Temporal contiguity between recalls predicts episodic memory performance

Per B. Sederberg
Department of Psychology and
Princeton Neuroscience Institute
Princeton University
Princeton, NJ 08540, USA

Jonathan F. Miller
Department of Psychology
University of Pennsylvania
Philadelphia, PA 19104, USA

Marc W. Howard
Department of Psychology
Syracuse University
Syracuse, NY 13244, USA

Michael J. Kahana
Department of Psychology
University of Pennsylvania
Philadelphia, PA 19104, USA

Manuscript under review.

One way to study the associative processes at work during episodic memory is to examine the order of participant responses, which typically reveal the strong tendency to transition between temporally contiguous or semantically proximal items. Here we assessed the correlation between participants' recall performance and their use of semantic and temporal associations to guide retrieval across nine delayed free-recall studies. The size of participants' temporal contiguity effects predicted their recall performance. High semantic proximity effects, on the other hand, did not predict better recall performance. These results suggest that participants who more effectively form and retrieve associations between items that occur nearby in time perform better on episodic recall tasks.

The free recall paradigm has provided a wealth of information regarding the nature of the associations at the core of episodic memory. Specifically, by quantifying the order in which items are recalled (rather than simply whether they are recalled), one can see the associative processes unfold as each recalled item cues the next. If the order of responses reflects the order in which items come to mind, then response order reflects the internal organization of memory for the list items (i.e., the associations between the items).

To quantify the effects of the temporal organization of list items, Kahana (1996) measured the conditional response probability as a function of serial position lag in the study-list (lag-CRP). Given that the participant has just recalled an item from serial position i, the lag-CRP indicates the probability that the next item recalled comes from serial position $i + \text{lag}$. Lag-CRP analyses have demonstrated that the *temporal contiguity effect*, a tendency for participants to transition between items that were presented in nearby serial positions in the study-list, and the *asymmetry effect*, a tendency for participants to recall items in the forward direction, are extremely robust properties of free recall (see Kahana, Howard, & Polyn, 2008, for a review).

Temporal associations are not the sole organizational factor revealed by response order. Howard and Kahana (2002b) extended the analysis of recall transitions, measuring the conditional response probability as a function of the semantic relationship between items (sem-CRP). Sem-CRP analyses revealed a *semantic proximity effect*, whereby participants tend to recall items that are semantically related to the just-recalled item.

The prevalence of temporal contiguity and semantic proximity effects indicate that participants rely on associations between items that occurred close together in time or were already semantically related prior to the task to guide retrieval. However, it does not imply that the processes that give rise to contiguity and proximity effects enable items to be recalled—it remains possible that the contiguity effect or the proximity effect (or both) merely affect the order in which recalls are uttered rather than the availability of items. Although recall order is, on the face of it, largely independent of the total number of recalls, the question at hand is whether temporal or semantic clustering predict recall performance. If associations between nearby items, whether in temporal or semantic space, reflect a fundamental process underlying episodic memory, then individual differences in response order should correlate with overall recall performance.

Here we examine the covariation between individual differences in probability of recall and temporal contiguity and

The authors acknowledge support from the Swartz Foundation and National Institutes of Health research grants MH055687 and MH061975 to MJK, MH069938 to MWH, and MH072138 and MH080526 to PBS. Correspondence concerning this article may be addressed to Per B. Sederberg (`persed@princeton.edu`).

semantic proximity effects. Previous work examining contiguity and proximity effects across conditions in which the probability of recall varies has shown a complex pattern of results. Consider the finding that is observed in multi-trial free recall. When a randomly-assembled list is presented in the same order for multiple recall trials, recall probability goes up with each subsequent trial. Howard, Addis, Jing, and Kahana (2007) showed that across trials, the temporal contiguity effect was enhanced, while the semantic proximity effect did not change. In this case, the contiguity effect covaried with recall probability across conditions while the semantic proximity effect did not. In contrast, consider the relationship between contiguity and proximity effects in continual-distractor free recall, in which a distractor interval follows every item in the list. Experiment 2 of Howard and Kahana (1999) included four conditions in which the duration of the distractor at the end of the list was held fixed, but the duration of the distractor between items varied from zero seconds (delayed free recall) to sixteen seconds (the same as the duration of the end-of-list distractor). As the length of the distractor between items was increased across conditions, the probability of recall decreased. The temporal contiguity effect did not change as the inter-stimulus interval was increased (Howard & Kahana, 1999). However, the semantic proximity effect decreased as the length of the distractor interval between condition was increased (Howard & Kahana, 2002b). In this case, the semantic proximity effect covaried with recall probability across conditions, whereas the temporal contiguity effect did not. One may also expect a correlation between semantic proximity effects and probability of recall based on work comparing recall of categorized lists to uncategorized lists (Bousfield, 1953; Pollio, Richards, & Lucas, 1969; Brown, Conover, Flores, & Goodman, 1991). At the extreme, in which lists are drawn from a single category, participants recall more items from the study-list but suffered memory impairments for the order that the items were presented (Greene & Crowder, 1984), suggesting that temporal contiguity effects may trade off with semantic effects under these circumstances.

In order to address the relationship between variability in temporal and semantic proximity effects and successful episodic recall, we conducted a meta-analysis across nine delayed free-recall studies. In the following sections we first quantify the temporal contiguity and semantic proximity effects exhibited by the participants in these studies. We then relate the degree that individual participants rely on temporal and semantic associations to drive responses to their overall recall performance. To provide an overview of our results, we find that participants in these studies exhibit strong temporal contiguity and semantic proximity effects. Those who exhibit a higher temporal contiguity effect recall more items, whereas individual differences in the semantic proximity effect do not significantly correlate with recall performance. In the general discussion we relate these results to other findings and to theories of episodic recall.

Method

Experiment Details

We combined results from nine delayed free-recall studies for the analyses in this manuscript.

Inclusion Criteria. We required each study to pass a strict set of criteria for inclusion.

First, we only included studies where we had access to the individual responses for each trial so that we could then analyze the recall transitions.

Second, to ensure an equal comparison between temporal and semantic associations on recall performance we only included delayed free-recall conditions, which have been shown to have consistent and significant temporal contiguity and semantic proximity effects (Howard & Kahana, 1999, 2002b). We excluded immediate free-recall studies because participants tend to exhibit an elevated temporal contiguity effect (i.e., a steep slope in the lag-CRP) that decreases with output position, likely due to the pronounced recency that simultaneously enhances temporal contiguity effects and decreases semantic proximity effects (Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Sederberg, Howard, & Kahana, 2008). Whereas the temporal contiguity effect is stable in continual-distractor free recall (Howard & Kahana, 1999), Howard and Kahana (2002b) demonstrated that the effect of semantic proximity decreases when a distractor is inserted between study-words, thus decreasing the likelihood that there would be any effect of semantic associations on recall performance. In addition to these issues with immediate and continual-distractor free recall, we wanted to avoid strong recency effects in general, which have been shown to introduce non-monotonicity into the lag-CRP (Farrell & Lewandowsky, 2008; Howard, Sederberg, & Kahana, in press).

Finally, we excluded any condition in which items had spaced repetitions within lists or that was performed by older participants, as we wanted to focus on normal recall that was unaffected by list manipulations or aging-related memory deficits. Although the words in the massed condition of Kahana and Howard (2005) and the items repeated across lists in the Zaromb et al. (2006) study are nominally repeated (see below), each unique word had an unambiguous serial position with respect to the current list and each paper reported typical lag-CRP functions for the repeated items.

Methodological Details of Included Experiments. The remaining nine experiments included a total of 510 participants and varied upon a number of methodological dimensions, such as list length, presentation rate, and distractor duration (see Table 1). Here we summarize the methods of each of the experiments we included. For all the experiments outlined below, participants gave vocal responses that were digitally recorded and processed offline to specify exact response onset times.

Howard and Kahana (1999), Exp. 1. In single sessions, 60 participants (three were excluded from the original study because of experimenter error) performed both immediate and delayed free-recall of 25 total lists (only the delayed free-recall lists were included in this study.) The first two lists for each participant were treated as practice, while the remaining 23 lists were randomly selected to be either immediate or delayed free-recall (i.e., participants performed different numbers of delayed free recall lists, ranging from 4 to 16 total lists, giving rise to 641 total lists across all participants.) Each list was composed of 12 randomly selected nouns from the Toronto Word Pool (Friendly, Franklin, Hoffman, & Rubin, 1982). Words were presented visually for 1000 ms each. While each word was on the screen, participants were required to perform a semantic orienting task, judging whether each word was "concrete" or "abstract" by pressing either the left or right control keys. After the presentation of the last item, participants performed true/false math problems of the form $A + B + C = D$, where A, B, and C are positive, single-digit integers, for 10 s. Then participants recalled the words on the just-studied list in any order during a 45 s recall period.

Howard and Kahana (1999), Exp. 2. Over the course of 10 sessions, 16 participants performed 4 variants of free recall (one delayed and three continual-distractor with varying durations of a distractor-filled inter-stimulus interval (ISI); only the delayed free-recall lists were included in this study.) Each list was composed of 12 nouns selected at random and without replacement from the Toronto Word Pool (Friendly et al., 1982). Words were presented visually for 1200 ms each. While each word was on the screen, participants were required to perform a semantic orienting task, judging whether each word was "concrete" or "abstract" by pressing either the left or right control keys. After the presentation of the last item, participants performed true/false math problems of the form $A + B + C = D$, where A, B, and C are positive, single-digit integers, for 16 s. Then participants recalled the words on the just-studied list in any order during a 60 s recall period.

Kahana et al. (2002), Exp. 2. In single sessions, 25 older and 25 younger participants performed delayed free recall of 23 lists (only the younger participants were included in the present study and the first three lists were treated as practice and removed from the analysis.) The 10 words in each list were presented visually for 1400 ms, followed by a 100 ms ISI. After the presentation of the last item, participants performed math problems of the form $A+B+C=?$, where A, B, and C are positive, single-digit integers, for 16 s before recalling the words on the just-studied list in any order during a 45 s recall period.

Kahana and Howard (2005). 65 participants performed delayed free recall of word lists with either massed or spaced repetitions of the list items (only the massed condition was included in the present study.) The 30 words were presented auditorally at a rate of one per 1500 ms, repeated three times in a row. For the purposes of the temporal contiguity analyses here, we redefined the serial position of each item as its position in the thirty-item list of unique words presented. That is, if a list started ABSENCE, ABSENCE, ABSENCE, HOLLOW, HOLLOW, HOLLOW . . . , the word ABSENCE was assigned "serial position" 1 and the word HOLLOW was assigned "serial position" 2. After the presentation of the last item, participants performed math problems of the form $A + B + C = ?$, where A, B, and C are positive, single-digit integers, until they answered 15 problems correctly in a row. After completing the self-paced distractor task, which took on average 45 ms, participants recalled the words on the just-studied list in any order during a 90 s recall period.

Bridge (2006). In single sessions, 119 participants performed free recall of 18 lists. Each list was made up of 25 nouns drawn randomly and without replacement from the Toronto Word Pool (Friendly et al., 1982). Words were presented visually for a maximum of 1100 ms each, with a 200 ms ISI. During each word presentation, participants were required to indicate if the word was "concrete" or "abstract" by pressing either the left or right control keys within the 1100 ms time limit. Once they made their response, the ISI period was initiated. After the presentation of the last item, participants performed math problems of the form $A+B+C=?$, where A, B, and C are positive, single-digit integers, for 30 s before recalling the words on the just-studied list in any order during a 60 s recall period. Note that 57 trials (never more than 6 from any participant) were excluded due to a combination of mechanical failure and experimenter error.

Sederberg et al. (2006). Across three separate testing sessions, 35 participants performed free recall of 48 lists. Lists were composed of 15 high-frequency nouns presented visually for 1600 ms with a 800–1200 ms blank ISI. After the presentation of the last item, participants performed math problems of the form $A + B + C = ?$, where A, B, and C are positive, single-digit integers, for 20 s before recalling the words on the just-studied list in any order during a 45 s recall period.

Zaromb et al. (2006), Exp. 1. In single sessions, 100 participants performed free recall of 16 lists, each of which contained 20 common nouns drawn from the Toronto Word Pool (Friendly et al., 1982). The lists were designed such that the first two lists were each composed of 20 unique words. The remaining 14 lists each contained up to 4 items repeated from 1, 2, 4, or 8 lists back, randomly selected from within that list. Words were presented visually for 1400 ms, followed by a 200 ms ISI. After the presentation of the last item, participants performed math problems of the form $A+B+C=?$, where A, B, and C are positive, single-digit integers, for 16 s before recalling the words on the just-studied list in any order during 90 s recall period.

Zaromb et al. (2006), Exp. 2. In single sessions, 63 participants performed free recall of 14 lists, each of which contained 20 common nouns drawn from a modified version of the Toronto Word Pool (Friendly et al., 1982) that had words with negative connotations removed. The lists were designed such that the first four lists were each composed of 20 unique words. Of the remaining 10 lists, 3 lists contained all new items and 7 lists contained 6 items repeated from 1, 2, and 3 lists back (i.e., two from one list back, two from two lists back, and two from three lists back), randomly selected from within that list. Words were presented visually for 1400 ms, followed by a 200 ms ISI. After the presentation of the last item, participants performed math problems of the form $A + B + C =?$, where A, B, and C are positive, single-digit integers, for 16 s before recalling the words on the just-studied list in any order during the 90 s recall period.

Sederberg et al. (unpublished). Across three separate testing sessions, 27 participants performed free recall of 48 total lists (16 per session). Each list was generated to ensure that words with varying degrees of semantic relatedness occurred at both adjacent and distant serial positions. Noun pairs from the word pool were divided into four groups of increasing semantic relatedness based on the word association spaces norms (WAS, Steyvers, Shiffrin, & Nelson, 2004), a computational measure of semantic similarity derived from free association norms (Nelson, McKinney, Gee, & Janczura, 1998). Two pairs of items from each of the four groups (i.e., 16 items per list) were selected without replacement for each list and arranged such that one pair occurred at adjacent serial positions and the other pair was separated by at least two other items. Each word was presented visually for 1000 ms with a 300–700 ms blank ISI. After the presentation of the last item, participants performed math problems of the form $A+B+C =?$, where A, B, and C are positive, single-digit integers, for 20 s before recalling the words on the just-studied list in any order during the 45 s recall period.

Analysis Details

The following sections provide the details of how we quantified temporal contiguity and semantic proximity effects and related them to recall performance.

Lag-CRP. The conditional response probability as a function of serial position lag (lag-CRP) was calculated according to the method introduced in Kahana (1996). For each participant, we initialized a set of numerators and denominators to zero, one for each possible transition lag. For example, if the list length was 12, there are 22 possible transition lags, from -11 to 11, excluding the lag of zero because we do not count transitions to the same word in the list. Then, for each list, we step through each recall transition, incrementing the numerator value matching the actual serial position lag of that transition and incrementing the denominators matching the set of all possible recall transitions. Transitions to and from intrusions and repetitions of already-recalled words are ignored. After incrementing the numerators and denominators for all of a participant's lists, the lag-CRP for that participant is simply the numerator divided by the denominator for each possible lag.

Sem-CRP. The conditional response probability as a function of semantic relatedness (sem-CRP) was calculated using a modification of the method outlined in Howard and Kahana (2002b). First, we determined semantic relatedness values, provided by the latent semantic analysis norms (LSA, Landauer & Dumais, 1997), for each pair of words in the wordpool. As with the lag-CRP analysis, we began by initializing a set of numerator and denominator bins for possible transitions. Instead of using 100 equal-sized bins, to remove the possibility that two or more possible transitions came from the same bin of semantic similarity values, we treated each similarity value as a separate bin. Then, for every transition in each list, the numerator bin corresponding to a word pair was incremented if, like in Howard and Kahana (2002b), the word was both in the current list and had not already been recalled. Similarly, the denominator bins were incremented for each possible transition to a non-recalled word on the list. We discarded transitions between words where the LSA-values were unknown. After processing every list in this fashion, we divided the numerators by the denominators for each bin, giving rise to the conditional response probability for each value of semantic relatedness.

For plotting and regression purposes, the probability values were reduced to 10 bins by taking the weighted sum of the values in each bin. Because there are more word pairs with low semantic relatedness values, we linearly increased the size of the bins such that the range of semantic relatedness values covered by the highest bin was four times that covered by the lowest bin. This has the effect of more evenly distributing the data that gives rise each point on the abscissa. To calculate the weighted sum, we first calculated the weight that each denominator value contributed to the total in the new, larger bin by dividing each denominator value in the bin by the sum of all of the denominator values in the bin. We then multiplied these weights by the probability values previously calculated for each value of semantic similarity and summed over these weighted probabilities to get the final bin value.

Temporal and semantic factors. Unlike the traditional conditional response probability analyses described above, which provide a measure of temporal contiguity or semantic proximity based on many recall transitions, the temporal and semantic factor analyses provide a non-parametric measure of temporal contiguity or semantic proximity for each recall transition, relative to the possible recall transitions available at any given time. For each transition, we rank all possible transitions in order of the negative absolute value of serial position lag (for the temporal factor) and in order of semantic relatedness (for the semantic factor). If there is a tie, the mean rank of the ties is shared between all members of the tie (for example, the ranks of the item transitions lags $[1, 2, 2, 3]$ would be $[1, 2.5, 2.5, 4]$.) To determine the temporal and semantic factors for that transition, we determine where the rank of the actual transition falls in the distribution of other

Table 1

Details about each experiment, including number of participants, list length, presentation rate, orienting task, distractor duration, and wordpool size. All experiments employed delayed free-recall paradigms. In Kahana & Howard (2005), each item was presented three times in succession, effectively acting as a 4500 ms presentation rate. Note that the total number of participants and lists for each experiment may vary from their published manuscripts. # Part.= number of participants. # Lists = number of lists each participant studied. LL = list length. PR = presentation rate, in ms. Orient = whether there was a required orienting task during study. Dist. Dur. = duration of the math distractor, either in time or number of consecutive correct answers. Pool = size of the wordpool from which the lists were drawn.

Study	# Part.	# Lists	LL	PR	Orient	Dist. Dur.	Pool
Howard & Kahana (1999), Exp. 1	60	641	12	1000	Yes	10 s	480
Howard & Kahana (1999), Exp. 2	16	647	12	1200	Yes	16 s	480
Kahana et al. (2002), Exp. 2	25	500	10	1400	No	16 s	482
Kahana & Howard (2005)	65	260	30	3×1500	No	15 correct	482
Bridge (2006)	119	2085	25	1100	Yes	30 s	482
Sederberg et al. (2006)	35	1680	15	1600	No	20 s	308
Zaromb et al. (2006), Exp. 1	100	1600	20	1400	No	16 s	482
Zaromb et al. (2006), Exp. 2	63	882	20	1400	No	16 s	455
Sederberg et al. (unpublished)	27	1296	16	1000	No	20 s	1156

ranks by means of the following equation, $(R-1)/(N-1)$, where R is the rank of the actual transition and N is the number of possible transitions. Consequently, each transition receives a number between 0.0 and 1.0 where factors greater than 0.5 indicate that the participant selected a related word and factors less than 0.5 indicate that the participant selected a unrelated word, relative to all possible valid transitions at that time. As with the conditional response probability analyses, we ignored any transition to or from an incorrect or repeated word.

Regressions. We performed standard least-squares regressions to determine the correlation between participants' percent recall and their mean temporal and semantic factors. In order to account for methodological differences between experiments, we normalized the percent recall for each participant in an experiment by first subtracting the mean percent recall across all participants for the specific experiment and then adding back the mean percent recall across all experiments. This measure of percent recall is used in every across-experiment analysis.

Permutation test. To ensure that differences in temporal factor did not arise from increases in recall of only one part of the list (as would be the case if a participant exhibited a high level of primacy or recency), we compared our results to a null distribution of temporal and semantic factors for each participant generated by means of a permutation test.

We shuffled the order of responses within and between each participant's lists 200 times, producing 200 sets of temporal factor values for each transition. In the first stage of shuffling the data, we looped through each recall (as identified by its serial position in the study-list) and randomly swapped it with the serial position of a recalled item in another list, making sure that neither item's recalled serial position already existed in the other list. This swapping procedure ensures that the total number of recalls of each serial position remains constant for each participant, yet the specific items recalled on each list are shuffled. After randomly swapping items between lists, we then shuffled the order of responses within each list to further ensure that the null distribution of recall transitions was what would be expected by chance.

We then applied the same analysis to the null distribution as we did to our actual data. For example, to assess whether the slope of the probability of recall as a function of temporal factor was significant, we generated 200 slopes, one for each permutation of our null distribution, and then compared the null distribution of slopes to the actual slope. If the actual slope was greater than all the slopes in the null distribution, this would be analogous to a p-value less than $1/200 = 0.005$.

Results

Temporal and semantic relatedness drives recall

Our first goal was to quantify the overall contiguity effects exhibited by the participants in these studies. Figure 1 reveals a robust temporal contiguity effect, both in recall probability (left) and in recall latency (right). Participants were more likely to transition to items studied at serial positions near to the just-recalled item. In addition, transitions to nearby items were significantly faster than to items from distant serial positions.

Participants also demonstrated a significant semantic proximity effect. As seen in Figure 2, participants exhibited an increased probability of transitioning to items that were semantically-related, as measured by latent semantic analysis (LSA, Landauer & Dumais, 1997), to the just-recalled item (left). Along with the increase in recall probability, transitions to semantically-related items were also faster than to less-related items (Figure 2, right). The semantic proximity effect is evident even at low LSA similarity values (below 0.3), where semantic relatedness is difficult to perceive by simple examination of word pairs (e.g., NUMBER and JOURNAL have a LSA-similarity of 0.11, while PONY and FOREHEAD have an LSA-similarity of 0.21).

Although traditional conditional response probability measures are well-suited for across-subject analyses, we

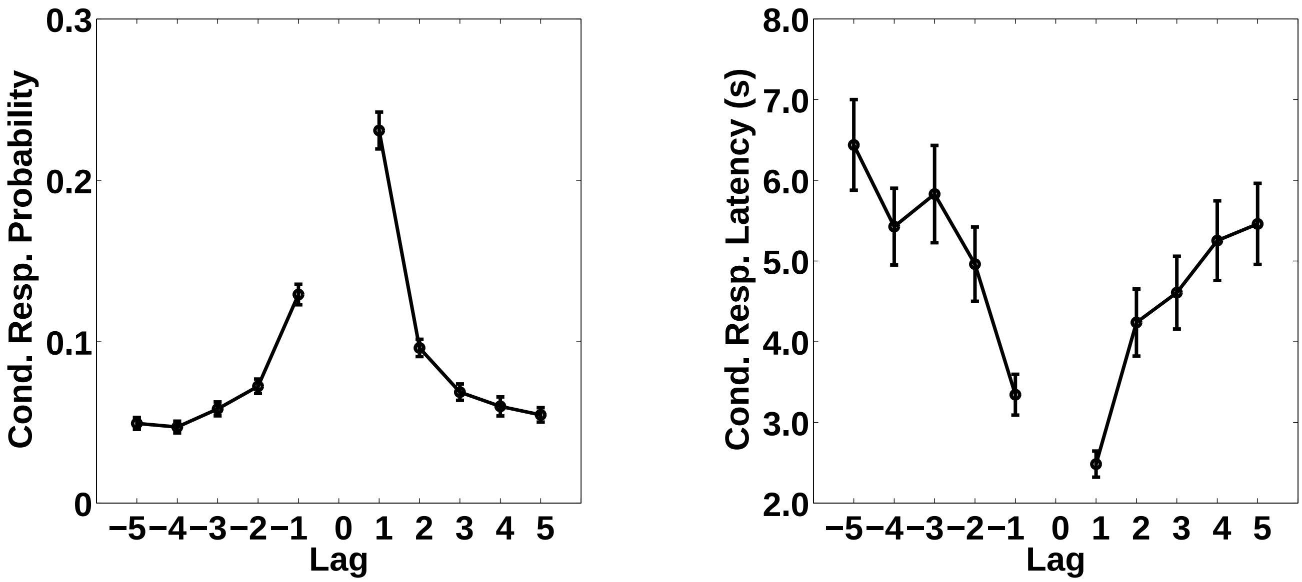

Figure 1. *Temporal contiguity effect.* Conditional response probability (left) and latency (right) as a function of serial position lag. Errors are 95% confidence intervals calculated across all participants.

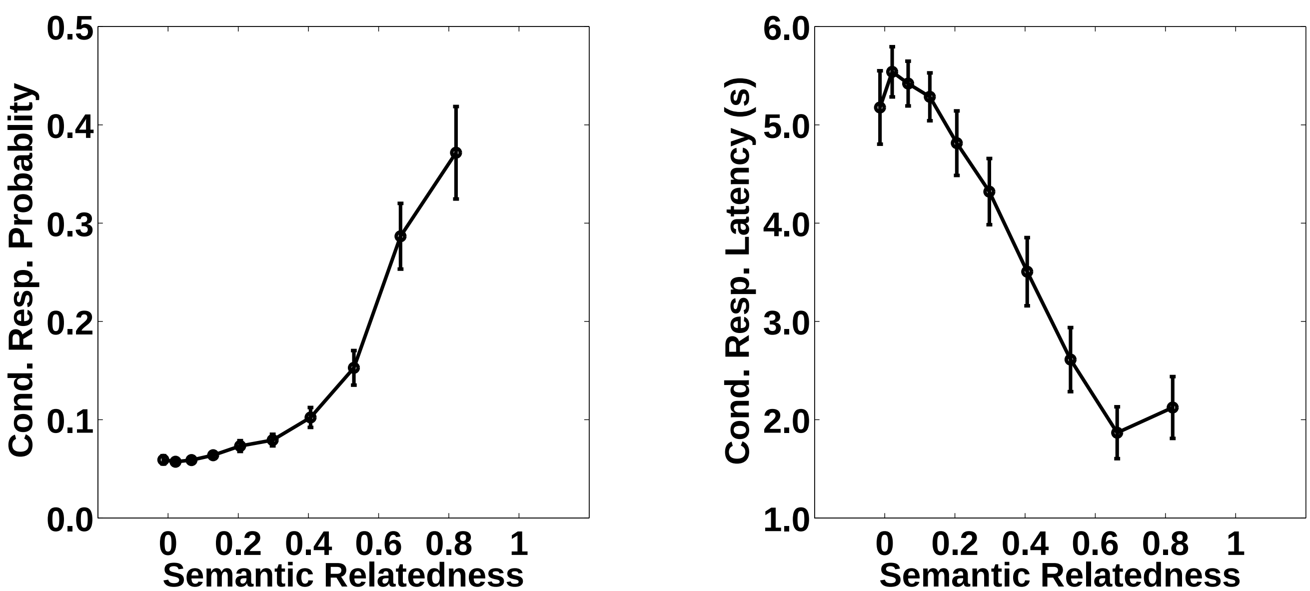

Figure 2. *Semantic proximity effect.* Conditional response probability (left) and latency (right) as a function of semantic relatedness calculated with the LSA measure of semantic relatedness between words. Errors are 95% confidence intervals calculated across all participants.

needed a stable measure of the temporal contiguity and semantic proximity effects for individual participants.[1] To this end, we employed a non-parametric measure of conditional response probability based on ranking the temporal contiguity and semantic proximity of each transition with respect to all possible transitions at that time (Polyn, Norman, & Kahana, 2009). The measures, called temporal and semantic factors, assign a value between 0.0 and 1.0 to each recall transition, taking into account all possible transitions to non-recalled words (i.e., all the unrecalled words at that time.) A factor of 1.0 indicates the participant transitioned to the most temporally or semantically proximal item, whereas a factor of 0.0 indicates that the participant transitioned to the least temporally or semantically proximal item. We can then average the factors across participant responses and test whether the distributions of temporal and semantic factors are different from 0.5, the value one would expect if there was no temporal contiguity or semantic proximity effect in either direction (i.e., what one would expect by chance).

As with the traditional conditional response probability analyses, the temporal and semantic factors indicated that participants relied on both temporal and semantic information to drive their responses. The mean temporal factor across all participants was 0.614 (t-test comparing the temporal factors to 0.5, $t = 31.171$, $df = 509$, $p < 10^{-5}$), while the mean semantic factor was 0.541 ($t = 16.681$, $df = 509$, $p < 10^{-5}$).[2] Although participants relied on both temporal and semantic associations to guide recall, there was no significant correlation between temporal and semantic factors across participants (Pearson's $r = 0.010$, $p = .839$).

Temporal and semantic factors are calculated based on individual recall transitions and are, therefore, independent from the total number of participant responses per list. However, contiguity effects are not immune to influence from changes in the level of recall due to primacy and recency effects. Participants tend to cluster together items in these areas, and, consequently, it is possible that the temporal factor, in particular, would be inflated due to participants recalling more items from the primacy and recency portions of the list.

To address this potential confound, we performed a permutation analysis whereby we shuffled participant responses both within and between lists, while ensuring that their resulting serial position curve remained unchanged (see Method). The null distribution of temporal factors calculated from the shuffled responses had a mean and standard error of 0.489 ± 0.0001 and all 200 shuffles were less than our actual temporal factor of 0.614. As such, participants still exhibited significant clustering based on temporal contiguity even when accounting for primacy and recency effects ($p < 0.005$, based on the number of permutations). As with the temporal factors, the mean semantic factor across participants was significantly greater than all the semantic factors in the null distribution, which had a mean and standard error of 0.501 ± 0.0001, verifying that participants significantly clustered recalls by semantic relatedness ($p < 0.005$).

Semantic and Temporal Factors and Recall Performance

Having demonstrated that participants performing a free-recall task rely on both temporal and semantic associations to drive their responses, we sought to determine whether individual differences in the degree to which participants relied on semantic and episodic associations related to overall recall performance.

Given that the studies included in this meta-analysis varied on several methodological dimensions (see Table 1), such as list length, presentation rate, and distractor duration, the mean recall probability varied across studies. To ensure that this variability did not corrupt our analysis, we normalized each subject's recall probability by subtracting the within-study mean and adding back the grand across-study mean. Thus, all recall probability values reported in the following figures and tables are normalized probabilities of recall.

To test for a correlation between participants' tendencies to transition to items from nearby serial positions and their recall performance, we plotted the mean proportion of items recalled as a function of mean temporal factors. The left panel of Figure 3 reveals the positive correlation between temporal factor and recall performance. The regression across participants split into deciles was highly significant (standardized $\beta = 0.883$, $p < .001$.)

As with the temporal factor calculations above, it is possible that the preceding analyses were biased by recency or primacy effects in the data. This could give rise to artificial increases in the temporal factor due to participants clustering items presented in the recency or primacy portions of the study-list. This potential confound is especially relevant to the current analysis because participants who exhibit increased recall performance as a result of of an increase in primacy may also exhibit an artificial increase in temporal contiguity, as measured by the temporal factor. Consequently, we performed the same regression analysis for all members of the null distribution calculated via the permutation procedure. Although the standardized β-values were greater than zero for the null-distribution (0.198 ± 0.012 SEM), the β for the actual data was greater than all values in the null distribution, indicating that there was a significant effect of temporal

[1] When calculating the lag- or sem-CRP for individual participants, the resulting data points have large variance, making it difficult to fit curves to estimate the magnitude of the effects.

[2] Although the mean semantic factor across participants was highly significant, the magnitude of the effect, compared to the temporal factor, was smaller than what might be expected based on the results of the traditional conditional response probability analyses. This is due to the fact that in the studies reported here, semantically-related words rarely occur on the same list, thus providing few opportunities for recall transitions to strongly-related words. Thus, a majority of the possible transitions during recall are between list-items that are either not related or only slightly related, which tend to exhibit much smaller output order effects. Given that each recall transition is weighted equally, this lessened the overall size of the semantic factor effect.

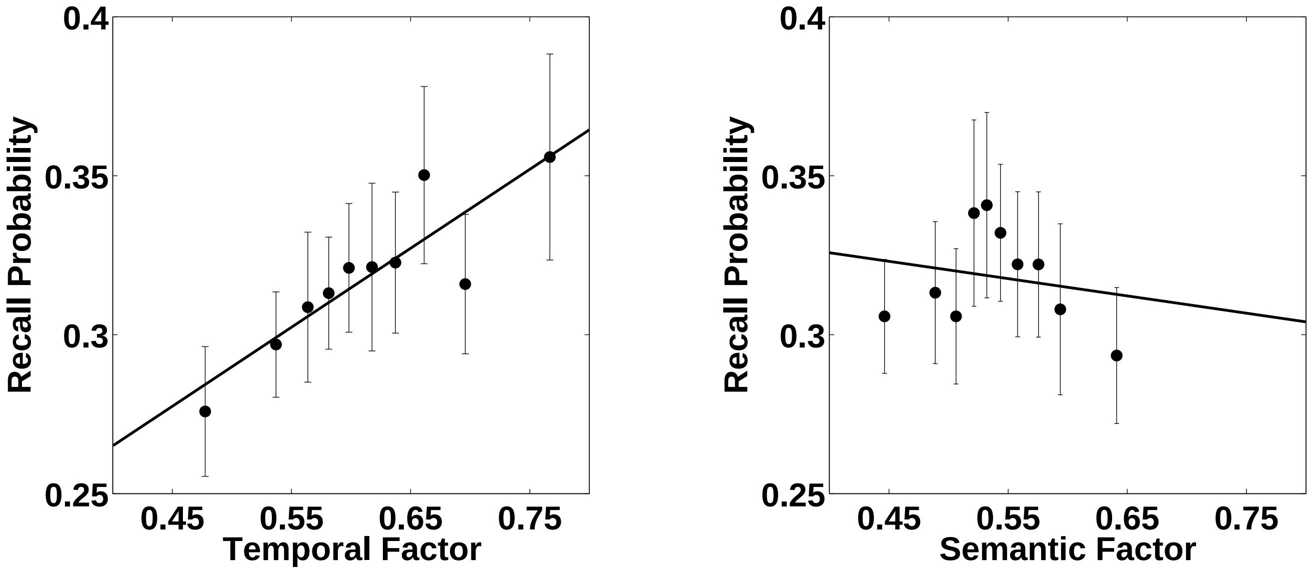

Figure 3. *Temporal/Semantic factor and recall performance.* Normalized probability of recall as a function of the temporal factor (left) and semantic factor (right). Each data point represents one decile of the total participants. Errors are 95% confidence intervals calculated across all participants in each decile. The line represents the regression fit to the data points.

factor on recall performance ($p < .005$).

The right panel of Figure 3 plots the same correlation analysis with the semantic factors for each participant. The regression fit reveals a negative trend across deciles, but it was not significant (standardized $\beta = -0.193$, $p = 0.592$) and, thus, the degree participants clustered recalls based on semantic relatedness between items did not predict overall recall performance.

The regression analyses performed for Figure 3 include individuals from all experiments split into deciles. We also performed a similar regression analysis across participants for each experiment independently (see Table 2). While not all the experiments demonstrated a significant positive correlation between temporal factor and recall performance, they each exhibited positive β coefficients for temporal factor predicting recall, explaining why the aggregated data exhibited such a strong effect. It is also clear from Table 2 that our finding of a robust correlation of probability of recall with the magnitude of the temporal factor does not depend on any individual study. Removing any individual study from the regression across participants split into deciles retains the highly significant positive correlation between recall performance and temporal factor (least significant result: standardized $\beta = 0.823$, $p < .003$).

General Discussion

Through a reanalysis of individual trial data from nine delayed free-recall conditions reported in previous studies, we have attempted to uncover the relationship between temporal contiguity and semantic proximity (as measured by response order) and overall recall performance. We first replicated the finding that participants performing delayed free-recall rely on both temporal and semantic associations to drive retrieval. By means of a recently-developed method of quantifying temporal and semantic proximity effects, we found that participants who exhibit greater temporal contiguity in the order of their responses recall more words. In contrast to the extremely robust correlation between the temporal proximity effect and probability of recall, increases in the semantic factor did not give rise to significant increases in participants' recall performance (Figure 3). In fact, there was a trend towards there being a decrease in probability of recall with more reliance on semantic information to guide retrievals. This distinction between the effects of temporal and semantic factors on recall is of theoretical interest if it can be taken at face value. However, there are some empirical caveats that must be considered before continuing on to discuss the theoretical implications.

Empirical caveats

We can not completely rule out the possibility that our failure to observe a correlation between the semantic factor and probability of recall could reflect weaker sensitivity of our semantic measures. Presumably temporal factors are precisely measured by comparison of serial position, whereas our estimate of semantic proximity relies on LSA. Perhaps noise in this estimate reduces the correlation between semantic factors and probability of recall. While there is no way to rule this out, there are several pieces of evidence arguing against it. First, LSA provided strong sensitivity in predicting recall transitions. Figure 2 demonstrates that there were systematic differences in transition probability and latency, even for relatively small values of LSA $\cos\theta$. Second, we have repeated our analyses using the word association space (WAS Steyvers et al., 2004), a computational measure of semantic similarity derived from free-association

Table 2
Regression results for each experiment. Each row contains the results from a regression predicting the recall performance of the participants in each study based on their temporal and semantic factors. The symbols indicate the significance of the standardized beta coefficients (+: $0.1 < p < 0.05$; ∗: $0.01 < p < 0.05$; ∗∗: $p < 0.01$).

Study	Temporal Factor	Semantic Factor
Howard & Kahana (1999), Exp 1	$\beta = 0.166$	$\beta = -0.210$
Howard & Kahana (1999), Exp 2	$\beta = 0.561^{*}$	$\beta = 0.059$
Kahana et al. (2002), Exp 2	$\beta = 0.391^{+}$	$\beta = 0.046$
Kahana & Howard (2005)	$\beta = 0.499^{**}$	$\beta = 0.034$
Bridge (2006)	$\beta = 0.083$	$\beta = 0.048$
Sederberg et al. (2006)	$\beta = 0.289$	$\beta = -0.005$
Zaromb et al. (2006), Exp 1	$\beta = 0.233^{*}$	$\beta = -0.043$
Zaromb et al. (2006), Exp 2	$\beta = 0.249^{+}$	$\beta = -0.040$
Sederberg et al. (unpublished)	$\beta = 0.418^{+}$	$\beta = 0.115$

norms (Nelson et al., 1998), in place of LSA and observed the same pattern of results.[3] It is certainly possible that all extant computational models of semantic similarity are insufficiently sensitive—perhaps the semantic relations between words that control recall transitions show substantive individual differences based on participants' idiosyncratic history of exposure to the items. It is also the case that the distribution of values of semantic proximity, which roughly follows an ex-Gaussian distribution (Howard & Kahana, 2002b) is different than the distribution of temporal contiguities.[4] Moreover, the recall-to-recall variability in the distribution of semantic proximities to non-recalled items fluctuates more widely than the variability in non-recalled serial position lags. It is conceivable that one or both of these factors results in reduced sensitivity of the semantic measure independently of noise in estimating semantic similarities between pairs of words.

It is also important to note that, with one exception, the studies analyzed in this manuscript did not explicitly control the number and strength of semantic associations between items within the lists. When selected randomly from a pool of high-frequency nouns, strong semantic associations between items are rare, making reliance on semantic relatedness between items an ineffective retrieval strategy. The lack of a positive correlation between semantic factor and recall performance is, therefore, not surprising given that one of the main reasons participants have trouble recalling is that they are unable to selectively target the list items as distinct from the rest of the items in their memory. Whereas relying on inter-item associations formed during encoding helps to focus recall on the list context, relying on preexisting semantic associations can lead to recalls from outside the target list.

Implications for models of free recall

The fact that participants who exhibit greater temporal contiguity also recall more words indicates that relying on associations formed between words that were presented nearby in time to drive recall is a fundamental aspect, if not the fundamental aspect, of successful episodic memory performance. Interestingly, this finding is not an unavoidable prediction of extant free recall models. These models have many components that could in principle lead to increased probability of recall in a variety of ways. Only some of these options would lead to the effects observed here. We describe several concrete examples in the context of two widely-used models of free recall, the search of associative memory model (SAM, Raaijmakers & Shiffrin, 1980, 1981) and the temporal context model (TCM, Howard & Kahana, 2002a; Sederberg et al., 2008; Polyn et al., 2009).

In SAM, direct item-to-item associations form between items while they co-inhabit the short-term store (STS) buffer. These newly-formed associations are responsible for temporal contiguity effects in SAM (Kahana, 1996; Sirotin, Kimball, & Kahana, 2005). While boosting the strength of these newly-learned associations would, all other things being equal, lead to enhanced probability of recall, there are numerous other factors in SAM that could also result in enhanced probability of recall. For instance, in SAM it is assumed that the items in the study list are associated with a context state that is unique to the list (Raaijmakers & Shiffrin, 1980, but see Mensink & Raaijmakers, 1988; Sirotin et al., 2005). Enhancing the magnitude of the context-to-item association would result in an increase in the probability of recall. Similarly, parameters that control the number of attempts in memory search would also tend to result in an increase in the probability of recall. Variability in either of these other factors, however, would not predict the results we observed because the increase in recall would occur without a corresponding increase in temporal contiguity.

In SAM, temporal contiguity effects result from increments in the long-term memory strength between items that co-occur in the buffer. Variability in this same matrix of direct item-to-item connections in long-term memory is responsible for semantic proximity effects (Sirotin et al., 2005). Because temporal and semantic proximity effects rely on the same component of the model, other factors that could control the number of recalled items ought to have similar effects on the sensitivity of the search process to temporal contiguity and semantic proximity. The fact that there was a large corre-

[3] These analyses were complicated by the fact that not all words in our experiments have been normed, meaning that those words must be excluded from our analyses because WAS values are not available for transitions involving them.

[4] The distribution of available lags is uniform (modulo removal of already-recalled items) on any given recall attempt. However, calculation of the temporal factor uses the absolute value of lag.

lation between the temporal proximity effect and the number of recalled items, but that we failed to observe a similar relationship between for semantic effects suggests, in the context of SAM, that the encoding of new item-to-item associations in STS is the major factor determining probability of recall. If other factors contribute, our results suggest that they must be weak relative to the effect of newly-learned associations due to co-occurrence in STS, that multiple such factors contribute but conflict with their tendency to amplify the importance of the item-to-item associations in determining output order, or that there are factors that affect probability of recall but that do not affect temporal and semantic proximity effects.

Similarly, the presence of a correlation between the temporal contiguity effect and probability of recall, and the lack of a correlation for the semantic proximity effect, places constraints on TCM. In order to understand these constraints, some background on the basis of temporal and semantic proximity effects in TCM is necessary. Unlike in SAM, neither temporal contiguity nor semantic proximity effects are due to direct item-to-item associations in TCM. Instead, the cue for recall is the current state of a temporal context vector. Behavioral associations between two items are a consequence of the effect the first item has on the state of context, and the overlap between that state of context and the encoding context of the second item. Temporal contiguity effects result from the similarity of the context recovered by a cue item with the encoding context of the target item. One might say that in TCM behaviorally observed associations between two items are a consequence of item-to-context-to-context-to-item associations.

Recent work has attempted to understand the role of semantic associations in generating retrievals within TCM. Sederberg et al. (2008) assumed that the context-to-item matrix that defines the optimal contextual cue for an item can be decomposed into a pre-experimental part that reflects the long-term history of the contexts in which each item is experienced, and a newly-learned part that reflects the context-to-item associations formed during study of the list. These two matrices can be understood to reflect the semantic and episodic contextual encoding, respectively. Polyn et al. (2009) set the entries in the pre-experimental context-to-item matrix to be equal to the LSA cosine distance between items, enabling them to account for semantic proximity effects simultaneously with a wide variety of temporal and source memory effects in immediate free recall. Other recent theoretical work (Shankar, Jagadisan, & Howard, in press; Howard, Shankar, & Jagadisan, submitted) has shown that it is possible to use temporal context derived from long-term exposure to language (modeled by training on a large corpus of naturally-occurring text) to assemble both a pre-experimental context-to-item matrix and a pre-experimental item-to-context matrix in a principled way. These additions have been shown to allow the model to learn semantic relationships between items, mediated by temporal context, that perform comparably well to LSA in capturing semantic similarity effects between words.

At its core, TCM proper determines rules for the changes in the state of context due to item presentations and retrievals and the way context cues for retrieval of the studied items. It is still necessary to creating a mapping between the degree to which each potential recall is activated by the context cue and the behaviorally-observed probability of recall. In recent applications, this mapping is accomplished *via* a set of competing accumulators that grow based on the strength of the item activations (Sederberg et al., 2008; Polyn et al., 2009). There are numerous changes to the accumulator retrieval process that would tend to amplify variability in the input into the accumulators and increase the probability of recalling an item. Because the output of TCM reflects both temporal and semantic factors, such changes in the accumulators would be expected to result in similar correlations between temporal and semantic factors and probability of recall. The fact that the correlation of these factors with probability of recall is different argues against variability in the sensitivity of the accumulators being a major factor determining probability of recall.

Moreover, the present findings suggest that the pre-experimental and newly-learned context-to-item matrices do not strategically trade off with one another on the scale of an individual participant (it still remains possible that temporal and semantic factors trade off during the recall of a single list.) If they did, one would expect a negative correlation between the magnitude of the temporal and semantic factors across participants (no such correlation was found). Consequently, the current findings are most consistent with the hypothesis that the dominant factors in determining the number of recalled items in TCM are the magnitude of new context-to-item associations formed during study and the ability to reconstruct contextual states from the study list during retrieval.

Link to older participants

Studies exploring the effects of aging on episodic memory provide additional evidence for the interaction between temporal and semantic associations and recall performance. Kahana et al. (2002) found that, while recency effects remain unchanged between younger and older participants performing free recall, both the overall level of recall and the magnitude of the contiguity effect is lower in older participants. This suggests that older participants lie upon the lower end of the continuum reported in the present study. As such, the same explanation for poorer performance in older participants could apply to the current results, as well. Howard, Wingfield, and Kahana (2006) were able to capture the decrease in contiguity in older adults with TCM simulations by decreasing the effectiveness of contextual retrieval. That is, in those simulations, older participants were less able to retrieve the context that was present when they studied a just-recalled item, giving rise to a weaker cue for the subsequent recall of nearby items than they would have had with full contextual retrieval. Failure to bind items within their episodic context and to retrieve those associations would give rise to both decreased performance and less contiguity (Naveh-Benjamin, 2000; Naveh-Benjamin, Hussain, Guez,

& Bar-On, 2003). Interestingly, a change in the magnitude of context-to-item associations across age groups would have been inconsistent with these data as it would have predicted a change in the recency effect across groups.

Other studies suggest that older participants supplement their lack of temporal associations with semantic associations between items. Golomb, Peelle, Addis, Kahana, and Wingfield (2008) studied free and serial recall in young and older adults and found that increased temporal contiguity, especially in serial recall, correlated with better performance (see also Lewandowsky, Brown, & Thomas, 2009, for a similar link between temporal contiguity and performance in young participants performing short-term serial recall of letters.) Critically, Golomb et al. (2008) found that older adults tend to rely on semantic information to guide retrieval, even when it hurts recall performance in serial recall. This suggests that older participants will rely on semantic cues to make up for the lack of an effective temporal cue to drive retrieval. Consequently, it is not surprising that older participants also tend to make more intrusions during recall tasks and that intrusions are typically semantically related to the last correct recall (Kahana, Dolan, Sauder, & Wingfield, 2005; Zaromb et al., 2006).

Taken together, these results further support the hypothesis that the ability to form and retrieve item-to-context associations correlates with both recall performance and the magnitude of the temporal contiguity observed between recall transitions. When participants fail to retrieve context, they rely more on semantic relatedness to drive recall, even when semantic associations are not an effective retrieval cue.

Conclusions

We have shown a positive correlation between the temporal contiguity that participants exhibit between recalls and their overall recall performance. Individual differences in the semantic proximity between recalls, however, did not significantly correlate with recall performance. These results suggest that forming and retrieving associations between items that occur nearby in time, possibly mediated by temporal context, is the essence of episodic memory. They further suggest that, to the degree that participants can bias their retrieval strategy during the course of recall, the more a participant relies on temporal associations to drive retrieval, the better they will perform.

References

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

Bridge, D. (2006). Memory and cognition: What difference does gender make? *Unpublished Honor's thesis, Syracuse University*.

Brown, S., Conover, J., Flores, L., & Goodman, K. (1991). Clustering and recall: Do high clusterers recall more than low clusterers because of clustering. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *17*, 710–721.

Davelaar, E. J., Goshen-Gottstein, Y., Ashkenazi, A., Haarmann, H. J., & Usher, M. (2005). The demise of short-term memory revisited: Empirical and computational investigations of recency effects. *Psychological Review*, *112*, 3–42.

Farrell, S., & Lewandowsky, S. (2008). Empirical and theoretical limits on lag recency in free recall. *Psychonomic Bulletin and Review*, *15*, 1236-1250.

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.

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.

Greene, R. L., & Crowder, R. G. (1984). Effects of semantic similarity on long-term recency. *American Journal of Psychology*, *97*, 441-449.

Howard, M. W., Addis, K. A., Jing, B., & Kahana, M. J. (2007). Handbook of latent semantic analysis. In T. K. Landauer, D. S. McNamara, S. Dennis, & W. Kintsch (Eds.), *LSA: A road towards meaning* (pp. 121–141). Mahwah, NJ: Laurence Erlbaum and Associates.

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*, 923–941.

Howard, M. W., & Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology*, *46*, 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., Sederberg, P. B., & Kahana, M. J. (in press). Reply to Farrell & Lewandowsky: Recency-contiguity interactions predicted by the temporal context model. *Psychonomic Bulletin & Review*.

Howard, M. W., Shankar, K. H., & Jagadisan, U. K. K. (submitted). Constructing semantic representations from a gradually-changing representation of temporal context. *Topics in Cognitive Science*.

Howard, M. W., Wingfield, A., & Kahana, M. J. (2006). Modeling the age-related associative deficit in self-initiated recall. *Psychonomic Bulletin & Review*, *13*(3), 439–445.

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

Kahana, M. J., Dolan, E. D., Sauder, C. L., & Wingfield, A. (2005). Intrusions in episodic recall: Age differences in editing of overt responses. *Journal of Gerontology: Psychological Sciences*, *60*, 92–97.

Kahana, M. J., & Howard, M. W. (2005). Spacing and lag effects in free recall of pure lists. *Psychonomic Bulletin & Review*, *12*, 159-164.

Kahana, M. J., Howard, M. W., & Polyn, S. M. (2008). Associative retrieval processes in episodic memory. In H. L. Roediger III (Ed.), *Cognitive psychology of memory. Vol. 2 of Learning and memory: A comprehensive reference, 4 vols. (J. Byrne, Editor).* 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*, 530–540.

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.

Lewandowsky, S., Brown, G. D. A., & Thomas, J. L. (2009). Traveling economically through memory space: Characterizing output order in memory for serial order. *Memory & Cognition*, *37*, 181–193.

Mensink, G.-J. M., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434-455.

Naveh-Benjamin, M. (2000). Adult-age differences in memory performance: Tests of an associative deficit hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*, 1170-1187.

Naveh-Benjamin, M., Hussain, Z., Guez, J., & Bar-On, M. (2003). Adult Age Differences in Episodic Memory: Further Support for an Associative-Deficit Hypothesis. *Learning, Memory*, *29*(5), 826–837.

Nelson, D. L., McKinney, V. M., Gee, N. R., & Janczura, G. A. (1998). Interpreting the influence of implicitly activated memories on recall and recognition. *Psychological Review*, *105*, 299-324.

Pollio, H. R., Richards, S., & Lucas, R. (1969). Temporal properties of category recall. *Journal of Verbal Learning and Verbal Behavior*, *8*, 529-536.

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.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, p. 207-262). New York: Academic Press.

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

Sederberg, P. B., Gauthier, L. V., Terushkin, V., Miller, J. F., Barnathan, J. A., & Kahana, M. J. (2006). Oscillatory correlates of the primacy effect in episodic memory. *NeuroImage*, *32*(3), 1422–1431.

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.

Shankar, K. H., Jagadisan, U. K. K., & Howard, M. W. (in press). Sequential learning using temporal context. *Journal of Mathematical Psychology*.

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.

Steyvers, M., Shiffrin, R. M., & Nelson, D. L. (2004). Word association spaces for predicting semantic similarity effects in episodic memory. In A. F. Healy (Ed.), *Cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer.* Washington, DC: American Psychological Association.

Zaromb, F. M., Howard, M. W., Dolan, E. D., Sirotin, Y. B., Tully, M., Wingfield, A., et al. (2006). Temporal associations and prior-list intrusions in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(4), 792–804.