The modulation and elimination of temporal organization in free recall

Min Kyung Hong, Jordan B. Gunn, Lisa K. Fazio, Sean M. Polyn

January 19, 2023

Abstract

Experiences occur in a continual succession, and the temporal structure of those experiences is often preserved in memory. The temporal contiguity effect of free recall reveals the temporal structure of memory: When a particular item is remembered, the next response is likely to come from a nearby list position. This effect is central to retrieved-context models, which propose temporal organization arises from the interaction of temporal context with the contents of memory. Across 6 experiments, we demonstrate methodological manipulations that dramatically modulate and even eliminate temporal organization in free recall. We find that temporal organization is strongly modulated by semantic structure, retrieval practice, and list length. Other factors such as orienting task, item structure, and retention interval duration have more subtle effects on temporal organization. In an accompanying set of simulations, we show that the modulation and elimination of temporal organization follows lawful patterns predicted by the Context Maintenance and Retrieval (CMR) retrieved-context model. We also find cases where CMR does not specifically predict the modulation of temporal organization, and in these cases our analysis suggests how the theory might be developed to account for these effects.

Introduction

Episodic memories associate the details of an experienced event with its temporal context (Tulving, 1983; Schacter, 1987), providing temporal structure to our past experience (Friendly, 1979; Underwood, 1969). The influence of this temporal structure on memory performance can be seen in the free-recall task, where a participant studies a list of items presented one at a time, and then recalls the items in whatever order they come to mind. The study period is a temporally extended experience, punctuated at regular intervals with nameable study items. This means that each recall response can be linked to a particular moment from the study experience, and memory search can be characterized in terms of the temporal organization of the remembered study items.

Organizational analyses examine the sequence of recall responses as a series of transitions between successively reported items. These analyses provide unique insight into the structure of stored memories by characterizing the features of past experience that cause memories to group together during search (Tulving, 1968; Bower, 1970; Postman & Hasher, 1972; Crowder, 1976; Puff, 1979; Kahana, 2012). The temporal continuity of memories can be seen in the temporal contiguity effect, the tendency for successively recalled items to come from neighboring positions in the study list (Kahana, 1996, 2012). This temporal organizational effect provides a laboratory analogue of one of the key features of episodic memory: mental time travel, the subjective revisiting of one's own past experience during reminiscence (Tulving, 2002; Kahana et al., 2008). As such, understanding the cognitive mechanisms supporting the temporal contiguity effect will provide insight into the nature of episodic memory.

We consider the temporal contiguity effect from the perspective of retrieved-context theories of memory search. Retrieved-context models are computational models that describe the cognitive machinery which gives rise to the temporal contiguity effect. By these models, the human cognitive system constructs an internal representation of temporal context that evolves gradually over the course of a study list. As studied items are encountered, they are associated with the current state of the context representation. During memory search, the context representation is part of a retrieval cue which supports memory targeting on the basis of the time of an item's occurrence. These core mechanisms work together to produce temporal organization of responses in the free-recall task.

In this paper we examine how the manipulation of several benchmark properties of free-recall experiments can modulate and eliminate the temporal contiguity effect. This modulation naturally raises questions regarding the generality of the empirical phenomenon of temporal organization, as well as questions regarding whether this modulation challenges the utility or validity of retrieved-context models. We begin by reviewing prior empirical work establishing the wide range of experimental conditions that give rise to the temporal contiguity effect. This review establishes the general robustness of the effect. We then present a series of experiments that replicate and extend prior work examining the manipulation of benchmark experimental factors that modulate (sometimes dramatically) temporal organization in free recall.

These experiments are accompanied by simulations using the Context Maintenance and Retrieval (CMR; Polyn et al., 2009) retrieved-context model of the free-recall task. We use these simulations to examine whether CMR can account for the modulation of temporal organization due to these experimental manipulations. Our simulations demonstrate that in many cases the modulation of temporal organization follows lawful patterns predicted by retrieved-context models. In these cases, CMR can account for the observed modulation without adding new representations or processes to the model. We also identify cases where the theory does not specifically predict the modulation of temporal organization. For these cases, our analysis suggests ways the theory can be modified or developed to capture these effects. In the discussion, we examine how other modeling approaches can benefit from the insights developed here. From the standpoint of theory development, we argue that understanding the boundary conditions of temporal organization in free recall will give us a better general understanding of the structure and dynamics of human episodic memory.

Temporal organization in free recall

Early measures of temporal organization in free recall treated the recall response sequence as a series of transitions between successively remembered items, often referred to as recall pairs (Asch & Ebenholtz, 1962; Mandler & Dean, 1969). These measures indicate the degree of correspondence between the study sequence (or input) and the response sequence (or output). To calculate *input-output correspondence* (sometimes called *input-output concordance*), one counts up the number of recall pairs that include two neighboring items from the study sequence, and divides this by the maximum number of possible matching pairs (usually the total number of recalled items minus 1). These measures were inspired by subjective organization (SO) and inter-trial repetition (ITR) statistics designed to calculate the degree of correspondence between two successive recall sequences in a multi-trial free recall task (Tulving, 1962; W. Bousfield et al., 1964). Input-output correspondence measures have been used in a variety of research domains (Wallace, 1970; Kintsch, 1970; Mandler et al., 1969; Postman, 1972). Some studies used these measures to examine how temporal organization is affected in certain special populations (Koh et al., 1971; Koh & Kayton, 1974). Other studies used these measures to examine how item-order information is affected by task or item characteristics (Dong, 1972; Nairne et al., 1991; Serra & Nairne, 1993; DeLosh & McDaniel, 1996).

Kahana (1996) introduced a novel kind of organizational analysis that gave new insight into temporal organization in free-recall tasks. This is the lag-based conditional response probability (or lag-CRP) analysis, in which each recall transition is assigned a lag specifying the number of serial positions separating the two studied items. For example, recall of item 8 followed by item 9 is a +1 lag transition. For each possible lag distance, the analysis counts up the number of times a transition of that lag was observed, divided by the number of times a transition of that distance was possible. For example, a transition might not be possible if the item at that distance was already recalled, or if that lag distance extends beyond the beginning or end of the study list.

The lag-CRP analysis provides a more fine-grained look at the temporal dynamics of recall relative to earlier measures. Whereas the input-output correspondence measure described above only counted +1 lag transitions, the lag-CRP analysis provides a conditional probability value for each possible lag, as seen in many figures throughout this paper (e.g., Figures 2B & 14B). This analysis allows for an examination of the relative probability of transitions of different lags, revealing what has come to be known as the *temporal contiguity effect*. Two features of temporal organization are clarified by this analysis. The first is an effect of proximity: Transitions between nearby items are more likely than transitions between distant items. The second is forward asymmetry: Short forward-going transitions are more likely than short backward-going transitions.

In the years since its introduction, the lag-CRP analysis has been used to demonstrate the near-ubiquity of the contiguity effect in studies of free recall (Kahana, 1996; Howard & Kahana, 1999; Howard et al., 2006; Kahana et al., 2008; Healey & Kahana, 2014). In a review of 34 free-recall experiments the contiguity effect appears under a wide variety of experimental conditions, and is robust to a large number of experimental manipulations (Healey et al., 2018). The temporal contiguity effect appears in recalls of items from the beginning, middle, and end of a study list. It appears throughout the recall sequence, and appears regardless of whether recall is spoken or written. It is unaffected by presentation rate, and appears for both visual and auditory presentation modalities. It is also unaffected by the presence of an effortful distraction task at the end of the list, and is similarly unaffected by an effortful distraction task before and after each study item (Lohnas & Kahana, 2014). The review does identify a number of factors that modulate the contiguity effect, including certain item characteristics (e.g., categorized vs. uncategorized items, paired associates vs. single words), presentation rate (e.g., slow vs. fast item presentation during study), presentation modality (e.g., visual vs. auditory), amount of practice with the free-recall task, age, and IQ. However, in their survey, none of these manipulations or circumstances were shown to eliminate the effect.

Healey & Kahana (2014) demonstrated the impressive consistency of the temporal contiguity effect in a large data set with many individuals performing the same free-recall task. Nearly every individual participant showed a clear contiguity effect, suggesting that this does not represent an idiosyncratic strategy only exhibited by some participants. The contiguity effect has also been linked to overall success at the free-recall task. Participants that show a stronger contiguity effect tend to recall more items overall, across a number of data sets (Sederberg et al., 2010). Golomb et al. (2008) showed that the decline in free-recall performance observed in healthy older adults (compared to younger adults) was associated with a diminished temporal contiguity effect. Palombo et al. (2018) showed that poor memory performance in patients with hippocampal amnesia is accompanied by a disrupted temporal contiguity effect. This finding is consistent with an fMRI neurorecording study by Kragel et al. (2015) demonstrating that fluctuations in hippocampal activity observed over the course of a recall period correspond with fluctuations in the strength of the temporal contiguity effect.

Retrieved-context theory and temporal organization.

The fine-grained temporal organization of the contiguity effect may reflect the operation of fundamental cognitive mechanisms that support memory search. However, it may also be the case that the temporal contiguity effect is not an obligatory empirical feature of free-recall performance. In order to establish the importance of temporal organizational effects for theories of human memory, we review the basic structure of retrieved-context models. These models were designed to explain the dynamics of temporal organization in free recall (Howard & Kahana, 2002a; Sederberg et al., 2008; Polyn et al., 2009; Lohnas et al., 2015; Healey & Kahana, 2014; Kahana et al., 2008; Kahana, 2012, 2020).

The first retrieved-context model, the Temporal Context Model (TCM; Howard & Kahana, 2002a) is implemented as a simplified connectionist model (Anderson et al., 1977) with two representational layers, one representing the features of studied items, and the other representing contextual features associated with those items. When an item is studied, its representation is activated on the item layer, and this is projected through the item-to-context weights, which retrieves an associated item-specific contextual representation. Due to an integration mechanism embedded in the context layer, this incoming activity only partially displaces the previous contextual state. This causes the representational state of the context layer to change gradually as the list progresses, creating a temporal code that is similar for neighboring items. A Hebbian learning mechanism binds the representation of the item to the context representation. This allows the context representation to be used to retrieve associated items, and allows a retrieved item representation to retrieve associated states of context. During memory search an iterative process gives rise to the temporal contiguity effect: The context representation prompts retrieval of a particular item, which then retrieves its associated temporal context. This retrieved temporal context updates the contextual retrieval cue, so the next response is likely to be a neighbor of the just-recalled item. These dynamics are described in more detail in Appendix B.

In cognitive models of memory search, temporal structure is often one of several kinds of associative structure that can give rise to organizational effects during memory search. In versions of the SAM model, item-to-item associations and list context associations are complemented by semantic associations representing one's prior knowledge about the studied items and their relations to one another (Sirotin et al., 2005; Kimball et al., 2007). These different influences combine during a competitive retrieval process, and SAM includes weighting parameters that can adjust the relative strength of the different kinds of associations. The Context Maintenance and Retrieval model (CMR; Polyn et al., 2009) is a retrieved-context model that includes multiple contextual layers containing different kinds of information. As such, the CMR framework is well suited for an exploration of the experimental factors that modulate temporal organization. For example, CMR can include source context or environmental context information in one of the contextual layers to target many items from a study list simultaneously (Polyn et al., 2009; Sederberg et al., 2011). CMR and CMR2 also incorporate semantic structure that competes with temporal structure during memory search (Polyn et al., 2009; Lohnas et al., 2015; Morton & Polyn, 2016). The CMR framework allows us to examine how semantic structure interacts with other methodological features of free recall to disrupt temporal organization.

Overview of Experiments and Simulations

In this paper, we identify methodological conditions necessary to eliminate, and recover, the temporal contiguity effect in the free-recall task. As mentioned above, a study list with strong semantic structure disrupts temporal organization. Another way to disrupt temporal organization is to review the studied items in a shuffled order, as in multi-trial free recall or retrieval practice tasks (Klein et al., 2005; Karpicke et al., 2014). We find that strong semantic structure and reordered practice are sufficient to eliminate the temporal organization of recall responses. Across several experiments, we manipulate these and other properties of the free-recall task, in order to characterize how each one modulates, or leaves unaffected, the temporal contiguity effect. These include study item structure (paired associates vs. single items), orienting task, list length, and retention interval duration. A full factorial manipulation of these methodological properties would be valuable, but infeasible. If 2 levels were chosen for each of the 6 factors listed above, this would define 64 distinct experimental conditions. As such, we have taken a more practical approach to the issue, examining these factors a few at a time, in an attempt to perform a rough survey of the relevant methodological space.

Figure 1B provides an overview of the key methodological features of each experiment described in this report. Figure 1A shows the strength of temporal organization observed in the different conditions of each experiment, using the percentile-rank temporal organization score. This analysis technique has been used in many studies since its introduction by Polyn et al. (2009) (Sederberg et al., 2010; Polyn et al., 2011; Lohnas & Kahana, 2014; Polyn et al., 2015; Morton & Polyn, 2017; Healey, 2018; Healey et al., 2018; Murty et al., 2018; Sahakyan & Kwapil, 2018; Mundorf et al., 2022). The temporal organization score uses the same basic data as the lag-CRP analysis: recall transitions marked with lag distance. A score of 1.0 indicates perfect temporal organization, and 0.5 indicates chance-level temporal organization. This and other temporal organization statistics are aggregated in Table 2 in Appendix A. This appendix also provides more detail regarding the analysis methods used to characterize recall organization in these experiments.

- *Semantic structure disrupts temporal organization.* In Experiment 1, we manipulate the semantic structure of the study list, from strong (12 items drawn from the same category) to weak (one item from each of 12 categories, or 12 items randomly selected from a large word pool). The lists are short (12 items), the retention interval is short (10 seconds), and the items are presented once. The temporal contiguity effect is observed in all conditions, and is substantially weakened when semantic structure is strong.

- *Semantic structure and reordered practice eliminate temporal organization.* In Experiment 2, we use the same study materials as Expt. 1, but construct longer lists where multiple categories are intermixed. A practice period follows the study period, in which all studied items are reviewed in a shuffled order. The lists are relatively long (90 items), the retention interval is relatively long (15 minutes), and the study items are presented as *category name–exemplar* paired associates. These methodological conditions completely eliminate temporal organization. We present an organizational analysis of archival data from M. A. Smith et al. (2013) which replicates the null temporal contiguity effect (and we contrast this with a second experiment of theirs where temporal organization is intact).

- *Keep semantic structure and remove reordered practice; temporal organization is absent.* Experiment 3 is structured like Experiment 2 in terms of study materials, list length, and delay. We remove the practice period from the design, and have participants perform a generation task during the study period to preserve overall recall performance. Temporal organization is eliminated again.

- *Weaken semantic structure and keep reordered practice; temporal organization is absent.* Experiment 4 modifies the study items to weaken their semantic structure but preserve their paired-associate structure. Items are practiced in a shuffled order, as in Expt. 2. The list length is intermediate (40 items) and the retention interval duration is long (15 minutes). The temporal contiguity effect is eliminated in all but one condition, where very weak temporal organization is observed.

- *Weaken semantic structure and remove reordered practice; temporal organization returns.* In Experiment 5, lists have weak semantic structure and we remove the practice period from the design. List length is relatively short (24 items) and retention interval is intermediate (3 minutes). We manipulate the item structure (paired associates vs. singly presented words), and the orienting task (retyping vs. semantic) to determine whether these factors likely influenced the results of the previous experiments. A strong temporal contiguity effect is observed in all conditions. There is an interaction of item structure and orienting task: Shifting from a retyping task to a semantic orienting task weakens the temporal contiguity effect for single words, but a similar shift does not weaken the temporal contiguity effect for paired associates.

- *List length modulates temporal organization, retention interval duration does not.* In Experiment 6, we manipulate both list length (20 or 40 items) and retention interval duration (3 minutes or 15 minutes), to determine whether these factors likely influenced results of the previous experiments. The study lists have weak semantic structure. A strong temporal contiguity effect is observed in all conditions. As list length increases, the probability of a +1 transition reliably drops, but not all measures of temporal organization agree that the temporal contiguity effect is getting weaker.

Retention interval duration does not affect temporal organization.

A set of simulation analyses accompany the experimental results. These simulations use the Context Maintenance and Retrieval (CMR) modeling framework (Polyn et al., 2009; Lohnas et al., 2015; Kragel et al., 2015) to demonstrate the model's predictions with regards to these methodological manipulations. In order to simulate participant-level variability in performance, the model was separately fit to 126 individual participants from the Penn Electrophysiology of Encoding and Retrieval Study (PEERS; Healey & Kahana, 2014). This creates 126 distinct parameter sets that capture individual variability in task performance. This allows us to simulate the effects of different methodological manipulations without fitting the model separately to each experimental data set. As such, the five simulations demonstrate how temporal organization can be modulated without adjusting the core parameters of CMR.

The model and simulation methods are described in detail in Appendix B. The data and study materials from these experiments will be made available through a project page hosted with the Open Science Foundation. The URL will be inserted and this text updated once the project page is made public. Analysis code and simulation code will be made available on GitHub (Organization: Vanderbilt University Computational Memory Lab, vucml), repository names will be inserted once they are made public.

- Simulation 1 demonstrates that temporally adjacent categorical associates can disrupt but not abolish the temporal contiguity effect, as examined in Expt. 1 (Figure 14).
- Simulation 2 demonstrates that widely spaced categorical associates can cause extreme temporal disruption as examined in Expts. 2 & 3 (Figure 15).
- Simulation 3 demonstrates that reviewing studied items in a different order disrupts but does not eliminate the temporal contiguity effect (Figure 16).
- Simulation 4 combines semantic structure and reordered practice to demonstrate that these two factors together are devastating to the temporal contiguity effect (Figure 17).
- Simulation 5 demonstrates that increasing list length disrupts temporal organization and can amplify the effectiveness of other disruptive factors (Figure 19).

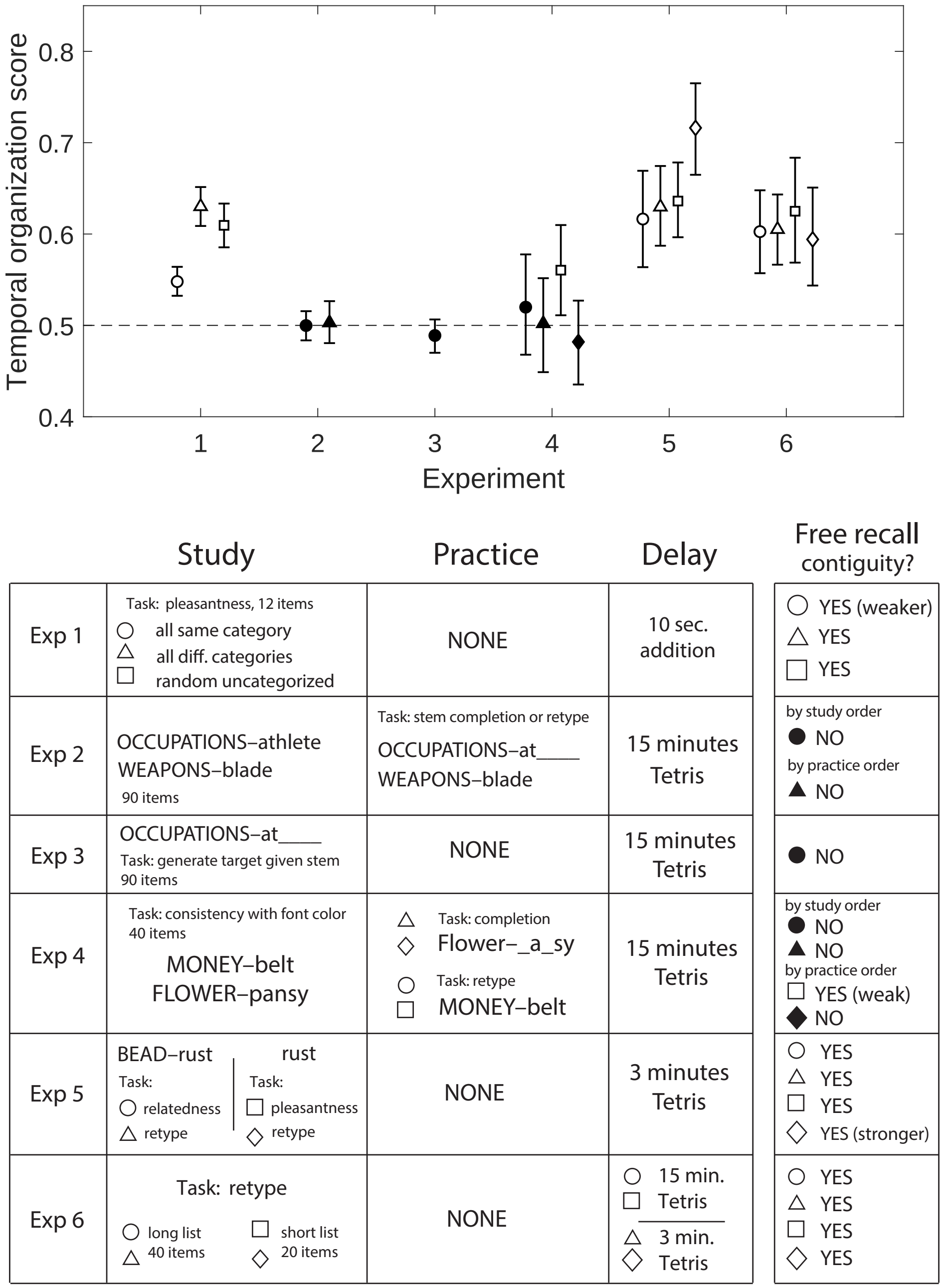

Figure 1: (A) An overview of temporal organizational scores from the 6 experiments. Error bars indicate bootstrapped 95% confidence intervals, with open symbols indicating statistically reliable above-chance temporal contiguity effects. The symbols can be used to find that condition's corresponding methodological details in the bottom panel. (B) A schematic overview of the 6 experiments. See individual experiment sections for full methodological details. Expts. 1, 5, & 6 show a strong and reliable temporal contiguity effect in each condition. Expts. 2–4 demonstrate a variety of experimental conditions that abolish the temporal contiguity effect.

Experiment 1: Semantic structure disrupts temporal organization

Semantic knowledge about studied materials exerts itself as semantic organization during free recall: Meaningfully related items tend to be remembered successively in the recall sequence (Deese, 1959; W. A. Bousfield, 1953). This can be seen even when the set of studied items are randomly chosen and nominally unrelated (Howard & Kahana, 2002b). In modern work, semantic relatedness is often characterized using distributional semantic models (Landauer & Dumais, 1997; Mikolov et al., 2013; Pennington et al., 2014). These are algorithms that process the co-occurrence patterns of words in a large corpus of text to extract representational vectors for the words within the corpus. The similarity structure of the vectors reflects the relatedness of the words, and these can be incorporated into a cognitive model to define the semantic associative strength between any pair of studied items. Temporal and semantic organization are not mutually exclusive; the same recall sequences can simultaneously exhibit both forms of organization (Howard & Kahana, 2002b; Polyn et al., 2009; Morton & Polyn, 2016). However, the two forms of organization can have a seemingly competitive effect. When a study list has strong semantic structure, such as when items from distinct taxonomic categories are intermixed, the contiguity effect can be reduced considerably relative to lists with weaker semantic structure (Polyn et al., 2011). The relative strength of each form of organization can be modulated by instructions to focus on the temporal structure or semantic structure of the study list, consistent with the proposal that these associative influences can be independently weighted (Healey & Uitvlugt, 2019).

The retrieved-context models CMR and CMR2 propose that semantic associations compete with temporal associations (Polyn et al., 2009; Lohnas et al., 2015; Morton & Polyn, 2016). This predicts that increasing the strength of semantic relations between studied items will disrupt the temporal contiguity effect. Here, we demonstrate this effect empirically. In a condition with strong semantic structure, a short list of 12 items are all drawn from the same taxonomic category (the *same-category* condition). We compare this with a condition with weaker semantic structure but the same study materials, where lists are composed of one item drawn from each of 12 categories (the *mixed-category* condition). We include a third condition with weak semantic structure, in which 12 items were randomly selected from a large uncategorized word pool (the *Toronto Noun Pool* condition). The Toronto Noun Pool has been used in many prior free-recall studies examining temporal organization (including, e.g., Howard & Kahana, 1999, which also used 12-item lists).

Method

Participants

41 adults participated in exchange for monetary payment or course credit (17 male, ages 18–28).

Materials

Study words were drawn from 32 distinct taxonomic categories from normed word pools developed by Battig & Montague (1969) and Van Overschelde et al. (2004). These categorized words were used on two types of trials. *Same-category* lists were composed of 12 items drawn from a single category, presented in a randomized order. *Mixed-category* lists also contained 12 items, one from each of the remaining categories (i.e., those not used to create same-category trials), presented in a randomized order. Assignment of a particular category to the mixed-category or same-category condition was counterbalanced across participants. *Toronto Noun Pool* lists contained 12 words drawn randomly from the Toronto Noun Pool (Friendly et al., 1982). These lists were filtered to ensure they didn't contain strongly semantically related items. The words were chosen such that any pair of items from a given list had a Word Association Spaces (WAS) semantic similarity score ≤ 0.50, measured using cosine distance.

Procedure

Participants performed 4 practice trials followed by 8 trials from each of the three experimental conditions, intermixed. On each trial, participants studied a list of 12 words. Each word was presented for 1 second, during which participants were asked to make a 2-choice pleasantness judgment ("Is the item generally pleasant or unpleasant?"). Each item was followed by an 0.5 sec ISI with a centrally located fixation cross. After all items on a list were presented, participants performed 10 seconds of an arithmetic distraction task. For this task, two digits were presented sequentially in the center of the screen, followed by an equals sign. A third number was then presented and participants indicated with a keypress whether the third number equaled the sum of the first two. A 45-sec free-recall period followed the end-of-list distraction period. Vocal recall responses were recorded by a microphone.

Results

Memory performance

Items from lists with strong semantic structure had greatly enhanced memorability relative to lists with weak semantic structure. Proportion recalled on same-category trials: $M = 0.72$, 95% bootstrapped confidence interval (CI) [0.69–0.74]; mixed-category: $M = 0.47$, CI [0.44–0.50]; Toronto: $M = 0.43$, CI [0.39–0.46]. The performance increase from mixed-category to same-category lists was large and statistically reliable (via a two-tailed paired-sample t-test, $t(40) = 19.24$, $p < 0.001$). Items drawn from the Toronto Noun Pool were slightly (but reliably) less well remembered than the items on the mixed-category lists ($t(40) = 3.35$, $p = 0.0018$), and substantially less well remembered than items from same-category lists ($t(40) = 15.84$, $p < 0.001$). Figure 2A shows recall performance as a function of serial position of the studied item. Moderate primacy and recency effects were observed for all three conditions.

Extra-list intrusions are a memory error in which a participant reports an item that did not appear on the study list, or on any prior list. On same-category trials, a large proportion of extra-list intrusions could indicate participants are using the category name to generate responses, without regard to the actual contents of the study list. There was no evidence for this kind of strategy, in that extra-list intrusions were very rare in all three conditions. Extra-list intrusions were a negligible percentage of the 6798 total responses: 0.85% of responses in the same-category condition (58 total extra-list intrusions across all participants), 0.37% in the mixed-category condition (25 total), and 0.79% in the Toronto Noun Pool condition (54 total). The rate of extra-list intrusions was very similar between the same-category condition, which had strong category structure, and the Toronto Noun Pool condition, which had no category structure.

Temporal organization

Figure 2B shows that reliable contiguity effects were observed in all three conditions, although the effect is markedly weaker for same-category trials. A percentile-rank temporal organization analysis quantifies how the contiguity effect is altered by the semantic composition of the study list. On same-category trials: $M = 0.55$, CI [0.53–0.56]; mixed-category: $M = 0.63$, CI [0.61–0.65]; Toronto: $M = 0.61$, CI [0.59–0.63]. The contiguity effect was substantially weaker for trials with stronger within-list semantic structure (same-category vs. mixed-category, $t(40) = -6.36$, $p < 0.001$; same-category vs. Toronto, $t(40) = -4.48$, $p < 0.001$; mixed-category vs. Toronto, $t(40) = 1.75$, $p = 0.09$). Table 2 compares these temporal organization scores with an input-output correspondence statistic that calculates the likelihood of a +1 lag recall transition. These measures are in agreement regarding the substantial weakening of temporal organization in the same-

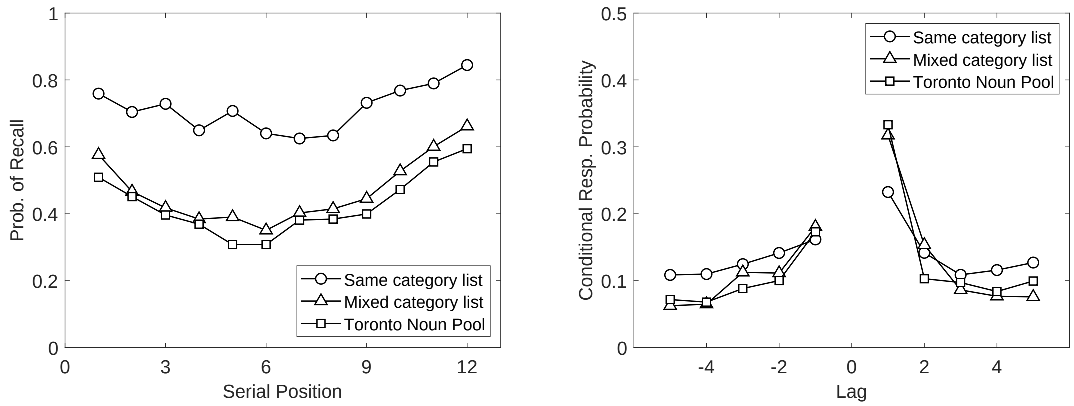

Figure 2: Experiment 1. Same: Study list with all 12 items from the same category. Mixed: One item from each of 12 categories. Toronto: 12 random items from the Toronto Noun Pool. A) Serial Position Curve analysis. All three conditions show moderate primacy and recency effects. A substantial improvement in performance is seen for the same-category lists at all serial positions. B) Lag-based conditional response probability analysis. Mixed and Toronto trials show similar contiguity effects, and the contiguity effect is substantially weakened on the same-category trials.

category condition relative to the other two conditions.

Discussion

Here, we see that strengthening the semantic structure of the study list substantially weakens the temporal contiguity effect, but the effect remains reliably present even in its weakened form. Sederberg et al. (2010) examined individual differences in free-recall performance across several experiments, demonstrating that participants who exhibit stronger temporal organization also recall more of the studied items (Sederberg et al., 2010). In those experiments, lists had weak semantic structure, with study words randomly chosen from a large pool. This experiment provides a counterexample (of sorts) to that finding, by demonstrating that strengthening semantic structure can simultaneously weaken temporal organization and improve recall performance. These results are not necessarily at odds with one another. When the semantic structure of a study list is weak, temporal organization is a good strategy for successfully remembering the study material. When semantic structure is strengthened, this increases the overall recall support for all studied items. This increased recall support improves overall performance, but the structure of the semantic associations interferes with the structure of the temporal associations, which weakens the behavioral signature of temporal organization. Simulation 1 demonstrates this trade-off between semantic and temporal structure within the CMR modeling framework.

Experiment 2: Strong semantic structure and reordered practice

Repeated exposure to study material is generally beneficial to memory (Madigan, 1969), and this benefit is greatly enhanced when the re-exposure involves testing or actively practicing retrieval of the material (the testing effect; Rowland, 2014). One theory of the testing effect, the Episodic Context Account, proposes that the beneficial effects of retrieval practice involve retrieval of the initial episodic context of the practiced item (Karpicke et al., 2014). This retrieval is proposed to strengthen the association between an item and its temporal context, which makes the item easier to retrieve in the future. Given that the temporal contiguity effect is thought to reflect the retrieval of an item's temporal context, one might expect a strengthening of the temporal contiguity effect following retrieval practice. However, we are unaware of clear empirical evidence for this hypothesis.

Whiffen & Karpicke (2017) examined how retrieval practice affects temporal organization. Participants studied two lists of items. During a retrieval practice period, participants were shown items and were prompted to indicate which list each was originally presented on. During a final free-recall period, participants in the retrieval practice condition more reliably organized their responses by the list of origin. This coarse form of temporal organization was strengthened by retrieval practice, but it is unclear whether a standard retrieval practice task involving retrieval of the study item itself would also strengthen temporal organization. It is also unknown whether the temporal contiguity effect as characterized by the lag-CRP analysis is strengthened by retrieval practice.

There is a potential challenge in evaluating the hypothesis that retrieval practice will strengthen temporal organization of a study list. To describe this challenge we consider a multi-trial free recall study reported by Klein et al. (2005). In one condition, the same set of items was presented in a different order on each trial, and each trial was followed by a free-recall period. Klein et al. showed that the temporal contiguity effect (based on the lag distances defined by the first presentation order) became progressively weaker across trials. This makes sense from the perspective of retrieved-context theory, as each trial introduces a new temporal ordering of the studied items. During memory search, these multiple temporal structures may compete with one another, making it seem like the fine-grained temporal structure of the list has been weakened. We explore this idea in the simulations following these experiments.

Direct retrieval of studied items can be prompted with a word-stem completion practice task, which involves presenting a few letters of the study item. Carpenter & DeLosh (2006) found that a more difficult word-stem practice task (i.e., one that presents fewer letters) enhances the memorability of the practiced item on a later test. We used a task design similar to M. A. Smith et al. (2013, Expt. 4), where study items were drawn from multiple taxonomic categories, and retrieval practice prompts involved giving both the category name and a few letters from the target word. We sought to determine if the improvement in performance associated with retrieval practice difficulty would also be associated with enhanced temporal organization. We found, however, no hint of temporal organization for any level of retrieval practice difficulty. We demonstrate that this null temporal contiguity effect is reliable with a re-analysis of M. A. Smith et al. (2013) Expt. 4, which also shows an absence of temporal organization. A different experiment from the Smith et al. series shows substantial temporal organization, allowing us to draw some conclusions regarding the importance of reordered practice for disrupting the effect.

We note that the free-recall task examined in this Experiment differs from a standard free-recall task in a few ways. Aside from the strong semantic structure and reordered practice period, the study list and the retention interval are both substantially longer than a standard free-recall experiment. Our simulations suggest that the effects of semantic structure and reordered practice may combine with list length to eliminate or completely obscure temporal organization. We return to the question of retention interval duration in the general discussion.

Method

Participants

Thirty-seven Vanderbilt University undergraduates participated in exchange for course credit (11 male, mean age 19.7 years). One participant was excluded due to a software error, leaving 36 participants in the analysis.

Materials

Each word pair consisted of a category name as a cue and an exemplar as a target (e.g., Birds–Finch), using the same categorized word pool as Expt. 1. These norms rank category exemplars by frequency of production; we excluded the top four exemplars from each category from the pool of target words, to reduce the likelihood that participants might guess the target word during the practice period (in place of episodic recall). The stimulus set included fifteen category names and six target items from each category. These 90 cue-target pairs were divided into three sets of 30 cue-target pairs (for assignment to each of the three kinds of practice task). Each set contained 15 categories with 6 exemplars per category, and each set was matched in terms of word length of the exemplars (which varied from 5–10 letters). Due to a coding error, one word pair appeared twice in two different conditions. This item was excluded from the analyses.

Procedure

Participants were given a brief overview of the experiment and performed a practice round with three items not included in the experimental list. Given the length of the study list and duration of the distraction periods, the experiment consisted of a single trial per participant: Initial presentation of the study list, followed by the practice period, followed by the distraction period, followed by a free-recall test.

The study list contained 90 category-exemplar word pairs. Each word pair (e.g., Bird–Finch) was presented for 5 seconds with an empty response box below it. Participants typed the target word (without its category name, e.g., Finch) in the response box. The presentation order of word pairs was randomized for each participant. A 5-min distraction period followed the study phase, during which participants played a video game (Tetris) on the testing computer.

During the practice phase the 90 study pairs were presented again, with varying levels of retrieval difficulty. The category label was presented, along with the associated exemplar's first two letters (high retrieval demands), the first four letters (low retrieval demands), or the entire word (restudy/retype). In the case of the two word-stem prompts, blanks were presented to indicate the number of missing letters. These three practice types were intermixed, with 30 pairs assigned to each type. In each case a blank response box was presented beneath the exemplar, and the participant typed the entire word (not just the missing letters). Each practice event lasted 8 seconds. No feedback was provided.

A second distraction period followed the practice period (15 minutes of Tetris on the testing computer). This was followed by the final free-recall period. Participants were asked to recall as many target words as possible, in any order and without a time limit. Participants typed each word in a blank response box. They pressed the "enter" key to submit the recall, which caused the typed response to disappear, leaving a blank response box for their next recall. When they decided they were finished recalling the target words, participants typed "done" in the response box, which ended the session.

Smith et al. 2013, Experiments 1 & 4

We include analyses from two archival experiments reported by M. A. Smith et al. (2013) (Smith Experiment 1 and 4). The current Expt. 2 was modeled after Smith Experiment 4. They used similar study materials (category name–exemplar paired associates). Study lists contained 60 items (6 items from each of 10 categories). Delay between initial study and practice period was 3 minutes (distraction task, Pac-Man), and the delay between the practice period and final free-recall period was 15 minutes (distraction task, Tetris). During the practice period, retrieval trials (cued recall given category name and 2-letter word stem) and restudy trials (retype the word) were intermixed. They also included a between-subject manipulation of whether retrieval practice was overt or covert. We included both overt and covert trials in our reanalysis.

We also include an analysis of Smith Experiment 1 for comparison. Smith Experiment 1 used similar study materials. Study lists contained 90 items drawn from the same set of categories (4–6 items per category), and the items were blocked by category. The study-to-practice delay was 3 minutes, as above. During the practice period participants performed category cued recall, either overtly (typing remembered category exemplars) or covertly. The delay from practice to final free recall was 15 minutes, as above.

Results

Memory performance

During the practice phase, participants were able to reliably retrieve the studied target words given the category name and cue letters. Proportion of the 2-letter cues successfully completed: $M = .86, SD = .08$, and 4-letter cues: $M = .97, SD = .05$. Participants were always accurate when typing in the target word on the restudy trials.

Participants performed well on the free-recall test, recalling on average about half of the studied items ($M = 0.48$ [0.42–0.54]). The likelihood of remembering a particular item was modulated by the type of retrieval practice it received (free recall given 2-letter cue practice: $M = 0.56, SD = .19$; free recall given 4-letter cue practice: $M = 0.52, SD = .22$; free recall given restudy: $M = 0.43, SD = .22$). A repeated measures analysis of variance (ANOVA) examined the effect of practice condition on recall. Free recall performance varied significantly across conditions ($F(2,70) = 18.18, MSE = .17, p < .001, \eta_p^2 = .34$). Contrasts showed that restudied words were remembered more poorly than words practiced with 2-letter cues ($t(70) = 5.92, p_{tukey} < .001$), and 4-letter cues ($t(70) = 3.94, p_{tukey} < .001$). There was not a statistically significant difference in free-recall performance for items practiced with 2-letter cues and 4-letter cues.

Figure 3 shows probability of recall as a function of the serial position of the studied pair. Two serial position curves are presented, one respective to the order of the items in the study phase, and the other respective to the order of the items in the practice phase. For both orderings, there was a slight trend towards a primacy advantage, otherwise items throughout the list were recalled similarly well. Given the long distraction periods following both phases, we did not expect to see a recency effect, and one was not observed.

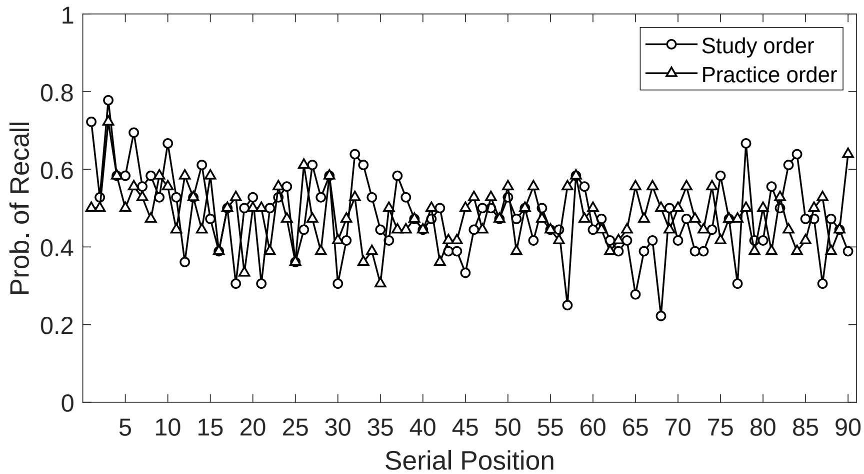

Figure 3: Serial position curve for Experiment 2. The overall flatness of the curve indicates that list position was not a strong predictor of recall probability, although a slight primacy advantage was observed. The circles indicate probability of recall according to study position; the triangles indicate probability of recall according to practice position.

Temporal organization.

We found no evidence for temporal organization during the free-recall period. Figure 4 presents two lag-CRP analyses, one relative to the initial study order, and one relative to the practice order. Both show completely flat curves, with no evidence of increased likelihood of short-lag recall transitions.

To characterize these flat curves, we constructed a linear regression model with two parameters, a slope and an intercept (i.e., conditional response probability = slope $* |lag| +$ intercept). This was fit separately to the forward-going lags (+1 to +5) and the backward-going lags (-1 to -5). We note that it would not be appropriate to apply this linear regression to the lag-CRP curves from Expts. 1, 5, & 6, as the lag-CRP values from those experiments show a non-linear curve better described by a power function (see Table 3). Appendix A discusses why the non-linear regression model is not appropriate for the experiments showing flat lag-CRP curves. Neither the backward (slope = 0.001, intercept = 0.013) nor forward (slope = 0.000, intercept = 0.018) slopes differed significantly from zero (backward slope: $t(35) = 0.80$, $p > 0.2$, forward slope: $t(35) = -0.18$, $p > 0.2$). This indicates that short-lag transitions are no more likely than longer-lag transitions.

A percentile-rank temporal organization analysis confirmed the lack of reliable temporal organization. The average temporal factor score (relative to study order: $M = 0.50$ [0.48–0.52], relative to practice order: $M = 0.50$ [0.48–0.53]) did not significantly differ from the chance level of 0.5 (both $ps > 0.2$). The p(+1) input-output correspondence statistic is not well suited to establishing chance-level temporal organization

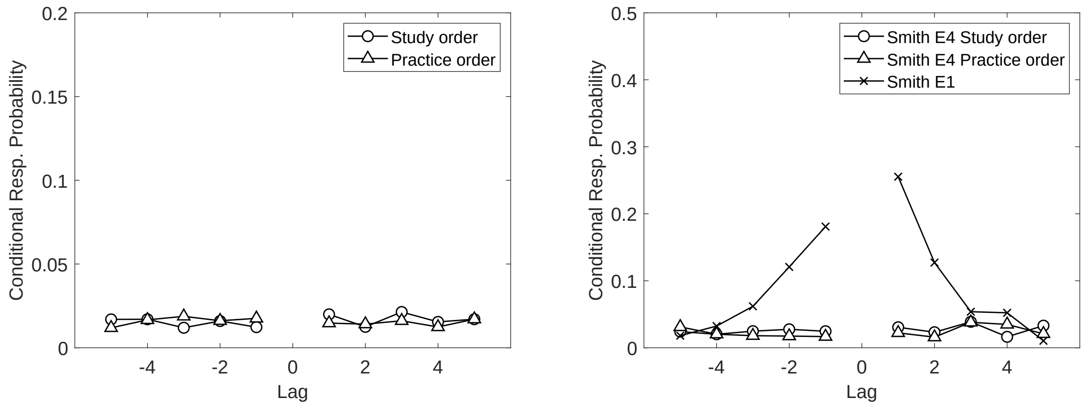

Figure 4: (A) Conditional Response Probability analysis for Experiment 2. The circles indicate probability of lag-transitions relative to the original study order. The triangles indicate probability of lag-transitions relative to the practice order. Neither analysis shows evidence of temporal organization. (B) Conditional Response Probability analysis for archival data from Smith et al. (2013) Experiments 1 & 4. Smith Expt. 4, which is structured like the current paper's Expt. 4, shows no evidence of temporal organization. We compare this to Smith Expt. 1, where participants perform free recall by category during the practice period, and there is robust temporal organization during final free recall.

(see Appendix A) but nevertheless the associated p(+1) likelihood is very low (see Table 2).

Figure 4B shows a set of lag-CRP analyses on archival data from M. A. Smith et al. (2013) (methods described above). Our Expt. 2 was modeled after Smith Expt. 4; both experiments have long lists with strong category structure and items from a given category are widely spaced. Both experiments also have a practice period where the items are presented in a different order. There is no evidence for temporal organization in Smith Expt. 4, either with regard to the study order or practice order. We contrast this with an analysis of Smith Expt. 1, which shows robust temporal organization and has two key methodological differences: Study items are blocked by category, and participants perform a recall-by-category task during the practice period (as described above). Percentile-rank temporal organization scores for the Smith experiments are reported in Appendix A, Table 2.

Polyn et al. (2011) also examined temporal organization in lists composed of items from multiple taxonomic categories. In their report, temporal organization seemed to be absent in the overall lag-CRP analysis. They ran a follow-up analysis which separately examined within-category and between-category recall transitions (see Appendix A for a detailed description of the analysis), and this analysis revealed reliable temporal organization for both within-category and between-category recall transitions. We ran this analysis on the current data, as shown in Figure 5. Unlike in Polyn et al. (2011), there was no evidence of temporal organization in either the within-category or between-category recall transitions. In the Polyn et al. (2011) study, there were only three categories in the study list, so the spacing of same-category items was much smaller

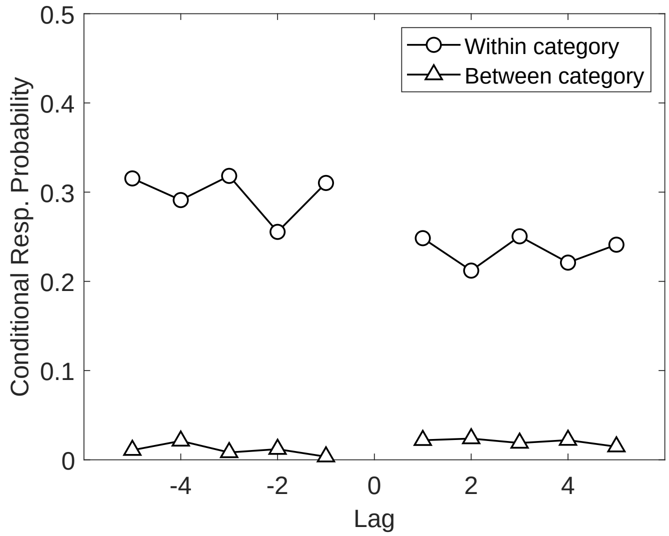

Figure 5: (A) Conditional Response Probability analysis for within-category and between-category transitions in Experiment 2. Both analyses are with regard to the original study order. The circles indicate probability of lag-transitions between items from the same category. The triangles indicate probability of lag-transitions between items from different categories. Neither analysis shows evidence of temporal organization.

than the current experiments. Our simulations suggest that temporal organization can be diminished but generally preserved when the spacing of same-category items is smaller, and that as same-category spacing increases, this becomes progressively more disruptive to temporal organization.

Semantic organization.

We observed strong and reliable category clustering during free recall, measured with an Adjusted Ratio of Clustering (ARC) score (Appendix A). On average, participants' mean ARC score was 0.70 [0.62–0.77]. This was dramatically and significantly greater than chance, as estimated with a permutation analysis with a thousand randomly generated recall sequences using our stimuli (permuted ARC $M = 0.0045$, observed ARC score was significant at $p < .001$ compared to the permutation distribution).

Discussion

Overall recall performance in Expt. 2 was very good (about 45 responses per participant), and retrieval practice improved item memorability relative to restudying items. This level of recall performance suggests that participants had no trouble targeting the coarse temporal context of the study and practice lists. Nevertheless, there was no evidence for the fine-grained temporal organization associated with the temporal contiguity effect. Prior work (including Expt. 1 of this report) has shown that strong semantic organization can weaken the contiguity effect, but has not previously been shown to completely abolish the effect (Polyn et al., 2011; Healey, 2018).

The reliable temporal contiguity effect observed in Expt. 1 suggests that the elimination of temporal organization is not simply due to the nature of the study items themselves. Our simulations suggest that both the shuffled practice order and widely spaced strong semantic associations likely contributed to the elimination of temporal organization. In Experiment 3, we demonstrate that removing the practice period (and thus only having one temporal ordering of the items in memory) does not cause the contiguity effect to re-emerge.

Our reanalysis of the data from M. A. Smith et al. (2013) Expt. 4 demonstrates that this nullification of temporal organization is reliable under similar methodological conditions. Furthermore, our reanalysis of Smith Experiment 1 shows that reliable temporal organization can emerge with a few key methodological differences. Two differences are most relevant. In Smith Expt. 1 study items are blocked by category. Our simulations suggest that this is less disruptive to temporal organization than having the different categories intermixed. Smith Expt. 1 also has participants perform category cued recall during the practice period. To the extent that participants produce responses that themselves exhibit temporal organization during the practice period, this will tend to preserve temporal organization during the final free recall period. In contrast, the practice periods of the current experiment and Smith Expt. 4 involve a complete reordering of the study materials.

Experiment 3: Strong semantic structure, no reordered practice

We hypothesize that in Experiment 2, the strong semantic structure of the study materials and the reordered presentation of the items during practice were important factors leading to a null contiguity effect. Experiments 3 and 4 retain many design features of Experiment 2. In this experiment we retain the strong semantic structure of the study materials, but there is only a single study period. We also introduce a generation task to the study list (described below) to ensure participants would have good memory for the study materials after the removal of the practice period Slamecka & Graf (1978).

Method

Participants

Thirty-one adults participated in exchange for course credit or a cash payment (9 male, mean age 23.5 years). One participant was excluded for failing to follow instructions, leaving 30 participants in the analysis.

Materials, Design, and Procedure

The materials and many aspects of the design were the same as those used in Experiment 2. 90 category-exemplar word pairs (15 categories with 6 items each) were presented during the study phase, followed by a 15-min Tetris filler task, followed by a free-recall test. The study phase for this experiment was modeled after the practice phase of Expt. 2, in that study items were presented as a category label plus one of three cue types: A 2-letter cue, a 4-letter cue, or a whole-word cue (i.e., no generation). 30 study items were assigned to each cue type and were randomly intermixed. For 2- and 4-letter generation cues, participants were instructed to type a word that would fit the given word stem and its category. For example, when given the cue: Occupations - At_ _ _ _ _, 'Attorney' or 'Athlete' are both acceptable answers. The number of blanks was chosen to be consistent with a target word from the category, but participants could provide another valid exemplar, even if the total number of letters didn't match the provided number of blanks. Generated responses were considered incorrect if they didn't match the stem letters (e.g., 'Astronaut') or were not from the target category (e.g., 'Atlanta'). For items with whole-word cues, participants were instructed to simply copy (retype) the entire target word (e.g., 'Athlete') in the response box.

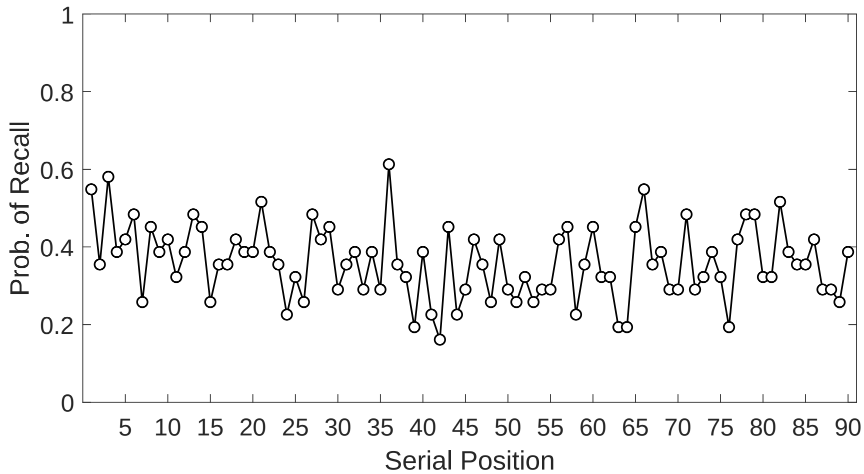

Figure 6: Serial Position Curve for Experiment 3. The overall flatness of the curve indicates that list position was not a strong predictor of recall probability.

Results

Memory performance

During the study phase, whole-word retyping was always successful, and generation was significantly better given a 4-letter cue ($M = .95, SD = .10$) than a 2-letter cue ($M = .72, SD = .13$). Free-recall performance was calculated as a proportion of the number of items the participant successfully generated during the study phase. Participants generally performed well on the free-recall test. Free recall performance was affected by the study task. For items generated from 2-letter cues: $M = 0.50$, $SD = .15$, 4-letter cues: $M = 0.45$, $SD = .16$, and retyped words: $M = 0.27$, $SD = .13$. A repeated measures ANOVA examined how study task (within-subjects: 2-letter cue, 4-letter cue, retype) affected free-recall performance, and showed a main effect of study task ($F(2,58) = 45, p < .001, \eta_p^2 = .068$). Participants recalled items generated from 2-letter and 4-letter cues similarly well, and both were recalled better than retyped words: 2-letter cue vs. retype, $t(58) = 9.03, p_{tukey} < .001$, 4-letter cue vs. retype, $t(58) = 7.03, p_{tukey} < .001$.

Figure 6 shows probability of recall as a function of serial position during the study phase. As in Expt. 2, items throughout the list were recalled similarly well. We did not see strong evidence for primacy or recency effects.

Temporal and semantic organization.

As in Expt. 2, lag-CRP curves were flat (Fig. 7). We characterized this with a linear regression model. The backward slope did not differ significantly from zero (slope = -0.002, intercept = 0.013; $t(30) = -1.03$,

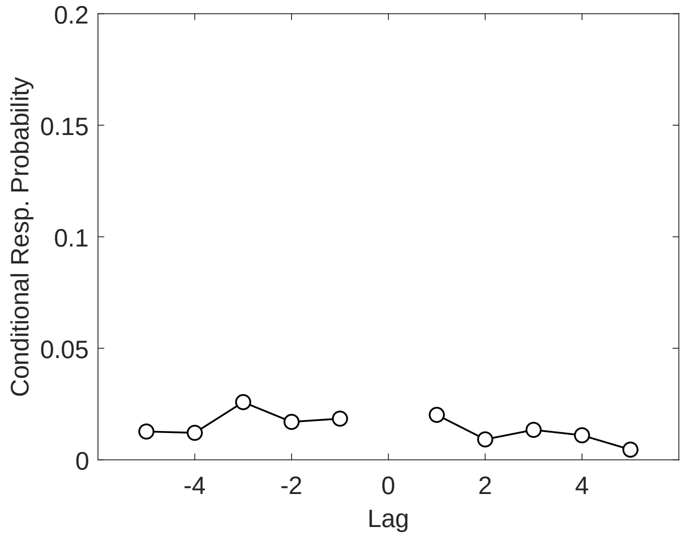

Figure 7: Experiment 3. Conditional Response Probabilities according to study order. Each point corresponds to the probability of making a transition of this serial position lag conditional on the availability of that item for recall. No evidence for a contiguity effect is present. See text for statistical details.

$p > 0.2$), but the forward slope was slightly but significantly negative, indicating a small effect of temporal organization (slope = -0.003, intercept = 0.018; $t(30) = -2.13$, $p < 0.05$). In contrast, according to the percentile-rank temporal organization score, temporal organization ($M = 0.49$ [0.47–0.51]) did not significantly differ from chance level of 0.5 ($t(29) = 1.05$, $p = .98$, $BF10 = .32$).

We examined whether temporal organization was stronger for items generated using a partial retrieval cue relative to those that were retyped. There was no evidence for temporal organization for either subset of items considered on their own. We examined whether temporal organization estimates were affected by masking out serial positions containing items that were not successfully generated (i.e., treating transitions to these positions as invalid for the purposes of determining possible transitions, see Appendix A). Both the lag-CRP analysis and temporal organization score analysis showed a null contiguity effect with this modification. With this modification, the marginally significant slope effect reported above was no longer significant. We ran the lag-CRP analysis restricted to either within-category or between-category transitions and found no evidence of temporal organization. Finally, we ran a variant of the temporal organization score analysis restricted to either within-category or between-category transitions and found no evidence of temporal organization.

As in Experiment 2, we observed strong and reliable category clustering during free recall, measured with ARC, $M = 0.65$ [0.56–0.73], $p < .001$ by a permutation test.

Discussion

As in Experiment 1, overall recall performance was good, suggesting that participants were able to focus their memory search on the temporal context defined by the study list. The weak temporal organization suggested by the regression analysis of the full data set was not apparent in any of our other analyses. As such we consider this experiment as providing another example of a null contiguity effect. As in Expt. 2, we observed strong semantic organization in the form of category clustering. Simulation 2 demonstrates how widely spaced categorical associates can be devastating for the temporal contiguity effect.

Experiment 4: Weak semantic structure, reordered practice

Whereas Experiment 2 had participants study items in two distinct orders, in Experiment 3 the items were only presented once. In both cases the temporal contiguity effect was absent. In this experiment, we retain the two study orders, but use study materials with weaker inter-item semantic associations, no category structure, and a shorter list length. As in the previous two experiments, we do not see the temporal contiguity effect. Aspects of this experiment were previously reported by Hong et al. (2019), which may be consulted for full methodological details. Here we focus on the organizational analysis, which was not previously reported.

Participants

Sixty-one adults (14 male; mean age 19.6 years) participated in exchange for cash payment. Participants were randomly assigned to a retrieval practice condition and a restudy condition.

Materials

Participants studied word pairs (cue-target pairs, e.g., Flower–pansy) selected from a pool of normed paired associate study items (Jacoby, 1996). In the study phase and the restudy practice condition, the cue and target words were presented intact. In the retrieval practice condition, the pairs were presented as an intact cue word and a fragment with some letters missing (e.g., Flower–_a_sy). The pool was designed by Jacoby (1996) such that two possible target words could complete the fragment, one typical (daisy), and one atypical (pansy). Our study pairs used the atypical target words to minimize the likelihood of guessing based on semantic associations in the absence of episodic recollection. Jacoby (1996) reported the average generation frequency for each item given the cue word as a prompt. For the atypical target words this generation frequency was .086 (vs. .623 for their typical alternatives). While the cue and target words within a pair had a reasonably strong semantic association, there was not any systematic semantic structure connecting the words from one word pair to the words in the other word pairs.

Procedure

The study phase included 40 cue-target pairs, whose order was randomized across participants. Each word pair was presented for 5 seconds (with a 1-second inter-presentation interval) in one of four font colors (blue, green, orange, or yellow), with ten cue-target pairs in each color. As the pair was presented, the participant performed a simple orienting task (pressing a labeled key indicating the font color of the pair). A

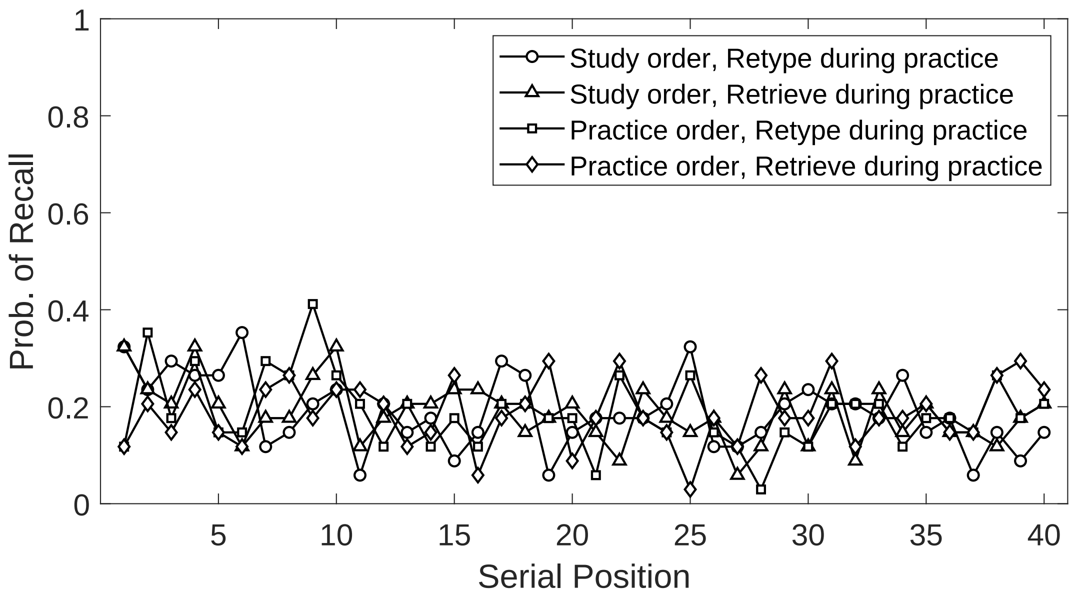

Figure 8: Serial Position Curves for Experiment 4. Overall recall performance was lower than in previous experiments. The curves were flat relative to both study order (circles) and practice order (triangles), and for the retype condition (open symbols) and retrieval practice condition (filled symbols), indicating that list position was not a strong predictor of recall probability.

5-min distraction period followed the study phase (Tetris). During the practice phase the 40 cue-target pairs were presented in a different order, and the practice task varied between subjects. In the restudy condition, participants simply typed the (fully presented) target word. In the retrieval practice condition, participants were shown the cue word and the target fragment, and were asked to complete the fragment. A 15-minute distraction period followed the practice phase (Tetris). This was followed by a free-recall test.

Results

Memory Performance

During the practice phase, participants in the retrieval practice condition successfully completed the fragment on about half of the trials ($M = .56, SD = .15$). Retyping was always successful. On the free-recall test, items that were successfully retrieved during the practice phase had an advantage relative to items in the restudy condition (successful free recall after retrieval practice success: $M = .29$, $SD = .13$ vs. after restudy: $M = .19$, $SD = .12$), $t(66) = 3.36$, $p = .001$. Items that were not successfully retrieved during the retrieval practice phase were poorly remembered during free recall ($M = 0.05$, $SD = 0.08$). Free recall for this set of items was reliably worse than for the items that were successfully retrieved during the practice phase ($t(33) = 11.62$, $p < .001$).

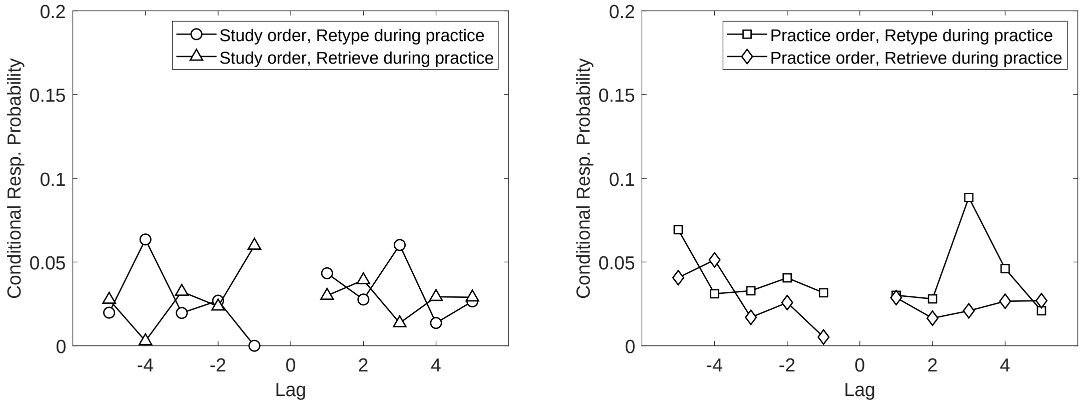

Figure 9: Conditional Response Probability analysis for Experiment 4. (A) Organizational analysis relative to the original study order. (B) Organizational analysis relative relative to the practice order. These lag-CRP analyses did not show evidence for temporal organization, but the temporal organization score analysis showed weak but significant temporal organization relative to the practice order, when the retyping task was used during practice. See text for statistical details.

Figure 8 depicts a serial position analysis of free-recall performance. As in Experiments 2 and 3, there was no evidence of either a recency or primacy effect.

Temporal organization.

As in Experiment 2, we carried out two lag-CRP analyses of temporal organization during free recall (Figure 9), one with respect to the original study order, and one with respect to the practice order. We carried out a linear regression on the forward and backward legs of the lag-CRP curves for both the study and practice orders, split by practice period task (retrieval practice or retyping; see Table 2). The slope was not reliably different from zero for any of these linear regression fits (all $ps > 0.05$, see Table 2 for mean slope estimates). In other words, the lag-CRP analysis provided no evidence for temporal organization.

For the most part, the percentile-rank temporal organization score showed the same thing. When this analysis was performed with respect to the practice order, for the participants who performed the retyping task during practice, there was a weak but statistically significant effect ($M = 0.56$ [0.51–0.61], $t(30) = 2.40$, $p = 0.02$). None of the other conditions showed reliable a temporal contiguity effect by this measure (all $ps > 0.2$, see Table 2 for mean values and confidence intervals).

Discussion

Experiment 2 establishes that strong semantic structure and reordered practice together can eliminate temporal organization. Expt. 3 shows that strong semantic structure on its own can eliminate temporal organization. Similarly, Expt. 4 shows that reordered practice on its own can eliminate temporal organization. The simulations following the experiments examine each of these factors more closely.

In this experiment we found a weak contiguity effect when temporal organization was calculated relative to the practice order, when a retyping task was used during practice. The weakness of the effect raises the possibility it is a false positive. However, we note that in Expt. 5 the retyping task also produces stronger temporal organization than a semantic orienting task. The effect of orienting task on temporal organization is revisited in the discussion.

Experiment 5: Weaken semantic structure, remove reordered practice

In Expt. 4 we weaken the semantic structure of the study materials and retain the reordered practice period, and temporal organization is not observed. In this experiment we demonstrate that with weak semantic structure and removal of the reordered practice period, temporal organization returns. We also evaluate whether the paired associate structure of the studied materials used in Expts. 2, 3, & 4 is an important factor for temporal organization, we manipulated item structure in this experiment. In two conditions items are presented as singletons, and in the other two they are presented as paired associates. In our previous experiments, participants only recalled the second terms of the paired-associate study items. The same technique was used in this experiment, for consistency. This technique makes the single-item presentation condition more directly comparable to the paired-associate presentation condition. For each type of item structure we also manipulated the orienting task used while studying the items, to determine whether emphasizing the semantic features of the studied items would affect temporal organization.

Participants

Sixty adults participated in exchange for course credit or $10 (11 male, mean age 19.2 years).

Materials

Each study list contained either 24 cue-target pairs (e.g., MICROWAVE - popcorn) or 24 singleton study words (e.g., popcorn). The stimuli were selected from a pool of 812 words used in the Penn Electrophysiology of Encoding and Retrieval Study (PEERS; Healey & Kahana, 2014). We calculated pairwise semantic relatedness scores for these words using the Word Association Spaces model (Steyvers et al., 2004; Kahana, 2000). In general, a WAS value of 0.7 = high similarity, 0.4 = medium similarity and 0.14 = low similarity (Steyvers et al., 2004). We selected cue-target word pairs whose WAS similarity value was 0.19 on average. This particular value (0.19) was determined based on the average cue-target association strength from Experiment 4 where cue words were paired with atypical associates as targets.

Procedure

The experiment used a 2 (study item type: single word, word pair) x 2 (orienting task type: typing, semantic) between-subjects design. An experimental session had 5 trials. A given trial consisted of a study period (a 24-item list), a distraction period (3 min playing Tetris), and a free-recall period (with typed responses). Between trials participants performed a filler task for 3 minutes (Tetris). The order of study item presentation was randomized for each participant. Each item was presented for five seconds with a one second inter-presentation interval. When a study item was presented, participants either performed a typing task (as in Expt. 6) or semantic orienting task. For the typing task, only the target word was typed for paired-associate items. For the semantic orienting task on paired-associate items, participants rated cue-target relatedness on a 4-point scale. For the singleton items the semantic task was to rate the pleasantness of the word on a 4-point scale. During the free-recall period participants were instructed to only type the target items in the paired-associate conditions.

Results

Memory performance

An ANOVA examined the effect of item type and orienting task on free-recall performance, in terms of the proportion of study items that were recalled. We observed a main effect of item type ($F(1,56) = 12.57$, $p < .001$): Singleton target words were better remembered than paired associate target words ($M_{singleton} = .41$ vs. $M_{paired} = .26$). There was not a main effect of orienting task on recall performance ($M_{retype} = .31$ vs. $M_{semantic} = .36$, $F(1,56) = 1.35$, $p = .25$). Post-hoc comparisons revealed that participants in the singleton-retype task condition ($M = .45$ [0.36–0.54]) performed significantly better compared to both paired-retype task ($M = .28$ [0.21–0.35], $t(56) = 6.67$, $p_{tukey} < .001$) and paired-semantic task ($M = .25$ [0.21–0.30], $t(56) = 7.65$, $p_{tukey} < .001$) conditions. However, free-recall performance in the singleton-retype condition ($M = .38$ [0.28–0.48]) was not reliably different than the singleton-semantic condition ($t(56) = 1.22$, $p_{tukey} = .62$).

Figure 10 shows recall performance as a function of serial position of the study item. A primacy effect was observed that spanned the first several list positions. No recency effect was observed. This was expected given the distraction-filled retention interval.

Temporal organization

Figure 11 shows that reliable contiguity effects were observed in all four conditions. The full set of temporal organization statistics are gathered in Table 2. A 2x2 ANOVA examined the effect of item type and orienting task on the percentile-rank temporal organization score. There was a main effect of item type on temporal organization ($F(1,56) = 4.55$, $p < 0.05$), but not a main effect of orienting task ($p > 0.2$), and there was not a reliable interaction between the two factors ($p > 0.2$). Contrasts showed that generally, there was reliably stronger temporal organization for singleton items compared to paired associates ($M_{singleton} = .67$ vs. $M_{pair} =$

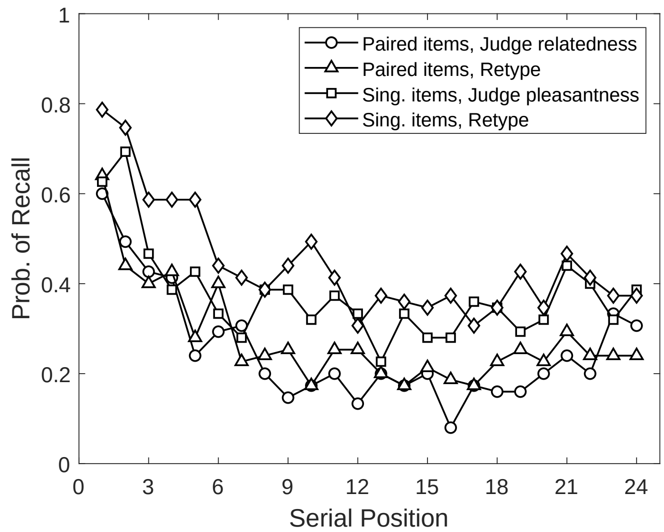

Figure 10: Serial position curve for Experiment 5. A substantial primacy effect was observed for all conditions. Singleton study items (triangles) were generally better recalled than paired-associate study items (circles). Within item type, participants showed similar levels of overall recall for the retyping orienting task (filled symbols) and the semantic orienting task (open symbols), see text for statistical details.

.62, $t(56) = 2.36$, $p_{tukey} = .022$). Within the singleton condition there was stronger temporal organization for the retyping orienting task as compared to the semantic (pleasantness) orienting task ($M_{retyping} = 0.72$ vs. $M_{pleasantness} = 0.64$, $t(28) = 2.33$, $p = 0.03$). Within the paired associates condition orienting task didn't reliably affect temporal organization ($M_{retyping} = 0.63$ vs. $M_{relatedness} = 0.62$, $t(28) = 0.36$, $p > 0.2$).

Discussion

The contiguity effect was observed in all four conditions. The results of this experiment help to determine which methodological details are most likely responsible for the dramatic modulation of temporal organization across these experiments. For example, it is unlikely that using Tetris as a distraction task was responsible for the elimination of temporal organization in earlier experiments, as it was also used here. It is also unlikely that the paired-associate item structure used in earlier experiments was critical for eliminating temporal organization, as we observed robust temporal organization here for both paired-associate conditions. It is possible that paired-associate items are processed differently than the singleton items, and we return to this point in the discussion.

Most studies examining temporal organization involve study lists where words are presented one at a time. Classic work proposed that the associative structures formed in paired-associates and free-recall tasks were theoretically similar (Postman, 1971; Dong, 1972). Davis et al. (2008) found that cross-pair

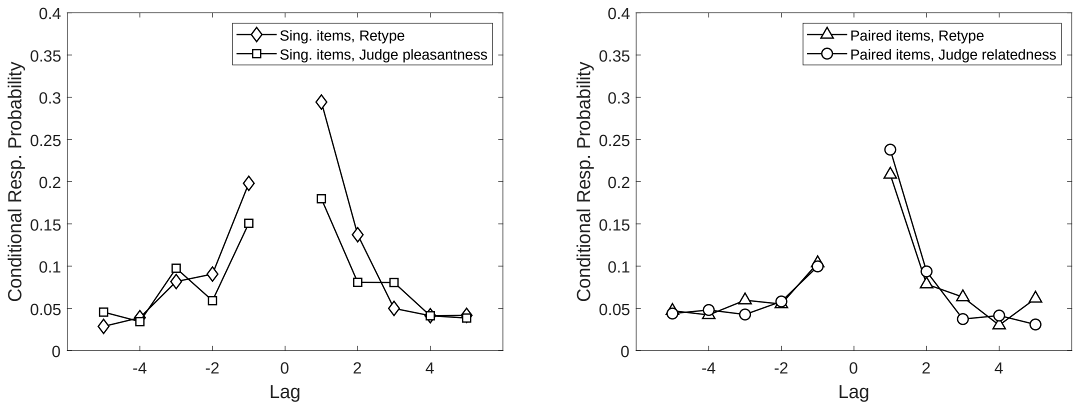

Figure 11: Conditional Response Probability analysis for Experiment 5. (A) For the singleton study items orienting task affected the contiguity effect, with the retyping task (diamonds) supporting a stronger contiguity effect than the semantic task (squares). (B) For the paired-associate study items, orienting task did not reliably affect the strength of the contiguity effect (retyping task: triangles, semantic task: circles). See text for statistical details.

intrusion errors in a cued-recall paired-associates task were more likely for temporally proximal pairs, an effect reminiscent of the temporal contiguity effect in free recall. In contrast, Osth & Fox (2019) found that error rates for rearranged lures in an associative recognition paired-associates task did not show an effect of temporal proximity. Campbell et al. (2014) found the same thing for younger participants, but for healthy older participants, error rates for rearranged lures were affected by temporal proximity.

It is possible that when participants are asked to study a list of word pairs, they focus on forming a strong association between the two items comprising the pair. This focus could come at the expense of forming a strong association to a gradually changing temporal representation, and thus could dampen the contiguity effect during a later free-recall period. We know of two studies that potentially speak to this question. Lehman & Malmberg (2013) and Cox & Criss (2020) each examined free-recall performance after participants studied a list of simultaneously presented paired associates. Both studies found strong temporal organization for items from the same pair. Lehman & Malmberg (2013) found very weak, if any, evidence for temporal organization between pairs. However, they also observed weak temporal organization in lists of singly presented items, raising the possibility that some other factor was weakening temporal organization in both conditions. Cox & Criss (2020) observed a contiguity effect in the between-pair recall transitions (their Figure C3, showing both an effect of proximity and a forward asymmetry), though the reliability of the effect was not characterized. We note that there is a potentially relevant difference between these two studies and the current experiment. In the current experiment participants were asked to only recall the second item of the paired associates, whereas in these other two studies participants were asked to recall both items from each pair. It is possible that shifting from within-pair to across-pair associations within a memory search is disruptive to temporal organization.

Experiment 6: List length and retention interval

Our final experiment examines the effects of both list length (20 vs. 40 items) and retention interval duration (3 vs. 15 minutes) on temporal organization. Some of our previous experiments used relatively extreme list lengths and retention intervals, motivating us to examine how these factors contribute to temporal organization. Prior work has demonstrated a progressive weakening of temporal organization as list length increased from 1 to 15 items (Ward et al., 2010), as measured by the likelihood of a +1 recall transition. A similar effect can be seen in a reanalysis of free-recall data from Murdock (1962) by Healey et al. (2018).

Temporal distinctiveness theories of memory propose that experiences are tagged with temporal information that places those events on a mental timeline (Glenberg et al., 1983; Brown et al., 2007). As these experiences recede into the past, temporal information is compressed logarithmically, causing more temporally distant events to crowd one another and become less distinctive, and therefore harder to remember. These temporal distinctiveness accounts have primarily been used to explain experimental modulation of the recency effect. However, the idea that temporal information is compressed with the passage of time raises the possibility that the temporal contiguity effect could become progressively weaker as time passes.

Method

Participants

Sixty-four subjects participated in exchange for course credit or $10 (17 male; mean age 21.5 years).

Materials

The study words were selected from a pool of 812 words used in the Penn Electrophysiology of Encoding and Retrieval Study (PEERS; Healey & Kahana, 2014). 120 words were chosen from this larger pool for use in this experiment.

Procedure

Each trial consisted of a study period, followed by a distraction-filled retention interval, followed by a free-recall test. Participants were randomly assigned to one of four conditions which crossed the two list lengths (20 and 40) the two retention interval durations (3 min and 15 min). During the study period a self-paced typing orienting task was used (type the word shown onscreen). Participants played Tetris during the retention interval. Free-recall responses were typed. Trials in the different conditions took substantially

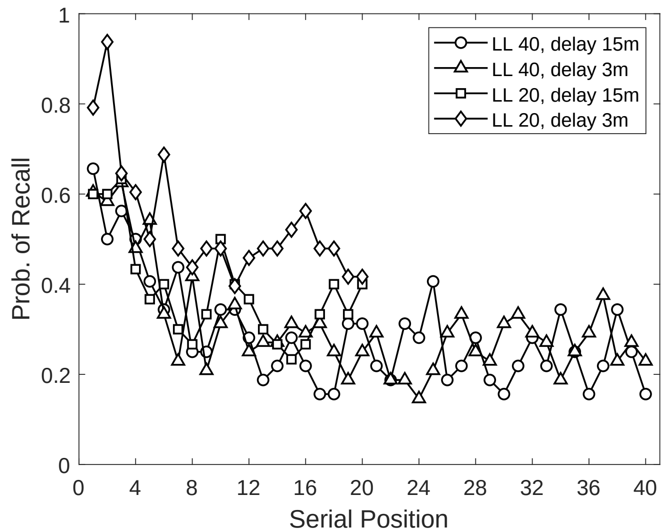

Figure 12: Serial position curves for Experiment 6. All conditions show a strong primacy effect. None of the conditions show a recency effect.

longer with the long retention intervals. A participant performed as many trials as could fit into a 1-hour experimental session, given their condition assignment.

Results

Memory performance.

Study items were best remembered with a short list length and short delay ($M = 0.54$, CI [0.44–0.64]]), followed by a short list length and longer delay ($M = .39$, CI [0.29–0.50]). Performance was lower with a long list length and short delay ($M = .31$, CI [0.24–0.38]), and lowest with a long list length and long delay ($M = .29$, CI [0.24–0.33]). A 2x2 ANOVA examined the effect of list length and retention interval on recall performance. There was a main effect of list length ($F(1,59) = 12.83$, $p < .001$), with shorter lists better remembered than longer lists. There was a marginal effect of retention interval duration ($F = 6.45$, $p = 0.074$), with slightly better memory after a shorter retention interval. The interaction of list length and retention interval was not significant ($F = 2.13$, $p = .15$).

Temporal organization.

Figure 13 shows that reliable temporal contiguity effects were observed in all four conditions. Table 2 reports three measures of temporal organization. The percentile-rank temporal organization scores suggest that neither list length or retention interval affect the strength of temporal organization. A 2x2 ANOVA

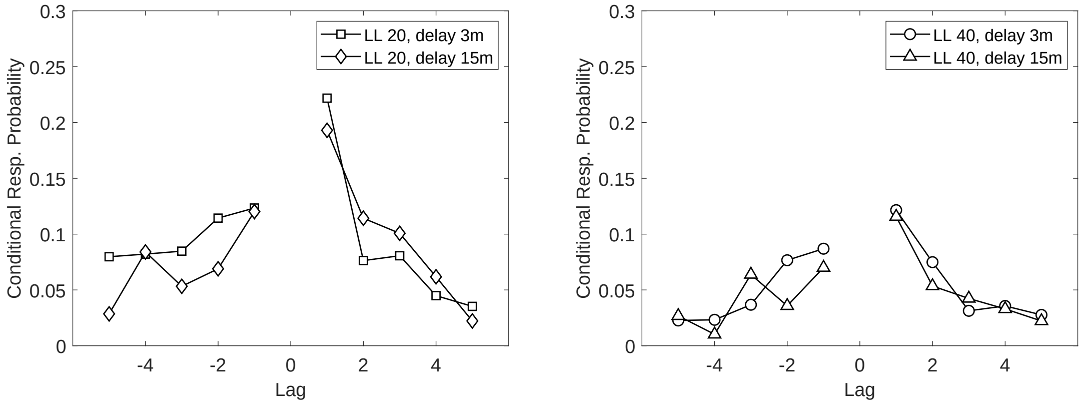

Figure 13: Conditional Response Probability analysis for Experiment 6. (A) With list length 20, increasing the duration of the retention interval did not substantially affect the strength of temporal organization. (B) The same is true with list length 40. However, for each level of retention interval, increasing the list length weakened the contiguity effect. See text for statistical details.

(list length x retention interval duration) confirms this. There was not a main effect of list length ($p > 0.2$), or retention interval ($p > 0.2$), and no significant interaction between the two factors ($F(1,58) = 0.42$, $p = 0.52$). In contrast, a p(+1) input-output correspondence analysis shows that +1 lag transitions are substantially less likely with a longer list length. A 2x2 ANOVA on this measure revealed a main effect of list length ($F(1,58) = 7.30$, $p < 0.01$), but not of retention interval ($p > 0.2$), and no significant interaction between the two factors ($p > 0.2$).

Thus, there is a discrepancy between the percentile-rank temporal organization score and the p(+1) input-output correspondence measure. The input-output correspondence analysis directly estimates the likelihood of seriation (recalling items in the same order as they were studied; Mandler & Dean, 1969) during free recall, whereas the temporal organization score provides a more holistic assessment of the strength of temporal organization by examining all lag distances. The two analyses have the potential to deviate, as they measure temporal organization in different ways. Here, we see that the two measures are affected differently by increasing list length. We elaborate upon the reason for this difference in Appendix A.

Discussion

The temporal contiguity effect was seen clearly in all four conditions. As list length is increased, the likelihood of a +1 transition decreases, but the percentile-rank temporal organization score does not. In Simulation 5, we demonstrate that CMR captures the weakening of temporal organization with increasing list length, and also show that the model predicts that the p(+1) measure is more sensitive to the weakening of temporal organization due to increasing list length. Finally, the strength of the temporal contiguity effect is not affected by an increase in retention interval duration. This is consistent with prior experimental and modeling work, which we review in the general discussion.

Simulation 1: Adjacent categorical associates weaken temporal organization.

CMR was designed to explore the interaction of different cue types during memory search in free recall (Polyn et al., 2009). The dynamics of the model are described in full in Appendix B. Briefly, temporal associations are formed when each item is studied. These associations link an item's features to the current state of temporal context. Items also contain semantic features that link meaningfully related study items to one another. These semantic associations can compete with the temporal associations that usually give rise to the temporal contiguity effect. With simulation 1, we simulate the basic design of Expt. 1. A semantic scale parameter controls the relatedness of study items to one another. Here, we use a simplified form of semantic category structure where same-category items all have a uniform degree of association, similar to how this was implemented in the CMR2 model (Lohnas et al., 2015, Simulation 4). Here, we simulate three levels of semantic relatedness. Low semantic strength (semantic scale parameter, s=0.0) corresponds to the mixed-category and Toronto conditions, and the model produces strong temporal organization (Fig. 14B). Medium and high semantic strength (s=0.5 and 1.0 respectively) correspond to the same-category condition, and demonstrate how temporal organization is progressively weakened as inter-item semantic associations become stronger (Fig. 14B).

For simulation 1, the same-category condition involves drawing all 12 study items from the same category. As such, all same-category items are adjacent. While this indeed disrupts temporal organization, it is less disruptive than the scenario examined in later simulations, where the same-category items are widely spaced throughout the study list, which contains items from many categories. The basic reason that this is less disruptive can be seen by comparing the schematic diagrams in Figs. 14A and 15A. When adjacent items are from the same category, the semantic and temporal support is combined, which increases support for both near and distant recall transitions. Due to the non-linear nature of the recall competition, this reduces some of the advantage for nearby items. However, the advantage for nearby items is not completely eliminated, as these items also benefit from being in the same category as the just-recalled item. As we will see later, when the same-category items are non-adjacent, there is very little relative support for nearby transitions, creating conditions where it is possible to eliminate the temporal contiguity effect.

Simulation 2: Widely spaced categorical associates cause extreme temporal disruption

Simulation 2 examines how strong category structure affects temporal organization when the list contains many different categories, and the same-category items are widely distributed throughout the list (Figure 15B). Here, we match the list length and category structure of Expt. 2, but do not include a reordered practice period. Figure 15B shows that increasing the strength of same-category associations devastates the temporal contiguity effect. This is because the semantic associations compete with the temporal associations. The wide spacing of the categorical associates puts the semantic associations in strong opposition to the temporal associations, so the same degree of semantic associative strength is more disruptive to temporal organization. We will see in Simulation 5 that increasing list length increases the disruptive effects of other methodological factors. Here, we simulate a list length of 90 and strong semantic associations, and the temporal contiguity effect is nearly completely eliminated, as was observed in Expt. 3.

Simulation 3: Reviewing study materials can disrupt temporal organization

Simulation 3 demonstrates that a second exposure to the study items (in the form of a retrieval practice period) disrupts temporal organization (Figure 16A), as was observed in Expts. 2 & 4. We use an intermediate list length (30 items) and remove semantic structure to establish the model's basic predictions more clearly. Temporal organization in Simulation 3 is stronger than Simulation 2 both because the list length here is shorter (30 vs. 90) and because semantic associations are turned off.

We simulate a condition where the study list is presented once (the *No practice* condition) to establish a baseline level of temporal organization (Fig. 16B). We also simulate a condition where the 30-item study list is presented again in a shuffled order. This substantially weakens, but does not eliminate, temporal organization. This can be seen in the two lag-CRP curves in Fig. 16B calculated relative to the original study order, and relative to the shuffled practice order. The practice order produces slightly weaker temporal organization. This is because the model retrieves study-period contextual information when the item is practiced. The study-period contextual information is integrated into the practice-period context and is strengthened by the Hebbian learning process that binds item and context representations. Because Expts. 2–4 so effectively eliminate temporal organization, the current set of experiments do not directly test this prediction regarding the relative strength of temporal organization in study and practice periods. We return to this point in the discussion. We note that in these simulations, we do not distinguish between different kinds of retrieval practice activities (e.g., restudying an item vs. cued recall of an item). This simplification and its implications also receive further attention in the discussion.

In a separate set of simulations (not reported), we examined a condition where the list is practiced in the same order as the original study list. The model predicts that under these circumstances retrieval practice will enhance temporal organization. This prediction is consistent with a previous study of multi-trial free recall (Klein et al., 2005). Klein et al. showed that when the same items were studied in the same order across multiple trials, the contiguity effect was strengthened.

Simulation 4: Combining semantic structure and review of study materials

Simulations 1 and 2 demonstrate the disruptive influence of semantic structure on temporal organization, and simulation 3 demonstrates the disruptive influence of a reordered practice period on temporal organization. Simulation 4 combines strong semantic structure and reordered practice to examine how these two factors interact. We simulate a list length of 30 to allow for comparison with Simulation 3 (the low semantic strength lag-CRP in Fig. 17A is the same curve shown in Fig. 16B labeled *No practice*).

Comparing Figures 17A & B shows how reordered practice affects temporal organization for different levels of semantic strength. Figure 17A shows lag-CRP curves when the items are studied once, and 17B shows lag-CRP curves when the study period is followed by a reordered practice period. The same general pattern is observed in both panels, with stronger semantic associations producing weaker temporal organization. However, when there are two conflicting temporal orderings already disrupting temporal organization, the disruptive effect of the categorical associates is somewhat blunted.

We note that across all of these simulations, temporal organization can be weakened substantially, but the model predicts that it will not be completely eliminated. Simulation results depicted in Fig. 18 demonstrate that even with a list length of 90 and reordered practice, a residual effect of temporal organization can still be seen. This is in contrast to the results of Expts. 2 and 3, where no evidence of temporal organization was apparent. Should this be considered a discrepancy between the model's predictions and the empirical data? One feature of the simulations with possible relevance to this discrepancy is that we simulate a few hundred trials per participant with the model (in order to obtain clean predictions), but in Expt. 2 for example, we only have one trial per participant, for a total of 37 trials worth of data. If we sub-sample our simulated trials to match our number of empirical trials, how often does the model predict a completely null temporal contiguity effect?

In Expt. 2, the percentile-rank temporal organization score was 0.50 with a confidence interval of 0.48–0.52. In contrast, our simulation with list length 90 and high semantic structure produced a percentile-rank temporal organization score of 0.547, with a confidence interval of 0.543–0.552 (across 126 simulated subjects). We sub-sampled the simulated trials to match the amount of empirical data 1000 times, each time sampling 40 trials and re-calculating the temporal organization score. The 95% confidence interval of the sampled distribution was 0.51–0.58, meaning that 95 percent of the samples produced a temporal organization score in this range. Only 6 of the 1000 samples produced a temporal organization score less than or equal to 0.50. This suggests that the model is over-predicting the expected degree of temporal organization relative to the observed data. However, it is something of a close call, as the upper end of the empirical confidence interval reaches into the lower end of the range of our sub-sampled estimates. In the discussion, we return to the question of whether the model is properly capturing the empirical absence of temporal organization.

Simulation 5: Increasing list length disrupts temporal organization

Figure 19 shows how increasing list length can attenuate the contiguity effect. Each response is produced following a recall competition among the memories of the studied items. Adding more items to the list introduces more items to the recall competition. After recalling a particular item, items from nearby study positions are supported by temporal associations. This support dwindles as temporal distance increases, but it reaches a non-zero asymptote. As such, adding more items to the list steadily increases the likelihood of a long-distance transition. This can be seen in Figure 19B, where two bins capture the aggregate likelihood of long distance forward-going (lag > 6) and backward-going (lag < -6) transitions. This binning analysis is described in more detail in Appendix A. Note that as list length increases, these long-distance bins include a progressively wider range of possible lag values. For example, with list length 15 the lag > 6 bin includes all lags from +6 to +14 (the longest possible forward-going transition), and with list length 30 the lag > 6 bin includes all lags from +6 to +29. Because each of these long-distance transitions has a small but non-zero probability, increasing the number of possible long-distance transitions necessarily decreases the likelihood of short-distance transitions.

In our simulations, we also noticed that increasing list length can amplify the disruptive effect of other factors. In our simulations with categorical associates, increasing the list length can increase the average lag separating same-category associates. With regards to reordered practice, items that are neighbors on the original study list will be assigned, on average, progressively more distant relative positions on the reordered list as list length increases.

In the empirical results of Expt. 6, we noted a dissociation between the percentile-rank temporal organization score, and the p(+1) input-output correspondence score. The p(+1) measure indicated that temporal organization was getting weaker, but the temporal organization score was seemingly unaffected. In order to examine how these measures are affected by list length in our simulated data, we calculated mean temporal organization scores and p(+1) scores for each list length, as reported in Table 1. The p(+1) measure is generally more sensitive to an increase in list length than the temporal organization score. Differences between these two measures are examined further in Appendix A.

Simulated list length	Temp. Org. Score	Δ	p(+1)	Δ
15	0.69	–	0.39	–
30	0.65	0.03	0.23	0.16
45	0.63	0.03	0.17	0.06
60	0.61	0.02	0.13	0.04
75	0.59	0.01	0.11	0.02
90	0.58	0.01	0.10	0.02

Table 1: Comparison of percentile-rank temporal organization score and p(+1) input-output correspondence measure as simulated list length is increased. For smaller list lengths the p(+1) measure is more sensitive to increasing list length, as indicated by the larger change (Δ) in this measure from one list length to the next.

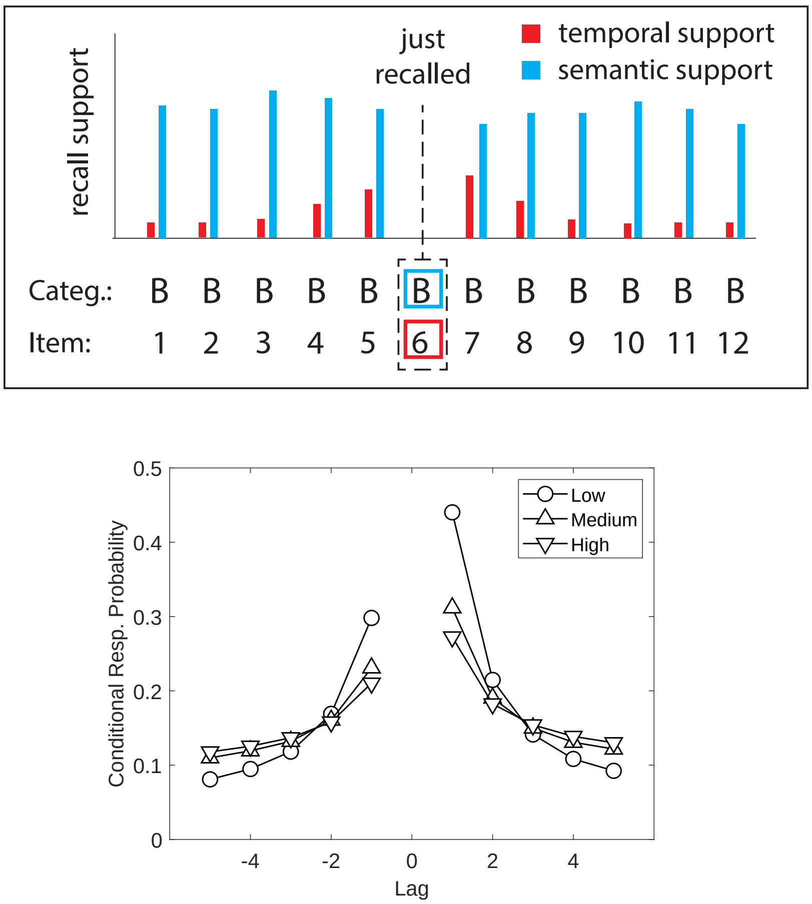

Figure 14: Simulation 1 schematic and results. (A) A schematic depiction of recall support for not-yet-recalled items, after recalling a mid-list item on a short list. When the list has weak categorical structure, only the red bars indicating temporal support are relevant. When the list has strong categorical structure, both red and blue bars are relevant, indicating that strong categorical semantic associations can overwhelm temporal associations. (B) Simulation results demonstrating how the strength of semantic associations affects probability of transitions to nearby list positions, as measured by a lag-CRP analysis. Strong within-category semantic associations increase the likelihood of transitions at all temporal distances, which weakens the temporal contiguity effect. A semantic scaling parameter alters the strength of within-category semantic associations from low (0.0) to medium (0.5) to high (1.0). See text and Appendix B for details.

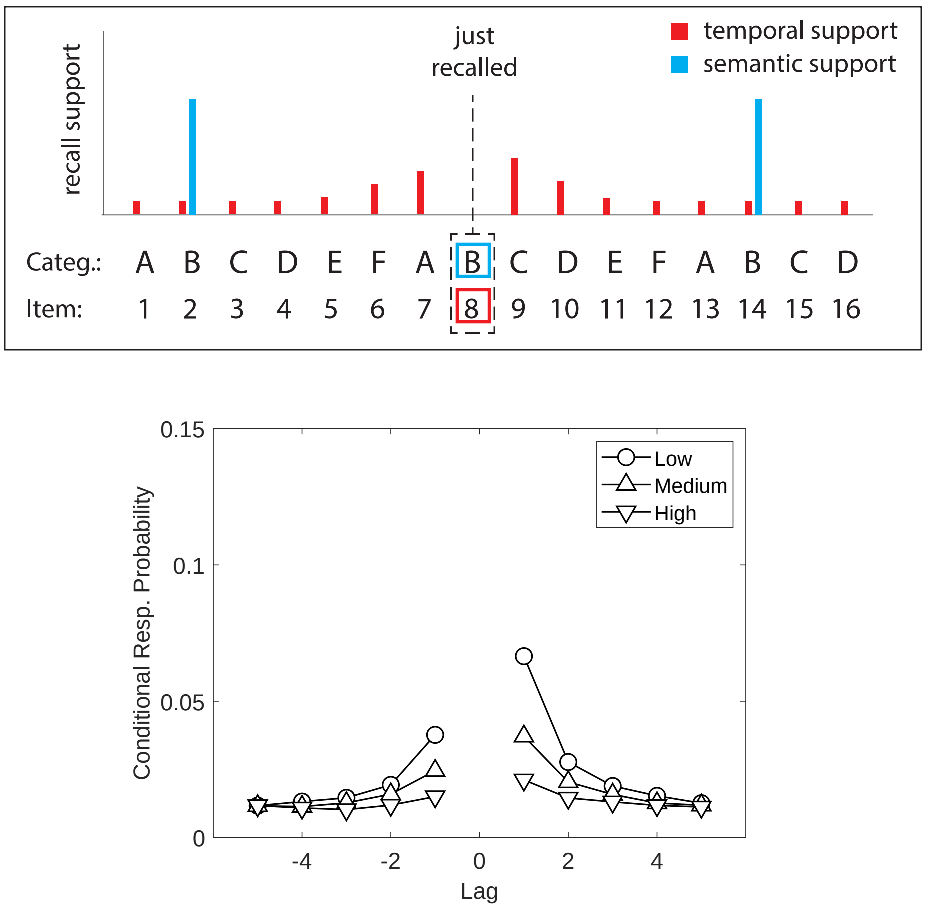

Figure 15: Simulation 2 schematic and results. (A) A schematic depiction of how widely spaced categorical associates can cause extreme temporal disruption. In the example, item 8 is recalled, and it is a member of category B. The just-recalled item's semantic associates are in relatively distant list positions, and this semantic support outshines the temporal support for item 8's neighbors. (B) As the strength of semantic associations is increased from Low (s=0.0) to Medium (s=0.5) to High (s=1.0), the temporal contiguity effect gets progressively weaker. See text and Appendix B for simulation details.

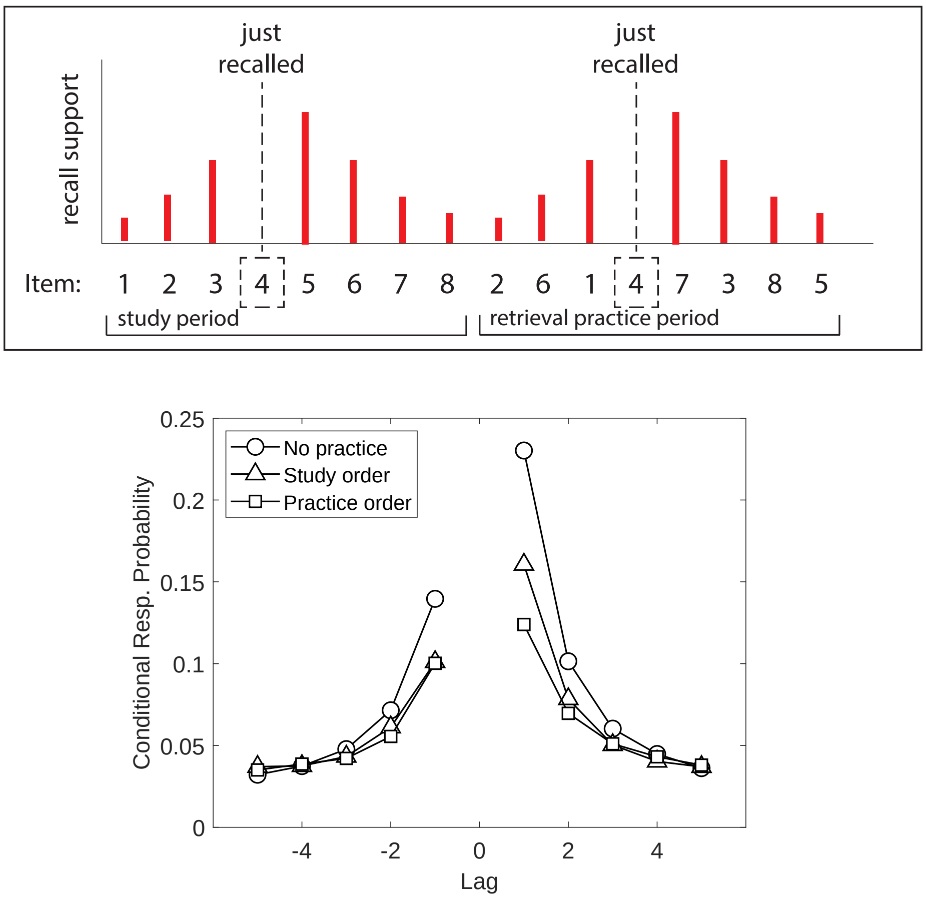

Figure 16: Simulation 3 schematic and results. (A) A schematic depiction of how a second exposure to the studied materials in a different order introduces competing temporal structures which can disrupt the temporal contiguity effect. Here, 8 items are studied in one order, and then in a shuffled order during a retrieval practice period. When item 4 is recalled, this provides temporal support for its original neighbors (items 3 and 5) as well as its neighbors during the retrieval practice period (items 1 and 7). (B) Simulation results with list length 30. The lag-CRP produced in a condition with no retrieval practice (No practice) provides a baseline. Temporal organization is weakened in a condition when the 30 items are practiced in a shuffled order. This allows us to produce two lag-CRP curves, one relative to the original study order (Study order) and one relative to the shuffled practice order (Practice order). See text for details.

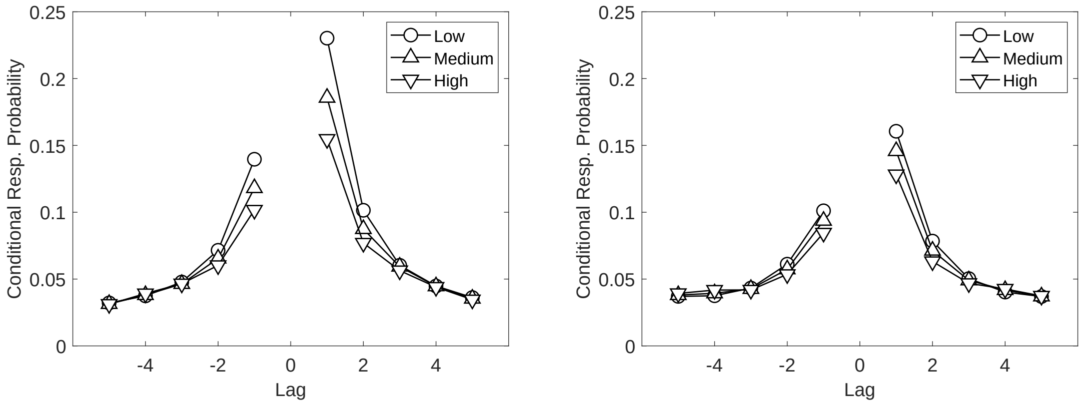

Figure 17: Simulation 4 results. (A) Simulation results with list length 30, with no practice period. As the strength of semantic associations increases, the temporal contiguity effect becomes progressively weaker. (B) Simulation results with list length 30, with a reordered practice period. Again, as the strength of semantic associations increases, the temporal contiguity effect becomes weaker for each level of semantic strength. Comparing the lag-CRP curves across the two panels, the interaction of temporal and semantic disruption can be seen.

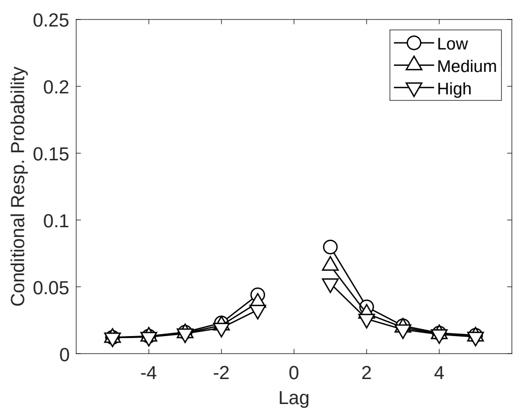

Figure 18: Simulation 4 results. Simulation results with list length 90, with a reordered practice period. The strength of semantic associations between categorical associates is varied from low, to medium, to high. While the temporal contiguity effect becomes very weak, it is never completely eliminated in these simulations.

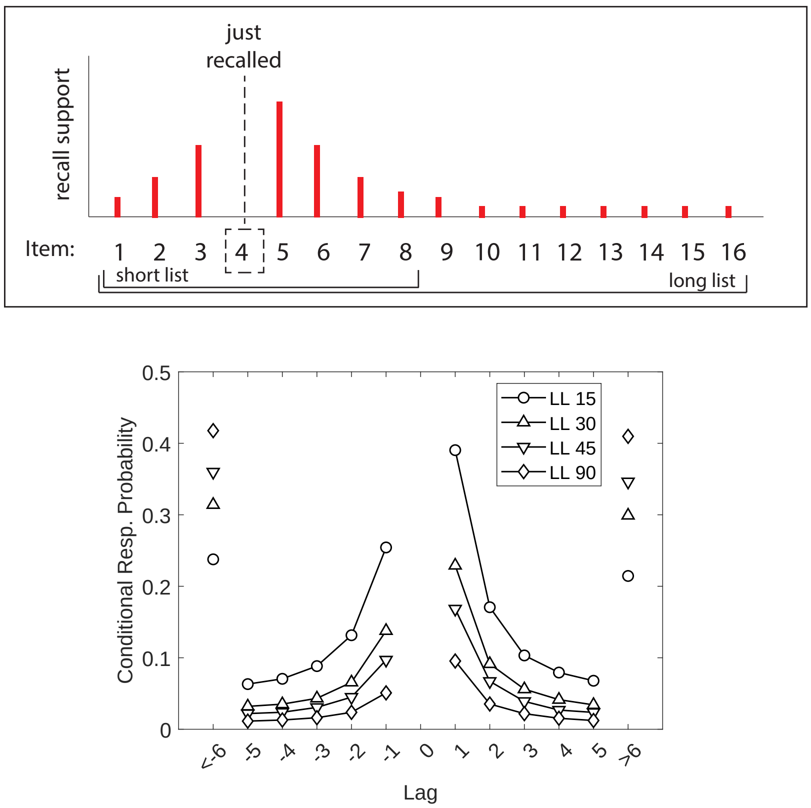

Figure 19: Simulation 5. (A) A schematic comparison of how increasing list length (from 8 to 16 items) can disrupt temporal organization. After item 4 is recalled, recall support for the surrounding study items is indicated in red. Because recall support asymptotes at a non-zero value with increasing temporal distance, increasing the list length steadily increases the likelihood of a long-range recall transition, even though support for any individual distant item is weak. (B) Simulation results showing how increasing list length affects simulated lag-CRP curves. As list length increases, the probability of nearby transitions drop steadily while the probability of long-distance transitions (lags < -6 and > 6) rise steadily.

General Discussion

The temporal contiguity effect is a form of temporal organization seen in the response sequences of the free-recall task. Successively recalled items tend to come from nearby list positions, and are more likely to be recalled in forward order. This effect is a relatively fine-grained measure of temporal organization, in that it shows how the local temporal neighborhood of a studied item influences the recall response sequence. Signs of coarser temporal organization can also be found in free recall, for example the tendency to group items on the basis of their list membership, and the ability to focus memory search on a particular target list.

Much of the prior work examining the temporal contiguity effect has focused on its ubiquity and robustness, as reviewed in the introduction. In six free-recall experiments we characterize the experimental factors that modulate, and in three cases eliminate, the temporal contiguity effect. No single factor is clearly responsible for the elimination of temporal organization in Experiments 2–4. Rather, each of these experiments seem to contain multiple factors acting against the behavioral expression of the temporal contiguity effect. These experiments also show that one can eliminate the temporal contiguity effect but still have good overall recall performance.

Retrieved-context models provide a framework for understanding the cognitive processes that give rise to temporal organization in free recall. Prior work with these models has focused on the ubiquity of the temporal contiguity effect, and the stability of the effect in the face of potentially disruptive experimental manipulations, like the presence of an effortful inter-item distraction task (Howard & Kahana, 2002a; Sederberg et al., 2008). We ran a set of simulations using the retrieved-context model CMR, examining whether the dramatic modulation of temporal organization observed in these experiments is consistent with the principles of retrieved-context models. We created a set of parameter estimates for individual participants in a large independent free-recall data set (Healey & Kahana, 2014). This allowed us to simulate the key experimental manipulations without adjusting the core parameters of the model. These simulations show that the dramatic modulation of temporal organization seen in these experiments follows lawful patterns captured by the model. Foremost amongst these disruptive factors are category structure, which pits semantic associations against temporal associations, and reordered practice, which introduces multiple conflicting sets of temporal associations into the recall competition. In the following sections, we review the major methodological manipulations explored in these experiments and simulations, and the theoretical conclusions we can draw from these results.

Semantic structure

Polyn et al. (2011) and Healey & Uitvlugt (2019) show that weak contiguity effects are a reliable consequence of using study materials with strong semantic structure. When semantic structure is placed in opposition to temporal structure, i.e., by placing semantically related items in non-neighboring list positions, this is highly disruptive to temporal organization. Our simulations (Figs. 14, 16) demonstrate that CMR can capture the disruption of the contiguity effect without altering the basic dynamics of the retrieved-context model. The outshining hypothesis (S. M. Smith & Vela, 2001), originally invoked to explain the influence of environmental context on memory performance, seems appropriate here. When an item is recalled, the model reactivates both temporal contextual information which would support retrieving a neighboring item, as well as semantic information which enhances the likelihood of retrieving a more temporally distant semantic associate. If the semantic support is strong enough, it may outshine the weaker temporal support.

The fact that recall intrusions are rare in these experiments, and recalled items overwhelmingly tend to come from the study list suggests that a coarse form of temporal targeting is still in operation in the absence of the more fine-grained contiguity effect. In other words, the participant isn't likely shifting to a purely semantic retrieval cue. A pure semantic cue would not honor the list structure at all, and would tend to give rise to recall of semantic associates that weren't on the study list. The false memory effect (as seen in the DRM paradigm) provides an example of how semantic cues, in the extreme, can break not only the fine-grained temporal organization of the contiguity effect, but even the coarse temporal organization that ensures responses come from the targeted list. One can create a list with multiple strong semantic associates of an unstudied critical item, and this reliably causes the unstudied item to intrude upon the recall sequence (Deese, 1959; Roediger & McDermott, 1995; Kimball et al., 2007).

List length

As the number of items on a list increases, temporal organization becomes progressively weaker (Ward et al., 2010; Healey et al., 2018). As such, the long list length of 90 for Experiments 2 and 3 likely contributed to the null contiguity effects observed in those experiments. In our simulations, increasing list length weakens the contiguity effect because there are more items competing to be recalled, and therefore more opportunities for long-distance recall transitions (Fig. 19A). When a long list length is combined with strong semantic structure, support for distant same-category items can outshine the support for nearby items from other categories.

Our simulations suggest that in the absence of strong semantic structure, simply increasing list length will not completely abolish the contiguity effect (Fig. 19B). Comparing the results of Expts. 6 & 4, we see that a list length of 40 can either produce a reliable contiguity effect or a null contiguity effect, depending on the other methodological details of the experiment. It is unclear whether an extreme list length with weak semantic structure would be enough to eventually abolish the temporal contiguity effect. This possibility could be tested in future work by examining temporal organization in extremely long lists, e.g., a few hundred items.

Retention interval

In the current studies, Experiment 6 demonstrates that the contiguity effect is stable when a distraction-filled delay period is extended from 3 to 15 minutes. Howard & Kahana (1999) showed that the temporal contiguity effect becomes weaker when a retention interval is extended from 0 to 10 seconds, but this difference may have been due to the fact that with 10 seconds delay there was a distraction task, but with 0 seconds delay there was no intervening task. Lohnas & Kahana (2014) found that the temporal contiguity effect was stable as a retention interval was lengthened from 8 to 16 seconds, and here both retention intervals contained the same distraction task. In our review of list-based free-recall studies examining temporal organization, we did not find any studies that used a retention interval exceeding 30 seconds, so Expt. 6 is a useful demonstration of the insensitivity of the temporal contiguity effect to longer retention intervals. In a naturalistic study of free recall where participants recalled the details of a museum tour, Diamond & Levine (2020) found that the contiguity effect retained its strength through much longer retention intervals of 2 days and 2 weeks.

The insensitivity of the temporal contiguity effect to retention interval duration is captured by retrieved-context theory. Sederberg et al. (2008) simulated delayed free recall with the retrieved-context model TCM-A. By their theory, the distraction-filled retention interval disrupts the context representation so it no longer resembles its end-of-list state. This allows the model to capture the fact that the recency effect is disrupted in delayed free recall. Once a studied item is successfully recalled, the contextual retrieval operation recovers the temporal context of the recalled item, which causes following responses to show intact temporal organization.

Retrieval practice and study order

In our two experiments involving retrieval practice (Expts. 2 & 4), we observed a null contiguity effect. It is possible that the second presentation of the items in a different order was sufficient to abolish the standard temporal contiguity effect. However, the lack of a contiguity effect in Experiment 3 (where there was no practice period) suggests that there are other contiguity-dampening effects simultaneously at play.

The relationship between retrieval practice and temporal organization is relevant to current theories regarding the nature of the memory benefit seen after retrieval practice (Roediger & Abel, 2022). The Episodic Context Account of retrieval practice proposes that when a participant practices retrieving an item, the original episodic context of the item is reactivated and strengthened, improving the item's memorability relative to an item that is simply restudied (Karpicke et al., 2014). Retrieved-context theory suggests that a stronger connection between an item and its episodic context will yield a stronger contiguity effect during free recall. As such, one might expect a stronger contiguity effect following recall of items that were practiced. However, our simulations point out a problem with assessing this hypothesis. The order in which items are practiced creates a second set of temporal associative structures in memory (Simulation 3). In our simulations, the two sets of temporal associations simultaneously influence memory search. If the items are practiced in the same order they are studied, this enhances temporal organization, but if they are practiced in a different order, it disrupts the contiguity effect. It may be difficult to disentangle a potential enhancement of temporal organization due to retrieval practice from the potential disruption of temporal organization due to the second presentation order. Our simulations suggest that the temporal contiguity effect for the study period may be reliably stronger than the one for the practice period. The model predicts that this difference is due to the retrieval of study-period episodic context.

Paired-associate structure of study items.

Our Expts. 2 & 4 used paired-associate study items, because this item structure facilitates retrieval practice tasks where one must prompt the participant to remember a particular study item. In Expt. 5 we observed reliable contiguity effects for the paired-associate conditions, and the contiguity effect for paired-associate study items was of similar strength to conditions with singleton study items. This suggests that the paired-associate structure of the study items on its own wasn't responsible for the null contiguity effects we observed.

Lehman & Malmberg (2013) and Cox & Criss (2020) found weak temporal organization during free recall after studying items presented with paired-associate structure. We discuss potentially important differences between our design and their design in the discussion section of Expt. 5. Generally speaking, retrieved-context theory needs development regarding the proper way to simulate paired-associate study materials. Cox & Criss (2020) describe an instance-based model that creates two memory traces for the items in a given pair. These traces include common associative features that functionally bind the two items together. These associative features are distinct from temporal contextual features that bind a given item to the broader list context and the local temporal neighborhood of the list. Adding these kinds of associative features to a retrieved-context model could allow the model to capture the potentially separable influences of within-pair and across-pair associations in free recall. These associative features could be represented by an independent sub-layer of context, similar to how temporal and source features were represented by independent contextual sub-layers in the original CMR paper (Polyn et al., 2009).

Orienting task and item characteristics

An orienting task is often used in free-recall tasks to ensure that participants pay attention to the study items. The nature of the orienting task can modulate the strength of the temporal contiguity effect, although usually not dramatically so. In intentional free-recall tasks participants know a memory test will follow the study list. In this case, if an orienting task is not specified, a robust temporal contiguity effect is usually seen. This is likely because participants engage in self-initiated strategic processing that supports temporal organization. When the orienting task explicitly requires evaluation of the semantic characteristics of an item, this tends to diminish the strength of temporal organization (Long & Kahana, 2017; Mundorf et al., 2022). This could be because the orienting task amplifies the influence of semantic associations, or it could be because it disrupts self-initiated processing of the items. Mundorf et al. (2022) proposed the disruption was due to the latter. They found that temporal organization was disrupted to a similar extent by an orthographic processing task, and that orthographic and semantic tasks produced similar levels of semantic organization. In incidental free-recall tasks, participants are not told about the upcoming memory test. As such, they may be less likely to engage in self-initiated strategic processing of the study items. A study by Nairne et al. (2017) reported null temporal contiguity effects in a number of incidental free-recall conditions. Other studies have confirmed that temporal organization is substantially weakened in incidental free recall relative to intentional free recall (Healey, 2018; Mundorf et al., 2021).

In Expt. 5 we found that an orienting task affected the strength of temporal organization for singly presented items but not for paired associates. Our modeling framework does not give insight into the reason for this difference, as the model implements encoding in an abstract way, without reference to the specifics of the orienting task. In recent work, we simulate orienting task more explicitly, and this work may eventually help us better understand how orienting task relates to temporal processing (Diachek et al., n.d.). Future model development can focus on bridging between retrieved-context models and other frameworks that simulate executive control processes more explicitly (e.g., Rougier & O'Reilly, 2002; Becker & Lim, 2003).

Finally, item characteristics may influence the strength of the temporal contiguity effect. It is possible that different types of study items are more or less appropriate for a participant's self-initiated processing, raising the possibility that item identity will interact with orienting task. McDaniel et al. (2011) examined the effect of orthographic distinctiveness of study items on temporal organization in free recall. Orthographically distinctive words having distinctive spelling or unusual letter combinations (e.g., lynx, methyl, knoll, calypso) were compared to common words. They found substantial contiguity effects in mixed lists containing both distinctive and common words, and in lists containing only common words. However, a null contiguity effect was observed in lists containing only orthographically distinct items (the *pure distinctive* condition). Bean et al. (2017) replicated the relevant conditions from McDaniel et al. (2011) with a larger sample size (338 participants as compared to 36 participants in the 2011 study). They found a reliable contiguity effect in the pure distinctive condition, but found that it was diminished relative to the other conditions. Work remains to determine how best to implement these experimental manipulations in a retrieved-context modeling framework.

Utility for other modeling frameworks

In this paper, we focus on retrieved-context models of free recall, as this class of models was explicitly developed to account for organizational phenomena in this task. However, we hope that the lessons of these simulations will prove useful for the development of other modeling frameworks to account for temporal organizational effects. Some of these models are very similar to CMR. The CRU model of serial recall tasks uses a context-based retrieval mechanism very similar to that of CMR (Logan, 2021). Howard et al. (2015) propose a generalization of retrieved-context theory in which the unitary context representation of TCM and CMR is replaced by a family of context representations, each with a different drift rate (Howard et al., 2015). To our knowledge, neither of these other frameworks directly simulates the semantic structure of study materials. As such, our simulations may provide useful guidance in developing those frameworks to simultaneously account for semantic and temporal organizational effects in memory search.

Other modeling frameworks have proposed alternative cognitive mechanisms that may underlie temporal organization in free recall. The Search of Associative Memory (SAM) model includes a short-term storage buffer which can be used strategically during the study period to maintain the representations of a subset of studied items (Shiffrin & Atkinson, 1969; Raaijmakers & Shiffrin, 1981). Neighboring items tend to occupy the buffer simultaneously, which allows the system to form direct item-to-item associations. When an item is retrieved during memory search, the item representation is used as part of the retrieval cue for the next retrieval attempt. The item-to-item associations support subsequent retrieval of a neighboring item. Several simulation studies have demonstrated that this mechanism can give rise to the contiguity effect in the SAM framework (Howard & Kahana, 1999; Davelaar et al., 2005; Sirotin et al., 2005). Lehman &

Malmberg (2013) propose an instance-based model of the free-recall task. When an item is studied, the stored memory trace contains both gradually changing contextual features, as well as features of other study items that were simultaneously maintained in a short-term storage buffer. Both of these mechanisms can give rise to the temporal contiguity effect. The empirical results presented here may be useful in developing these alternative frameworks to account for the modulation of temporal organization in free recall. The SAM model contains both episodic and semantic associative structures, and these can compete with one another during memory search. These competitive temporal-semantic dynamics would likely allow SAM to account for the effects of semantic structure on temporal organization observed here. SAM can also likely account for the disruption of temporal organization due to multiple presentation orders. As in CMR, a second presentation of the studied items would create a second set of inter-item associations that would compete with the first set during memory search.

Some other modeling frameworks capture temporal organization using mechanisms that are quite different from CMR. For example, Farrell (2012) proposes a connectionist model in which positional codes can be replayed in sequential order to give rise to +1 lag transitions. This model accounts for the decrease in the likelihood of +1 lag transitions with increasing list length, and similarly to CMR, this decrease is due to increased levels of competition from the other studied items. Farrell's model was not used to directly simulate the effects of semantic relatedness on memory search performance, though he acknowledges that a set of control elements could be embedded in the model to capture the category structure of studied items. Given that recall is a competitive process, these semantic control elements would likely compete with the positional elements to disrupt temporal organization. In the end it would be valuable to explore these proposals with actual simulations, to determine how well these alternative frameworks can account for the modulation of temporal organization in free recall.

Conclusions

The primary empirical contribution of this paper is to establish the methodological conditions necessary to eliminate and recover temporal organization in free recall. Expts. 2, 3, & 4 eliminate temporal organization, while Expts. 1, 5, & 6 preserve it. A comparison of these experiments suggests important roles for semantic structure, reordered practice, and list length. Orienting task modulates the strength of temporal organization inconsistently, and to a lesser degree. Substantial modulations of retention interval do not affect temporal organization. Presenting items singly, or as paired associates, does not substantially affect temporal organization, with some caveats noted above.

The primary theoretical contribution of the paper comes from the simulations, which help to clarify our proposals regarding the cognitive mechanisms underlying these dramatic shifts in the strength of temporal organization. CMR has distinct sets of associations containing semantic structure and temporal structure. When the structure of the study list puts these associations in opposition, by widely spacing categorical associates throughout the list, this is extremely disruptive to the behavioral manifestation of temporal organization. There was no need, in our simulations, to remove, diminish, or modify the processes creating temporal structure. The introduction of strong competitive semantic structure was enough to outshine the influence of temporal structure during free recall. CMR predicts that a reordered practice period will be disruptive to temporal organization, as this adds a second set of temporal associations that compete with the original set. CMR also predicts that increasing list length will progressively diminish temporal organization, as distant items have a small but non-zero likelihood of retrieval, which adds up as list length increases.

The empirical work suggests that either semantic structure on its own (Expt. 3) or reordered practice on its own (Expt. 4) can eliminate temporal organization. However, Simulation 3 suggests that without strong semantic structure, one should still observe reliable temporal organization relative to both the study order and practice order. The semantic structure of the list in Expt. 4 was not particularly strong, making the failure to observe temporal organization somewhat mysterious. The model suggests that of the three experiments that did not show temporal organization, Expt. 4 should be the closest to the boundary of methodological conditions that will and won't produce observable temporal organization in free recall. As such, it can serve as a beach head of sorts for follow-up investigations attempting to better characterize this boundary in methodological space. A first step could be to examine methodological manipulations that improve recall performance, e.g., by reducing list length, increasing presentation time, or providing participants with more practice on the task. This future work will help clarify the contributions of these interacting cognitive mechanisms to memory search behavior and temporal organization.

Acknowledgments

This work was supported by a grant from the National Science Foundation to SMP (#1756417). Special thanks to Megan Sumeracki for providing data from the Smith et al. (2013) experiments. Corresponding author: Sean Polyn.

Author Note

The data and study materials from these experiments will be made available through a project page hosted with the Open Science Foundation (URL will be inserted once the upload is completed). Analysis code and simulation code will be made available on GitHub (Organization: Vanderbilt University Computational Memory Lab, vucml), repository names will be inserted once they are made public. These studies were not preregistered.

	Temp. Org. Score [CI]	p(+1) [CI]
Experiment 1		
Same	0.55 [0.53–0.56]	0.23 [0.20–0.26]
Mixed	0.63 [0.61–0.65]	0.32 [0.29–0.35]
Toronto	0.61 [0.59–0.63]	0.33 [0.29–0.38]
Experiment 2		
Study order	0.50 [0.48–0.52]	0.02 [0.01–0.03]
Practice order	0.50 [0.48–0.53]	0.01 [0.01–0.02]
Smith et al. (2013)		
E4 Study order	0.52 [0.50–0.53]	0.03 [0.02–0.04]
E4 Practice order	0.53 [0.51–0.55]	0.02 [0.01–0.03]
E1 Study order	0.81 [0.77–0.84]	0.26 [0.20–0.32]
Experiment 3		
Generate	0.49 [0.47–0.51]	0.02 [0.01–0.03]
Experiment 4		
Study order, retrieve	0.50 [0.45–0.55]	0.03 [0.01–0.06]
Study order, retype	0.52 [0.47–0.58]	0.04 [0.01–0.09]
Pract. order, retrieve	0.48 [0.43–0.53]	0.03 [0.01–0.05]
Pract. order, retype	0.56 [0.51–0.61]	0.03 [0.01–0.06]
Experiment 5		
Single, pleas.	0.64 [0.60–0.68]	0.18 [0.12–0.25]
Single, retype	0.72 [0.66–0.76]	0.29 [0.21–0.38]
Paired, relat.	0.62 [0.56–0.67]	0.24 [0.15–0.34]
Paired, retype	0.63 [0.59–0.67]	0.21 [0.14–0.28]
Experiment 6		
LL 20, RI 3	0.59 [0.54–0.65]	0.22 [0.16–0.30]
LL 20, RI 15	0.63 [0.57–0.68]	0.19 [0.11–0.28]
LL 40, RI 3	0.61 [0.57–0.64]	0.12 [0.08–0.17]
LL 40, RI 15	0.60 [0.56–0.65]	0.12 [0.07–0.17]

Table 2: Comparison of the percentile-rank temporal organization score and the p(+1) input-output correspondence statistic across all experiments. Brackets indicate bootstrapped 95% confidence intervals. See text for details.

Appendix A: Organizational analysis methods and comparison

Temporal organizational analysis

All of the temporal organization analysis methods are carried out on the same underlying response sequences. To prepare these response sequences for analysis, each recall response is labeled either as valid (one of the studied items on the most recent list), or invalid (a repeated report of one of the studied items, a prior-list intrusion, an extra-list intrusion, or any non-recall-related utterance). Valid recall responses are labeled with the serial position of the remembered item. Then the recall sequence is converted into a series of transitions (*from* one response and *to* the next). Transitions between valid responses are labeled with the lag of the transition, which is the serial position of the *from* response minus the serial position of the *to* response. These valid transitions enter into the temporal organization analysis. Transitions between valid

	Power curve (a,b) (backward), (forward)
Experiment 1	
Same	(0.17, 0.22), (0.22, 0.47)
Mixed	(0.19, 0.64), (0.32, 1.02)
Toronto	(0.18, 0.77), (0.33, 1.41)
Experiment 5	
Single, pleas.	(0.15, 2.43), (0.18, 0.89)
Single, retype	(0.20, 1.16), (0.30, 1.04)
Paired, relat.	(0.10, 1.82), (0.25, 1.76)
Paired, retype	(0.10, 2.21), (0.21, 2.19)
Experiment 6	
LL 20, RI 3	(0.12, 0.36), (0.22, 1.51)
LL 20, RI 15	(0.12, 1.82), (0.21, 2.22)
LL 40, RI 3	(0.10, 1.91), (0.13, 1.10)
LL 40, RI 15	(0.08, 3.20), (0.12, 1.95)
	Linear (slope) (backward), (forward)
Experiment 2	
Study order	(0.001), (0.000)
Practice order	(-0.001), (0.000)
Smith et al. (2013)	
E4 Study order	(-0.001), (0.000)
E4 Practice order	(0.003), (0.002)
E1 Study order	(-0.04), (-0.06)
Experiment 3	
Generate	(-0.002), (-0.003)
Experiment 4	
Study order, retrieve	(-0.001), (-0.003)
Study order, retype	(0.008), (-0.005)
Pract. order, retrieve	(0.010), (0.001)
Pract. order, retype	(0.007), (0.000)

Table 3: Least squares power curve ($a|lag|^b$) fits are provided for Expts. 1, 5, & 6 and linear (slope$|lag|+$ intercept)) fits are provided for Expts. 2–4. Forward-going fits are to conditional response probabilities for lags +1 to +5, and backward-going fits are to lags -1 to -5. See text for details.

responses and invalid responses are left out of the analysis. In experiments involving retrieval practice, the items are studied in two distinct orders (the study order and the practice order). To calculate temporal organization for each phase of the experiment, the items are simply labeled with the serial positions from that phase and lag is calculated accordingly. Each analysis is carried out at the level of the individual participant, so across-participant statistics can be calculated.

Lag-based Conditional Response Probability (lag-CRP)

The lag-CRP analysis defines a set of lag bins, representing all possible lag values associated with valid recall transitions. These range from $-(\text{list length} - 1)$ to $+(\text{list length} - 1)$, skipping lag 0 (as repeated responses are labeled invalid). Each recall transition is associated with an observed lag, and a set of possible lags to any study items that have not yet been reported. To determine the conditional response probability associated with each lag bin, one counts up the number of times a transition of a given lag was observed (the numerator), and divides this by the number of times a transition of that lag was possible (the denominator), where each count is aggregated across all valid transitions for that participant.

One of the advantages of the lag-CRP analysis is the ability to examine the pattern of probabilities associated with different lag transitions. This richness is theoretically beneficial, but comes with an associated challenge. Often, hypothesis testing requires a unidimensional summary measure specifying the strength of temporal organization in a given data set, task condition, or group of participants. Over the years, a number of different attempts have been made to do this. Howard & Kahana (1999) calculated the correlation between the integer lag values associated with each transition bin, and the probabilities associated with those bins. The correlation could be run on subsets of the lag bins, e.g., lags +2 through +6 to show that there was a reliable advantage for shorter-lag transitions even when transitions to immediate neighbors were excluded. Howard (2004) summarized the contiguity effect with two ratios. For the forward direction, this was a ratio of the probability of +1 lag transitions and +2 lag transitions (and -1 and -2 lag transitions for the backward direction). A few studies fit a power function ($a|lag|^{-b}$) to the forward and backward legs of the lag-CRP curve, reporting the b parameter as a measure of the sharpness of the temporal contiguity effect (Kahana et al., 2002; Klein et al., 2005).

The p(+1) statistic reported in Table 2 is similar to the classic input-output correspondence statistic reviewed in the introduction. The p(+1) statistic reports the value from the lag +1 bin of the lag-CRP analysis. The main difference between this version and the classic version is in how each statistic is normalized. Those older statistics calculated the proportion of all recall transitions with a +1 lag. Any p(+1) statistics reported in this paper uses the lag-CRP-style normalization. In practice, this is a relatively minor correction to the denominator of the probability calculation, relative to the classic version.

We note in Expt. 2 that the p(+1) statistic is not well suited to establishing chance-level temporal organization (unlike the percentile-rank score described below). This is because the p(+1) statistic involves a single lag bin. To establish chance-level performance it is necessary to show that short-lag transitions are not more likely than longer-lag transitions.

We include some analyses which involve calculating the likelihood of a recall transition encompassing a range of lag values (Simulation 5). This involves constructing a bin capturing all long-distance lag transitions greater than a certain value (as depicted in Figure 19B, lags < -6 and lags > 6). For a given recall transition, if the observed lag falls within one of these aggregate bins, this increased the count of the numerator by one. Similarly, for any recall transition where a transition to that bin was possible, the denominator was increased by one.

Within- and between-category lag-CRP analyses

Generally, lag is calculated with regard to the number of valid study items separating the *from* and *to* items of a given recall transition. For the standard lag-CRP, all study items are valid, so the lag is simply the difference of the two items' serial positions, as described above. However, for a subset of analyses in this paper, we calculate a lag-CRP restricted to either within-category transitions, or between-category transitions. For the within-category analysis, only study items from the same category as the just-recalled item are considered valid. For the between-category analysis, only study items from a different category as the just-recalled item are considered valid.

For example, consider the study sequence A1 B2 B3 B4 A5 B6, where the letter indicates category and the number indicates serial position. A recall transition from A1 to A5 would normally be assigned a lag of +4, but for the within-category analysis, it would be assigned a lag of +1, as the other-category B items are ignored. A recall transition from A1 to B6 would normally be assigned a lag of +5, but for the between-category analysis it would be assigned a lag of +4, as the same-category A items are ignored.

Percentile-rank temporal organization score.

Polyn et al. (2009) introduced an alternative measure of temporal organization, the percentile-rank temporal organization score (also called *temporal factor* in some papers), to characterize the strength of temporal organization with a single number.

Each recall transition is assigned a percentile rank by comparing its lag to relative to the set of lags associated with all possible transitions at that moment. For each recall transition, the set of possible lags (determined as described above for the lag-CRP analyses) are rank ordered, based on their absolute value. Each lag is assigned a percentile rank (stepping from 1.0, indicating the smallest/nearest lag value, to 0.0 indicating the largest/most distant lag value). Ties are resolved by assigning each tied lag value (e.g., -1 and +1) the mean percentile rank of the two tied values. The percentile rank assigned to the observed lag is then pulled from this rank ordering. The final score is simply the average rank of all valid recall transitions.

An average score of 0.5 indicates chance-level organization with respect to lag, and 1.0 indicates perfect temporal organization. If a recall sequence is dominated by short-lag transitions, the average rank will be close to 1.0, indicating strong temporal organization. If items are recalled in a random order, without regard to lag, the rank will be close to the chance value of 0.5.

The current set of experiments provide a useful opportunity to examine the relative sensitivity and idiosyncrasies of these different measures of temporal organization. In Expt. 6 we noted a discrepancy between the temporal organization score and the p(+1) input-output correspondence measure. The p(+1) measure indicated that temporal organization was weaker with increased list length, but the percentile-rank temporal organization score indicated that temporal organization was unaffected by increased list length.

To explain this discrepancy, we note that the percentile-rank measure is designed to detect temporal organization relative to chance-level performance where items are chosen randomly from the list without regard to serial position. Each recall transition is assigned a rank value based on its lag, and this rank value is affected by list length. On longer lists, a larger set of lag values enter the percentile ranking procedure for any given recall transition. This causes the rank values assigned to short-lag transitions to increase (relative to a transition of the same lag on a shorter list). This is a necessary part of the algorithm: As list length increases, the average lag between two items selected randomly from the list will necessarily increase. As such, a short-lag transition becomes progressively more notable in terms of its deviation from chance performance.

Semantic organizational analysis.

The Adjusted Ratio of Clustering (ARC) score quantifies semantic organization related to the category structure of the studied material (Roenker et al., 1971). Each item is assigned a category label, and the analysis counts up the number of recall transitions associated with a category repetition. This is divided by the maximum possible number of category repetitions, and normalized by the expected (chance) number of category repetitions. For the full mathematical details, see Roenker et al. (1971), or consult the analysis code associated with the software resources described below. ARC scores can range in value from -1.0 to 1.0, where 0 indicates chance-level category clustering and 1.0 indicates perfect clustering.

Software resources

A few resources exist to help an interested researcher calculate temporal organization in free-recall experiments. The Episodic Memory Behavioral Analysis in MATLAB (EMBAM) toolbox (hosted at github.com by VUCML, project name: EMBAM) provides a set of MATLAB functions to carry out a variety of analyses on free-recall data. The Psifr (Psychological free recall) analysis toolbox (hosted at github.com by mortonne, project name: psifr) provides a set of Python functions to carry out similar free-recall analyses. Both toolboxes contain functions capable of carrying out the analyses described in this report, and many more.

	Description	Mean (st. dev.)
β_{enc}	Encoding contextual integration rate	0.881 (0.07)
β_{rec}	Recall contextual integration rate	0.886 (0.114)
γ	M^{FC} strength of expt. assoc.	0.184 (0.229)
α	M^{CF} baseline support during recall	0.526 (0.321)
δ	M^{CF} strength of pre-expt. assoc.	14.708 (19.669)
ϕ_s	Primacy scaling	18.843 (20.458)
ϕ_d	Primacy decay	32.636 (30.351)
β_{start}	Reactivation of start-of-list context	0.169 (0.23)
θ_s	Recall termination scaling	0.007 (0.01)
θ_r	Recall termination growth rate	0.407 (0.127)
τ	Recall support softmax-like scaling	5.105 (2.942)

Table 4: Key parameters of the CMR model with a short functional description of each. These are accompanied by the mean and standard deviation of the best-fitting values across 126 simulated participants.

Appendix B: Description of the CMR model

The Context Maintenance and Retrieval (CMR) retrieved-context model is a simplified connectionist model with two vector space representational layers, one representing the features of studied items (F) and another representing the contextual features retrieved when items are presented or remembered (C). These two layers influence one another via two associative matrices M^{FC} (item-to-context) and M^{CF} (context-to-item). Each study item is assigned an orthonormal vector representation f_i on F such that each unit on F corresponds to an item and is activated when that item is presented. The contextual layer C is initialized with a representational state c_0 orthogonal to the contextual features associated with the studied items.

We distinguish between two kinds of associations, each of which represents bindings between item features and contextual features. Pre-experimental associations are created when the model is initialized and represent associations formed during prior experience. Experimental associations are formed during the study period. For the item-to-context matrix M^{FC}, the diagonal elements represent pre-experimental associations, and are initialized as $1-\gamma$. As such, the initialized M^{FC} acts as a scaled identity matrix that links each item unit with a corresponding context unit. The learning rate of experimental associations on M^{FC} is γ. As such, γ controls the relative strength of pre-experimental and experimental associations on M^{FC}.

For the context-to-item matrix M^{CF}, the diagonal elements represent one kind of pre-experimental association, which are initialized as δ, and the off-diagonal elements represent another kind of pre-experimental association which are initialized as α. The α parameter provides a baseline level of support for all items in the recall competition. The learning rate of experimental associations on M^{CF} is influenced by the primacy associative gradient mechanism below, which gives early items a boost but decays to a learning rate of 1.0.

As such, δ influences the relative strength of pre-experimental and experimental associations on M^{CF}.

The model has two modes of operation, predictive and generative (Polyn, 2022). The predictive mode is used during the maximum-likelihood parameter fitting process described below. In both modes, each recall event involves calculating the probability of recalling each studied item (Eq. 10). These probabilities are stored in a vector $P(i)$, where i indexes study items. For each recall attempt, these values determine the probability of each potential recall response and recall termination.

In the predictive mode, the model assigns a likelihood to each recall response by selecting the $P(i)$ value corresponding to each observed response. These $P(i)$ values are converted to log-likelihoods and summed up to obtain the overall log-likelihood score for a particular participant, for a particular parameter set. In this mode, the data determines which item wins the recall competition, and the model simply logs the predicted likelihood of that event. The generative mode is used to create simulated recall sequences for Simulations 1–5. In the generative mode, the $P(i)$ values are used to guide a stochastic selection process.

The study period

When an item i is studied, its corresponding feature representation f_i is activated on F. The item representation is projected through M^{FC} with a matrix multiplication to provide input to the context layer:

$$c_i^{IN} = M^{FC} f_i \tag{1}$$

which is normalized to have length 1. The contextual state c_i is then updated according to:

$$c_i = \rho_i c_{i-1} + \beta c_i^{IN} \tag{2}$$

with β set to β_{enc}. The parameter ρ is set according to:

$$\rho_i = \sqrt{1 + \beta^2 \left[\left(c_{i-1} \cdot c_i^{IN} \right)^2 - 1 \right]} - \beta \left(c_{i-1} \cdot c_i^{IN} \right) \tag{3}$$

which enforces the length of c_i to 1. The item representation and context representation are then associated with one another via an outer-product-based Hebbian learning process which forms experimental associations according to:

$$\Delta M_{exp}^{FC} = \gamma c_i f_i' \tag{4}$$

and

$$\Delta M^{CF}_{exp} = \phi_i f_i c'_i \tag{5}$$

where γ controls the strength of experimental associations on M^{FC}_{exp} and ϕ_i enforces a primacy effect by scaling the strength of associative connections on M^{CF}_{exp} based on the serial position of the studied item, according to:

$$\phi_i = \phi_s e^{-\phi_d (i-1)} + 1 \tag{6}$$

This function decays as serial position increases, with ϕ_s modulating the strength of primacy and ϕ_d modulating the rate of decay.

The free-recall period

Between the study period and free-recall period, we allow the model to partially retrieve the start-of-list contextual state c_0. This helps the model capture the primacy effect. In the following equations i indexes study items by their serial position, and j indexes the output position of successive recall attempts.

$$c_{start} = \rho_j c_j + \beta_{start} c_0 \tag{7}$$

with ρ calculated as in Eq. 3. With each recall attempt j, the current state of context is used as a cue to attempt the retrieval of some studied item. The semantic identity of the previously recalled item also influences the recall competition, as in the item-semantics model CMR variant described by Morton & Polyn (2016). Semantic associations are embedded in an associative matrix M^{FF} which is multiplied by a semantic scaling parameter s, as shown below. The associative strengths in M^{FF} reflect the category structure of studied items, as described below in *Experiment simulation details*. An activation a is calculated for each item according to:

$$a = M^{CF} c_j + s M^{FF} f_{j-1}. \tag{8}$$

For the first recall attempt, no items have been recalled yet, so the f_{j-1} term is treated as a vector of zeros, which transiently inactivates the second term of Eq. 8. Each item is assigned a minimum activation of 10^{-7}.

To determine the probability of a given item being recalled, we first calculate the probability of recall termination, an event that returns no item and ends memory search. This termination probability P_{stop} varies as a function of output position j:

$$P_{stop}(j) = \theta_s e^{j\theta_r}. \tag{9}$$

Parameters θ_s and θ_r control the scaling and rate of increase of the exponential recall termination function. The probability of recalling a given item $P(i)$ is calculated as:

$$P(i) = (1 - P_{stop}(j)) \frac{a_i^\tau}{\sum_k^N a_k^\tau} \tag{10}$$

where τ is a scaling parameter that can accentuate differences in activation amongst items, as in a softmax rule. These probabilities are used differently depending on whether the model is in predictive or generative mode, as described above. When an item is recalled, that item's representation is reactivated on F. This prompts the retrieval of its associated contextual state via Eq. 1. This retrieved contextual information is integrated into the contextual representation via Eq. 2, with β set to β_{rec}. The recall cycle (Eqs. 8–10) iterates until recall terminates.

Experiment simulation details

The CMR model was fit to individual participant data reported by Healey & Kahana (2014). For each of the 126 participants, we used a differential evolution optimization procedure to find best-fitting maximum-likelihood parameter settings (Storn & Price, 1997). This process uses the predictive mode of the model, as described above. We then used these best-fit parameters to simulate the various experimental manipulations examined in Simulations 1–5, using the generative mode of the model.

Simulation 1. Categorical associates in neighboring positions. The category identity of study items was used to set same-category and between-category associative strength in M^{FF}. Same-category associations were set to 1.0, and between-category associations were set to 0.0. In Eq. 8 these associations are scaled by s, which was set to 0, 0.5, and 1.0 for the low, medium, and high levels of semantic association described in the paper. *Simulation 2: Categorical associates in distant positions.* The implementation here is the same as Simulation 1, with the only differences being that multiple categories of items appear on the same list, and in widely spaced positions. *Simulation 3: Reordered practice.* The practice period is simulated identically to the study period, using the dynamics described above. When an item is presented a second time, Eq. 1 causes the contextual input to include influences of both pre-experimental associations and the Hebbian associations formed during the study period. In these simulations, different kinds of retrieval practice are simulated identically to study events. The potential for different cognitive operations to be engaged during different kinds of retrieval practice is a potentially fruitful avenue for future work. *Simulation 4: Combination of categorical associates and reordered practice.* This simulation uses identical methods as Simulation 3, but introduces categorical structure as in Simulation 2, and manipulates the semantic scale parameter to low, medium, and high values as above. *Simulation 5: Modifying list length.* This simulation does not add any new simulation methods, we simply modify the simulated list length and examine the effect on recall dynamics.

References

Anderson, J. A., Silverstein, J. W., Ritz, S. A., & Jones, R. S. (1977). Distinctive features, categorical perception, and probability learning: Some applications of a neural model. *Psychological Review*, *84*, 413–451.

Asch, S. E., & Ebenholtz, S. M. (1962). The principle of associative symmetry. *Proceedings of the American Philosophical Society*, *106*, 135–163.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut Category Norms. *Journal of Experimental Psychology Monograph*, *80*(3 Pt. 2), 1–46.

Bean, R. R., Bernhardt, L. V., Brimmeier, L. L., Smith, R. L., Zerka, J. N., Uitvlugt, M. G., & Healey, M. K. (2017). The Temporal Contiguity Effect is modulated, but not eliminated, by orthographic distinctiveness. *Mid-Michigan Symposium for Undergraduate Research Experiences*.

Becker, S., & Lim, J. (2003). A computational model of prefrontal control in free recall: Strategic memory use in the California verbal learning task. *Journal of Cognitive Neuroscience*, *15*, 821–832.

Bousfield, W., Puff, C., & Cowan, T. (1964). The development of constancies in sequential organization during repeated free recall. *Journal of Verbal Learning and Verbal Behavior*, *3*(6), 489–495.

Bousfield, W. A. (1953). The occurrence of clustering in the recall of randomly arranged associates. *Journal of General Psychology*, *49*(2), 229–240.

Bower, G. H. (1970). Organizational factors in memory. *Cognitive Psychology*, *1*(1), 18–46.

Brown, G. D., Neath, I., & Chater, N. (2007). A Temporal Ratio Model of Memory. *Psychological Review*, *114*(3), 539–576.

Campbell, K. L., Trelle, A., & Hasher, L. (2014). Hyper-binding across time: Age differences in the effect of temporal proximity on paired-associate learning. *Journal of Experimental Psycholgy: Learning, Memory, & Cognition*, *40*(1), 293–299.

Carpenter, S. K., & DeLosh, E. L. (2006). Impoverished cue support enhances subsequent retention: Support for the elaborative retrieval explanation of the testing effect. *Memory & cognition*, *34*(2), 268–276.

Cox, G. E., & Criss, A. H. (2020). Similarity leads to correlated processing: A dynamic model of encoding and recognition of episodic associations. *Psychological Review*, *127*(5), 792.

Crowder, R. G. (1976). *Principles of learning and memory*. Hillsdale, NJ: Lawrence Erlbaum and Associates.

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*(1), 3–42.

Davis, O. C., Geller, A. S., Rizzuto, D. S., & Kahana, M. J. (2008). Temporal associative processes revealed by intrusions in paired-associate recall. *Psychonomic Bulletin & Review*, *15*(1), 64–69.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, *58*(1), 17–22.

DeLosh, E. L., & McDaniel, M. A. (1996). The role of order information in free recall: Application to the word-frequency effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *22*(5), 1136.

Diachek, E., Brown-Schmidt, S., & Polyn, S. M. (n.d.). Items outperform adjectives in a computational model of binary semantic classification. *Submitted*.

Diamond, N. B., & Levine, B. (2020). Linking detail to temporal structure in naturalistic-event recall. *Psychological Science*, *31*(12), 1557–1572.

Dong, T. (1972). Cued partial recall of categorized words. *Journal of Experimental Psychology*, *93*(1), 123.

Farrell, S. (2012). Temporal clustering and sequencing in working memory and episodic memory. *Psychological Review*, *119*(2), 223–271.

Friendly, M. (1979). Methods for finding graphic representations of associative memory structures. In C. R. Puff (Ed.), *Memory organization and structure* (pp. 85–129). New York: Academic Press.

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982). The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods and Instrumentation*, *14*, 375–399.

Glenberg, A. M., Bradley, M. M., Kraus, T. A., & Renzaglia, G. J. (1983). Studies of the long-term recency effect: Support for a contextually guided retrieval theory. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *12*, 413–418.

Golomb, J. D., Peelle, J. E., Addis, K. M., Kahana, M. J., & Wingfield, A. (2008). Effects of adult aging on utilization of temporal and semantic associations during free and serial recall. *Memory & Cognition*, *36*(5), 947–956.

Healey, M. K. (2018). Temporal contiguity in incidentally encoded memories. *Journal of Memory and Language*, *102*, 28–40.

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.

Healey, M. K., Long, N. M., & Kahana, M. J. (2018). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, *26*(3), 699–720.

Healey, M. K., & Uitvlugt, M. G. (2019). The role of control processes in temporal and semantic contiguity. *Memory & Cognition*, *47*(4), 719–737.

Hong, M., Polyn, S., & Fazio, L. (2019). Examining the episodic context account: Does retrieval practice enhance memory for context? *Cognitive Research: Principles and Implications*, *4*(1), 1–9.

Howard, M. W. (2004). Scaling Behavior in the temporal context model. *Journal of Mathematical Psychology*, *48*, 230–238.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *25*(4), 923–941.

Howard, M. W., & Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299.

Howard, M. W., & Kahana, M. J. (2002b). When does semantic similarity help episodic retrieval? *Journal of Memory and Language*, *46*, 85–98.

Howard, M. W., Kahana, M. J., & Wingfield, A. (2006). Aging and contextual binding: Modeling recency and lag-recency effects with the temporal context model. *Psychonomic Bulletin & Review*, *13*(3), 439–445.

Howard, M. W., Shankar, K. H., Aue, W. R., & Criss, A. H. (2015). A distributed representation of internal time. *Psychological Review*, *122*(1), 24–53. doi: 10.1037/a0037840

Jacoby, L. L. (1996). Dissociating automatic and consciously controlled effects of study/test compatibility. *Journal of Memory and Language*, *35*(1), 32–52.

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*(1), 103–109.

Kahana, M. J. (2000). Contingency analyses of memory. In E. Tulving & F. I. M. Craik (Eds.), *Oxford handbook of human memory* (pp. 323–384). England: Oxford Press.

Kahana, M. J. (2012). *Foundations of Human Memory* (1st ed.). New York, NY: Oxford University Press.

Kahana, M. J. (2020). Computational models of memory search. *Annual review of psychology*, *71*, 107–138.

Kahana, M. J., Howard, M. W., & Polyn, S. M. (2008). Associative processes in episodic memory. In H. L. Roediger III & J. Byrne (Eds.), *Cognitive Psychology of Memory, Vol. 2 of Learning and Memory: A Comprehensive Reference* (pp. 476–490). Oxford: Elsevier.

Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag recency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *28*(3), 530–540.

Karpicke, J. D., Lehman, M., & Aue, W. R. (2014). Retrieval-Based Learning: An Episodic Context Account. In *Psychology of learning and motivation - advances in research and theory* (Vol. 61, pp. 237–284). Academic Press Inc.

Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM Model of False Recall. *Psychological Review*, *114*(4), 954–993.

Kintsch, W. (1970). *Learning, memory, and conceptual processes*. New York, N.Y.:John Wiley & Sons.

Klein, K. A., Addis, K. M., & Kahana, M. J. (2005). A comparative analysis of serial and free recall. *Memory & Cognition*, *33*(5), 833–839.

Koh, S. D., & Kayton, L. (1974). Memorization of ”unrelated” word strings by young nonpsychotic schizophrenics. *Journal of Abnormal Psychology*, *83*(1), 14–22.

Koh, S. D., Vernon, M., & Bailey, W. (1971). Free-recall learning of word lists by prelingual deaf subjects. *Journal of Verbal Learning and Verbal Behavior*, *10*(5), 542–547.

Kragel, J. E., Morton, N. W., & Polyn, S. M. (2015). Neural activity in the medial temporal lobe reveals the fidelity of mental time travel. *The Journal of Neuroscience*, *35*(7), 2914–2926.

Landauer, T. K., & Dumais, S. T. (1997). Solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psychological Review*, *104*, 211–240.

Lehman, M., & Malmberg, K. J. (2013). A buffer model of memory encoding and temporal correlations in retrieval. *Psychological Review*, *120*(1), 155–189. doi: 10.1037/a0030851

Logan, G. D. (2021). Serial order in perception, memory, and action. *Psychological Review*, *128*(1), 1.

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

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015, 4). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review*, *122*(2), 337–363. doi: 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.

Madigan, S. A. (1969). Intraserial repetition and coding processes in free recall. *Journal of Verbal Learning and Verbal Behavior*, *8*(6), 828–835.

Mandler, G., & Dean, P. J. (1969). Seriation: Development of serial order in free recall. *Journal of Experimental Psychology*, *81*(2), 207–215.

Mandler, G., Pearlstone, Z., & Koopmans, H. S. (1969). Effects of organization and semantic similarity on recall and recognition. *Journal of Verbal Learning and Verbal Behavior*, *8*(3), 410–423.

McDaniel, M. A., Cahill, M., Bugg, J. M., & Meadow, N. G. (2011). Dissociative effects of orthographic distinctiveness in pure and mixed lists: An item-order account. *Memory and Cognition*, *39*(7), 1162–1173.

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient Estimation of Word Representations in Vector Space. *arXiv preprint arXiv:1301.3781*.

Morton, N. W., & Polyn, S. M. (2016). A predictive framework for evaluating models of semantic organization in free recall. *Journal of Memory and Language*, *86*, 119–140. doi: 10.1016/j.jml.2015.10.002

Morton, N. W., & Polyn, S. M. (2017). Beta-band activity represents the recent past during episodic encoding. *NeuroImage*, *147*, 692–702. doi: 10.1016/j.neuroimage.2016.12.049

Mundorf, A., Lazarus, L. T., Uitvlugt, 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*(8), 1264–1287.

Mundorf, A., Uitvlugt, M. G., & Healey, M. K. (2022). Does depth of processing affect temporal contiguity? *Psychonomic Bulletin & Review*, 1–11.

Murdock, B. B. (1962). Direction of recall in short term memory. *Journal of Verbal Learning and Verbal Behavior*, *1*(5), 119–124.

Murty, V. P., McKinney, R. A., Dubrow, S., Jalbrzikowski, M., Haas, G. L., & Luna, B. (2018). Differential patterns of contextual organization of memory in first-episode psychosis. *Schizophrenia*, *4*(1), 3. doi: 10.1038/s41537-018-0046-8

Nairne, J. S., Cogdill, M., & Lehman, M. (2017, 4). Adaptive memory: Temporal, semantic, and rating-based clustering following survival processing. *Journal of Memory and Language*, *93*, 304–314. doi: 10.1016/j.jml.2016.10.009

Nairne, J. S., Riegler, G. L., & Serra, M. (1991). Dissociative effects of generation on item and order retention. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*(4), 702.

Osth, A. F., & Fox, J. (2019). Are associations formed across pairs? A test of learning by temporal contiguity in associative recognition. *Psychonomic Bulletin & Review*, *26*(5), 1650–1656.

Palombo, D. J., Di Lascio, J. M., Howard, M. W., & Verfaellie, M. (2018, 9). Medial temporal lobe amnesia is associated with a deficit in recovering temporal context. *Journal of Cognitive Neuroscience*, *22*(7), 1–13.

Pennington, J., Socher, R., & Manning, C. D. (2014). Glove: Global vectors for word representation. In *Proceedings of the empirical methods in natural language processing* (Vol. 12, pp. 1532–1543).

Polyn, S. M. (2022). Assessing neurocognitive hypotheses in a likelihood-based model of the free-recall task. In B. U. Forstmann & B. Turner (Eds.), *Introduction to Model-Based Cognitive Neuroscience.* New York: Springer.

Polyn, S. M., Erlikhman, G., & Kahana, M. J. (2011). Semantic cuing and the scale-insensitivity of recency and contiguity. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *27*(3), 766–775.

Polyn, S. M., McCluey, J. D., Morton, N. W., Woolard, A. A., Luksik, A. S., & Heckers, S. (2015). Temporal context and the organizational impairment of memory search in schizophrenia. *Cognitive Neuropsychiatry*, *20*(4), 296–310.

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009, 1). A Context Maintenance and Retrieval Model of Organizational Processes in Free Recall. *Psychological Review*, *116*(1), 129–156. doi: 10.1037/a0014420

Postman, L. (1971). Organization and interference. *Psychological Review*, *78*(4), 290–302. doi: 10.1037/h0031031

Postman, L. (1972). A pragmatic view of organization theory. In *Organization of memory.* Academic Press.

Postman, L., & Hasher, L. (1972). Conditions of proactive inhibition in free recall. *Journal of Experimental Psychology*, *92*, 276–284.

Puff, C. R. (1979). Memory Organization Research and Theory: The State of the Art. In C. R. Puff (Ed.), *Memory organization and structure* (pp. 3–17). New York: Academic Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review*, *88*(2), 93–134.

Roediger, H. L., & Abel, M. (2022). The double-edged sword of memory retrieval. *Nature Reviews Psychology*, 1–13.

Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *21*(4), 803–814.

Roenker, D. L., Thompson, C. P., & Brown, S. C. (1971). Comparison of measures for the estimation of clustering in free recall. *Psychological Bulletin*, *76*(1), 45–48.

Rougier, N. P., & O'Reilly, R. C. (2002). Learning Representations in a Gated Prefrontal Cortex Model of Dynamic Task Switching. *Cognitive Science*, *26*, 503–520.

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a meta-analytic review of the testing effect. *Psychological bulletin*, *140*(6), 1432–1463. doi: 10.1037/a0037559

Sahakyan, L., & Kwapil, T. R. (2018). Moving beyond summary scores: Decomposing free recall performance to understand episodic memory deficits in schizotypy. *Journal of Experimental Psychology: General*, *147*(12), 1919.

Schacter, D. L. (1987). Memory, amnesia, and frontal lobe dysfunction. *Psychobiology*, *15*, 21–36.

Sederberg, P. B., Gershman, S. J., Polyn, S. M., & Norman, K. A. (2011, 6). Human memory consolidation can be explained using the Temporal Context Model. *Psychonomic Bulletin & Review*, *18*(3), 455–468. doi: 10.3758/s13423-011-0086-9

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), 893–912.

Sederberg, P. B., Miller, J. F., Howard, M. W., & Kahana, M. J. (2010). The temporal contiguity effect predicts episodic memory performance. *Memory & Cognition*, *38*(6), 689–699.

Serra, M., & Nairne, J. S. (1993). Design controversies and the generation effect: Support for an item-order hypothesis. *Memory & Cognition*, *21*(1), 34–40.

Shiffrin, R. M., & Atkinson, R. C. (1969). Storage and retrieval processes in long-term memory. *Psychological Review*, *76*(2), 179–193.

Sirotin, Y. B., Kimball, D. R., & Kahana, M. J. (2005). Going beyond a single list: Modeling the effects of prior experience on episodic free recall. *Psychonomic Bulletin & Review*, *12*(5), 787–805.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*, 592–604.

Smith, M. A., Roediger, H. L., & Karpicke, J. D. (2013). Covert retrieval practice benefits retention as much as overt retrieval practice. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *39*(6), 1712–1725.

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: a review and meta-analysis. *Psychonomic Bulletin and Review*, *8*(2), 203–220. doi: 10.3758/BF03196157

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.), *Experimental cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer* (pp. 237–249). Washington, DC: American Psychological Association.

Storn, R., & Price, K. (1997, 12). Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. *Journal of Global Optimization*, *11*(4), 341–359. Retrieved from `http://link.springer.com/article/10.1023/A%3A1008202821328` doi: 10.1023/A:1008202821328

Tulving, E. (1962). Subjective organization in free recall of “unrelated” words. *Psychological Review*, *69*(4), 344–354.

Tulving, E. (1968). Theoretical issues in free recall. In T. R. Dixon & D. L. Horton (Eds.), *Verbal behavior and general behavior theory* (pp. 2–36). Englewood Cliffs, N. J.: Prentice-Hall.

Tulving, E. (1983). *Elements of Episodic Memory*. New York: Oxford. doi: 10.1098/rstb.1983.0060

Tulving, E. (2002). Episodic memory: from mind to brain. *Annual Review of Psychology*, *53*(1), 1–25. doi: 10.1146/annurev.psych.53.100901.135114

Underwood, B. J. (1969). Attributes of memory. *Psychological Review*, *76*(6), 559–573.

Van Overschelde, J. P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language*, *50*(3), 289–335. doi: 10.1016/j.jml.2003.10.003

Wallace, W. P. (1970). Consistency of Emission Order in Free Recall. *Journal of Verbal Learning and Verbal Behavior*, *9*, 58–68.

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. doi: 10.1037/a0020122

Whiffen, J. W., & Karpicke, J. D. (2017). The role of episodic context in retrieval practice effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(7), 1036.