Mem Cognit. Author manuscript; available in PMC 2013 July 31.

# Recall termination in free recall

**Jonathan F. Miller**,
University of Pennsylvania

**Christoph T. Weidemann**, and
Swansea University

**Michael J. Kahana**
University of Pennsylvania

## Abstract

Although much is known about the dynamics of memory search in the free recall task, relatively little is known about the factors related to recall termination. Reanalyzing individual trial data from 14 prior studies (1,079 participants, 28,015 lists), and defining termination as occurring when a final response is followed by a long non-response interval, we observe that termination probability increases throughout the recall period and that retrieval is more likely to terminate following an error than a correct response. Among errors, termination probability is higher following prior-list intrusions and repetitions than following extra-list intrusions. To verify that this pattern of results can be seen in a single study, we report a new experiment in which 80 participants contributed recall data from a total of 9,122 lists. This experiment replicated the pattern observed in the aggregate analyses of prior studies.

We are commonly faced with the situation of trying to recall a set of items learned in a given context without regard to the order in which the items were experienced. In the laboratory, this type of memory is studied using the free recall task: after studying a list of items (typically words) participants are asked to recall the list items in any order. The unconstrained nature of the free recall task provides a rich source of information on the nature of retrieval cues used during memory search. The analysis of the dynamics of free recall has revealed the importance of semantic relatedness and temporal contiguity in guiding recall (e.g., Howard & Kahana, 2002; Kahana, 1996). Such analyses have also played an important role in developing and testing theories of memory retrieval (e.g., Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Laming, 2010; Kimball, Smith, & Kahana, 2007; Polyn, Norman, & Kahana, 2009; Sederberg, Howard, & Kahana, 2008).

Although much attention has been paid to the way people make transitions from one response to the next, two other components of the recall process are also of critical importance: recall initiation and recall termination. Recall initiation was a major focus of Deese and Kaufman's classic (1957) study of the serial position effect in free recall. They documented the relation between the recency effect (superior recall of end of list items) and participants' tendency to initiate recall with items from the end of the list. Subsequent analyses of recall initiation have enriched our understanding of the recency effect in both immediate free recall and free recall following various distractor schedules (Bhatarah, Ward, & Tan, 2008; Davelaar et al., 2005; Laming, 1999, 2010; Sederberg et al., 2008).

---

Correspondence concerning this article should be addressed to Michael J. Kahana, 3401 Walnut St. Suite 303C, Philadelphia, PA 19104, USA; kahana@psych.upenn.edu; fax: 215-746-6848.

Much less, however, is known about the factors responsible for recall termination. In a study of inter-response times (IRTs) in free recall, Murdock and Okada (1970) found that the IRT prior to the final correct response tended to be approximately 8–10 seconds regardless of how many items the participant recalled. They also showed that IRTs increased exponentially with output position (i.e., the position of a response in the sequence of recalls; see also, Wixted & Rohrer, 1994; Polyn et al., 2009). This suggests that participants may terminate recall following a long period in which no new items are successfully retrieved.

Whereas Murdock and Okada inferred recall termination based on the final correct response given in a fixed recall period, a more recent study by Dougherty and Harbison (2007) assessed recall termination by asking participants to press a key when they could not remember any additional items. Dougherty and Harbison found that the duration between the last successful retrieval and the termination response (exit latency) decreased as the total number of items recalled increased. Further, they showed that variability in exit latencies was closely related to participants' decisiveness, with participants who scored high on a decisiveness scale terminating recall more quickly.

An important feature of the recall process not considered in these previous studies concerns the nature of the responses themselves. Although most recalled items are correct responses (i.e., items studied on the target list), participants also occasionally commit errors, recalling items studied on an earlier list but not on the current list (prior-list intrusions), recalling items not studied on the current list or any earlier list (extra-list intrusions), or repeating already recalled items. It is known that errors tend to occur late in recall (Roediger & McDermott, 1995; Kimball et al., 2007) and that they elicit subsequent errors (Zaromb et al., 2006). As such, one might hypothesize that whatever process contributes to recall errors may also play a role in recall termination. To test this hypothesis, we asked whether the conditional probability of stopping differed following various types of recall events. Because these events occur with different frequencies during the recall process, we compute these conditional probabilities separately as a function of output position. For this purpose, we have carried out secondary analyses of raw trial-by-trial data culled from 1,079 participants across 14 large free recall experiments comprising a total of 28,015 recall trials (for a description of each experiment, see Sederberg, Miller, Howard, & Kahana, 2010). By pooling raw data from so many trials, we were able to look at relatively rare events that happen during recall and to see how these events predicted recall termination. To foreshadow our results, we find that retrieval is more likely to terminate following recall errors than following correct responses, and that this effect appears consistently throughout the recall period. The increased tendency to terminate recall after committing an error varies significantly across the three types of recall errors that we studied: prior-list intrusions, extra-list intrusions, and repetitions. After reanalyzing these prior datasets, we further validated our results by showing that the same pattern of increased termination following errors can be seen in a single new experiment, reported below.

## Meta-analysis Methods

We reanalyzed individual trial data from the 14 experiments listed in Table 1. Our criteria for inclusion was stringent. First, we limited our secondary analysis to studies for which we could obtain individual trial data for each participant. Second, we required those data to include information on the order of individual responses on each trial, including errors. Third, we excluded studies for which the nature of recall errors was not classified according to the three key categories: prior list intrusions, extralist intrusions, and repetitions. Finally, we further limited our analyses to studies reporting the timing of individual responses. Nonetheless, we were able to include data from 10 experimental conditions reported in 7 published articles, and an additional 4 studies reported in working papers. In each of the included studies, lists consisted of between 10 and 25 common words (often nouns selected from the Toronto noun pool, see Friendly, Franklin, Hoffman, & Rubin, 1982) and recall was vocal, with speech being digitized and latencies recorded. The Appendix provides brief descriptions of the methods used in each of the experiments we analyzed.

In free recall tasks, participants are typically given a fixed amount of time to recall the list items. As such, these studies do not tell us when recall actually terminates. For example, one may ask whether the recall period has terminated while the participant is still actively recalling words, whether the participant has given up early in the recall interval, or whether the participant is trying hard to recall items, but nothing is coming to mind. Another possibility is that recall terminates because participants have already recalled all of the list items. However, this almost never happens with the long lists used in these (and most) free recall studies. What we do know, on the basis of recall latencies, is that participants make most of their responses early in the recall period and that the time between successive recalls increases approximately exponentially with output position (Murdock & Okada, 1970; Polyn et al., 2009; Rohrer & Wixted, 1994).

In the present study, we define recall termination as occurring when the time between the last recalled item and the end of the fixed recall period was both longer than every interresponse time on the current trial and exceeded a criterion of 12 s. This value was chosen to exceed the mean exit latency of 10 s reported by Dougherty and Harbison (2007) in an open-ended retrieval period. Out of 28,015 trials, 18,829 met these criteria (67.21%). In the included trials, there were a total of 127,240 responses. 111,211 (87.40%) were correct, 3,589 (2.82%) were repetitions, 6,000 (4.72%) were prior-list intrusions, and 6,440 (5.06%) were extra-list intrusions. 41% of prior-list intrusions were correctly recalled on their initial presentation list. We carried out a parallel set of analyses without excluding any trials and obtained nearly identical results.

## Meta-analysis Results

Figure 1A shows the conditional probability of recall termination following correct responses and each type of recall error as a function of output position (for the first eight output positions during recall)[1]. We define recall errors as either a repetition of an already recalled item, a prior-list intrusion (PLI), or an extra-list intrusion (ELI). We determined each participants' probability of recall termination by dividing, separately for each output position and response type, the number of responses that were the final response in a trial by the total number of responses of that type. When calculating the mean probabilities for each response type and output position, participants' data were weighted according to the number of responses they contributed. To assess differences in the probability of recall termination following the four response types, we calculated bias-corrected and accelerated bootstrap 95% (two-tailed) confidence intervals (Efron & Tibshirani, 1993) for all six possible differences at each output position (see Figure 1B). We consider differences with confidence intervals that do not include zero to be significant.

Across the first eight output positions, which subsume the majority of recall data across these experiments, participants were more likely to terminate recall following PLIs than following either ELIs or correct responses. For later output positions (5–8), participants were also very likely to terminate recall following repetitions: termination probability

---

[1] Because these analyses were conducted across a wide range of experiments, varying in list length and other parameters, we had to restrict our analyses to a fixed number of output positions that would subsume the majority of the recall data across these studies (in this case, the first eight recalled items). Although participants occasionally recalled more than eight items, especially in experiments involving longer lists and slower presentation rates, it would be very difficult to show those data in these aggregate analyses. In Experiment 1, described below, we provide data on later output positions in a study involving lists of 16 items.

following repetitions was similar to the PLIs, and exceeded that of both ELIs and correct responses. Recall termination following ELIs was generally intermediate between PLIs and correct responses. Recall termination was significantly more likely following ELIs than correct responses (for output positions 3–8) and significantly less likely than recall termination following PLIs (at all output positions) or repetitions (output positions 5–8). The pattern of results seen in the Figure is thus quite reliable in our large sample of data: people are more likely to terminate recall following errors than correct responses, and among errors, recall termination following PLIs and repetitions is generally higher than following ELIs. One exception is the significantly lower probability of terminating recall following repetitions than following PLIs at output positions 3 and 4.

To further evaluate these results, we determined for each participant the earliest output position after which recall stopped, and then we aggregated the corresponding probabilities for the different response types across participants (e.g., if one participant always recalled at least 4 items and another always recalled at least 6 items, we aggregated the probabilities for output positions 4 and 6 for these two participants, respectively). We aligned output positions both at the individually determined first and last stopping position, and in both cases, we observed an ordering of probabilities of stopping that was consistent with that shown in Figure 1: probabilities of stopping tended to be lowest following correct recalls, larger after ELIs, and even larger after PLIs and repetitions. Repeating all of the above analyses without excluding trials based on our recall termination criteria yielded virtually identical results.

Although the results of the meta analysis seem clear, some readers may not be at ease with analyses aggregated across so many diverse datasets. We therefore sought to validate these results in a single large experiment. Fortuitously, at the time of this writing we were in the midst of conducting a large scale study on the electrophysiological correlates of memory encoding and retrieval in free recall (Long, Miller, & Kahana, 2011). With 80 participants having completed 7 experimental sessions each involving free recall of 16 study-test lists we had sufficient power to assess whether the patterns observed in the meta-analysis would replicate in a single study.

## Experiment

### Methods

**Participants—**80 participants performed a free recall experiment consisting of one practice session and 6 subsequent experimental sessions. Participants were consented according the University of Pennsylvania’s IRB protocol and were compensated for their participation. Each session lasted approximately 1.5 hours.

**Procedure—**Each session consisted of 16 lists of 16 words presented one at a time on a computer screen. Each study list was followed by an immediate free recall test and each session ended with a recognition test. Half of the sessions (randomly chosen) included a final free recall test before recognition in which participants recalled words from any of the lists from the session. This experiment was part of larger study that included EEG recordings and further manipulations of the recognition and recall periods (as described in Long et al., 2011).

Items were either presented concurrently with a task cue, indicating the judgment that the participant should make for that word, or with no encoding task. The two encoding tasks were a size judgment (“Will this item fit into a shoebox?”) and an animacy judgment (“Does this word refer to something living or not living?”), and the current task was indicated by the color and typeface of the presented item. There were three conditions, control lists (no task),

task lists (all items were presented with the same task), and task shift lists (items were presented with either task). List and task order was counterbalanced both across sessions and participants. Additionally, using the results of a prior norming study, only words that were clear in meaning and that could be reliably judged in the size and animacy encoding tasks were included in the pool.

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

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

## Results

Before reporting on recall termination following various types of errors, we first show results of more standard analyses applied to this dataset. Standard serial position effects are observed, with marked recency as expected in any immediate free recall task, and a moderately strong primacy effect extending about 4 or 5 serial positions into the list (Figure 2A). Related to the recency effect, participants exhibit a strong tendency to begin recall with one of the last few items—a tendency that slowly dissipates across subsequent recalls (Figure 2B).

The dynamics of free recall are largely characterized by the contiguity (or lag recency) effect, and by the semantic proximity effect: recall of an item tends to be followed by recall of a neighboring or similar item. The contiguity effect in this experiment, as shown in Figure 2C, shows the usual forward asymmetry (Kahana, 1996). The semantic proximity effect in this experiment, shown in Figure 2D, is similar whether semantic relatedness is defined by WAS similarity or Latent Semantic Analysis (Landauer & Dumais, 1997). Because the results described above were only minimally affected by the different encoding conditions we report all analyses collapsed across these conditions.

Recall termination effects were analyzed in the same manner as in the meta-analyses described above. Because there were very few trials with fewer than 4 correct responses (3.7%) or more than 12 correct responses (26%), we limited our analyses to output positions 4 through 12. Of the 9,122 trials, 6,527 met our inclusion criteria for selecting trials where participants were likely to have terminated recall. In the included trials, there were a total of 67,671 responses. 64,348 (95.09%) were correct, 1,570 (2.32%) were repetitions, 563 (0.83%) were prior-list intrusions, and 1,190 (1.76%) were extra-list intrusions.

As shown in Figure 3A, the tendency to terminate recall was greater following PLIs, ELIs and repetitions than following correct responses. Furthermore, the ordering of termination probabilities was identical to that in the aggregate analyses, being highest following PLIs, next highest following repetitions, lower following ELIs, and lowest following correct responses. With the exception of the comparison between ELIs and correct responses, each of the other comparisons was statistically significant in the predicted direction for a majority of output positions between positions 4 and 12 (see Figure 3C). Additionally, we performed the previously described aligned output position analysis on these data, and we observed an ordering of probabilities matching those shown in Figure 3A. We also repeated the analyses without excluding any trials. As shown in Figures 3B and 3D, these results are nearly identical to those that used our trial exclusion criteria.

A somewhat unusual feature of the present study, and also of several studies in the meta-analysis described above, is the high level of experience that participants obtained with the free recall task. One may therefore wonder whether these results reflect strategies that develop through extensive practice, or whether they are typical of the results that would be obtained with less highly practiced participants. We address this question by separately analyzing data from first and last sessions of the reported experiment (sessions 1 and 7). As shown in Figure 4, recall termination was more likely after incorrect than correct responses for both the first and last sessions. Additional analyses revealed that the order of the probabilities for correct responses, ELIs, repetitions, and PLIs matched those shown in Figures 1 and 3 for both session 1 and session 7.

## Discussion

Understanding recall termination is particularly important because whatever accounts for recall termination determines the total number of items that are ultimately recalled. Although previous research has revealed a great deal about how people initiate recall and how they transition between successively recalled items, much less is known about the correlates of recall termination.

Through a reanalysis of individual trial data from 14 experiments in previous studies as well as from a newly reported study, we found that termination is consistently more likely to occur after an error than after a correct recall, and that this tendency to terminate recall following an error depends on the kind of error that is made. Recall termination is most likely to follow prior-list intrusions and repetitions of already recalled items, less likely to follow extra-list intrusions, and least likely to follow correct responses.

Models of free recall in which retrieval of an item serves as a cue for the next response (e.g., Howard, Kahana, & Wingfield, 2006; Kimball et al., 2007; Metcalfe & Murdock, 1981; Polyn et al., 2009; Raaijmakers & Shiffrin, 1980; Sederberg et al., 2008) suggest that the increased tendency to terminate recall following errors may reflect a fundamental memory process. Specifically, these models assume that neighboring items are associated during study, and that recall of an item tends to retrieve items studied in proximate list positions. In this way, the models account for the well known contiguity effect, which is seen in people's strong tendency to successively recall items studied in neighboring list positions (Kahana, 1996). By the same logic, these models predict that intrusions will tend to be poor cues for subsequent correct recalls. For example, the recall of an item presented on an earlier list is likely to be a good cue for other items from the prior list, which compete with current list items. Zaromb et al. (2006) provide empirical support for this proposition. They found that participants were significantly more likely to commit prior-list intrusions following other prior-list intrusions, and further, that such intrusions tended to come from the same prior list. They also found that when an item on the current list was presented on an earlier list, recall of that item was more frequently followed by a PLI than by recall of a current-list item. Thus, PLIs tend to retrieve contextual information that is inappropriate to the current list, and therefore lead to further recall errors and recall termination. By the same logic, ELIs are also poor recall cues. Although such responses do not evoke specific competition from recently studied items, they are nonetheless poor retrieval cues insofar as their associated temporal context will not serve as an effective cue for current list items.

To the extent the specific competition from recent (prior) list responses is greater than the competition associated with extralist associations (as discussed above) one would expect the probability of termination to be greater following PLIs than following ELIs, as we have observed. It is somewhat less obvious, however, why participants are nearly as likely to terminate recall following repetitions as following PLIs. Such items do not harbor strong associations to items on earlier lists and are not strongly associated with temporal contexts that are unrelated to other list items. On the other hand, the fact that repetitions were previously recalled suggests that the list items that were effective at cueing these repeated items were likely to have already been recalled as well. This account also suggests an explanation for the lower probability of terminating recall following repetitions at early output positions when only few items have been recalled, where it is likely that items cued by the repeated item are still available for recall and thus the detrimental effects of repetitions should be limited to later output positions. One might expect that repeating an already recalled item at later output positions would activate other recalled items, which would simply consume retrieval time without leading to another correct response (and thereby leading to increased termination probabilities). The idea that resampling and rejecting previously recalled items consumes retrieval time, and thus predicts recall termination in a fixed interval task, forms the basis of accounts of the exponential growth of inter-response times in free recall (Murdock & Okada, 1970; Rohrer & Wixted, 1994).

The finding that people are more likely to terminate recall following errors than following correct responses, even when controlling for output positions and recall time, adds to a growing body of evidence that recall of an item evokes contextual information previously associated with that item, and that this contextual information can serve to either support or hinder subsequent recalls (see Kahana, Howard, & Polyn, 2008, for a review). Whereas earlier evidence for this process was based solely on recall transitions, the present study suggests that recall termination depends on the loss of appropriate retrieval cues.

Although the observed pattern of results suggests a causal relation between recall errors and recall termination, one cannot strictly rule out the possibility that these results arise from some other endogenous aspect of the recall process giving rise to both recall error and recall termination. Future research will be able to better adjudicate between these theoretical accounts by testing sophisticated process models of recall that can simultaneously fit data on recall initiation, recall transitions, and recall termination. The serious consideration of recall termination data in these models will in turn enable the models to speak more clearly to the memory mechanisms that underlie recall impairments in both healthy aging and neurological disease (e.g., Dubois & Albert, 2004; Golomb, Peelle, Addis, Kahana, & Wingfield, 2008; Grober, Lipton, Hall, & Crystal, 2000; Kahana et al., 2002).

## Acknowledgments

The authors gratefully acknowledge support from National Institutes of Health grant MH55687. We thank Marc Howard and Per Sederberg for contributing data from their unpublished studies. We also thank Michael Dougherty and Isaiah Harbison for providing us with data from their study of response termination in free recall.

## Appendix: Details of Included Experiments

**Howard and Kahana (1999), Exp. 1.** In single sessions, 62 participants (one was excluded from the original study because of experimenter error) performed both immediate and delayed free-recall of 25 total lists. 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). Each list was composed of 12 randomly selected nouns from the Toronto Word Pool (Friendly et al., 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 either immediately began recall, or 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. 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). 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. 1.** In single sessions, 28 older and 31 younger participants performed immediate free recall of 33 lists. The 10 words in each list were presented visually for 1400 ms, followed by a 100 ms ISI. Immediately following the presentation of the last item, participants began recalling the words on the just-studied list in any order during a 45 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. 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), Massed lists.** 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, 48 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. The number of subjects reported here is higher than the 35 reported in Sederberg et al. (2006), as additional data were collected subsequent to the article’s publication.

**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, 42 older and 63 younger 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.

**Howard et al. (2007), Control lists.** In single sessions, 294 participants (one was excluded from the original study because of experimenter error) performed immediate free recall of 48 lists, each composed of 10 randomly selected nouns from the Toronto Word Pool (Friendly et al., 1982). Half of the lists were in the experimental condition and repeated items within a list, whereas the half in the control condition contained no repeated item presentations. Only the control lists are used here. Each item was presented both visually and auditorily for a maximum of 1200 ms, with a 500 ms ISI. Participants were required to indicate if the word was “concrete” or “abstract” by pressing either the left or right control keys within the 1200 ms time limit, after which the ISI period was initiated. After the last item, participants began recalling the words on the just-studied list in any order during the 30 s recall period.

**Polyn et al. (2009).** Across two separate testing sessions, 45 participants performed immediate free recall of 34 total lists (17 per session). Each list was composed of 24 items selected from the word association spaces norms (WAS, Steyvers et al., 2004). For each item, participants were asked to either make a size judgement ("Will this item fit in a shoe-box?") or an animacy judgement ("Is this item living or non-living"). Items were presented visually for 3000 ms, with an 800 ms ISI, and participants indicated their response during this time via a keypress. After the final item, participants began recalling the words on the just-studied list in any order during the 90 s recall period.

**Polyn et al. (unpublished) A.** Across three separate testing sessions, 38 participants performed immediate free recall of 48 total lists (16 per session). Each list was composed of a series of 24 photographs drawn from three categories: famous people, well-known landmarks, and common objects. For each category, participants were asked to rate on a scale from 1 to 4 how much they liked the celebrity, wanted to visit the landmark, or how often they encountered the object. Photographs were presented visually for 3500 ms, with an 800 ms ISI, and participants indicated their rating during this time via a keypress. After the final item, participants began recalling the words on the just-studied list in any order during the 90 s recall period.

**Polyn et al. (unpublished) B.** Across four separate testing sessions, 42 participants performed immediate free recall of 48 total lists (12 per session). Each list was composed of 24 items selected from the word association spaces norms (WAS, Steyvers et al., 2004). For each item, participants were asked to either make a size judgement ("Will this item fit in a shoebox?") or an animacy judgement ("Is this item living or non-living"). Items were presented visually for 3000 ms, with an 800 ms ISI, and participants indicated their response during this time via a keypress. After the final item, participants began 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, 37 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 et al., 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.

## References

Bhatarah P, Ward G, Tan L. Examining the relationship between free recall and immediate serial recall: The serial nature of recall and the effect of test expectancy. Memory & Cognition. 2008; 36(1):20–34. [PubMed: 18323059]

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

Davelaar EJ, Goshen-Gottstein Y, Ashkenazi A, Haarmann HJ, Usher M. The demise of short-term memory revisited: Empirical and computational investigations of recency effects. Psychological Review. 2006; 112:3–42. [PubMed: 15631586]

Deese J, Kaufman RA. Serial effects in recall of unorganized and sequentially organized verbal material. Journal of Experimental Psychology. 1957; 54:180–187. [PubMed: 13475644]

Dougherty M, Harbison J. Motivated to Retrieve: How Often Are You Willing to Go Back to the Well When the Well Is Dry? Journal of Experimental Psychology: Learning, Memory, and Cognition. 2007; 33(6):1108.

Dubois B, Albert M. Amnestic mci or prodromal alzheimer's disease? Lancet Neurology. 2004; 3(4): 246–248. [PubMed: 15039037]

Efron, B.; Tibshirani, RJ. An introduction to the bootstrap. New York: Chapman & Hall; 1993.

Friendly M, Franklin PE, Hoffman D, Rubin DC. The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. Behavior Research Methods and Instrumentation. 1982; 14:375–399.

Golomb JD, Peelle JE, Addis KM, Kahana MJ, Wingfield A. Effects of adult aging on utilization of temporal and semantic associations during free and serial recall. Memory & Cognition. 2008; 36(5): 947–956. [PubMed: 18630201]

Grober E, Lipton R, Hall C, Crystal H. Memory impairment on free and cued selective reminding predicts dementia. Neurology. 2000; 54(4):827–832. [PubMed: 10690971]

Howard MW, Kahana MJ. Contextual variability and serial position effects in free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition. 1999; 25:923–941.

Howard MW, Kahana MJ. When does semantic similarity help episodic retrieval? Journal of Memory and Language. 2002; 46:85–98.

Howard MW, Kahana MJ, Wingfield A. Aging and contextual binding: Modeling recency and lag-recency effects with the temporal context model. Psychonomic Bulletin & Review. 2006; 13(3): 439–445. [PubMed: 17048728]

Howard MW, Venkatadass V, Norman KA, Kahana MJ. Associative processes in immediate recency. Memory & Cognition. 2007; 35(7):1700–1711. [PubMed: 18062547]

Kahana MJ. Associative retrieval processes in free recall. Memory & Cognition. 1996; 24:103–109. [PubMed: 8822162]

Kahana MJ, Howard MW. Spacing and lag effects in free recall of pure lists. Psychonomic Bulletin & Review. 2005; 12:159–164. [PubMed: 15948289]

Kahana, MJ.; Howard, MW.; Polyn, SM. Associative retrieval processes in episodic memory. In: Roediger, HL., III, editor. Cognitive psychology of memory. Vol. 2 of Learning and memory: A comprehensive reference, 4 vols. (J. Byrne, Editor). Oxford: Elsevier; 2008.

Kahana MJ, Howard MW, Zaromb F, Wingfield A. Age dissociates recency and lag recency effects in free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2002; 28:530–540.

Kimball DR, Smith TA, Kahana MJ. The fSAM model of false recall. Psychological Review. 2007; 114(4):954–993. [PubMed: 17907869]

Laming D. Testing the idea of distinct storage mechanisms in memory. International Journal of Psychology. 1999; 34:419–426.

Laming D. Serial position curves in free recall. Psychological Review. 2010; 117(1):93. [PubMed: 20063965]

Landauer TK, Dumais ST. Solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. Psychological Review. 1997; 104:211–240.

Long NM, Miller JF, Kahana MJ. Unpublished technical report. 2011 (Tech. Rep.).

Metcalfe J, Murdock BB. An encoding and retrieval model of single-trial free recall. Journal of Verbal Learning and Verbal Behavior. 1981; 20:161–189.

Murdock BB, Okada R. Interresponse times in single-trial free recall. Journal of Verbal Learning and Verbal Behavior. 1970; 86:263–267.

Nelson DL, McKinney VM, Gee NR, Janczura GA. Interpreting the influence of implicitly activated memories on recall and recognition. Psychological Review. 1998; 105:299–324. [PubMed: 9577240]

Polyn SM, Norman KA, Kahana MJ. A context maintenance and retrieval model of organizational processes in free recall. Psychological Review. 2009; 116(1):129–156. [PubMed: 19159151]

Raaijmakers, JGW.; Shiffrin, RM. SAM: A theory of probabilistic search of associative memory. In: Bower, GH., editor. The psychology of learning and motivation: Advances in research and theory. Vol. Vol. 14. New York: Academic Press; 1980. p. 207-262.

Roediger HL, McDermott KB. Creating false memories: Remembering words not presented in lists. Journal of Experimental Psychology: Learning, Memory and Cognition. 1995; 21:803–814.

Rohrer D, Wixted JT. An analysis of latency and interresponse time in free recall. Memory & Cognition. 1994; 22:511–524. [PubMed: 7968547]

Sederberg PB, Gauthier LV, Terushkin V, Miller JF, Barnathan JA, Kahana MJ. Oscillatory correlates of the primacy effect in episodic memory. NeuroImage. 2006; 32(3):1422–1431. [PubMed: 16814568]

Sederberg PB, Howard MW, Kahana MJ. A context-based theory of recency and contiguity in free recall. Psychological Review. 2008; 115(4):893–912. [PubMed: 18954208]

Sederberg PB, Miller JF, Howard WH, Kahana MJ. The temporal contiguity effect predicts episodic memory performance. Memory & Cognition. 2010; 38(6):689–699. [PubMed: 20852233]

Steyvers, M.; Shiffrin, RM.; Nelson, DL. Word association spaces for predicting semantic similarity effects in episodic memory. In: Healy, AF., editor. Cognitive psychology and its applications: Festschrift in honor of Lyle Bourne, Walter Kintsch, and Thomas Landauer. Washington, DC: American Psychological Association; 2004.

Wixted JT, Rohrer D. Analyzing the dynamics of free recall: An integrative review of the empirical literature. Psychonomic Bulletin and Review. 1994; 1:89–106.

Zaromb FM, Howard MW, Dolan ED, Sirotin YB, Tully M, Wingfield A, et al. Temporal associations and prior-list intrusions in free recall. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2006; 32(4):792–804.

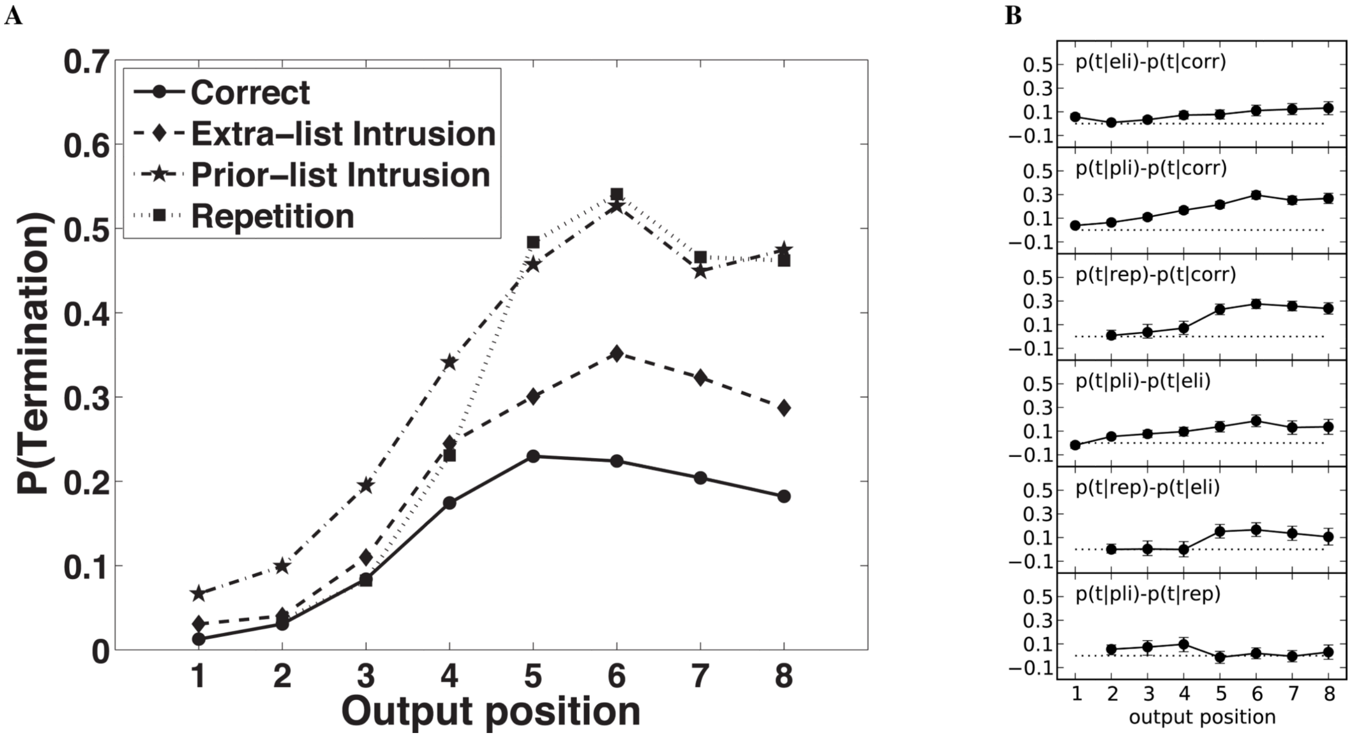

**Figure 1.**

**A.** Termination probability following correct recalls (corr), extra-list intrusions (ELI), prior-list intrusions (PLI), and repetitions (rep). Aggregate data from 14 free recall experiments.

**B.** Differences in the probability of termination, $p(t)$, between the various response types and corresponding 95% (two-tailed) confidence intervals (CI; determined by bias corrected and accelerated non-parametric bootstrap, Efron & Tibshirani, 1993). Dashed lines indicate zero difference (CIs that do not include zero indicate statistically significant differences).

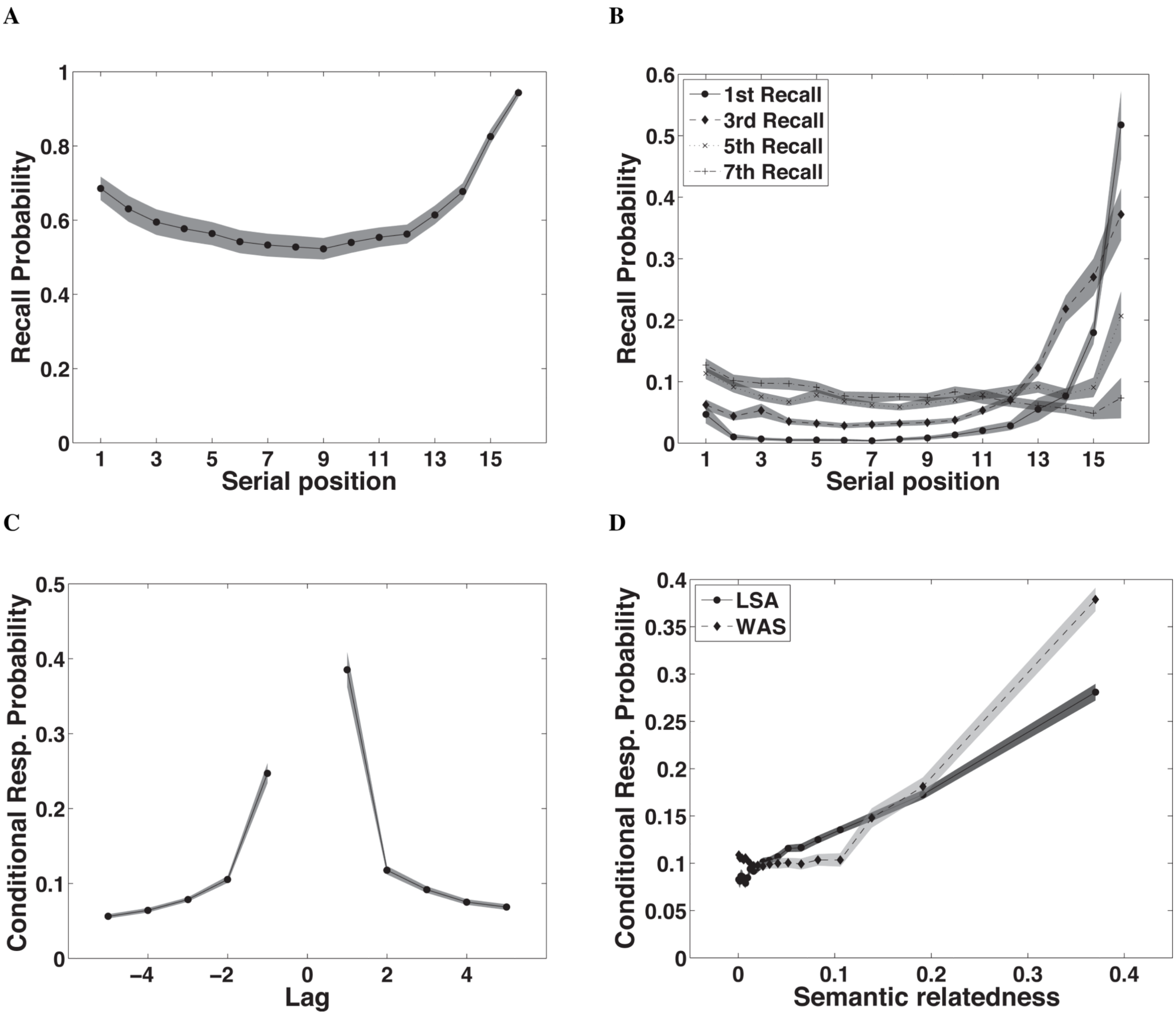

**Figure 2.**

Serial position, temporal contiguity, and semantic contiguity effects for the data from Experiment 1. Shaded regions are 95% confidence intervals. **A.** The probability of recall for items in each serial position. **B.** The probability of recalling presented items in output positions 1, 3, 5, and 7. The probability for output position one represents the probability of first recall (PFR) **C.** The lag-conditional response probability (lag-CRP) shows the conditional probability of recalling items presented in serial position $i + lag$, where $i$ is the serial position of the just recalled item. **D.** The semantic-conditional response probability (semantic-CRP) shows the conditional probability of recalling items from a given level of semantic relatedness.

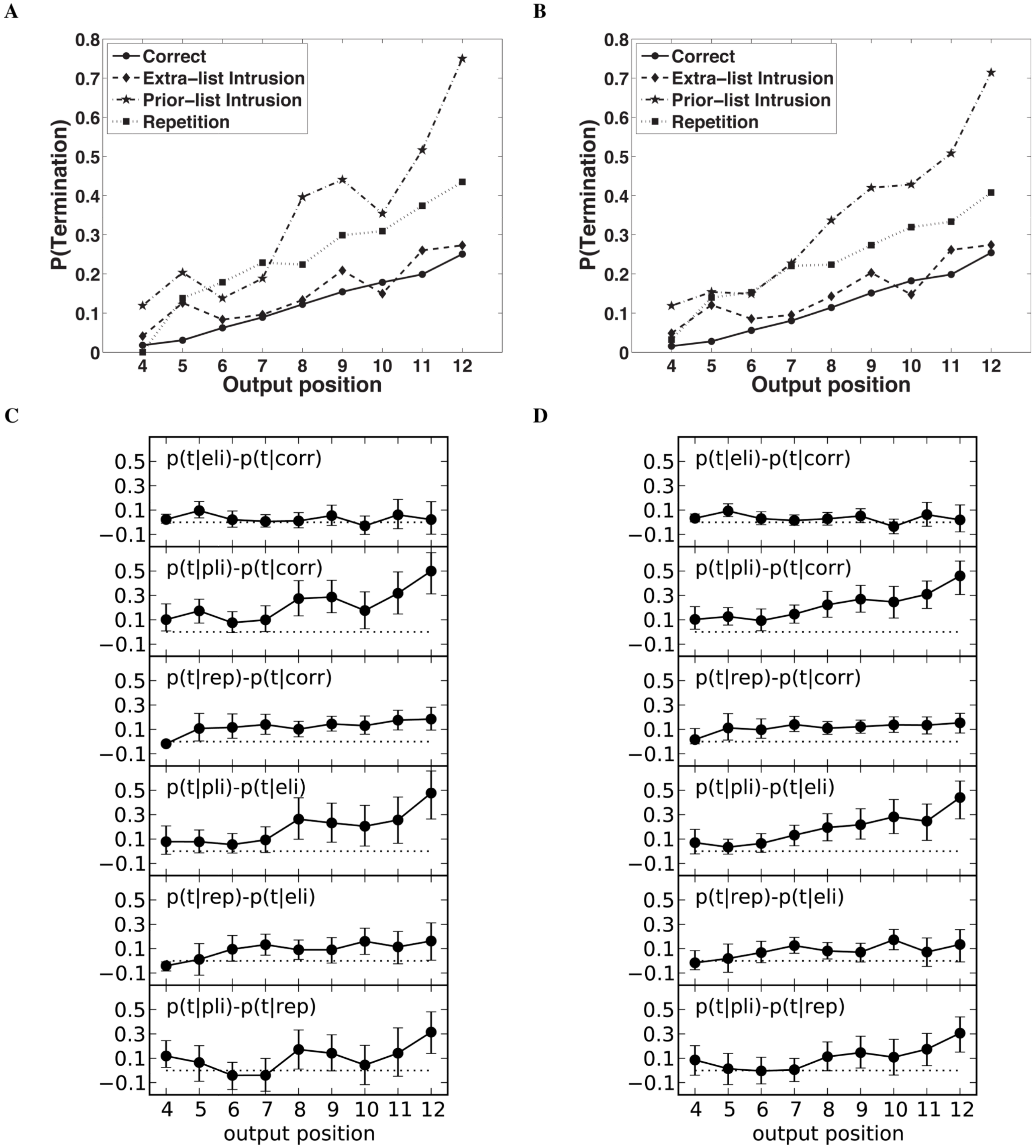

**Figure 3.**

**A.** Termination probability following correct recalls (corr), extra-list intrusions (ELI), prior-list intrusions (PLI), and repetitions (rep). Data from Experiment 1. **B.** Termination probability following correct recalls, extra-list intrusions, prior-list intrusions, and repetitions. Data from Experiment 1 with no trials excluded. **C and D.** Differences in the probability of termination, $p(t)$, between the various response types and corresponding 95% (two-tailed) confidence intervals (CI; determined by bias corrected and accelerated non-parametric bootstrap, Efron & Tibshirani, 1993). Dashed lines indicate zero difference (CIs that do not include zero indicate statistically significant differences). Panel C corresponds to data from panel A, and panel D corresponds to data from panel B.

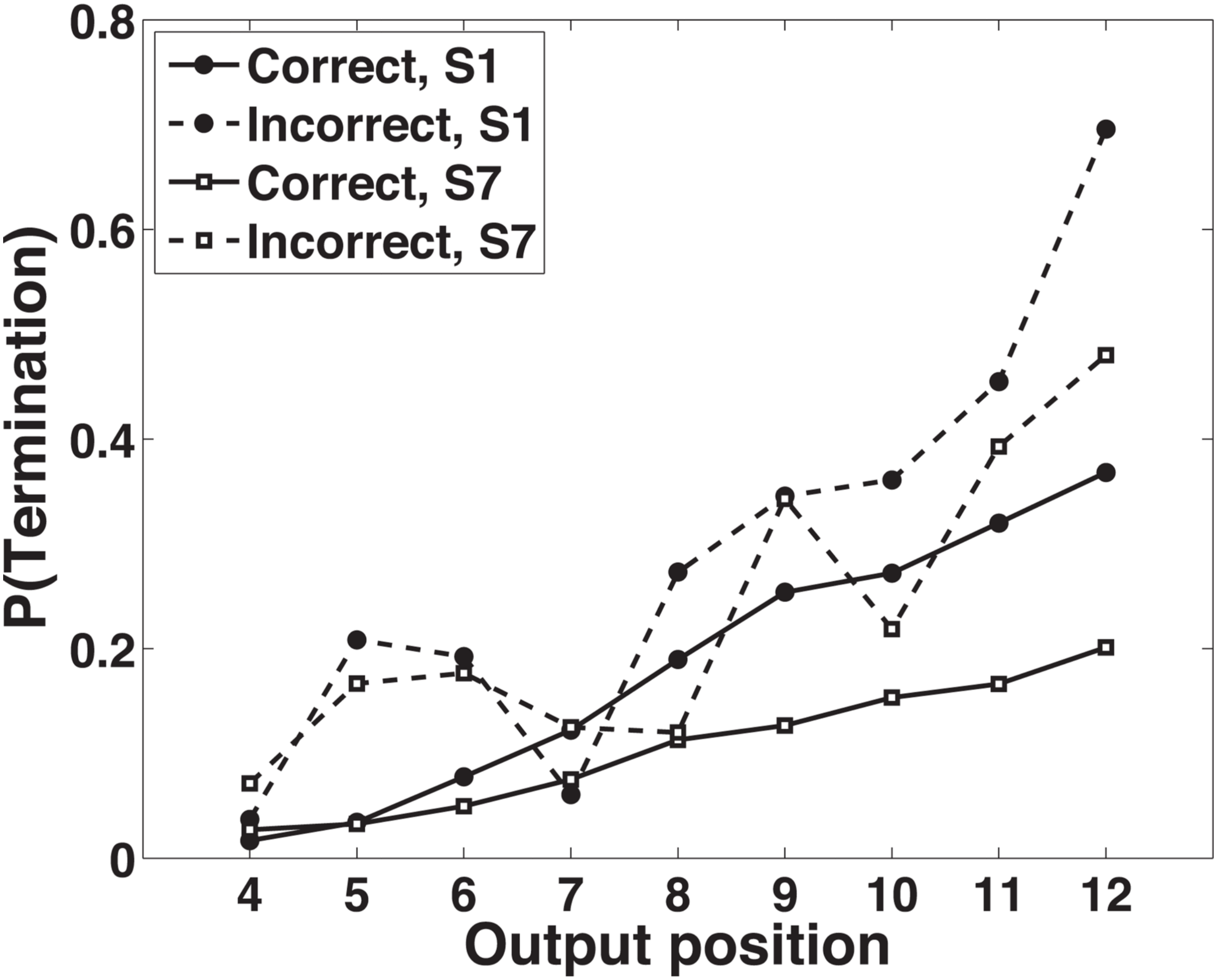

**Figure 4.**

Termination probability following correct recalls and incorrect responses. Data are from the first session (filled cricles) and the last session (open squares) of Experiment 1.

## Table 1

Summary of experimental conditions. N indicates sample size, LL indicates list length, PR indicates presentation rate, Dist. indicates duration of an end-of-list arithmetic distractor task (0 indicates no distractor), and RP indicates recall period.

| **Study** | ***N*** | **LL** | **PR (s)** | **Dist. (s)** | **RP (s)** |
|---|---|---|---|---|---|
| Howard & Kahana (1999) Exp. 1 | 62 | 12 | 1.0 | 0 or 10 | 45 |
| Howard & Kahana (1999) Exp. 2 | 16 | 12 | 1.2 | 16 | 60 |
| Kahana et al. (2002) Exp. 1 | 59 | 10 | 1.4 | 0 | 45 |
| Kahana et al. (2002) Exp. 2 | 50 | 10 | 1.4 | 16 | 45 |
| Kahana & Howard (2005) | 65 | 30 | 4.5 | 45 | 90 |
| Bridge (2006) | 119 | 25 | 1.1 | 30 | 60 |
| Sederberg et al. (2006) | 48 | 15 | 1.6 | 20 | 45 |
| Zaromb et al. (2006) Exp. 1 | 100 | 20 | 1.4 | 16 | 90 |
| Zaromb et al. (2006) Exp. 2 | 105 | 20 | 1.4 | 16 | 90 |
| Howard et al. (2007) Control | 293 | 10 | 1.2 | 0 | 30 |
| Polyn et al. (2009) | 45 | 24 | 3.0 | 0 | 90 |
| Polyn et al, unpublished A | 38 | 24 | 3.5 | 0 | 90 |
| Polyn et al, unpublished B | 42 | 24 | 3.0 | 0 | 90 |
| Sederberg et al., unpublished | 37 | 16 | 1.0 | 20 | 45 |
| Total | 1,079 | - | - | - | - |