Title

Emphasizing associations from encoding affects free recall at retrieval

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 44(44)

Authors

Hong, Bryan
Pace-Tonna, Carleigh A
Barense, Morgan
et al.

Publication Date

2022

Peer reviewed

Emphasizing associations from encoding affects free recall at retrieval

Bryan Hong (bryan.hong@mail.utoronto.ca)
Department of Psychology, University of Toronto
bryan.hong@mail.utoronto.ca

Carleigh A Pace-Tonna
Department of Psychology, University of Toronto
carleigh.pacetonna@mail.utoronto.ca

Morgan D. Barense
Department of Psychology, University of Toronto
Rotman Research Institute
morgan.barense@utoronto.ca

Michael L. Mack
Department of Psychology, University of Toronto
michael.mack@utoronto.ca

Abstract

Paradigms using the free recall of word lists have furthered our understanding of the organizational structure of memory by elucidating the role of contextual associations on memory search. We adapted the traditional word list-learning paradigm to investigate whether emphasizing contextual associations between items influences subsequent retrieval. Specifically, we introduced a review period between encoding and recall of word lists where items were repeated to highlight either the temporal or semantic associations at encoding. We found that temporal review led to stronger temporal clustering compared to a semantic or control review, and semantic review led to stronger semantic clustering compared to a temporal or control review. Moreover, participants recalled more list items when semantic associations were emphasized, with the degree of semantic clustering at recall predicting memory performance. These results demonstrate that emphasizing contextual associations during a repeated viewing after initial encoding can affect subsequent memory organization and recall.

Keywords: episodic memory; semantic memory; free recall; context

Introduction

Reviewing items using repetition has long been known to be an effective strategy to improve later memory (Ebbinghaus, 1913; Hintzman, 1976). Two primary accounts proposed to explain the mnemonic benefits of repetition are contextual variability and study-phase retrieval (Cepeda et al., 2006; Delaney et al., 2010; Maddox, 2016). The contextual variability account (Glenberg, 1979; Melton, 1970) posits that the context associated with each item presentation is encoded alongside a given item. Each repetition provides a different context, which diversifies the available cues at later retrieval. The study-phase retrieval account (Hintzman et al., 1975; Thios & D'Agostino, 1976) proposes that each repetition reactivates the initial memory for the item at encoding. This process of reactivation then strengthens memory for the retrieval cue for this item.

These accounts are not necessarily mutually exclusive from one another, with retrieved context models of memory search having been proposed to integrate the mechanisms from both contextual variability and study-phase retrieval accounts to explain repetition effects in memory (Lohnas & Kahana, 2014). Like contextual variability accounts, retrieved context models assume that a slowly changing internal representation of context is encoded alongside items in memory (Howard & Kahana, 1999, 2002a; Polyn et al., 2009; Polyn & Cutler, 2017). Akin to study-phase retrieval accounts, the associated contextual state from initial encoding can be reinstated when an item is repeated—the contextual representation associated with the first presentation then influences the contextual representation associated with the repeated item. This retrieved context account has been used to successfully predict previous findings from studies investigating these repetition effects in memory (Lohnas & Kahana, 2014). This framework has also been used to explain several robust behavioural effects seen in the free recall of word lists, such as the temporal contiguity effect, where items that are encoded close by in time tend to be recalled together, and the semantic proximity effect, where items that are semantically similar to one another tend to be recalled together (Healey et al., 2018; Kahana, 2020).

Previous evidence has shown that the degree of contextual reinstatement at recall is predictive of overall memory performance. Evidence has shown that the retrieval of temporal associations, but not semantic associations, from encoding is predictive of free recall performance on a word list-learning task (Healey et al., 2014; Healey & Uitvlugt, 2019; Sederberg et al., 2010). However, the words typically used in free recall paradigms are largely selected at random, so the semantic associations at encoding may not have been as strong of a retrieval cue as the temporal associations in these situations. Word lists constructed using items drawn from distinct semantic categories tend to be better remembered relative to word lists constructed using random items, which may be attributed to the strong long-standing semantic associations that overpower the relatively arbitrary temporal associations between items at encoding (Healey & Uitvlugt, 2019; Polyn et al., 2011). When items are semantically related, evidence suggests that the pattern of contextual reinstatement flips, with reinstatement of semantic associations, but not temporal associations, predicting recall performance (Healey & Uitvlugt, 2019).

Retrieval of contextual associations is thought to occur spontaneously—however, evidence suggests that certain factors can influence the degree of contextual reinstatement observed during recall. For example, participants tend to show greater temporal contiguity with increased task experience over the course of a free recall experiment, suggesting that participants may be learning to leverage these

In J. Culbertson, A. Perfors, H. Rabagliati & V. Ramenzoni (Eds.), *Proceedings of the 44th Annual Conference of the Cognitive Science Society*.

temporal associations to scaffold memory as task experience increases (Healey et al., 2018). Conversely, temporal contiguity effects are reduced when participants are learning items with incidental encoding and a surprise free recall task (Healey, 2018). Furthermore, when asked to pay attention to order, participants tend to show a stronger temporal contiguity effect and when asked to pay attention to meaning, participants tend to show a stronger semantic proximity effect—this also affects their free recall performance, with participants directed to focus on the semantic relationships between items having higher overall performance than those directed to focus on temporal relationships or no relationships at all (Healey & Uitvlugt, 2019).

In this study, we aimed to extend this previous work by investigating whether the reviewing items in a particular order could emphasize different contextual associations and furthermore, whether this would affect subsequent memory recall. According to retrieved-context models, the reinstatement of an item's initial encoding context during a repeated presentation allows the item to be associated with multiple contexts (Lohnas & Kahana, 2014). We hypothesized that we could bias participants to leverage different types of associations by structuring the order of items at a repeated viewing to emphasize different contextual associations. Furthermore, we hypothesized that this bias would affect their overall recall performance.

We had participants complete a modified free recall paradigm where participants encoded and recalled word lists made up of groups of items from distinct semantic categories with an intermediate review phase. To make the temporal associations between items more meaningful, items were presented superimposed on a series of images depicting first-person navigation along a real-world route. To manipulate the saliency of contextual associations during this review phase, participants were presented with the same words from encoding in an order that either emphasized the temporal associations from encoding, the existing semantic associations, or neither (in a random order, as a baseline condition). We predicted that participants would demonstrate a stronger temporal contiguity effect after reviewing content that emphasizes the temporal context when compared to the semantic or baseline conditions, and a stronger semantic proximity effect after reviewing content that emphasizes the semantic context when compared to the temporal or baseline conditions. Overall recall accuracy was predicted to be highest after a semantic review period, but also higher after a temporal review period compared to a random review period.

Methods

Participants

Participants were recruited online using *Prolific* (https://prolific.co/) with pre-screening filters to ensure that participants were between the ages of 18-35 years old, fluent in English, had normal or corrected-to-normal vision, and were using a computer to complete the study. All participants provided informed consent online prior to the study and received monetary compensation for their participation. The study was approved by the Research Ethics Board at the University of Toronto.

An a priori power calculation revealed that we would need 18 participants in each condition to achieve 95% power based on effect sizes from an initial pilot study. To increase our ability to detect an effect in the event that our pilot study overestimated the effect size, participants were recruited until we had complete data from at least 25 participants in each experimental condition. We collected data from a sample of 81 participants with a mean age of 24.64 years (SD = 4.72 years, range = 18-36 years, 51 men/30 women). Of our 81 participants, 29 participants were assigned to the temporal review condition, 25 participants were assigned to the semantic review condition, and 27 participants were assigned to the random review condition (see *Task design* below for more detail).

Word pool

Participants studied words drawn from a large-scale study that collected feature norms for 541 different concepts (McRae et al., 2005). These feature norms are the production frequencies of responses on a feature generation task, where participants were asked to provide as many semantic features as they could for a corresponding concrete concept.

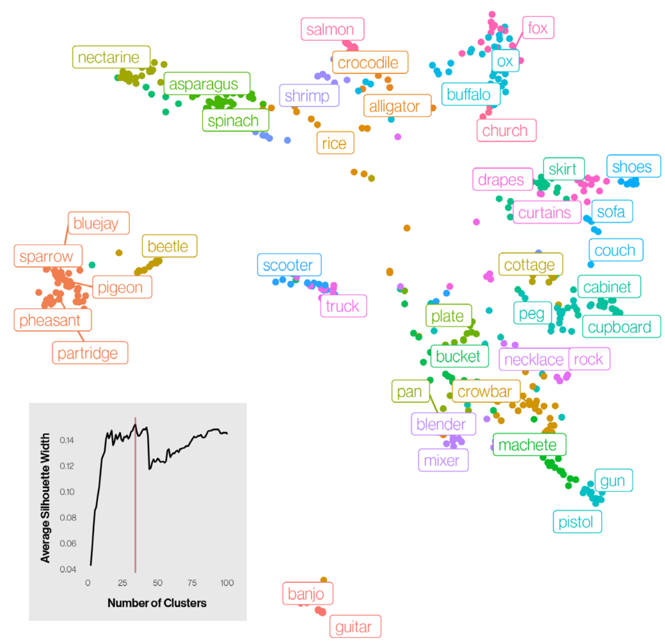

Figure 1: Uniform manifold approximation and projection of the k-medoids clustering analysis on the feature norm production frequencies from McRae et al. (2005). Each colour denotes a cluster identified in the final clustering solution. Inset plot shows the average silhouette width as a function of the number of clusters—the red line denotes the highest average silhouette width.

To identify categories of semantically related words in this subpool, we conducted a k-medoids clustering analysis with the Partitioning Around Medoid algorithm on the cosine dissimilarity matrix between each word pair using the *cluster* package in R (Maechler et al., 2021). This cosine dissimilarity matrix was computed in McRae et al. (2005) by taking the word × feature production frequency matrix and calculating the cosine dissimilarity between each word × feature vector. To identify an optimal number of clusters, we ran independent k-medoids analyses setting the number of clusters between 2 and 100—we then identified the solution with the highest average silhouette width (Rousseeuw, 1987). From this, we arrived at a final solution of 34 clusters of semantically similar words in our subpool (Figure 1). Each cluster had an average of 15.09 words (SD = 8.66 words).

Task design

Participants completed eight trials of a word list learning task, with each trial being broken up into an encoding phase, a review phase, and a recall phase (Figure 2). Participants were randomly assigned to one of three experimental conditions, described below, to avoid carryover effects across conditions.

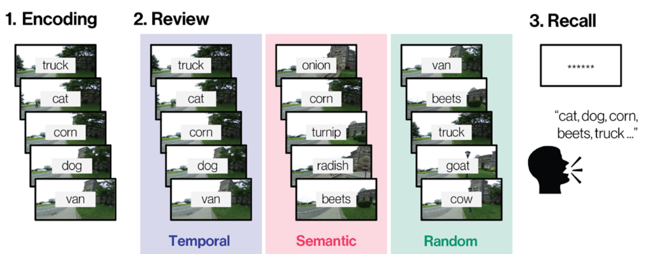

Figure 2: Schematic of task paradigm. Each trial consisted of an encoding, review (temporal/semantic/random conditions), and recall phase.

Encoding phase During the encoding phase, participants first learned a word list composed of 15 words made up of five words from three of the identified clusters—the words corresponding to the medoids of the selected clusters had no overlapping features to ensure that they were distinct categories.

Words were presented sequentially for 1200ms each in a pseudorandom order, with no more than two adjacent words coming from the same cluster. These words were presented overlaid a continuous series of first-person photographs navigating along real-world spatial routes to further strengthen the temporal associations between items during encoding—the images from each list were all associated with the same route, with the first word of each list being presented over an image at the start of the route and the last word of each list being presented over an image at the end of the route to convey stepwise navigation along the route.

Review phase Participants were then presented with the same words from encoding, with the order of presentation being manipulated depending on the experimental condition assigned. Participants could be assigned to (1) a temporal review condition, where words were ordered in the original temporal sequence to emphasize the temporal associations, (2) a semantic review condition, where words were ordered by semantic categories to emphasize the semantic associations, or (3) a random review condition, where words were ordered in a new random temporal sequence as a baseline measure.

Importantly, all words were presented with the same background image as the encoding phase so that the background images in a temporal review phase would be presented in an order that depicted the same continuous stepwise navigation seen during encoding—they had no meaningful order during a semantic or random review phase.

Recall phase Participants were presented with a row of asterisks with an auditory beep to indicate the beginning of the recall phase. Participants were then given 45 seconds to verbally recall as many words as they could remember from the just-studied trial list—participants were not asked to recall the words in a specific order.

Statistical analyses

We computed the lag-conditional responses probabilities (lag-CRP; Kahana, 1996), semantic-conditional response probability (sem-CRP; Howard & Kahana, 2002), and temporal and semantic factor scores (Polyn et al., 2009) during recall using the *psifr* package in Python (Morton, 2020).

Lag-CRP curves provide a measure of temporal clustering during recall via the probability of recalling item *j* after recalling item *i* as a function of lag, or the number of items between *i* and *j* at encoding—for example, if item *i* was at the 10^{th} word presented during encoding, a lag of +1 indicates that item *j* was the 11^{th} word presented during encoding while a lag of -1 indicates that item *j* was the 9^{th} word presented during encoding. Similarly, sem-CRP curves provide a measure of semantic clustering during recall via the probability of recalling item *j* after recalling item *i* as a function of the cosine between their feature norm production frequencies (McRae et al., 2005)—all cosine values were treated as separate bins to transition between (Sederberg et al., 2010).

To quantify the differences in the lag-CRP curves across the different review conditions, an individual linear regression model was fit for each participant's lag-CRP curve (Diamond & Levine, 2020). This was done separately for the positive and negative lags, up to and including an absolute lag value of 5 to compare the steepness of the positive and negative components of the lag-CRP curves across the review conditions (Howard et al., 2007; Sadeh et al., 2015). Similar to the lag-CRP curves, an individual linear regression model was fit to capture the steepness of each participant's sem-CRP curve (Howard & Kahana, 2002b).

Temporal and semantic factor scores (Polyn et al., 2009) provide a single value that captures the average strength of temporal and semantic associations between transitions at recall, respectively. For each item recalled, the absolute lag (for temporal factor scores) or semantic similarity (for semantic factor scores) of all possible remaining transitions is calculated and the percentile rank of the actual transition is taken. This is then averaged across all transitions made, with a final score ranging between 0 and 1. A factor score of 0 indicates that the furthest item was always chosen, and a factor score of 1 indicates that the nearest item was always chosen—chance clustering is indicated by a factor score of 0.5. The temporal factor score does not take into account the direction of the transitions made.

Comparisons of these metrics across review conditions were analyzed using one-way independent-measures ANOVAs in R, with effect sizes being obtained using the *effectsize* package (Ben-Shachar et al., 2020). Post-hoc tests with the Tukey method were conducted with the *emmeans* package (Lenth et al., 2021).

Additionally, we fit a 2-level multilevel generalized logistic model using the *lme4* package in R (Bates et al., 2015) to predict recall of each individual word from a participant's temporal factor score, semantic factor score and review condition to investigate whether the review conditions affected the relationship between these retrieval dynamics and overall recall performance. To test whether review condition moderated the relationship between the factor scores and recall, we also modeled the interactions between review condition and both factor scores. Likelihood-ratio tests were conducted to obtain p-values for fixed effects related to review condition using the *afex* package (Singmann et al., 2021). Interactions were probed by comparing the simple slopes (Aiken et al., 1991) with the *emmeans* package (Lenth et al., 2021).

Results

Recall accuracy

There was a significant main effect of review condition on the overall number of items recalled ($F(2, 78) = 3.701$, $p = .0291$, $\eta^2 = .09$). Participants in the semantic review condition recalled significantly more items than participants in the random review condition ($t(78) = 2.681$, $p = .0241$, $d = 0.744$). There was no significant difference in the number of items recalled between participants in the temporal review condition and participants in the semantic review condition ($t(78) = -1.810$, $p = .173$, $d = -0.494$), or participants in the random review condition ($t(78) = 0.935$ $p = .620$, $d = 0.250$).

Temporal contiguity at retrieval

In addition to overall recall performance, we assessed the degree of temporal contiguity at retrieval by examining the lag-CRP curves (Figure 3A) and the average temporal factor scores (Figure 3B) for each review condition.

The canonical lag-CRP shape is most evident for participants in the temporal review condition, with both the lag recency effect, as seen by the higher CRP for more recent lags compared to later lags, and a forward asymmetry, as seen by the steeper curve for positive lags compared to negative lags, being observed (Healey et al., 2018; Howard & Kahana, 1999). However, there was no clear evidence of the temporal contiguity effect for participants in the semantic or random review conditions.

For positive lags, there was a significant main effect of review condition on the steepness of the lag-CRP curves ($F(2, 78) = 7.195$, $p = .00136$, $\eta^2 = .16$). Participants in the temporal review condition had significantly steeper slopes for positive lags than participants in the semantic review condition ($t(78) = -3.152$, $p = .0064$, $d = -0.860$), or participants in the random review condition ($t(78) = -3.349$, $p = .0035$, $d = -0.896$). There was no significant difference in the steepness of the slopes for positive lags between participants in the semantic review condition and participants in the random review condition ($t(78) = -0.127$, $p = .991$, $d = -0.0353$). For negative lags, there was no significant main effect of review condition on the steepness of the lag-CRP curve ($F(2, 78) = 1.034$, $p = .36$, $\eta^2 = .03$), with participants having comparable slopes for negatives lags across the temporal, semantic, and random review conditions.

Consistent with the findings from the lag-CRP curves, there was a significant main effect of review condition on the temporal factor scores at recall ($F(2, 78) = 11.45$, $p < .001$, $\eta^2 = .23$). Participants in the temporal review condition had significantly greater temporal factor scores than participants in the semantic review condition ($t(78) = 3.929$, $p < .001$, $d = 1.072$), or participants in the random review condition ($t(78) = 4.263$, $p < .001$, $d = 1.140$). There was no difference in temporal factor scores between participants in the semantic review condition and participants in the random review condition ($t(78) = 0.243$, $p = .968$, $d = .0675$). Temporal factor scores were significantly greater than chance for participants in the temporal review condition ($p < .001$), but not for those in the semantic or random review conditions (both p's > .05).

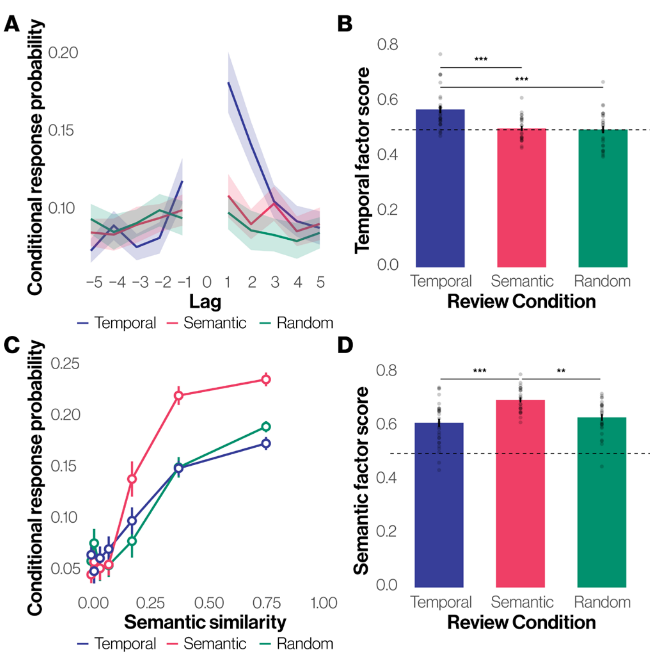

Figure 3: (A) Lag-CRP curves, (B) temporal factor scores, (C) sem-CRP curves, and (D) semantic factor scores for recall across random, semantic, and temporal review conditions. Bands and error bars indicate standard error. Dots represent individual participants. Dashed lines

represent chance clustering. Asterisks indicate statistically significant differences across review conditions: ** $p < .01$, *** $p < .001$

Semantic proximity at retrieval

We next assessed the degree of semantic proximity at retrieval by examining the sem-CRP curves (Figure 3C) and the average semantic factor scores (Figure 3D) for each review condition.

The shape of the sem-CRP curves is quite similar across participants in each review condition, with participants in all three conditions showing a greater probability to transition between items that are semantically related—these results are consistent with previous findings investigating sem-CRP curves with other metrics of semantic similarity (Howard et al., 2007; Kahana, 2020; Morton & Polyn, 2016; Sederberg et al., 2010).

There was a significant main effect of review condition on the steepness of the sem-CRP curves ($F(2, 78) = 7.043$, $p = .00154$, $\eta^2 = .15$), with participants in the semantic review condition having significantly steeper sem-CRP curves than participants in the temporal review condition ($t(78) = -3.677$, $p = .0012$, $d = -1.004$), and participants in the random review condition ($t(78) = 2.579$, $p = .0313$, $d = 0.716$). There was no significant difference in the steepness of the sem-CRP curves between participants in the temporal review condition and participants in the random review condition ($t(78) = -1.076$, $p = .532$, $d = -0.288$).

We also found corroborating results with the semantic factor scores, with a significant main effect of review condition on the semantic factor scores at recall ($F(2, 78) = 11.71$, $p < .001$, $\eta^2 = .23$). Participants in the semantic review condition had significantly greater semantic factor scores than participants in the temporal review condition ($t(78) = -4.683$, $p < .001$, $d = -1.278$), or participants in the random review condition ($t(78) = 3.513$, $p = .0021$, $d = 0.975$). There was no difference in temporal factor scores between participants in the semantic review condition and participants in the random review condition ($t(78) = -1.133$, $p = .497$, $d = -0.303$). Semantic factor scores were significantly greater than chance for participants across all three review conditions (all p's < .001).

Retrieval dynamics and memory performance

We observed a significant main effect of both semantic and temporal factor scores on overall recall (Figure 4), with a higher degree of semantic clustering being related to higher overall recall ($b = 0.314$, $SE = 0.0711$, $z = 4.420$, $p < .001$), and a higher degree of temporal clustering being related to higher overall recall ($b = 0.302$, $SE = 0.0696$, $z = 4.335$, $p < .001$).

Interestingly, we did not find a main effect of review condition on overall recall ($\chi^2(2) = 0.15$, $p = .927$). However, the relationship between semantic clustering and recall was moderated by review condition ($\chi^2(2) = 6.17$, $p = .046$)—this effect was driven by a significant relationship between semantic factor scores and recall for participants in the semantic review condition ($b = 0.611$, $SE = 0.162$, 95% CI [0.294, 0.927]), but not in the temporal ($b = 0.161$, $SE = 0.0948$, 95% CI [-0.0251, 0.346]), or random review condition ($b = 0.172$, $SE = 0.102$, 95% CI [-0.0285, 0.372]). However, this was not the case for temporal clustering, ($\chi^2(2) = 5.97$, $p = .051$). Although the interaction between temporal factor scores and review condition was not statistically significant, we conducted a simple slopes analysis to investigate any trends that may warrant exploration in future studies. Similarly to the pattern in the simple slopes of semantic factor, there was only a relationship between temporal factor scores and recall for participants in the semantic review condition ($b = 0.585$, $SE = 0.157$, 95% CI [0.276, 0.893]), but not in the temporal ($b = 0.173$, $SE = 0.101$, 95% CI [-0.0253, 0.371]), random review condition ($b = 0.148$, $SE = 0.0933$, 95% CI [-0.0347, 0.331]).

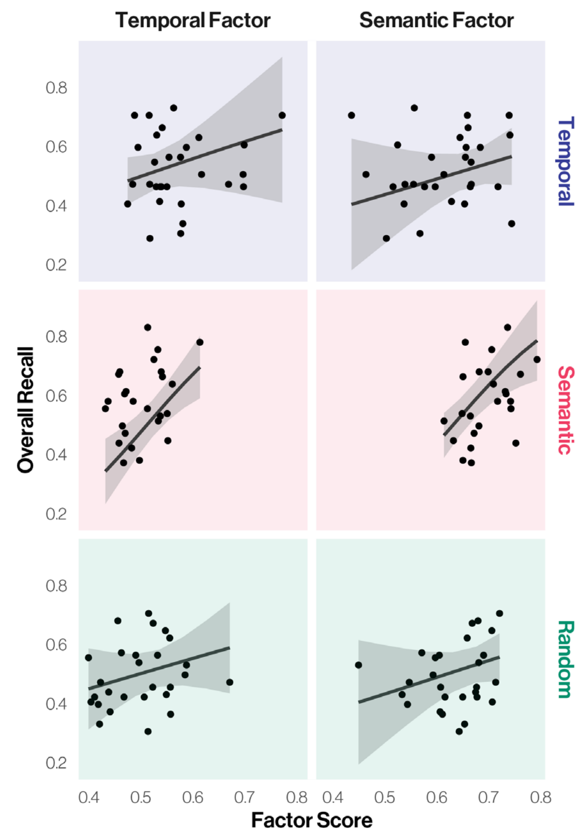

Figure 4: Scatterplots between factor scores and overall recall across participants in each review condition. Lines denote the estimated marginal means of the linear trend, with bands indicating the standard error.

Discussion

The present study aimed to investigate (1) whether memory retrieval could be implicitly biased towards stronger temporal or semantic organization via an intermediate review period, and (2) whether this shift in memory performance would be associated with better memory. The clustering at recall across trials with each of the different review types provided evidence that the order of reviewed items affected the contextual associations reinstated at recall. This manipulation affected overall recall accuracy, with participants correctly remembering more words when they had a semantic review period compared to those with a random review period—however, there was no significant difference in recall accuracy between participants with a temporal review period compared to participants with a semantic or random review period. We found stronger encoding order maintenance for participants in the temporal review condition compared to those in the semantic or random review conditions. The canonical shape for the lag-CRP curve, a hallmark for temporal context reinstatement, was most evident for participants in the temporal review condition, but not for participants in the random or semantic review conditions. Turning to semantic associations, semantic clustering was seen for all participants, but participants in the semantic review condition demonstrated this to a higher degree compared to those with a temporal or random review period.

We found that participants exhibited changes to memory organization when an intervening review period emphasized either the temporal or semantic associations from the encoding phase of a given list. These results expand on work showing that participants exhibit stronger temporal clustering or semantic clustering when asked to explicitly attend to the temporal order or semantic meaning of the items during encoding, respectively (Healey & Uitvlugt, 2019). This highlights the automaticity of encoding these aspects of memory, with participants being able to pick up on both sets of cues even when not explicitly directed to them. Future work can investigate the degree of memory reorganization depending on a participant's awareness of the semantic categories present in the word lists.

The patterns in overall recall performance across all participants are consistent with the increased recall accuracy observed for word lists with items that are drawn from distinct semantic categories (Healey & Uitvlugt, 2019; Polyn et al., 2011). Although all participants showed some degree of semantic clustering, further emphasizing these semantic associations led to better overall memory. Notably, we found no difference in memory between participants in the temporal review condition and participants in the semantic or random review conditions. This suggests that the temporal associations afforded by situating each item sequentially along a spatial route may not overcome the benefit afforded by the existing semantic associations between items. These results support the idea that memory benefits the most when participants are oriented towards the type of associations that can provide the strongest potential cues at recall (Healey & Uitvlugt, 2019). Although no benefits were seen in overall memory for participants in the temporal review condition, other aspects of memory may be improved after a review period that emphasizes the temporal associations at study. For example, emphasizing these temporal associations is predicted to improve performance on a serial recall task, where participants are asked to recall items in the same order that they were learned. In addition, it should also support the rich re-experiencing of the original encoding event, which could potentially be captured by asking participants to indicate the background image associated with each item.

One possible mechanism that may be driving this benefit in performance for participants with a semantic review condition is prediction error, when an occurrence defies expectations (Exton-McGuinness et al., 2015; Henson & Gagnepain, 2010; Sinclair & Barense, 2019). When a prediction error occurs, the existing memory trace is thought to become reactivated and labile to updating (Hupbach et al., 2007). In the current study, participants in the semantic and random review periods may experience a prediction error when reviewing items because the reviewed order does not match the one previously seen at encoding. In both cases, the reactivated item becomes more susceptible to modification by nearby items—however, this modification would only be beneficial in the semantic review condition because it emphasizes the existing semantic associations in the lists, while the associations in the random review condition would be arbitrary. These findings suggest that prediction error during a review period may drive new learning by drawing attention to existing associations that may otherwise go unnoticed.

The current results could inform the design of review paradigms that are tailored to memory organization strategies at retrieval. Reinstatement of temporal and semantic contextual information is thought to be linked to separate cortical networks (Kragel et al., 2021). Future work could identify whether participants are preferentially relying on the temporal or semantic aspects of context to guide memory retrieval. From this, we can investigate whether emphasizing the aspects of context that are congruent with the employed strategy during encoding produces any significant boosts in later memory recall.

In the current study, we used a modified free recall paradigm to provide evidence that memory organization at retrieval can be biased by emphasizing different aspects of study context during an implicit intermediate review period. The results demonstrate both the automaticity and the malleability of the organizational properties of memory that help facilitate recall. The results of this study help elucidate strategies that can be used to leverage these properties to ultimately maximize later memory.

Acknowledgements

The authors thank the anonymous reviewers for their helpful feedback, Morris Moscovitch for insightful discussions, Belinda Hoang for helping to create experimental stimuli and run pilot studies, and Juliana Adema and Gaeun Son for their tips on running the study online. This research is supported by the Natural Sciences and Engineering Research Council (NSERC) Postgraduate Scholarship to BH (396158181), the Ontario Graduate Scholarship to BH, the NSERC Discovery Grant to MLM (RGPIN-2017-06753), the Canada Foundation for Innovation and Ontario Research Fund (36601) to MLM, and the Brain Canada Future Leaders in Canadian Brain Research Grant to MLM.

References

Aiken, L. S., West, S. G., & Reno, R. R. (1991). *Multiple regression: Testing and interpreting interactions*. SAGE.

Bates, D., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48.

Ben-Shachar, M., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *The Journal of Open Source Software*, *5*, 2815. https://doi.org/10.21105/joss.02815

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Delaney, P. F., Verkoeijen, P. P. J. L., & Spirgel, A. (2010). Spacing and testing effects: A deeply critical, length, and at times discursive review of the literature. In B. H. Ross (Ed.), *Psychology of Learning and Motivation* (Vol. 53, pp. 63–147). Elsevier. https://doi.org/10.1016/S0079-7421(10)53003-2

Diamond, N. B., & Levine, B. (2020). Linking detail to temporal structure in naturalistic-event recall. *Psychological Science*, *31*(12), 1557–1572. https://doi.org/10.1177/0956797620958651

Ebbinghaus, H. (1913). *Memory: A contribution to experimental psychology*. Teachers College, Columbia University.

Exton-McGuinness, M. T. J., Lee, J. L. C., & Reichelt, A. C. (2015). Updating memories—The role of prediction errors in memory reconsolidation. *Behavioural Brain Research*, *278*, 375–384. https://doi.org/10.1016/j.bbr.2014.10.011

Glenberg, A. M. (1979). Component-levels theory of the effects of spacing of repetitions on recall and recognition. *Memory & Cognition*, *7*(2), 95–112. https://doi.org/10.3758/BF03197590

Healey, M. K. (2018). Temporal contiguity in incidentally encoded memories. *Journal of Memory and Language*, *102*, 28–40. https://doi.org/10.1016/j.jml.2018.04.003

Healey, M. K., Crutchley, P., & Kahana, M. J. (2014). Individual differences in memory search and their relation to intelligence. *Journal of Experimental Psychology: General*, *143*(4), 1553–1569. https://doi.org/10.1037/a0036306

Healey, M. K., Long, N. M., & Kahana, M. J. (2018). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, 1–22. https://doi.org/10.3758/s13423-018-1537-3

Healey, M. K., & Uitvlugt, M. G. (2019). The role of control processes in temporal and semantic contiguity. *Memory and Cognition*, 1–19. https://doi.org/10.3758/s13421-019-00895-8

Henson, R. N., & Gagnepain, P. (2010). Predictive, interactive multiple memory systems. *Hippocampus*, *20*(11), 1315–1326. https://doi.org/10.1002/hipo.20857

Hintzman, D. L. (1976). Repetition and memory. In G. H. Bower (Ed.), *Psychology of Learning and Motivation* (Vol. 10, pp. 47–91). Academic Press. https://doi.org/10.1016/S0079-7421(08)60464-8

Hintzman, D. L., Summers, J. J., & Block, R. A. (1975). Spacing judgments as an index of study-phase retrieval. *Journal of Experimental Psychology: Human Learning and Memory*, 31–40.

Howard, M. W., Addis, K. M., Jing, B., & Kahana, M. J. (2007). Semantic structure and episodic memory. In T. K. Landauer, D. S. McNamara, S. J. Dennis, & W. Kintsch (Eds.), *Handbook of Latent Semantic Analysis* (pp. 121–141). Psychology Press.

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

Howard, M. W., & Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*(3), 269–299. https://doi.org/10.1006/jmps.2001.1388

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

Hupbach, A., Gomez, R., Hardt, O., & Nadel, L. (2007). Reconsolidation of episodic memories: A subtle reminder triggers integration of new information. *Learning & Memory*, *14*(1–2), 47–53. https://doi.org/10.1101/lm.365707

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

Kahana, M. J. (2020). Computational models of memory search. *Annual Review of Psychology*, *71*(1), 107–138. https://doi.org/10.1146/annurev-psych-010418-103358

Kragel, J. E., Ezzyat, Y., Lega, B. C., Sperling, M. R., Worrell, G. A., Gross, R. E., Jobst, B. C., Sheth, S. A., Zaghloul, K. A., Stein, J. M., & Kahana, M. J. (2021). Distinct cortical systems reinstate the content and context of episodic memories. *Nature Communications*, *12*(1), 4444. https://doi.org/10.1038/s41467-021-24393-1

Lenth, R. V., Buerkner, P., Herve, M., Love, J., Riebl, H., & Singmann, H. (2021). *emmeans: Estimated Marginal Means, aka Least-Squares Means* (Version 1.6.1) [R Package]. https://cran.r-project.org/web/packages/emmeans/emmeans.pdf

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

Maddox, G. B. (2016). Understanding the underlying mechanism of the spacing effect in verbal learning: A case for encoding variability and study-phase retrieval. *Journal of Cognitive Psychology*, *28*(6), 684–706. https://doi.org/10.1080/20445911.2016.1181637

Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., Hornik, K., Studer, M., Roudier, P., Gonzalez, J., Kozlowski, K., Schubert, E., & Murphy, K. (2021). *cluster: Finding*

Groups in Data (Version 2.1.2) [R Package]. https://cran.r-project.org/web/packages/cluster/cluster.pdf

McRae, K., Cree, G. S., Seidenberg, M. S., & McNorgan, C. (2005). Semantic feature production norms for a large set of living and nonliving things. *Behavior Research Methods*, *37*(4), 547–559. https://doi.org/10.3758/BRM.40.1.183

Melton, A. W. (1970). The situation with respect to the spacing of repetitions and memory. *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 596–606. https://doi.org/10.1016/S0022-5371(70)80107-4

Morton, N. W. (2020). Psifr: Analysis and visualization of free recall data. *Journal of Open Source Software*, *5*(54), 2669. https://doi.org/10.21105/joss.02669

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. https://doi.org/10.1016/j.jml.2015.10.002

Polyn, S. M., & Cutler, R. A. (2017). Retrieved-context models of memory search and the neural representation of time. *Current Opinion in Behavioral Sciences*, *17*, 203–210. https://doi.org/10.1016/j.cobeha.2017.09.007

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*, *37*, 766–775. https://doi.org/10.1037/a0022475

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156. https://doi.org/10.1037/a0014420

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7

Sadeh, T., Moran, R., & Goshen-Gottstein, Y. (2015). When items 'pop into mind': Variability in temporal-context reinstatement in free-recall. *Psychonomic Bulletin and Review*, *22*(3), 779–790. https://doi.org/10.3758/s13423-014-0746-7

Sederberg, P. B., Miller, J. F., Howard, M. W., & Kahana, M. J. (2010). The temporal contiguity effect predicts episodic memory performance. *Memory and Cognition*, *38*(6), 689–699. https://doi.org/10.3758/MC.38.6.689

Sinclair, A. H., & Barense, M. D. (2019). Prediction error and memory reactivation: How incomplete reminders drive reconsolidation. *Trends in Neurosciences*, *42*(10), 727–739. https://doi.org/10.1016/j.tins.2019.08.007

Singmann, H., Bolker, B., Westfall, J., Aust, F., Ben-Shachar, M. S., Højsgaard, S., Fox, J., Lawrence, M. A., Mertens, U., Love, J., Lenth, R. V., & Christensen, R. H. B. (2021). *afex: Analysis of Factorial Experiments* (Version 1.0-1) [R Package]. https://cran.r-project.org/web/packages/afex/afex.pdf

Thios, S. J., & D'Agostino, P. R. (1976). Effects of repetition as a function of study-phase retrieval. *Journal of Verbal Learning and Verbal Behavior*, *15*(5), 529–536. https://doi.org/10.1016/0022-5371(76)90047-5