0.0000, 0.0144], p = 0.0495, d = 0.044$).

Another impact of serial position effects may manifest from differences in recall probability, as recall is more likely for items from earlier and later serial positions (Kahana, 2012; Murdock, 1962, Figure A2). Further, in analyses of recall by lag, not all serial positions are possible at all lags, raising a concern that some serial positions may contribute to some lags more than others. For the ERS analysis in control lists, we verified that this analysis reflects a fair representation across serial positions by calculating the mean number of items contributing to each lag and serial position in Figure 2D. Figure B4 shows the mean number of serial positions contributing to each lag, whether the transition was from that serial position as a recalled item (left) or to that serial position as a studied item (right). Transitions are somewhat less likely from recency positions, but this is less surprising given that we exclude the first three output positions, and immediate free recall generally begins with these recency items (Healey & Kahana, 2014; Howard & Kahana, 1999).

In two-task lists, we alleviated concerns that primacy items may dominate recall effects by recalculating the critical significant findings in two-task lists but excluding transitions to or from items in the first event of each list. With this exclusion, the lag-CRPs were qualitatively similar to those including items from the first event, with greater recall probability for transitions to a same-event neighbor, whether from a preboundary item (Figure B5B) or from a boundary item

(Figure B5C). Although the control lag-CRP is less intuitive to calculate for this analysis because there are no first event items in control lists, we nonetheless interpret the relatively similar numbers and qualitative pattern of results as evidence that items from the first event did not drive this effect in two-task lists.

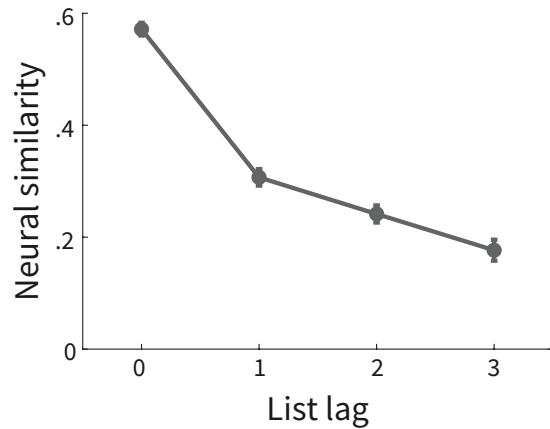


Figure B1

Neural similarity of feature vectors for item pairs from adjacent serial position numbers and the same task in control lists, as a function of list-lags. Data are averaged across all possible values of i and $i + 1$ (1 and 2, 2 and 3, ..., 15 and 16) for item pairs in control lists studied with the same task (size task, animacy task, or no task). Neural similarity of these item pairs is plotted as a function of list-lag, where list-lag=0 reflects 2 items from the same list, list-lag=1 reflects 2 items from successive lists such as lists 3 and 4. Consistent with properties of temporal context but inconsistent with properties of positional codes, neural similarity decreases as a function of list-lag. Error bars represent Loftus and Masson (1994) 95% confidence intervals.

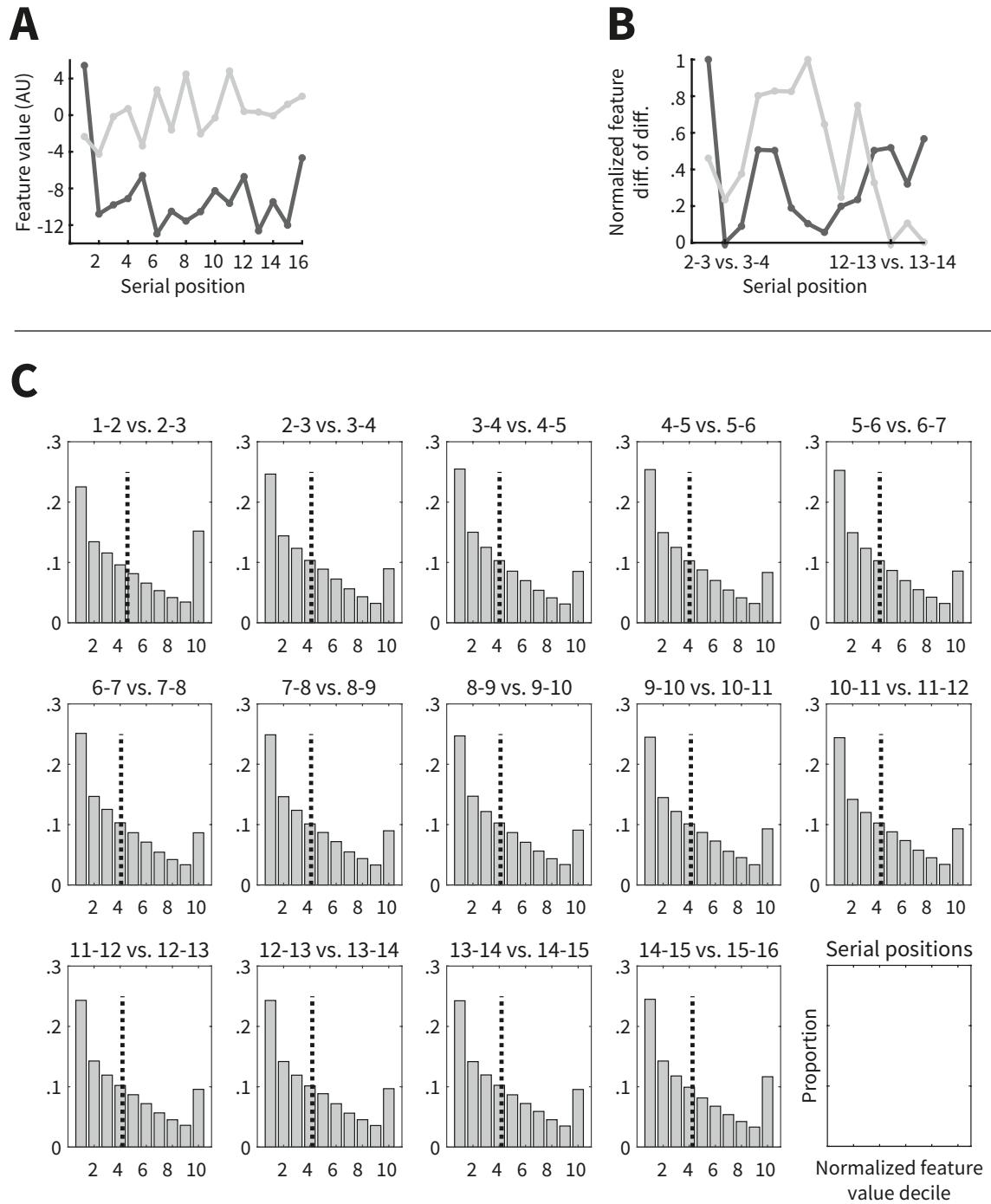


Figure B2

Histogram of feature value differences contributing to each feature vector. **A.** Examples of feature vector values for two control lists. **B.** The absolute value of the difference of differences, rescaled from 0 to 1. These feature value lists were chosen to be representative of more or less autocorrelation in primacy positions (light line and dark line, respectively). **C.** Histogram of normalized (i.e. absolute value and rescaled) difference of differences for feature vectors across lists, at each set of serial position values. These values were averaged across feature values across lists and participants. Dashed black lines indicate the mean across feature values when averaged by list and participant.

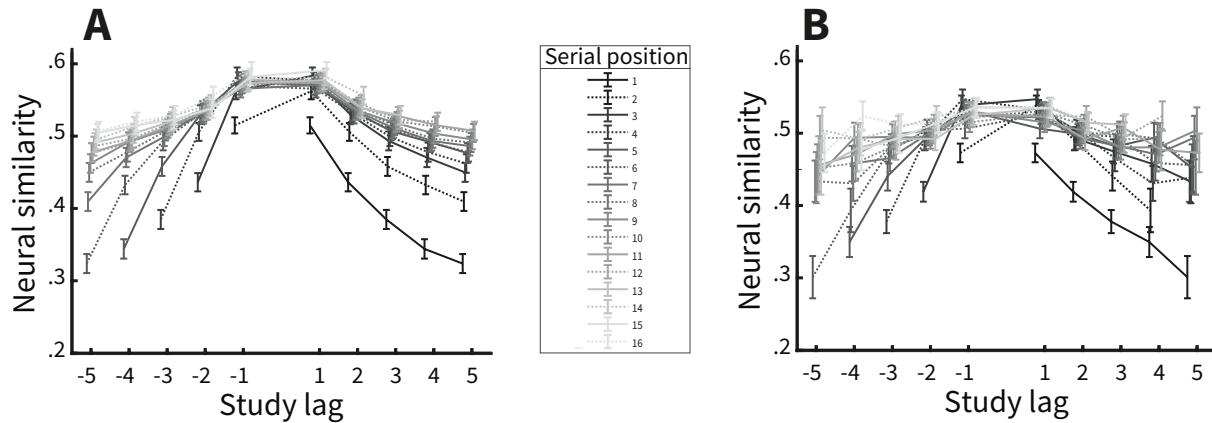


Figure B3

Neural similarity during study as a function of lag, divided by serial position. Error bars represent ± 1 standard error of the mean. A. Control lists. B. Items studied with the same task in two-task lists.

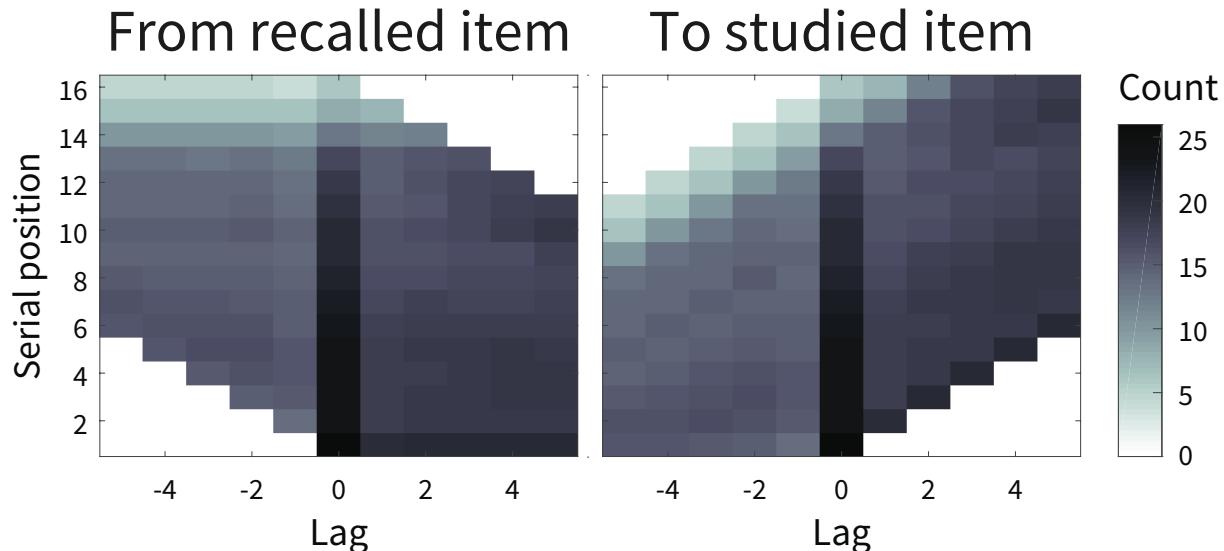


Figure B4

Number of serial positions contributing to ERS values at each lag in control lists (Figure 2D). Left: Counts of the ERS values contributing to each lag at each serial position of the recalled item. Right: Counts of the ERS values contributing to each lag at each serial position of the encoded item.

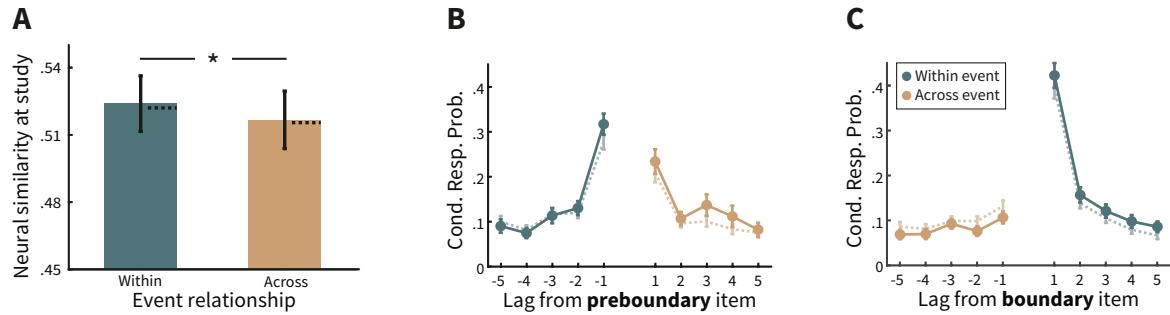


Figure B5

Analyses in two-task lists excluding items from the first event. **A. Neural similarity at study.** Collapsed across preboundary and boundary items, neural similarity is greater between two neighboring items within the same event than two items across different events, even when excluding items from the first event in each list (see also Figure 3). Error bars represent ± 1 standard error of the mean. * $p < .05$. **B, C. Recall transitions in two-task lists.** Even when excluding items from the first event, participants are more likely to recall items not separated by an event boundary, and more likely to be within the same event (darker teal lines) than items separated by an event boundary and from a different event (lighter orange lines), whether from a preboundary item (**B**) or from a boundary item (**C**). For more distant items with darker teal lines, items may be from the following event. Cond. Resp. Prob. = Conditional response probability. Error bars represent Loftus and Masson (1994) 95% confidence intervals. See also Figure 5. In all panels, dashed lines indicate the values when items from the first event are included in the analysis.

Appendix C

Time window of context reinstatement

We evaluated which time window would reflect the strongest context reinstatement based on negative ERS lags. In 100 ms increments, we evaluated time windows beginning from -1000 ms to -500 ms relative to the participant's recall vocalization, ranging in duration from 300 ms to 800 ms. We found that context reinstatement was strongest for the recall time window of -1000 to -600 ms, and used this time window in all analyses. Although choosing among several time windows may seem to present a selection bias, the purpose of time window selection was not to identify the existence of the effect of context reinstatement. Based on Manning et al. (2011), we sought to replicate context reinstatement in scalp EEG. Due to differences in recording techniques, we anticipated that the time window might differ between scalp and intracranial EEG. We chose such a time window to then evaluate how event boundaries in two-task lists impact feature vectors with properties of temporal context. Further, because we examined time windows with overlapping timepoints, and because the data at each timepoint incorporates the same electrodes, there was a cluster of significant *t*-values for time windows with overlapping timepoints to the time window with the strongest effect of context reinstatement. Although we chose the time window based on negative lags, there was a cluster around this window for positive lags as well (Figure C1).

In addition to examining which time window exhibited the greatest effect out of all considered windows, we also determined the significance of the selected time window when compared to a null distribution. For each subject and time window, we calculated the two ERS values which were used in our determination of context reinstatement: (1) lag=-1; (2) the average of lags -3, -4, -5. For each subject, we shuffled the lag labels of the ERS values at each time window, and calculated the maximum *t* value across all windows. We did this for 1000 shuffles of the ERS values, thus acquiring a null distribution of 1000 *t* values. The actual *t* value fell on the null distribution with one-tailed $p = .031$, suggesting that context reinstatement is greater than expected by chance across time windows.

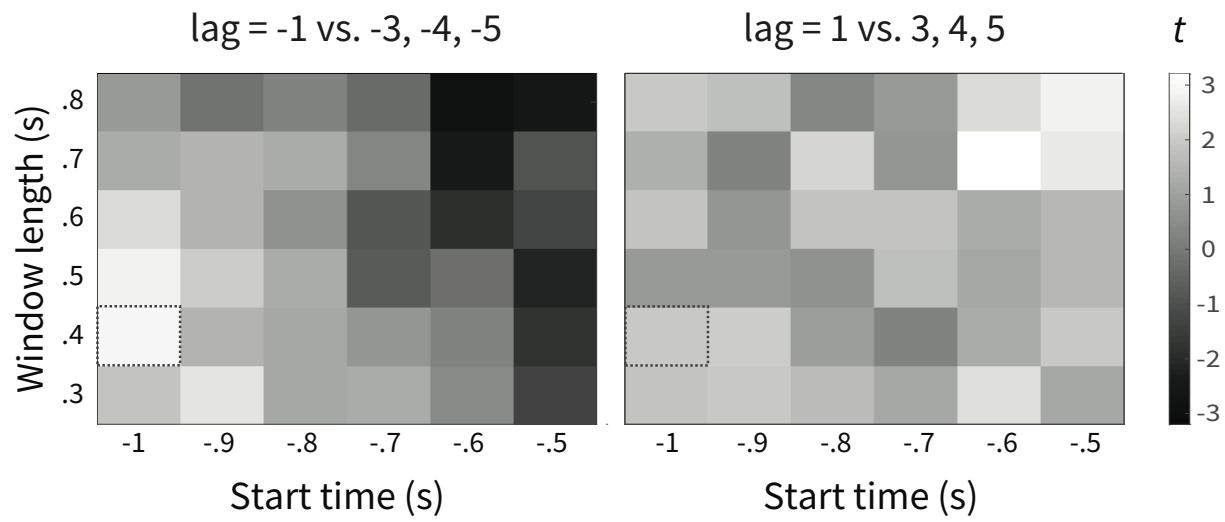


Figure C1

Size of the effect of neural similarity by lag, at each evaluated time window. Dashed boxes correspond to the time window used in all analyses.