The temporal contiguity effect predicts episodic memory performance

PER B. SEDERBERG
Ohio State University, Columbus, Ohio

JONATHAN F. MILLER
University of Pennsylvania, Philadelphia, Pennsylvania

MARC W. HOWARD
Syracuse University, Syracuse, New York

AND

MICHAEL J. KAHANA
University of Pennsylvania, Philadelphia, Pennsylvania

One way to study the associative processes at work during episodic memory is to examine the order of participant responses, which reveal the strong tendency to transition between temporally contiguous or semantically proximal items on the study list. 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 the participants' temporal contiguity effects predicted their recall performance. When interpreted in terms of two models of episodic memory, 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. Sample code may be downloaded as a supplement for this article from http://mc.psychonomic-journals.org/content/supplemental.

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 the items come to mind, the 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+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 (2002) extended the analysis of recall transitions, measuring the conditional response probability as a function of the semantic relation 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 indicates that participants rely on both newly learned temporal and preexisting semantic associations 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, 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, individual differences in response order should correlate with overall recall performance.

P. B. Sederberg, sederberg.1@osu.edu

Here, we examine the covariation between individual differences in probability of recall and temporal contiguity and semantic proximity effects. Previous work that examined contiguity and proximity effects across conditions in which the probability of recall varied has shown a complex pattern of results. Consider the finding that is observed in multitrial 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, whereas the semantic proximity effect did not change. In this case, the contiguity effect covaried with recall probability across conditions, whereas the semantic proximity effect did not. In contrast, consider the relation 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 0 sec (delayed free recall) to 16 sec (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, yet the temporal contiguity effect did not change (Howard & Kahana, 1999). In contrast, the semantic proximity effect decreased as the length of the distractor interval between conditions was increased (Howard & Kahana, 2002). 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 on the basis of work in which recall of categorized lists was compared with that of uncategorized lists (Bousfield, 1953; Brown, Conover, Flores, & Goodman, 1991; Pollio, Richards, & Lucas, 1969). At the extreme, in which lists are drawn from a single category, participants recalled more items from the study list but suffered memory impairments for the order in which 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 relation 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 to which individual participants relied on temporal and semantic associations to drive responses to their overall recall performance. To provide an overview of our results, we find that the participants in these studies showed strong temporal contiguity and semantic proximity effects. The participants who exhibited a higher temporal contiguity effect recalled more items. Although individual differences in the semantic proximity effect did not significantly correlate with recall performance, this is likely because of the low reliability of the measure to quantify semantic transitions in individual participants in these data. In the Discussion section, 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 article.

Inclusion Criteria

We required each study to pass a strict set of criteria for inclusion. First, we included only studies in which 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 included only delayed free recall conditions, which have been shown to have consistent and significant temporal contiguity and semantic proximity effects (Howard & Kahana, 1999, 2002). 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 because of 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 (2002) 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 nonmonotonicity into the lag-CRP (Farrell & Lewandowsky, 2008; Howard, Sederberg, & Kahana, 2009).

Finally, we excluded any condition in which items had spaced repetitions within lists or that was performed by older participants, because 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 were nominally repeated (see below), each unique word had an unambiguous serial position with respect to the current list, and in each article, typical lag-CRP functions for the repeated items were reported.

Methodological Details of the Included Experiments

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

Howard and Kahana (1999), Experiment 1. In single sessions, 60 participants performed both immediate and delayed free recall of 25 total lists. Three participants were excluded from the original study because of experimenter error, and only the delayed free recall lists were included in this study. The first two lists for each participant were treated as practice, whereas the remaining 23 lists were randomly selected to be either immediate or delayed free recall (i.e., the participants performed different numbers of delayed free recall lists, ranging from 4 to 16 total lists, giving rise to 641 total lists across participants). Each list was composed of 12 randomly selected nouns from the Toronto Word Pool (Friendly, Franklin, Hoffman, & Rubin, 1982). The words were presented visually for 1,000 msec each. While each word was on the screen, the participants were required to perform a semantic orienting task, judging whether each word was concrete or abstract by pressing

Table 1
Details About Each Experiment

Experiment	N	No. of Lists	List Length (Words)	Presentation Rate (msec)	Orient	Distractor Duration	Pool Size
Howard & Kahana (1999), Experiment 1	60	641	12	1,000	Yes	10 sec	480
Howard & Kahana (1999), Experiment 2	16	647	12	1,200	Yes	16 sec	480
Kahana, Howard, Zaromb, & Wingfield (2002), Experiment 2	25	500	10	1,400	No	16 sec	482
Kahana & Howard (2005)	65	260	30	$3 \times 1,500$	No	15 correct	482
Bridge (2006)	119	2,085	25	1,100	Yes	30 sec	482
Sederberg et al. (2006)	35	1,680	15	1,600	No	20 sec	308
Zaromb et al. (2006), Experiment 1	100	1,600	20	1,400	No	16 sec	482
Zaromb et al. (2006), Experiment 2	63	882	20	1,400	No	16 sec	455
Sederberg, Miller, Barnathan, & Kahana (2010)	27	1,296	16	1,000	No	20 sec	1,156

Note—Delayed free recall paradigms were employed in all of the experiments. In Kahana and Howard (2005), each item was presented three times in succession, effectively acting as a 4,500-msec presentation rate. The total number of participants and lists for each experiment may vary from those in the published articles. No. of Lists, number of lists that each participant studied; Orient, whether there was a required orienting task during study; Distractor Duration, duration of the math distractor, either in time or number of consecutive correct answers; Pool Size, size of the word pool from which the lists were drawn.

either the left or right control key. After the presentation of the last item, the participants performed *true–false* math problems of the form $A + B + C = D$, where A, B, and C were positive, single-digit integers, for 10 sec. The participants then recalled the words on the just-studied list in any order during a 45-sec recall period.

Howard and Kahana (1999), Experiment 2. Over the course of 10 sessions, 16 participants performed four variants of free recall: one delayed and three continual-distractor with varying durations of a distractor-filled interstimulus interval (ISI). Only the delayed free recall lists were included in the present study. Each list was composed of 12 nouns selected at random and without replacement from the Toronto Word Pool (Friendly et al., 1982). The words were presented visually for 1,200 msec each. While each word was on the screen, the participants were required to perform a semantic-orienting task, judging whether each word was concrete or abstract by pressing either the left or the right control key. After the presentation of the last item, the participants performed *true–false* math problems of the form $A + B + C = D$, where A, B, and C were positive, single-digit integers, for 16 sec. The participants then recalled the words on the just-studied list in any order during a 60-sec recall period.

Kahana, Howard, Zaromb, and Wingfield (2002), Experiment 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 1,400 msec, followed by a 100-msec ISI. After the presentation of the last item, the participants performed math problems of the form $A + B + C = ?$, where A, B, and C were positive, single-digit integers, for 16 sec before recalling the words on the just-studied list in any order during a 45-sec recall period.

Kahana and Howard (2005). Sixty-five 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 auditorially at a rate of 1 every 1,500 msec, 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 30-item list of unique words presented. That is, if a list started with *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, the participants performed math problems of the form $A + B + C = ?$, where A, B, and C were 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 sec, the participants recalled the words on the just-studied list in any order during a 90-sec 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). The words were presented visually for a maximum of 1,100 msec each, with a 200-msec ISI. During each word presentation, the participants were required to indicate whether the word was concrete or abstract by pressing either the left or the right control key within the 1,100-msec time limit. Once they had made their response, the ISI period was initiated. After the presentation of the last item, the participants performed math problems of the form $A + B + C = ?$, where A, B, and C were positive, single-digit integers, for 30 sec before recalling the words on the just-studied list in any order during a 60-sec recall period. Note that 57 trials (never more than 6 from any participant) were excluded because of 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. The lists were composed of 15 high-frequency nouns presented visually for 1,600 msec with an 800- to 1,200-msec blank ISI. After the presentation of the last item, the participants performed math problems of the form $A + B + C = ?$, where A, B, and C were positive, single-digit integers, for 20 sec before recalling the words on the just-studied list in any order during a 45-sec recall period.

Zaromb et al. (2006), Experiment 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 2 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. The words were presented visually for 1,400 msec, followed by a 200-msec ISI. After the presentation of the last item, the participants performed math problems of the form $A + B + C = ?$, where A, B, and C were positive, single-digit integers, for 16 sec before recalling the words on the just-studied list in any order during a 90-sec recall period.

Zaromb et al. (2006), Experiment 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 4 lists were each composed of 20 unique words. Of the remaining 10 lists, 3 contained all new items and 7 contained 6 items repeated from 1, 2, and 3 lists back. Specifically, the repeated items on the final 7 lists contained 2 items from 1 list back, 2 from 2 lists back, and 2 from 3 lists back, randomly selected from within those lists. Note that even though there were across-list repetitions, each word on a list received an unambiguous serial position. The words were presented visually

for 1,400 msec, followed by a 200-msec ISI. After the presentation of the last item, the participants performed math problems of the form $A + B + C = ?$, where A, B, and C were positive, single-digit integers, for 16 sec before recalling the words on the just-studied list in any order during the 90-sec recall period.

Sederberg, Miller, Barnathan, and Kahana (2010). In an unpublished experiment, 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 on the basis of 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 2 other items. Each word was presented visually for 1,000 msec, with a 300- to 700-msec blank ISI. After the presentation of the last item, the participants performed math problems of the form $A + B + C = ?$, where A, B, and C were positive, single-digit integers, for 20 sec before recalling the words on the just-studied list in any order during the 45-sec recall period.

Analysis Details

The following sections provide the details of how we quantified temporal contiguity and semantic proximity effects and related them to the participants' 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 0, one for each possible transition lag. For example, if the list length was 12, there were 22 possible transition lags, from −11 to 11, excluding the lag of 0 because we do not count transitions to the same word in the list. Then, for each list, we stepped 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 were ignored. After incrementing the numerators and denominators for all of a participant's lists, the lag-CRP for that participant was 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 (2002). First, we determined semantic relatedness values, provided by the latent semantic analysis norms (LSA; Landauer & Dumais, 1997), for each pair of words in the word pool. 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, as in Howard and Kahana (2002), the word both was in the current list and had not already been recalled. Similarly, the denominator bins were incremented for each possible transition to a nonrecalled 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 to 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 nonparametric 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 ranked 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 among all members of the tie (e.g., the ranks of the item transition lags [1, 2, 2, 3] would be [1, 2.5, 2.5, 4]). To determine the temporal and semantic factors for that transition, we determined where the rank of the actual transition fell in the distribution of other 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 received 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 an 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.[1]

Reliability Analysis

To test the reliability of the temporal and semantic factor calculations for individual participants, we performed a split-half reliability analysis on the data. First, for each participant, we randomly split their recall lists into two sets. If a participant had an odd number of lists, we randomly selected which set the remaining list would join. We then averaged all of the temporal or semantic factor values within each of the two sets for each participant, leaving us with two mean factor values for each participant, one for each half of their data. Finally, we correlated the mean factor values for each half of the data across participants with a Pearson's correlation and corrected the resulting ρ values with the Spearman–Brown prediction formula $(2\rho/[1 + \rho])$ to determine the reliability of doubling the length of the split-half data. We repeated this randomized split-half procedure 2,000 times to determine a stable distribution of reliability scores. Reliability scores near 0 imply that the measure is not stable and can be the source of null results.

Regressions

We performed standard least-squares regressions to determine the correlation between the 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 of the participants for the specific experiment and then adding back the mean percent recall across all of the experiments. This measure of percent recall is used in every across-experiments analysis.
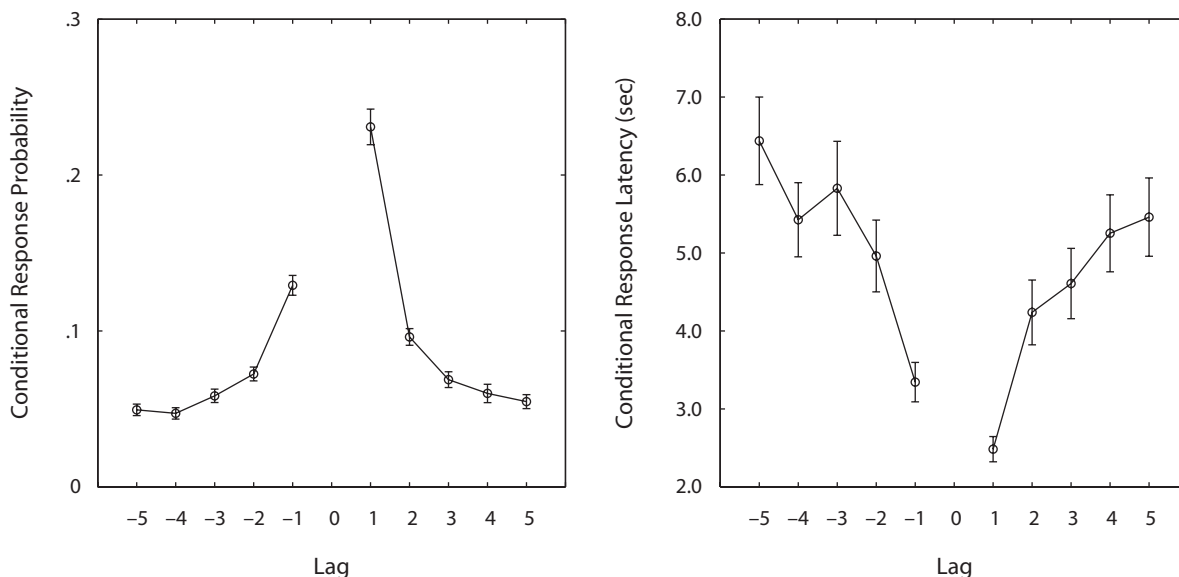
Figure 1. Temporal contiguity effect. Conditional response probability (left) and latency (right) as a function of serial position lag. Error bars represent 95% confidence intervals calculated across all of the participants.

Permutation Test

To ensure that differences in temporal factor values 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 with a null distribution of temporal and semantic factors for each participant generated with 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 ensured that the total number of recalls of each serial position remained constant for each participant, yet the specific items recalled on each list were 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 value was significant, we generated 200 slopes, one for each permutation of our null distribution, and then compared the null distribution of slopes with the actual slope. If the actual slope was greater than all of the slopes in the null distribution, this would be analogous to a *p* value of less than 1/200, or .005.

RESULTS

Temporal and Semantic Relatedness Drives Recall

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

The participants also demonstrated a significant semantic proximity effect. As is shown in Figure 2, the participants exhibited an increased probability of transitioning to items that were semantically related, as measured by LSA (Landauer & Dumais, 1997), to the just-recalled item (left panel). Along with the increase in recall probability, transitions to semantically related items were also faster than to less-related items (Figure 2, right panel). 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 an LSA similarity of 0.11, whereas *pony* and *forehead* have an LSA similarity of 0.21).

Although traditional conditional response probability measures are well suited for across-subjects analyses, we needed a measure of the temporal contiguity and semantic proximity effects for individual participants.[2] To this end, we employed a nonparametric measure of conditional response probability based on a ranking of 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 and 1 to each recall transition, taking into account all possible transitions to nonrecalled words (i.e., all of the unrecalled words at that time). A factor of 1 indicates that the participant transitioned to the most temporally or semantically proximal item, whereas a factor of 0 indicates that the participant transitioned to the least temporally or semantically proximal item (see the Method section). We could then average the factors across participant responses and test whether the distributions of temporal and semantic factors were different from 0.5, the value one would expect if there was no temporal contiguity or semantic proximity
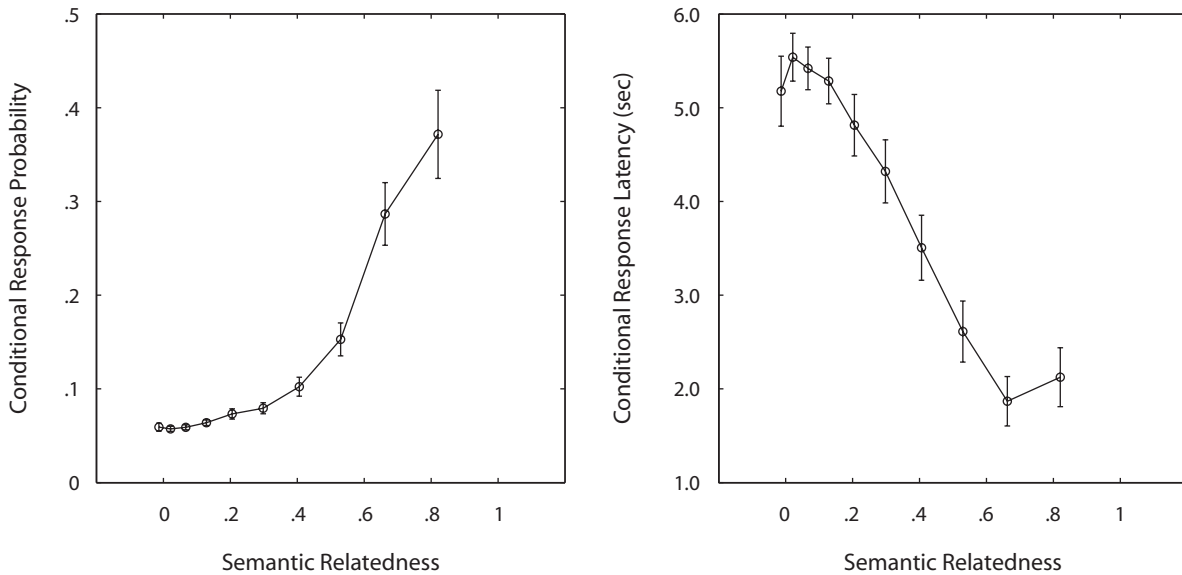
Figure 2. Semantic proximity effect. Conditional response probability (left) and latency (right) as a function of semantic relatedness calculated with the LSA measure of semantic relatedness between words. Error bars represent 95% confidence intervals calculated across all participants.

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 the participants relied on both temporal and semantic information to drive their responses. The mean temporal factor across all participants was .614 [t test comparing the temporal factors to .5, $t(509) = 31.2, p < .0001$], whereas the mean semantic factor was .541 [$t(509) = 16.7, p < .0001$].[3] Although the 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 $r = .010, p > .25$).

The temporal and semantic factors were calculated on the basis of individual recall transitions and were, therefore, independent of 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. The participants tended to cluster together items in these areas, and, consequently, it is possible that the temporal factor, in particular, would be inflated because of the 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 the Method section). The null distribution of temporal factors calculated from the shuffled responses had a mean of .49 and all 200 shuffles were less than our actual temporal factor of .614. As such, the participants still exhibited significant clustering on the basis of temporal contiguity even when accounting for primacy and recency effects ($p < .005$, given the number of permutations). As with the temporal factors, the mean semantic factor across par-

ticipants was significantly greater than all of the semantic factors in the null distribution, which had a mean of .50, verifying that the participants significantly clustered recalls by semantic relatedness ($p < .005$).

The 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 participant's recall probability by subtracting the within-study mean and adding back in the grand across-studies mean. Thus, all recall probability values reported in the present figures and tables are normalized probabilities of recall.

To test for a correlation between the 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 partitioned into deciles was highly significant (standardized $\beta = .88, p < .001$; see Table 2 for the regression results for each experiment without the participants grouped into deciles).

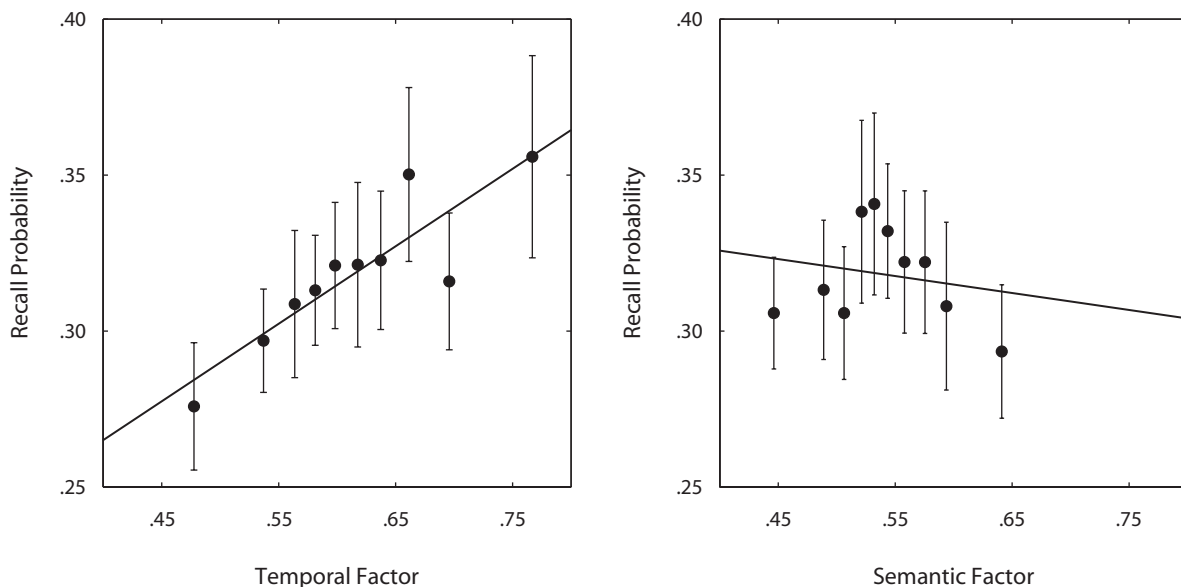As with the temporal factor calculations above, it is possible that the preceding analyses were biased by re-

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

cency or primacy effects in the data. This could give rise to artificial increases in the temporal factor due to the participants' clustering items presented in the recency or primacy portions of the study list. This potential confound is especially relevant to the present analysis, because participants who exhibit increased recall performance as a result of an increase in primacy may also exhibit an artificial increase in temporal contiguity, as measured by the temporal factor.[4] 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 0 for the null distribution ($M = .20$, $SEM = .01$), the β for the actual data was greater than all values in the null distribution, indicating that there was a significant effect of the temporal factor on recall performance ($p < .005$).

In the right panel of Figure 3, the same correlation analysis is plotted with the semantic factors for each participant. The regression fit reveals a negative trend across deciles, but it was not significant (standardized $\beta = -.19$, $p > .25$), suggesting that the degree to which the participants clustered recalls on the basis of semantic relatedness between items did not predict overall recall performance. It is possible, however, that this null result was due to the unreliability of the semantic factor calculation for individual participants in the included studies. We performed a split-half reliability analysis (see the Method section) on the mean semantic factors for each participant and found that the semantic factor values had low reliability (Spearman–Brown $\rho = .08$). The low semantic factor reliability is likely due to two main factors: As mentioned earlier, with the exception of one study that specifically

Table 2
Regression and Reliability Analysis Results for Each Experiment

| | Temporal Factor | | Semantic Factor | |
Study	β	Reliability	β	Reliability
Howard & Kahana (1999), Experiment 1	.17	.28	−.21	.00
Howard & Kahana (1999), Experiment 2	.56*	.66	.06	.00
Kahana, Howard, Zaromb, & Wingfield (2002), Experiment 2	.39†	.41	.05	.23
Kahana & Howard (2005)	.50**	.49	.03	.00
Bridge (2006)	.08	.44	.05	.00
Sederberg et al. (2006)	.29	.90	−.01	.00
Zaromb et al. (2006), Experiment 1	.23*	.61	−.04	.12
Zaromb et al. (2006), Experiment 2	.25†	.41	−.04	.23
Sederberg, Miller, Barnathan, & Kahana (2010)	.42†	.86	.12	.56
Total	.26**	.52	−.05	.08

Note—The reliability of participants' recall performance across all studies was .95. All standard errors on the reliability values were at least one order of magnitude smaller than the least significant digit reported.   † $.1 < p < .05$.   * $.01 < p < .05$.   ** $p < .01$.

varied the number of semantically related items on each list, the lists that we included in the present analysis rarely contained semantically related words, a problem that was magnified in some participants, who had relatively few lists that met the inclusion criteria. In support of this conclusion, the Sederberg et al. (2010) study, which had semantically related words on each list and 48 lists per participant, exhibited much more reliable semantic factor values (Spearman–Brown $\rho = .56$, which is on par with the mean reliability of the temporal factor values across all participants: Spearman–Brown $\rho = .52$).

The regression analyses performed for Figure 3 included individuals from all of the experiments, partitioned into deciles. We also performed a similar regression analysis across participants for each experiment independently (see Table 2). Although not all of the experiments demonstrated a significant positive correlation between temporal factor value and recall performance, they each exhibited positive β coefficients for the temporal factor predicting recall, which explains 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 retained the highly significant positive correlation between recall performance and temporal factor value (least significant result: standardized $\beta = .82$, $p < .003$).

DISCUSSION

Through a reanalysis of individual trial data from nine delayed free recall conditions, we have attempted to uncover the relation between temporal contiguity and semantic proximity (as measured by response order) and overall recall performance. By means of a recently developed method of quantifying temporal and semantic proximity effects, we found that participants who made recall transitions between words that were studied in nearby list positions also recalled more words. In contrast to this robust correlation between the temporal contiguity effect and probability of recall, increases in the semantic factor did not give rise to significant increases in participants' recall performance. These results suggest that retrieving information on the basis of associations formed between items that occurred in temporal proximity is a fundamental process underlying episodic memory performance. If it could be taken at face value, the distinction between the effects of the temporal and semantic factors on recall would be of great theoretical interest. However, there are some empirical caveats that must be considered before discussing the theoretical implications of what such a result would be.

Empirical Caveats

The split-half reliability analysis revealed that the semantic factor is less stable than the temporal factor, which argues against placing much import on the differences between the temporal and semantic factors in predicting recall performance. Here, we explore the possible sources of the low semantic factor reliability and what conclusions we can draw from the null semantic factor results, given this low reliability.

In addition to the possible sources of low semantic factor reliability outlined in the results—that the lists included in the present study rarely contained semantically related words and that some participants had relatively few lists that met the inclusion criteria—we cannot rule out the possibility that the low semantic factor reliability relative to the temporal factor could reflect weaker sensitivity of our semantic measures. Whereas temporal factors are precisely measured by means of serial position lag between nonrecalled words, our estimate of semantic proximity between possible recalls relied on LSA.[5] LSA, as well as other standard measures of semantic relatedness such as WAS (Steyvers et al., 2004), cannot account for individual differences in semantic relations between words. If interindividual variability in semantic relationships affects recall transitions, all of the standard measures would be blind to this effect. In addition, the distribution of semantic proximity values between possible recalls is roughly ex-Gaussian (Howard & Kahana, 2002). Whereas the distribution of temporal contiguity values because of serial position lag is consistent from list to list (and, to a large extent, from recall to recall), the semantic factor calculation is subject to additional variability caused by the distribution of semantic values available from list to list and from recall to recall.

In all but one of the studies analyzed here, lists were chosen randomly with respect to the semantic relationships between the items. The Sederberg et al. (2010) study included semantically related words from each quartile of the WAS distribution on each list and presented each participant with 48 lists. This study exhibited a much more reliable semantic factor value (Spearman–Brown $\rho = .56$) than the other studies. Nonetheless, the regression β for the semantic factor predicting recall did not approach significance ($p > .25$). Although it is clear that more data from subsequent studies will be needed to verify the null relation between the semantic factor and recall performance, the results of the Sederberg et al. (2010) study suggest two conclusions: first, that the lack of reliability of the semantic measure is largely due to the variability in list composition, and second, that even when this variability is controlled, there is still not positive evidence for a relationship between the effect of semantic transitions on recall performance.

Implications for Models of Free Recall

In retrospect, the failure to observe a relationship between the semantic factor and recall probability, to the extent to which one can take that null result at face value, makes sense. The challenge of free recall is to selectively target the list items as distinct from the rest of the items in memory. Whereas relying on interitem associations formed during encoding helps to focus recall on the list context, relying on preexisting semantic associations could also lead to recalls from outside the target list, interfering with correct recalls.

Although we found a highly significant correlation between individual differences in the temporal contiguity effect and recall performance, this finding is not a necessary prediction of extant free recall models. One way to increase the number of recalled items would be to increase the number of recall attempts made. For instance, in the search of associative memory (SAM; Mensink & Raaijmakers, 1988; Raaijmakers & Shiffrin, 1980; Sirotin, Kimball, & Kahana, 2005) model of recall, parameters that control the number of sampling attempts before terminating search would tend to increase the probability of recall. However, because the effect of temporal contiguity would be present at each of these sampling attempts, manipulating the threshold for terminating recall would not be expected to result in a correlation between temporal factors and the probability of recall.

The finding that participants' utilization of temporal contiguity predicts probability of recall, whereas their utilization of semantic proximity does not, also places strong constraints on models of free recall. In models of episodic retrieval, there is often a stage of the retrieval process in which the output of the memory system is used to recover a recallable item. In SAM, this stage corresponds to the recovery phase. In recent implementations of the temporal context model (TCM; Polyn et al., 2009; Sederberg et al., 2008), the output of the memory system provides the input to a set of leaky accumulators (Usher & McClelland, 2001) that model the retrieval process. Because both temporal and semantic factors provide input to these retrieval phases, the absence of an effect of semantic proximity on probability of recall suggests that the efficacy or sensitivity of postmemory retrieval, or recovery, is not a major factor in determining the probability of recall. If it were, both temporal and semantic factors would be expected to correlate with recall probability.

Consequently, the present findings are most consistent with the hypothesis that the strength and accessibility of the interitem associations formed during encoding are a dominant factor in determining the number of recalled items. In SAM, direct item-to-item associations form between items while they coinhabit the short-term store buffer. These newly formed associations are responsible for temporal contiguity effects (Kahana, 1996; Sirotin et al., 2005), and increases in the strength of these associations would also give rise to increased probability of recall. In TCM, the same effect is achieved by increasing the magnitude of newly learned 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 in which the effects of aging on episodic memory were explored provide additional evidence for the interaction between temporal and semantic associations and recall performance. Kahana et al. (2002) found that, although 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 are 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 present 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, because 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, as measured by the conditional probability of transitioning to the item in the next serial position, correlated with better performance in serial recall (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). Although the correlation between forward transitions and serial recall performance is to be expected, even with relative-order scoring, the critical result reported by Golomb et al. is that older adults tend to rely on semantic information to guide retrieval, even when it hurts recall performance. 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.

CONCLUSION

We have shown a positive correlation between participants' temporal contiguity effect 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 a fundamental process underlying episodic memory performance. They further suggest that, to the degree to which 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. J. (2006). *Memory and cognition: What difference does gender make?* Unpublished honors thesis, Syracuse University, Syracuse, NY.

Brown, S. C., Conover, J. N., Flores, L. M., & Goodman, K. M. (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 & 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 & 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**, 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: Erlbaum.

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

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

Howard, M. W., Sederberg, P. B., & Kahana, M. J. (2009). Reply to Farrell and Lewandowsky: Recency–contiguity interactions predicted by the temporal context model. *Psychonomic Bulletin & Review*, **16**, 973-984.

Howard, M. W., Wingfield, A., & Kahana, M. J. (2006). Aging and contextual binding: Modeling recency and lag recency effects with the temporal context model. *Psychonomic Bulletin & Review*, **13**, 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. *Journals of Gerontology: Psychological Sciences*, **60B**, P92-P97.

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 J. Byrne (Series Ed.) & H. L. Roediger III (Vol. Ed.), *Learning and memory: Vol. 2. Cognitive psychology of memory* (pp. 467-490). Oxford: Elsevier.

Kahana, M. J., Howard, M. W., Zaromb, F., & Wingfield, A. (2002). Age dissociates recency and lag recency effects in free recall. *Journal of Experimental Psychology: Learning, Memory, & 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, & 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. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **29**, 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 & 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**, 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.

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**, 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**, 893-912.

Sederberg, P. B., Miller, J. F., Barnathan, J. A., & Kahana, M. J. (2010). Unpublished raw data.

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**, 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.), *Experimental cognitive psychology and its applications* (pp. 237-249). Washington, DC: American Psychological Association.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, **108**, 550-592.

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

NOTES

1. Sample code to calculate temporal and semantic factors for free recall data is available in the online supplemental materials.

2. 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.

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

4. Whereas increases in the probability of recall in primacy or recency portions of the serial position curve would give rise to above-chance temporal factor values, increases in both primacy and recency relative to the baseline recall level would give rise to below-chance values. We calculated the expected temporal factor values on the basis of the transition probabilities for each participant's serial position curve and, just as we reported for the null distribution of the temporal factors above, found below-chance temporal factors for the data in the presently addressed

studies. This is because of the increased probability of making long-distance transitions between the primacy and recency portions of the list, which drives down the temporal factor value.

5. As an alternative, we have repeated our analyses using the WAS (Steyvers et al., 2004), a computational measure of semantic similarity derived from free-association norms (Nelson et al., 1998), and we observed the same pattern of results. Note that not all of the words in our experiments have been normed, meaning that recall transitions involving the words missing from the WAS were excluded from our analyses.

SUPPLEMENTAL MATERIALS

Sample code to calculate temporal and semantic factors for free recall data may be downloaded as a supplement for this article from http://mc.psychonomic-journals.org/content/supplemental.