Forward and Backward Recall

Dougherty, M.R., Halpern, D., and Kahana, M.J.
University of Pennsylvania

Abstract

Although recall proceeds most naturally in the order of encoding, people can also recall items in reverse order. Prior studies raise the question of whether recall differs as a function of direction. We reexamine this classic question by studying the dynamics of forward and backward serial recall. Consistent with prior work, we find similar levels of recall accuracy in forward and backward recall. However, analyses of recall dynamics demonstrated higher accuracy for initiation and lower accuracy for transitions in backward as compared to forward recall. Pre-cuing subjects to the direction of recall eliminated the difference in recall initiation and reduced, but did not erase, effects seen in recall transitions. We show that backward recall benefits and suffers from the recency effect: By promoting access to end of list items, recency facilitates initiation, but it hinders correct transitions by promoting fill-in-errors following omissions.

Introduction

The directional vs. symmetric nature of associative memory has puzzled theorists and experimentalists since the genesis of our field (see Ekstrand, 1966; Kahana, 2002, for reviews). Whereas people have great difficulty reciting a well-learned sequence in reverse order (e.g., Tulving, 1985), following a single learning trial, subjects can often recall a list in reverse order at similar accuracy levels to their forward recall (e.g., Madigan, 1971). Recall asymmetries also differ between recall of pairs and lists. When learning novel pairs of items, the order in which subjects encode a pair has virtually no effect on the accuracy of cued recall (Ekstrand, 1966; Kahana, 2002). Yet, in probed recall of serial lists, subjects exhibit a modest but consistent advantage in recalling the item following a probe word as compared with the preceding item (Kahana & Caplan, 2002). According to some models, subjects encode meaningful pairs holistically, which gives rise to symmetric retrieval (Asch &

Trial-level response data, analysis scripts and experiment code may be obtained from the public data repository hosted at `http://memory.psych.upenn.edu`. The authors express their gratitude to Alice Healy for her very helpful comments on an earlier version of this manuscript. We also thank Jesse Pazdera and Connor Keane for assistance with programming the experiment. Correspondence concerning this article should be addressed to Michael J.Kahana, kahana@psych.upenn.edu.

Ebenholtz, 1962; Ceraso, 1967; Murdock, 1982; Rizzuto & Kahana, 2001), whereas associative mechanisms supporting longer sequences possess an inherent forward asymmetry (e.g. Howard & Kahana, 2002; Murdock, 1997; Kahana, 2020). According to another view, the asymmetric retrieval of sequential information largely derives from a procedural memory system whose role emerges with extensive practice (e.g. Bunsey & Eichenbaum, 1996) as in the acquisition of motor skills (Knowlton & Schorn, in press).

Although a large literature has compared forward and backward recall under varying conditions (see Donolato, Giofrè, & Mammarella, 2017), the landscape of empirical findings has not yet brought us to a consensus regarding the degree to which common vs. distinct cognitive operations underlie forward vs. backward recall. Here, we revisit this question by applying a data analytic approach that has proven valuable in the analysis of free recall, but has rarely been applied to serial recall. This approach emphasizes the dynamics of the retrieval process, decomposing the serial position analysis into the components of initiation, transition, and termination (Kahana, Diamond, & Aka, in press; Miller, Weidemann, & Kahana, 2012). Initiation refers to the tendency to begin recall at a particular list position, and transition refers to the tendency to make subsequent recall transitions as a function of the relations between the recalled items. We specifically consider how the temporal separation, or lag, between successively recalled items predicts transitions. We also consider the question of termination by examining the probability that a subject will terminate recall following certain retrievals.

In a typical study of immediate serial recall, subjects know, prior to the start of each encoding list, that they will recall the list in forward order. However, in studies comparing backward and forward recall, subjects may either know the expected order of recall prior to the encoding period, or they may only learn of the direction of recall just prior to the retrieval phase. As subjects' expectations regarding the direction of recall may influence the manner of recall (Murdock, 1962; Hintzman, 2015), we introduced a within subject cueing manipulation. On half of the trials, we informed subjects in advance of the direction of subsequent recall. On the other half of trials, subjects only learned of the recall direction immediately after studying the list. To the extent that pre-cuing subjects to the direction of subsequent recall would lead them to optimize their encoding strategy, we would expect to see significant differences in recall dynamics between pre-cued and post-cued trials. Recall direction was also manipulated within-subject, such that subjects recalled in forward order for half of the trials and backward order for the remaining half.

Because encoding and retrieval strategies may depend on list length, we manipulated this variable both within and between subjects. Each subject completed 48 trials in total, with 12 trials assigned to each of the four combinations of cuing condition (pre- vs. post-) and recall direction (forward vs. backward). Lists comprised common words, with variable list lengths manipulated both within and between subjects such that each subject studied lists of three lengths. Group one studied lists of length six, nine and 12; group two studied lists of length seven, nine and 11; group three studied lists of length eight, nine and 10. This design sought to achieve several objectives: (1) varying list length within subject to reduce expectancy of list termination and any list-length specific encoding strategies, (2) to probe a wide range of list lengths while at the same time gathering a large amount of data at a single list length (nine) to allow for fine grained analyses of recall transitions in that critical condition.

Method

Subjects

COVID-related restrictions led us to deploy our experiment via Amazon Mechanical Turk (MTurk), an online platform for large-scale data collection (Mason & Suri, 2012). To qualify for our study, subjects were required to be from the United States and possess at least a 95% approval rating on MTurk. Additionally, the task could not be accessed from mobile devices or tablets. A total of 1341 subjects (570 male, 425 female, 346 unreported) completed Experiment 1 for compensation of $7.50. We excluded 29 subjects whose data were impacted by technical issues, and an additional 742 subjects who met at least one of the three exclusionary criteria defined in Appendix A. The remaining 570 subjects (303 male, 267 female) contributed to our analyses, and ranged in age from 19.0 to 73.0 years, with a mean age of 37.57 years ($SD = 10.59$). Prior to participating, subjects completed a consent form that was approved by the University of Pennsylvania Institutional Review Board.

Procedure

The word pool used in this study was identical to that used in Experiment 4 of the Penn Electrophysiology of Encoding and Retrieval Study (Kahana, Aggarwal, & Phan, 2018), with a few exceptions. Twenty words with at least one homophone in the English language were excluded from the pool, as well as words comprising eight or more letters. To limit the confusability of list items, each word in a given list was constrained to begin with a different letter of the alphabet and no two items in a single list could have a cosine similarity greater than 0.3 as determined by Google's Word2Vec algorithm. The word pool was open, each trial using unique and unrepeated words.

After providing informed consent, and before beginning the experiment, we gauged subjects attention to our instructions with a simple task. Each subject was instructed to carefully read a description of memory prior to proceeding. A sentence inserted within this description instructed subjects to reply "never" to a subsequent survey question about their own memory quality, regardless of their subjective answer. Only subjects who responded correctly advanced to the actual experiment.

After a brief test of their computer's audio output, subjects read task instructions describing when to recall in the forward direction (appearance of a right-pointing arrow), when to recall in the backward direction (appearance of a left-pointing arrow), and when these prompts would appear (either before or after all words were presented). Subjects were instructed to picture each word as it was presented and only focus on the most recently presented word. Following these instructions, subjects performed a series of four practice trials. These trials introduced subjects to our experimental manipulations (forward and backward recall, pre-cued and post-cued). All four practice trials were nine-word lists.

The experiment prompted subjects to hold down the Z and P keys to start each trial. Following trial initiation, a screen was displayed for 2000ms while three tones played to signal the beginning of word presentations. In the pre-cued condition, an arrow indicating direction of recall appeared on screen during this time. In the post-cued condition, the screen remained black. Words were then visually presented at a rate of 1000ms per item with a 500ms inter-stimulus interval. Following presentation of the final word, a screen was again displayed for 2000ms while three tones played, this time to indicate the beginning of the recall period. The screen remained black in the pre-cued condition, while in the post-cued condition, an arrow appeared on screen indicating the direction of recall. Following this, a text box appeared on screen for subjects to type recalled words. Subjects submitted responses by pressing the space bar, enter, comma, or semicolon keys. The text box was cleared after each word was submitted, preventing subjects from looking back at previously-recalled words. Recalls were spell-checked and scored automatically by an algorithm identical to that used by Healey (2018) in a similar typed recall task (see Appendix B for a full description). The recall period concluded either after 8000ms with no recall, or once two times the list length in seconds had past (e.g. 12000ms for list length six), whichever came first. Following each recall period, subjects participated in a math task to retain engagement, in which subjects added three numbers together and submitted their answer with the enter key, after which a new addition problem appeared. After 12 seconds of this math task, a new trial began.

All experimental data may be downloaded from our public data repository `http://memory.psych.upenn.edu/Data_Archive`.

Results

We begin our analyses by investigating the effects of direction, cuing, and serial position on correct recall. We considered an item to be correctly recalled if its serial position immediately followed that of the just-recalled item in forward recall, or immediately preceded that of the just-recalled item in backward recall. The first response is a special case, and is considered correct only if it is recalled in the correct absolute position (the first list item in forward recall and the last list item in backward recall). Thus, if subjects studied the list "seat, broom, vase, palm, rope, mouse" and recalled the sequence "seat, broom, vase, rope, mouse", the items "seat", "broom", "vase", and "mouse" would be considered correct. Solway, Murdock, and Kahana (2012) advocated for this relative-order scoring method in serial recall of long lists. Here we first report data based on lists of nine items, as all list length manipulations included this condition. We report on the effects of list length in a subsequent section.

When instructed to recall in forward order, subjects exhibited the standard, primacy-dominated serial position effect (Figure 1). When instructed to recall in reverse order, the serial position effect reverses, with strong recency replacing the primacy effect. Numerous prior studies have documented this primary-recency reversal in backward recall (e.g. Madigan, 1971; Donolato et al., 2017; Li & Lewandowsky, 1993).

When subjects could anticipate the order of recall prior to encoding (pre-cued condition), they exhibited superior performance in both forward and backward recall (Figure 1a, b). This difference is particularly pronounced in early output positions, i.e. early list items in forward recall and late list items in backward recall. This observation led us to conduct our subsequent analyses as a function of expected output position rather than serial position, where expected output position = 1 indicates recall of an item in the first serial position in forward recall and an item in the last serial position in backward recall. Expected output position in backward recall is the reverse of an item's serial position.

Utilizing expected output position rather than serial position in these analyses more directly demonstrates similarities and differences in retrieval dynamics underlying forward

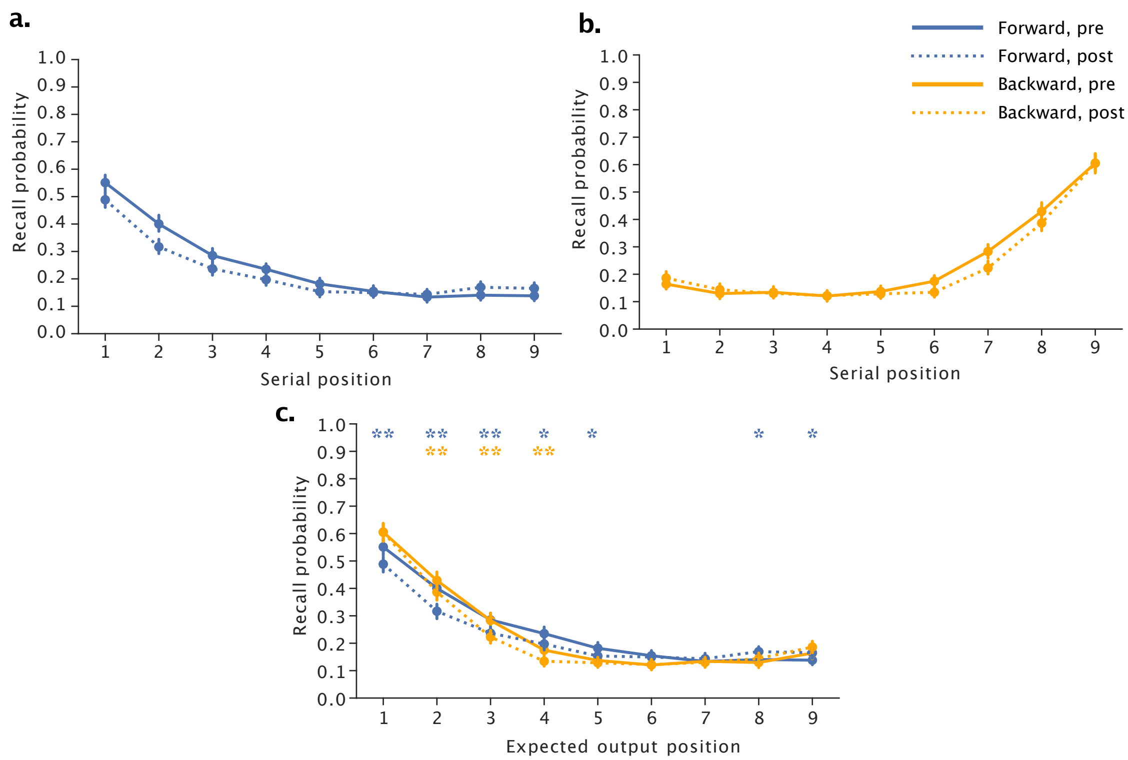

Figure 1. **Cueing effects on forward and backward recall**. Panels a and b illustrate the probability of correctly recalling an item as a function of its serial position, calculated using the relative order scoring method described in Solway et al. (2012). Panel a compares forward recall for pre- and post-cueing. Panel b compares backward recall for pre- and post-cueing. Panel c illustrates the probability of correctly recalling an item as a function of its expected output position, calculated using the same scoring method as above. Significance markers indicate the following: $\star$ indicates $p < 0.05$, $\star\star$ indicates $p < 0.001$ after conducting an FDR test for multiple comparisons. Error bars indicate 95% confidence intervals, calculated using 1000 iterations of bootstrapped resampling of the data.

and backward recall. Statistical tests of the data shown in Figure 1c demonstrated a significant main effect of cueing condition ($F(1, 569) = 23.37$, $MSE = 1.60$, $p < 0.001$), and a significant interaction between expected output position and cueing condition ($F(8, 569) = 22.29$, $MSE = 1.18$, $p < 0.001$), confirming the above observations. A closer examination of Figure 1c indicates that the beneficial effects of cuing occurred at somewhat different output positions in forward and backward recall, as reflected in a significant three-way interaction of cuing condition, output position, and recall direction was significant ($F(1, 8, 569) = 4.21$, $MSE = 0.10$, $p < 0.001$). Whereas pre-cuing benefits forward recall in early output positions, it appears to harm forward recall in later output positions, possibly because of recency effect that is accentuated in the post-cued condition.

In line with other studies using similar methodologies (e.g., Madigan, 1971; Guèrard, Saint-Aubin, Burns, & Chamberland, 2012; Li & Lewandowsky, 1995; Thomas, Milner, & Haberlandt, 2003), we do not observe a significant main effect of recall direction ($F(1, 569) = 0.00$, $MSE = 0.00$, n.s.). Figure 2a shows the serial position curve calculated using the Solway et al (2012) method of scoring, with results collapsed across cuing conditions. One

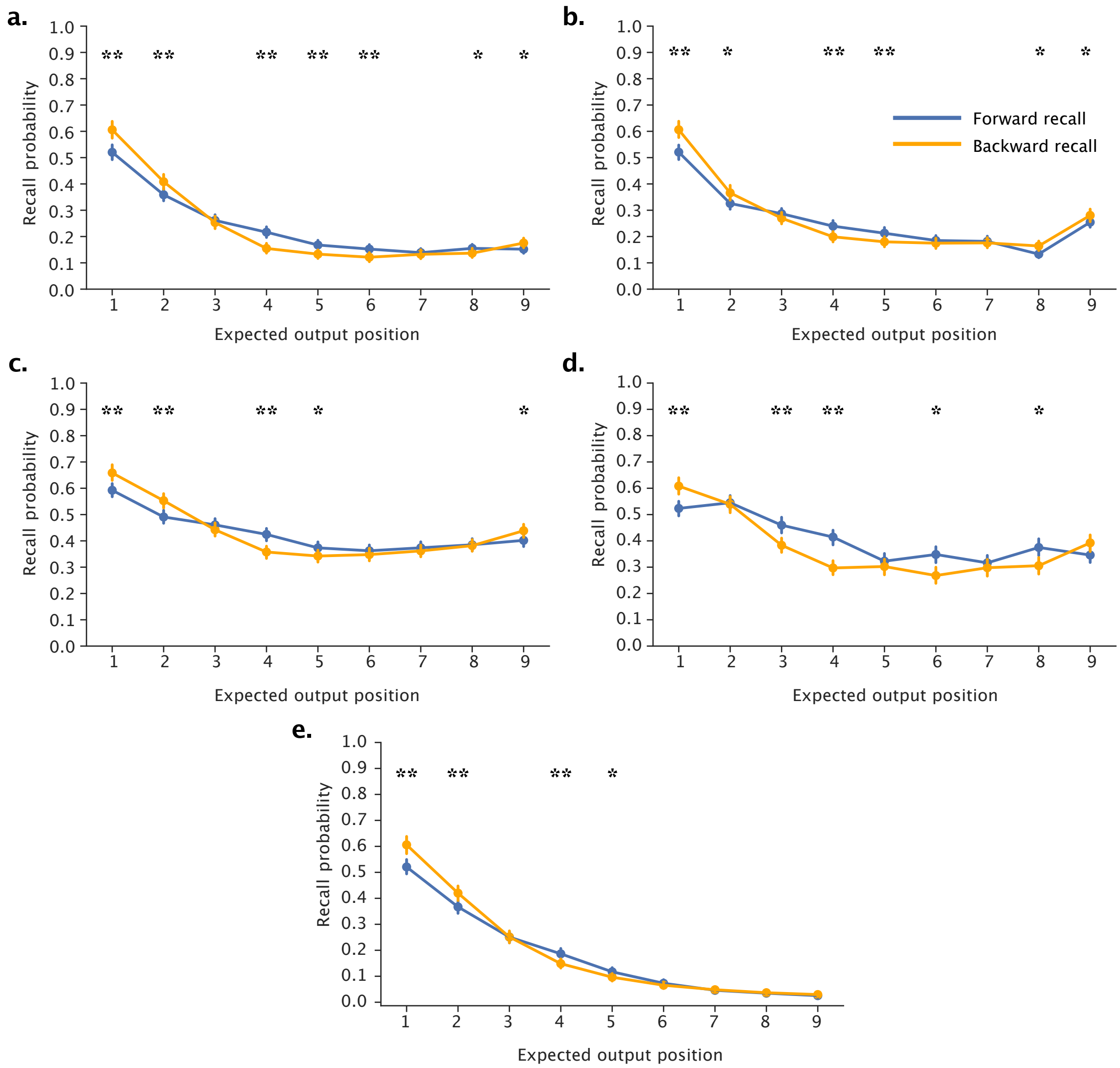

Figure 2. **Expected output position effects in forward and backward recall using four different scoring methods.** The graph illustrates the probability of correctly recalling an item as a function of its expected output position (serial position for forward recall or reverse serial position for backward recall), where correctness was determined using **a.** the relative order method of Solway et al (2012), **b.** the relative order method of Drewnowski and Murdock (1980), **c.** free recall scoring, **d.** conditional relative order scoring, as described in the main text, and **e.** strict positional scoring. Significance markers indicate the following: $\star$ indicates $p < 0.05$, $\star\star$ indicates $p < 0.001$ after conducting an FDR test for multiple comparisons. Error bars indicate bootstrapped 95% confidence intervals.

may ask whether the concordance between forward and backward recall revealed by this particular scoring method would also emerge with other scoring procedures. To answer this question, we directly compared forward and backward recall (irrespective of cuing condition) using an additional four different scoring techniques: Figure 2b follows Drewnowski and Murdock (1980)'s technique of scoring an item as correct if it follows an item that preceded it in the input list and precedes an item that followed it. The first and last recalls are scored as correct only if recalled in the same position that they were presented. Figure 2c shows results using the free recall scoring method, in which a recall is scored as correct regardless of its order in the output sequence. Figure 2d illustrates results using a "conditionalized" version of the Solway et al (2012) relative-order scoring method. This method adjusts for the fact that a given item can only be considered correct if its predecessor has already been recalled. As a result, the probability of correct recall should be conditioned on this event, resulting in a different denominator than in Solway et al (2012). Lastly, Figure 2e demonstrates a traditional strict positional scoring method, where each item is considered correct if it was recalled in the exact position that it was presented in forward recall, or the exact opposite position in backward recall. All five methods exhibit an overall equivalence between foward and backward recall. We do, however, observe a significant interaction between expected output position and recall direction ($F(8, 569) = 22.63$, $MSE = 1.18$, $p < 0.001$) using the Solway et al (2012) method of scoring, with a small backward recall advantage in early output positions and a small forward recall advantage in middle output positions. This result is similarly observable across all scoring methods.

The striking similarity in overall performance between forward and backward recall would seem to suggest a common mechanistic basis to serial recall, independent of order. Arguing against a strong form of this view, however, numerous studies have uncovered dissociations between forward and backward recall, indicating that forward and backward recall diverge in their degree of reliance on at least some cognitive processes. To help elucidate these processes we turned to an analysis of recall dynamics. As in prior work (e.g., Howard & Kahana, 1999; Kahana, Mollison, & Addis, 2010) we first investigate the accuracy of recall initiation (i.e., reporting the first item in the first output position in forward recall and the last item in the first output position in backward recall). Subjects initiated recall more accurately in backward than in forward recall (see Figure 3, $F(1, 561) = 30.97$, $MSE = 4.42$, $p < 0.001$). Cue timing had a significant effect on correct initiation in forward recall, with post-cuing reducing subjects tendency to initiate accurately ($t(568) = -4.99$, $p < 0.001$) compared to the pre-cued condition. There was no effect of cueing on initiation accuracy in backward recall ($t(568) < 1$, n.s.). Post-hoc tests revealed that the benefit to initiation in backward recall exists in both the pre-cued ($t(568) = 3.31$, $p < 0.01$) and post-cued conditions ($t(568) = 6.38$, $p < 0.001$). This difference between correct initiation rates as a function of recall direction is to be expected, given the effect of recency in immediate recall. We now turn our attention to effects of recall direction and cuing on recall transitions.

In both free and serial recall tasks, subjects exhibit a strong temporal clustering effect (Kahana, 2020) – following recall of an item studied in serial position i, the next recalled item is likely to come from a neighboring list position (e.g., $i \pm lag$ for small values of lag). Because transitions differ in their availability, and because availability changes dynamically throughout the recall process[1], analyses of temporal clustering must condition on the avail-

[1] Subjects very rarely recall previously recalled items, averaging 0.078 repetition occurrences per list.

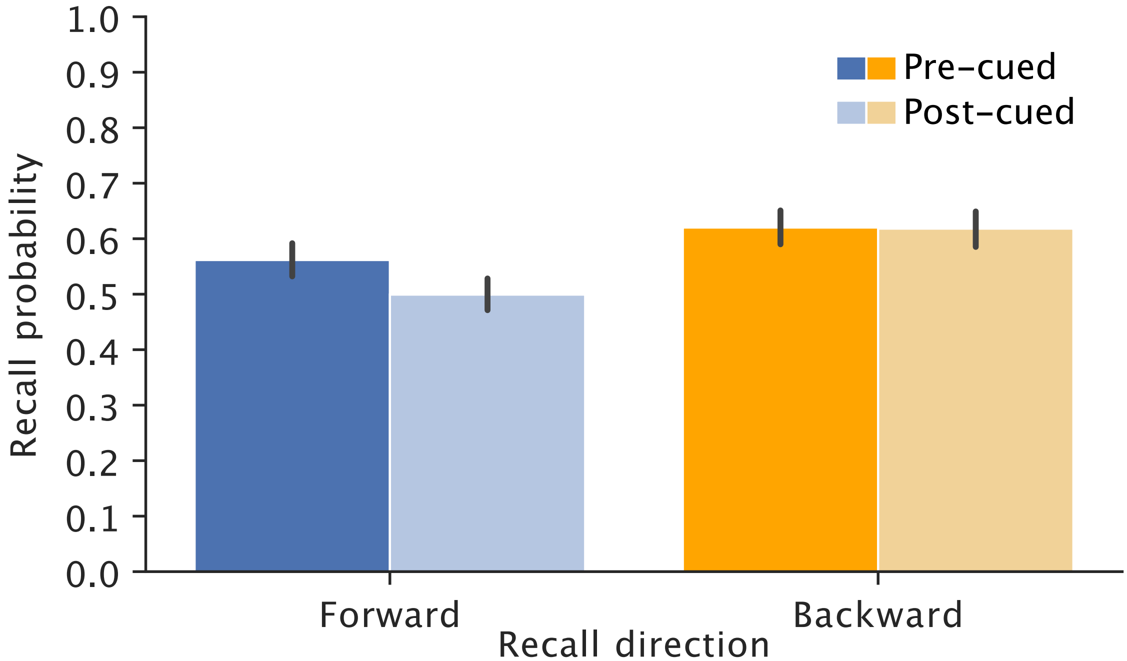

Figure 3. **Probability of correct first recall.** The four bars illustrate the probability of initiating recall with the first presented item in forward recall and the last presented item in backward recall, when recall direction is pre- vs. post-cued. Error bars indicate bootstrapped 95% confidence intervals.

ability of transitions to different lags. Figure 4a illustrates the temporal clustering effect separately for forward and backward recall and under pre- and post-cuing conditions. For the example list ABCDEFGHI (where each letter denotes a unique word in the study list), transitions from D to E in forward recall (or E to D in backward recall) would have a lag of +1. A "fill in" transition from D to C in forward recall (or F to G in backward recall) would have a lag of -1.

The high probability of $lag = +1$ transitions indicates that among all possible transitions, subjects were most likely to make a correct response (as defined by relative order scoring; see Figure 4a). Although we failed to observe a reliable difference in overall forward and backward recall accuracy (see Figure 2a) the conditional probability of a $lag = +1$ transition exhibited a small but reliable forward recall advantage ($F(1, 511) = 17.12$, $MSE = 0.76$, $p < 0.001$) as well as an advantage of pre-cuing ($F(1, 511) = 19.68$, $MSE = 0.36$, $p < 0.001$).

In addition to correct transitions, analyses of order errors can elucidate retrieval dynamics. Whereas $lag = +1$ signifies correct relative-order transitions, $lag = -1$ transitions reflect subjects' tendency to reverse direction and fill in a missing item. Statistical analyses of transitions of $lag = -1$ showed a main effect of recall direction ($F(1, 488) = 76.02$, $MSE = 3.29$, $p < 0.001$), with subjects more likely to reverse the direction of recall in backward recall than in forward recall. Pre-cuing subjects with the direction of re-

Therefore, we assume that these items do not play a significant role in the competition for retrieval.

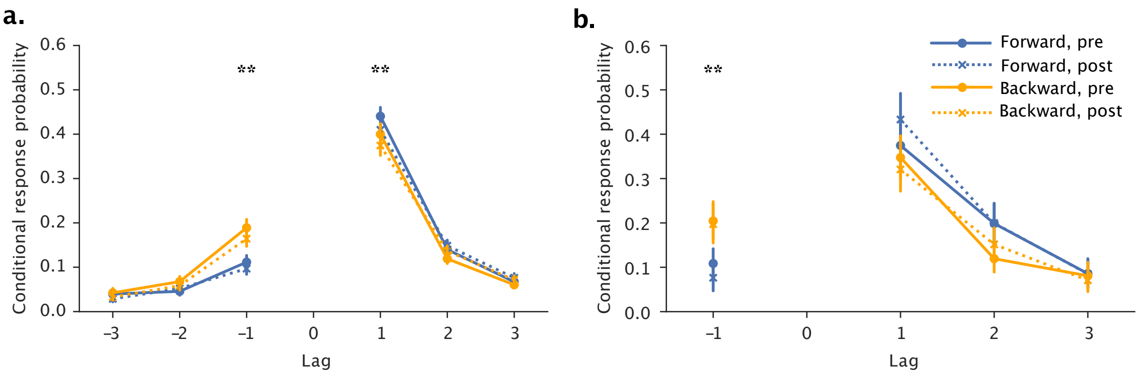

Figure 4. **Temporal clustering in serial recall.** Probability of recalling an item as a function of its lag from the just-recalled item, or the difference between its predecessor's expected output position and its expected output position. A lag of +1 marks a correct transition in both forward and backward recall, and equivalence of the conditions indicates symmetric forward and backward recall. Panel a illustrates conditional response probabilities for each of the four experimental conditions: forward-pre-cued, forward-post-cued, backward-pre-cued, and backward post-cued. Panel b illustrates the same conditions as in a, but only for recalls following the first order error of $lag = +2$. Significance markers indicate a difference as a function of recall direction with the following: $\star\star$ indicates $p < 0.001$. Error bars indicate bootstrapped 95% confidence intervals.

call also led to an increase in the probability of $lag = -1$ transitions ($F(1, 488) = 11.41$, $MSE = 0.25$, $p < 0.001$). These two factors, however, did not exhibit a reliable interaction ($F(1, 488) = 0.90$, $MSE = 0.02$, n.s.).

As opposed to transitions of $lag = +1$, subjects can only make transitions of $lag = -1$ after skipping at least one item in the presented list during recall. In addition, if a subject has made multiple errors, some items may not be available without repeating an item (which is rare in serial recall). Therefore, to more directly interpret the difference in fill-in errors as a function of recall direction, we examined transitions made after subjects committed the first error of $lag = +2$ in a recall sequence (Figure 4b), i.e. skipping over one item in the input sequence during recall. In this analysis, transitions of $lag = -1$ correspond to filling in a skipped over item, while transitions of positive lag correspond to continuing recall in the instructed direction. Analyses of transitions following the very first order error provided an avenue to illuminate the difference between so called "fill-in" and "in-fill" transitions (defined as transitions that either fill in an omitted item or continue on in the direction of recall after an omission, e.g., Osth and Dennis (2015); Surprenant, Kelley, Farley, and Neath (2005); Kahana et al. (2010)). Here, we see that fill-in transitions occur more frequently in backward than in forward recall ($F(1, 74) = 4.55$, $MSE = 0.21$, $p < 0.05$). We did not observe a main effect of cue condition ($F(1, 74) = 2.05$, $MSE = 0.08$, n.s.), or an interaction between cue condition and recall direction ($F(1, 250) = 0.18$, $MSE = 0.01$, n.s.). One potential explanation for the increased probability of fill-in transitions in backward recall is the effect of recency, as the fill-in item will be more recent than in-fill items. Consistent with this idea, in forward recall, following the first order error, subjects are more likely to make transitions of in-fill transitions (i.e. positive lags of any size) relative to backward ($t(241) = 2.16, p < 0.05$). However, if we split the analysis by $lag = +1$ and $lag > +1$, we do not observe significant differences ($t(234) = 1.45$, n.s.).

If recency affects the relative availability of transitions, we would also expect the highest probability of $lag < 0$ transitions at early output positions in backward recall, as the most recently presented items are the earliest in the recall sequence. To test this hypothesis we replicated the conditional response probability analysis of all transitions (Figure 4a) separately at each output position in recall. This analysis demonstrates that the probability of making a fill-in transition is higher in backward than in forward recall across nearly all output positions (Figure 5) [2]. Consistent with the effect of recency, we found the most reliable difference in fill-in transitions at output position 3. Additionally, we repeated the conditional probability analysis (Figure 4b) separately at each output position only including recalls following the first order error (Figure 6). This analysis confirms the finding of the largest difference in the probability of fill-in transitions occurring earliest in output.

One could argue that the difference in probability of fill-in errors between forward and backward recall could result from a difference in the probability of terminating recall. According to this reasoning, subjects may terminate recall rather than fill in skipped over items in forward recall. Figure 7 shows the probability of terminating recall immediately following the first order error in forward and backward serial recall. No differences emerged as a function of recall direction ($t(375) = 1.27$, $n.s.$). This finding serves as evidence against the hypothesis that termination underlies the differential fill-in rates between forward and backward recall.

List Length Effects

To prevent subjects from adopting list-length specific strategies, we varied list length between six and 12 words. We also varied the range of list lengths across three subject groups who studied and recalled lists of six, nine and 12 items, seven, nine and 11 items, or eight, nine and 10 items. The list lengths within subject varied such that subjects never studied two consecutive lists of the same length. As we collected the largest quantity of data in the list length nine condition, we presented those results in the preceding sections. Here, we report on how each of the key features of the data varied as a function of list lengths ranging from six to 12.

Figure 8a shows that the near-equivalence of overall forward and backward recall performance appears be consistently across all seven list lengths. To test for any effect of list length, as well as the list length condition that each subject was grouped into, we ran a $3 \times 3 \times 2$ mixed-design ANOVA, with recall direction as a within-subject variable, relative list length as a within-subject variable (such that each list length within a group was analyzed as a short, medium, or long list), and the grouping of list length as a between-subject variable. We found an expected main effect of relative list length (recall probability declined with increasing list length ($F(2, 1134) = 395.68$, $MSE = 2.65$, $p < 0.001$)), but we did not find a main effect of recall direction ($F(1, 567) = 1.22$, $MSE = 0.04$, n.s.)

[2]The only exception being expected output position two, for which these transitions would be relatively rare.

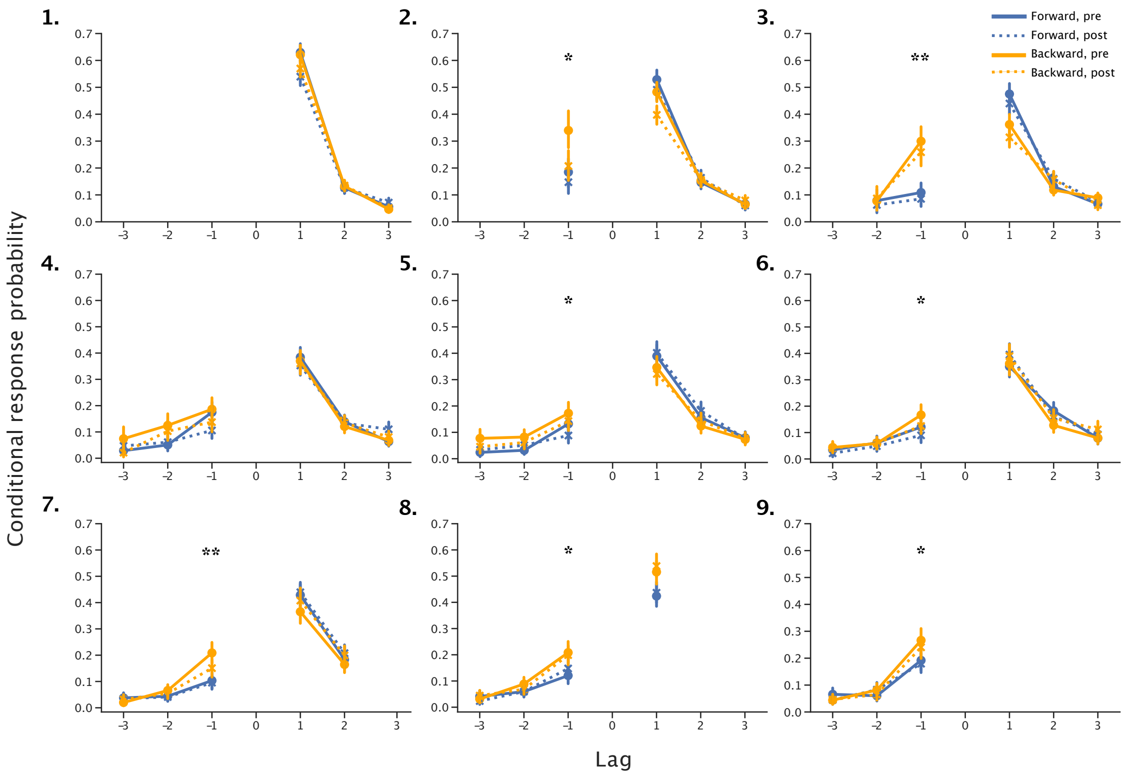

Figure 5. **Temporal clustering in serial recall for each expected output position.** Probability of recalling an item as a function of its lag (distance, in items) from the just-recalled item at each expected output position. For backward recall, we reversed the definition of lag so that it matches the expected output order. Thus, a lag of +1 marks a correct transition in both forward and backward recall, and equivalence of the conditions indicates symmetric forward and backward recall. Each panel illustrates conditional response probabilities for each of the four experimental conditions: forward-pre-cued, forward-post-cued, backward-pre-cued, and backward post-cued. Significance markers indicate a difference between forward and backward probability of fill-in errors at each output position; $\star$ indicates $p < 0.05$, $\star\star$ indicates $p < 0.001$ after conducting an FDR test for multiple comparisons. Error bars indicate bootstrapped 95% confidence intervals.

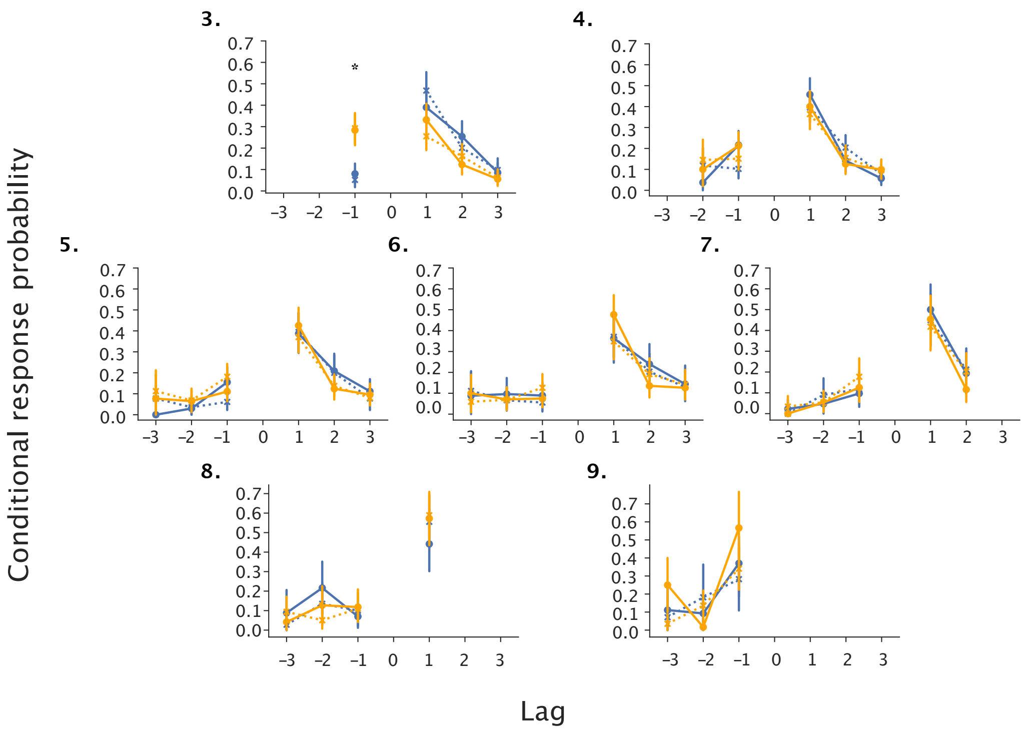

Figure 6. **Temporal clustering in serial recall for each expected output position following the first order error.** Probability of recalling an item as a function of its lag (distance, in items) from the just-recalled item at each expected output position following the first order error. For backward recall, we reversed the definition of lag so that it matches the expected output order. Thus, a lag of +1 marks a correct transition in both forward and backward recall, and equivalence of the conditions indicates symmetric forward and backward recall. Given that this analysis only considers transitions made after the first order error, the third expected output position is the earliest in the sequence with data. Each panel illustrates conditional response probabilities for each of the four experimental conditions: forward-pre-cued, forward-post-cued, backward-pre-cued, and backward post-cued. Significance markers indicate a difference between forward and backward probability of fill-in errors at each output position; ⋆ indicates $p < 0.05$, ⋆⋆ indicates $p < 0.001$ after conducting an FDR test for multiple comparisons. Error bars indicate bootstrapped 95% confidence intervals.

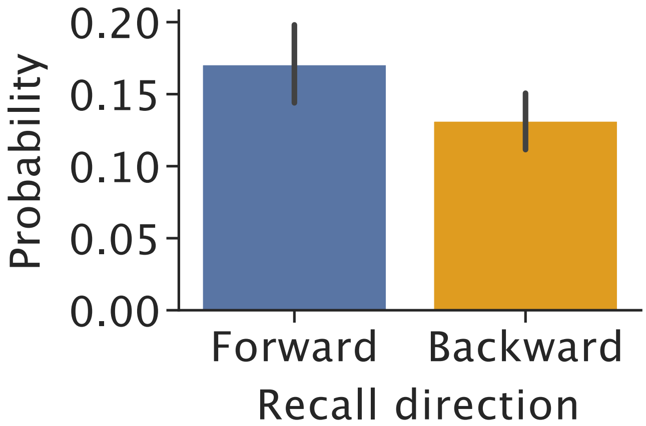

Figure 7. **Recall termination in serial recall following the first order error**. Each bar illustrates the probability of terminating recall immediately after committing the first order error. Error bars indicate bootstrapped 95% confidence intervals.

or a main effect of list length grouping ($F(2, 567) = 1.28$, $MSE = 0.23$, n.s.). We did, however, find a significant interaction between relative list length and recall direction, as evident in the slightly higher performance in forward serial recall at the shortest list lengths ($F(2, 1134) = 21.86$, $MSE = 0.08$, $p < 0.001$).

Figure 8b shows that subjects more often initiate recall accurately in backward than in forward recall for nearly all list lengths. Conducting the same ANOVA model as described above on recall initiation revealed a main effect of relative list length ($F(2, 1132) = 43.88$, $MSE = 0.85$, $p < 0.05$), a main effect of recall direction ($F(1, 566) = 23.70$, $MSE = 4.10$, $p < 0.001$) and a significant interaction between these variables ($F(2, 1132) = 32.76$, $MSE = 0.36$, $p < 0.001$). In addition, we found a significant three way interaction between recall direction, relative list length, and list length grouping ($F(4, 1132) = 3.82$, $MSE = 0.07$, $p < 0.01$). This interaction indicates that the degree to which correct initiation differ as a function of recall direction differs depending on the relative list length that an individual subject encounters, however, the degree to which this relative list length results in differences depends upon the grouping a subject is in. For example, the difference between correct initiation in forward and backward recall is significantly different at list lengths six, nine, and 12, but the difference is relatively consistent at list lengths eight, nine, and 10.

Finally, we examined subjects tendency to "fill in" following omission errors in forward and backward recall as a function of list length. Based on the prior literature, we expected to find a significant effect of list length on fill in errors, with greater fill in rates occurring for shorter list lengths. Once again, we conducted an ANOVA on these variables. We found a main effect of relative list length ($F(2, 1012) = 8.41$, $MSE = 0.11$, $p < 0.001$), as well as a main effect of recall direction ($F(1, 506) = 92.06$, $MSE = 4.11$, $p < 0.001$). In addition, we found a significant interaction between relative list length and list length grouping ($F(4, 1012) = 8.85$, $MSE = 0.12$, $p < 0.001$), which demonstrates that the probability of making a fill-in transition differs as a function of relative list length depending on the list length grouping that a subject is in. For example, the probability of making a fill-

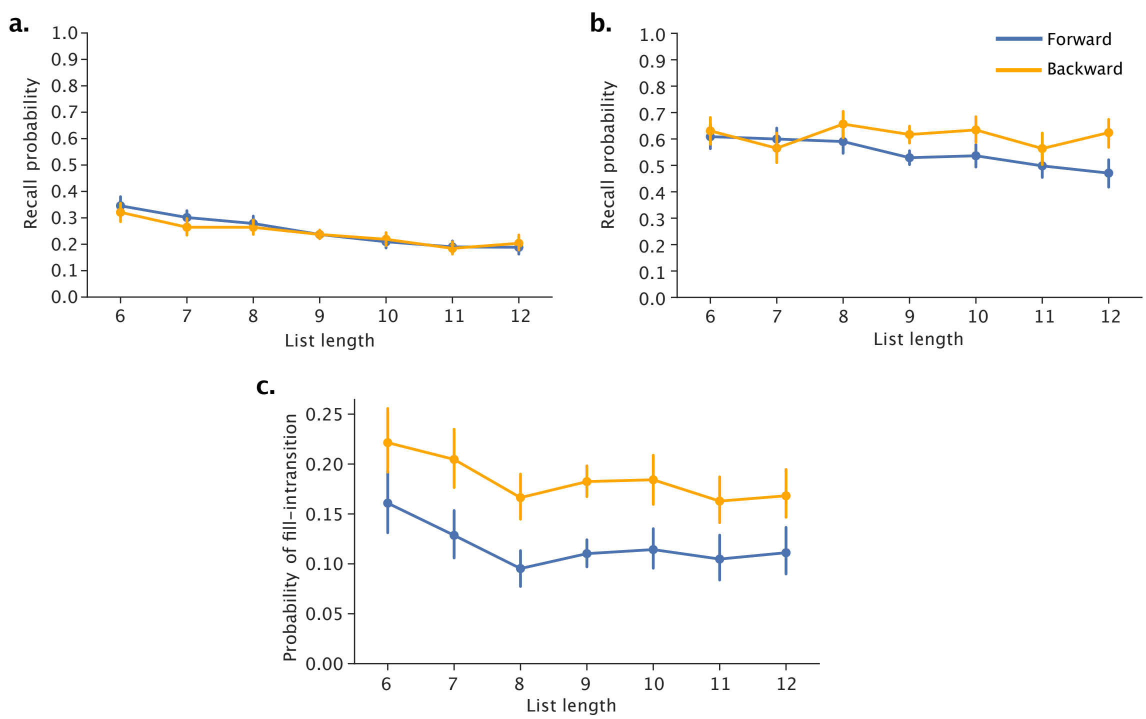

Figure 8. **Effects of list length on overall recall performance, the accuracy of recall initiation, and the forward-bias in recall transitions. a.** The overall probability of correct recall at each list length, scored using the methods described in Solway et al. (2012). **b.** The probability of correct first recall for each list length. **c.** The difference between the conditional response probability of committing a fill-in transition after the first order error in backward and forward recall for each list length. Error bars indicate bootstrapped 95% confidence intervals.

in transition at list length six is very different from that at list length 12, but the probabilities at list length eight and 10 do are not very different. The degree of difference between forward and backward fill-in probability, however, does not depend on this interaction.

Our examination of list lengths ranging from six to 12 generally supports conclusions drawn from our more detailed analysis of the list-length nine condition. With the exception of the shortest lengths, we see similar accuracy for forward and backward recall and more accurate initiation in backward recall (Figure 8a,b). The latter finding should not surprise readers in light of the diminishing recency of the first list item in forward recall. Across all list lengths, backward recall shows a higher likelihood of committing fill-in errors than forward recall, further supporting the idea that recency acts as the mechanism by which forward and backward recall dynamics differ (Figure 8c).

Discussion

The present study sought to re-examine questions regarding directionality of serial recall and the extent to which forward and backward serial recall reflect distinct cognitive processes. We report findings from 570 subjects who completed an online serial recall experiment that included within-subject manipulations of recall direction, cueing, and list length. In the pre-cued condition, subjects learned the expected direction of recall prior to the start of the list; in the post-cued condition, directional instructions appeared immediately after the last study item. By drawing list items from a large word pool (open set), each list contained a unique set of words, thus minimizing interlist interference.

Whereas previous studies using consonants, digits, or closed sets of words found superior performance in forward recall (Liu & Caplan, 2020; Li & Lewandowsky, 1993; Bireta et al., 2010; Hinrichs, 1968; Farrand & Jones, 1996), studies using open sets of words have often failed to find a forward advantage (Madigan, 1971; Guèrard et al., 2012; Li & Lewandowsky, 1995; Thomas et al., 2003). Consistent with these studies, we did not observe reliable differences in overall accuracy for forward and backward recall. This result was consistent across all list lengths tested, ranging from six to 12 items (see Figure 8a). Analyses of recall probability as a function of serial position demonstrated a higher recall probability for end-of-list items in backward recall (recency) and a lower probability for mid-list items, consistent with prior research (Donolato et al., 2017; Li & Lewandowsky, 1993). Pre-cuing subjects to the direction of recall resulted in superior recall performance in both forward and backward recall, particularly for items early in output (Figure 1c). Subjects may differentially employ rehearsal strategies in such pre-cued trials in a way that is consistent across recall directions. To better understand these directional serial position effects, which appeared across multiple methodologies used to determine recall accuracy (see Figure 2), we turned to analyses of recall initiation and dynamics, as discussed below.

In prior free-recall studies, a decomposition of the retrieval process into initiation, transitions, and termination has proven valuable in uncovering the cognitive processes involved in memory search (Kahana, 2020). To further elucidate differences in recall dynamics as a function of recall direction, we applied such analyses to our data. Consistent with the finding of enhanced recency in backward recall, we also found more accurate backward recall initiation for all but the shortest list lengths. This finding was more pronounced in the post-cued condition, but was evident in both cuing conditions.

Following omission errors, the most likely response was a correct transition as defined by our relative-order scoring criteria. In forward recall, we observed a small but reliable advantage for these correct transitions. Transitions in the incorrect direction occurred more frequently in backward than in forward recall, especially in early output positions. Analyses of transitions following the first order error provided an avenue to further analyze this finding, illuminating the difference between so called "fill-in" and "in-fill" transitions (defined as transitions that either fill in an omitted item or continue on in the direction of recall after an omission, e.g., Osth and Dennis (2015)). Whereas we found a higher probability of making an in-fill transition than a fill-in transition in both forward and backward recall[3], the probability of making a fill-in transition was significantly higher in backward than in forward recall. This result appeared consistently at all list lengths. We

[3]Subject's greater tendency to commit in-fill than fill-in responses in our experiment align with a detailed study of such errors provided by Osth and Dennis (2015). Whereas fill-in errors occur more frequently when subjects study short lists with low-rates of omissions, in-fill errors predominate in cases with high-omission rates and when the experiment does not encourage subjects to mark cases where they knowingly skipped over items (as in the present study).

also considered the possibility that subjects would differ in their probability of terminating recall immediately following the first order error. As shown in Figure 7, the probability of stopping did not differ on the basis of recall direction.

Our comparison of recall initiation and recall transitions supports two main conclusions. First, we see recency as a fundamental mechanism differentiating forward and backward recall. Recency benefits backward recall by making the end-of-list items highly accessible, with this effect appearing most strongly when subjects know that they will be recalling the last items first (Figure 3). But after making an early omission error in backward recall, recency serves as a source of negative interference, leading subjects to reverse direction and thereby making it harder to reach the earlier, and less recent, list items (Figures 4-6). The advantage of backward initiation over forward becomes more pronounced as list length increases (Figure 8b), however, the difference in making reversals (fill-in transitions) remains constant across list lengths (Figure 8c). These differences account for the small, yet significant, interaction between list length and recall direction, with forward recall performance declining slightly more as list length increases (Figure 8a).

In addition to the effect of recency, we see a small but significant advantage for forward recall when conditionalizing on a correct prior recall. This forward asymmetry disappears when measuring overall recall levels likely because subjects have greater difficulty successfully initiating in forward recall. The forward advantage observed in conditional measures of recall transitions aligns with earlier findings of a small but reliable forward asymmetry in probed recall of serial lists, particularly in situations that allow for compound cueing (e.g., Kahana & Caplan, 2002). One possibility is that this reflects a chunking mechanism with an inherent forward asymmetry (e.g., Farrell, 2012; Murdock, 1995). Another possibility is that this asymmetry arises from contextual retrieval dynamics used for modeling free recall (Howard & Kahana, 2002; Kahana, 2020).

The convenience of recruiting subjects through an online platform allowed us to efficiently gather data from $> 27,000$ trials of serial recall (48 trials $\times$ 570 subjects) at a time when COVID-restrictions greatly limited our ability to conduct in-person experiments. Although this large dataset afforded the power for our conditional output-order analyses, we nonetheless recognize several important limitations that should be addressed in future work. First, we chose to conduct all of our key comparisons within subject. As a result, subjects knew that they would be asked to recall lists in both forward and backward order, and under varied cuing and list length conditions. As such, our results may not fully generalize to conditions where subjects can optimize their encoding and retrieval strategies for recall in a particular direction. This would tend to reduce differences between forward and backward recall relative to what you might expect in a between subject comparison. Further, the effectiveness of precuing may have been greater had subjects been able to rely on being precued on all trials of the experiment. We considered addressing these limitations by conducting between-subject followup experiments. However, facing very high rates of data exclusion ($> 50\%$ in our study) and the likelihood that these rates would differ between groups, we feared that this would introduce a serious confound into any between-subject comparisons. Second, we designed our experiment to mimic the vocal recall procedure used in many recall experiments, but with typed rather than spoken responses. This procedure does not provide subjects with a convenient way of marking omissions, or backtracking to fill in items whose positions come to mind after making a subsequent recall. As noted by

Osth and Dennis (2015) this leads to a preponderance of in-fill as compared with fill-in errors. Future work should compare the present results with those obtained using recall procedures that allow for the marking of omissions.

The present findings buttress recent arguments for a unified theoretical analysis of serial and free recall (Grenfell-Essam, Ward, & Tan, 2017). Whereas models of free recall have emphasized recency-sensitive retrieval processes (Lohnas, Polyn, & Kahana, 2015; Sederberg, Howard, & Kahana, 2008; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Farrell, 2012) models of serial recall emphasize primacy mechanisms (e.g., Burgess & Hitch, 1999, 2006; Page & Norris, 1998; Farrell & Lewandowsky, 2004). The present comparison of forward and backward recall implicates recency as a potentially important explanatory factor in the analysis of serial order memory, both in considering the difficulty of recall initiation and in identifying potential sources of interference following omissions. Our data also help to resolve a long-standing puzzle regarding why forward asymmetry effects that appear consistently in free recall, frequently fail to show up in comparisons of forward and backward serial recall. Our data indicate that a recency advantage in backward recall initiation obscured a small but highly reliable forward asymmetry seen in transitions.

References

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

Bireta, T. J., Fry, S. E., Jalbert, A., Neath, I., Surprenant, A. M., Tehan, G., & Tolan, G. A. (2010). Backward recall and benchmark effects of working memory. *Memory & Cognition*, *38*(3), 279-291. doi: 10.3758/MC.38.3.279

Bunsey, M., & Eichenbaum, H. B. (1996). Conservation of hippocampal memory function in rats and humans. *Nature*, *379*(6562), 255-257.

Burgess, N., & Hitch, G. J. (1999). Memory for serial order: A network model of the phonological loop and its timing. *Psychological Review*, *106*(3), 551–581.

Burgess, N., & Hitch, G. J. (2006). A revised model of short-term memory and long-term learning of verbal sequences. *Journal of Memory and Language*, *55*, 627–652.

Ceraso, J. (1967). The interference theory of forgetting. *Sci Am*, *217*(4), 117–21.

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. doi: 10.1037/0033-295X.112.1.3

Donolato, E., Giofrè, D., & Mammarella, I. C. (2017, May). Differences in verbal and visuospatial forward and backward order recall: a review of the literature. *Frontiers in Psychology*, *8*, 1-12. doi: 10.3389/fpsyg.2017.00663

Drewnowski, A., & Murdock, B. B. (1980). The role of auditory features in memory span for words. *Journal of Experiment Psychology: Human Learning and Memory*, *6*(3), 319-332.

Ekstrand, B. R. (1966). Backward associations. *Psychological Bulletin*, *65*, 50–64.

Farrand, P., & Jones, D. (1996). Direction of report in spatial and verbal serial short-term memory. *The Quarterly Journal of Experimental Psychology*, *49A*(1), 140-158. doi: 10.1080/713755611

Farrell, S. (2012). Temporal clustering and sequencing in short-term memory and episodic memory. *Psychological Review*, *119*(2), 223-271. doi: 10.1037/a0027371

Farrell, S., & Lewandowsky, S. (2004). Modelling transposition latencies: Constraints for theories of serial order memory. *Journal of Memory and Language*, *51*, 115-135.

Grenfell-Essam, R., Ward, G., & Tan, L. (2017). Common modality effects in immediate free recall and immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(12), 1909-1933. doi: 10.1037/xlm0000430

Guèrard, K., Saint-Aubin, J., Burns, S. C., & Chamberland, C. (2012). Revisiting backward recall and benchmark memory effects: a reply to bireta et al. (2010). *Memory & Cognition*, *40*, 388-407. doi: 10.3758/s13421-011-0156-2

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

Hinrichs, J. V. (1968). Prestimulus and poststimulus cuing of recall order in the memory span. *Psychonomic Science*, *12*, 261-262.

Hintzman, D. L. (2015). Is memory organized by temporal contiguity? *Memory & cognition*, 1–11.

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *25*(4), 923–941. doi: 10.1037/0278-7393.25.4.923

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

Kahana, M. J. (2002). Associative symmetry and memory theory. *Memory & Cognition*, *30*, 823–840.

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

Kahana, M. J., Aggarwal, E. V., & Phan, T. D. (2018). The variability puzzle in human memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *44*(12), 1857–1863. doi: 10.1037/xlm0000553

Kahana, M. J., & Caplan, J. B. (2002). Associative asymmetry in probed recall of serial lists. *Memory & Cognition*, *30*(6), 841-849.

Kahana, M. J., Diamond, N. B., & Aka, A. (in press). Oxford handbook of human memory. In M. J. Kahana & A. D. Wagner (Eds.), (Vol. 1, chap. Laws of Human Memory). Oxford University Press.

Kahana, M. J., Mollison, M. V., & Addis, K. M. (2010). Positional cues in serial learning: The spin list technique. *Memory & Cognition*, *38*(1), 92-101.

Knowlton, B. J., & Schorn, J. M. (in press). Oxford handbook of human memory. In M. J. Kahana & A. D. Wagner (Eds.), (2nd ed., chap. Procedural and Motor Learning). Oxford, U. K.: Oxford University Press.

Li, S., & Lewandowsky, S. (1993). Intralist distractors and recall direction: Constraints on models of memory for serial order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *19*, 895-908.

Li, S., & Lewandowsky, S. (1995). Forward and backward recall: Different retrieval processes. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *21*, 837-847.

Liu, Y. S., & Caplan, J. B. (2020, July). Temporal grouping and direction of serial recall. *Memory & Cognition*, *48*, 1295-1315. doi: 10.3758/s13421-020-01049-x

Lohnas, L. J., Polyn, S. M., & Kahana, M. J. (2015). Expanding the scope of memory search: Modeling intralist and interlist effects in free recall. *Psychological Review*, *122*(2), 337–363. doi: 10.1037/a0039036

Madigan, S. A. (1971). Modality and recall order interactions in short-term memory for serial order. *Journal of Experimental Psychology*, *87*(2), 294-296.

Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior Research Methods*, *44*(1), 1–23. doi: 10.3758/s13428-011-0124-6

Miller, J. F., Weidemann, C. T., & Kahana, M. J. (2012). Recall termination in free recall. *Memory & Cognition*, *40*, 540–550.

Murdock, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482-488. doi: 10.1037/h0045106

Murdock, B. B. (1982). A theory for the storage and retrieval of item and associative information. *Psychological Review*, *89*, 609-626.

Murdock, B. B. (1995). Developing TODAM: Three models for serial-order information. *Memory & Cognition*, *23*, 631-645.

Murdock, B. B. (1997). Context and mediators in a theory of distributed associative memory (TODAM2). *Psychological Review*, *104*, 839-862. doi: 10.1037/0033-295X.104.4.839

Osth, A. F., & Dennis, S. (2015). The fill-in effect in serial recall can be obscured by omission errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *41*, 1447–1455.

Page, M. P. A., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review*, *105*, 761-781. doi: 10.1037/0033-295X.105.4.761-781

Rizzuto, D. S., & Kahana, M. J. (2001). An autoassociative neural network model of paired-associate learning. *Neural Computation*, *13*, 2075-2092.

Sederberg, P. B., Howard, M. W., & Kahana, M. J. (2008). A context-based theory of recency and contiguity in free recall. *Psychological Review*, *115*(4), 893–912. doi: 10.1037/a0013396

Solway, A., Murdock, B. B., & Kahana, M. J. (2012). Positional and temporal clustering in serial order memory. *Memory & Cognition*, *40*(2), 177–190.

Surprenant, A. M., Kelley, M. R., Farley, L. A., & Neath, I. (2005, April). Fill-in and infill errors in order memory. *Memory*, *13*(3-4), 267-273. doi: 10.1080/09658210344000396

Thomas, J. G., Milner, H. R., & Haberlandt, K. F. (2003, March). Forward and backward recall: Different response time patterns, same retrieval order. *Psychological Science*, *14*(2), 169-174. doi: 10.1111/1467-9280.01437

Tulving, E. (1985). How many memory systems are there? *American psychologist*, *40*(4), 385.

Appendix A

We excluded subjects who completed the full length of the experiment based on three factors. Firstly, we asked subjects whether they took notes to boost their performance upon completion of the experiment. Subjects who answered "yes" or who did not provide an answer were excluded from analysis. This accounted for 648/742 exclusions. Subjects who reported any language other than English as their native language were excluded, accounting for 64/742 exclusions. All subjects who did not recall any words were also excluded, accounting for the remaining 30 exclusions.

Appendix B

We automatically spell-checked and scored recalls using an algorithm identical to that used by Healey (2018) in a similar typed-recall task. We considered submitted words to be misspelled if they did not match any previously presented word and did not appear in Webster's Second International Dictionary (https://libraries.io/npm/web2a). We corrected misspellings to the most similar previously presented word (based on Damerau-Levenshtein distance; Damerau, 1964) if the misspelling was closer to that word than to 90% of the words in the dictionary.