# Forward and Backward Recall Dynamics

Dougherty, M.R., Halpern, D., and Kahana, M.J.

University of Pennsylvania

## Author Note

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. This research was supported by NIH grant MH55687 to M.J.K. Correspondence concerning this article should be addressed to Michael J.Kahana, kahana@psych.upenn.edu.

# Abstract

Although possible to recall in both forward and backward order, recall proceeds most naturally in the order of encoding. Prior studies ask whether and how forward and backward recall differ. We reexamine this classic question by studying recall dynamics while varying the predictability and timing of forward and backward cues. Although overall accuracy did not differ by recall direction, recall dynamics highlight key distinctions. Forward recall exhibits a modest advantage for correct transitions following errors, independent of cueing predictability and list length. Without consistent directional cueing, participants initiate backward recall more accurately, but this effect reverses with predictable directional cues. Following omissions, participants commit more fill-in errors in backward recall. Our findings implicate an asymmetric, cue-dependent retrieval process underlying forward and backward recall, with relative contributions of primacy and recency depending on directional predictability.

# Forward and Backward Recall Dynamics

## Introduction

The directional vs. symmetric nature of associative memory has puzzled theorists and experimentalists since the genesis of our field (Murdock, 1956; Asch & Ebenholtz, 1962). Whereas people have great difficulty reciting a well-learned sequence in reverse order (e.g., Tulving, 1985), following a single learning trial, participants can often recall a list in reverse order at similar accuracy levels to forward order (e.g., Madigan, 1971). Similar levels of overall recall, however, do not preclude differences in the way individuals retrieve sequences in reverse order or whether they encode sequences differently when expecting forward vs. backward recall. A large literature comparing forward and backward recall under varying conditions has yielded a complex landscape of empirical findings, with some variables exerting differential influence on forward and backward recall, and other variables showing similar or identical effects (see Donolato, Giofrè, & Mammarella, 2017, for a review). The present paper approaches the comparison of forward and backward recall by examining the dynamics of retrieval, quantifying participants' tendency to correctly initiate recall and their ability to successfully transition among studied items as a function of their relative position, or lag, within the list (Kahana, Diamond, & Aka, in press). This approach has proven valuable in the analysis of free recall, but has only recently been applied to forward serial recall (Ward, Tan, & Grenfell-Essam, 2010; Solway, Murdock, & Kahana, 2012; Spurgeon, Ward, Matthews, & Farrell, 2015). We also consider the role that foreknowledge of recall order plays in determining recall dynamics (e.g., Guitard & Saint-Aubin, 2022). To the extent that successful recall in a given direction depends on encoding processes, providing participants with foreknowledge of recall direction could impact both recall performance and retrieval dynamics.

Analyses of recall transitions estimate conditional probabilities: given that a participant just recalled an item from serial position $i$, we can ask how likely it is that their next response will be an item studied in serial position $i + \text{lag}$, conditional on the availability of that item. Such analyses may further condition on the position of the item in the output sequence. Estimating conditionals requires more data than estimating means, and depending on the correlational structure of the conditional distributions, it could require significantly more data. Therefore, we used an online platform to obtain data from 1077 participants who each recalled 48 lists. To describe the motivations of our specific experiments, we first provide a brief review of relevant theoretical issues raised in serial order memory literature.

## Theoretical Background

Lewandowsky and Farrell (2008) and Hurlstone (in press) both provide extensive reviews of the major models concerning serial order memory. Both note that several prominent models designed to account for immediate serial recall cannot perform backward recall at all, or at least cannot do so without adding auxiliary assumptions (Farrell, 2006; Burgess & Hitch, 1999; Botvinick & Plaut, 2006). In general, the creators of these models devised accounts of the rapid forward-ordered readout characterizing immediate recall of short lists (short-term memory) with the implicit assumption that participants recruit other memory mechanisms to solve the problem of backward recall (e.g., Li & Lewandowsky, 1993, 1995). For example, some have argued that participants recall in reverse order by successively recalling a set of items in forward order and peeling off the final item (e.g., Page & Norris, 1998; Thomas, Milner, & Haberlandt, 2003).

**Chaining and Positional Coding Theories.** Students of memory trained in the early classic theories of serial learning will readily envision using the same associative structures to account for both forward and backward serial recall. The two most prominent early models employ mechanisms of associative chaining or positional coding (Kahana, 2012). According to chaining models, each studied item associates with its predecessor alone (simple chaining, e.g., Lewandowsky and Murdock (1989)) or with multiple preceding items as a decreasing function of their recency (compound chaining, Murdock

(1995); Solway et al. (2012), see Osth and Hurlstone (in press) for a nice analysis of these models). In backward recall, chaining models posit that participants use the last item, assumed to be accessible at the start of recall, to retrieve the next to last item, which in turn cues subsequent items (the cue being either a single item, or a weighted sum of previously recalled items). In forward recall, chaining models require an additional process to access the first item in the list, which is presumably no longer accessible in memory. Lewandowsky and Murdock (1989) assume a start of list cue that participants can access at the time of test (c.f., Logan, 2021).

Chaining models predict monotonic recency in backward recall and monotonic primacy in forward recall (Solway et al., 2012). They also predict a greater difficulty with initiation in forward recall owing to the possible failure of the initiation mechanism (which is specific to forward recall). Some chaining models assume symmetric forward and backward associations (e.g., convolution in Murdock's (1982) TODAM model), whereas other models allow for asymmetric forward-biased retrieval, as in matrix models (see Rizzuto and Kahana (2001) and Caplan (2005) for details) and some neural network implementations (Hertzog, Dixon, Hultsch, & MacDonald, 2003).

Positional coding models assume the existence of an abstract representation of ordinal list position, or a temporal code, that associates with each item during serial learning (Ladd & Woodworth, 1911; Conrad, 1965). At test, participants can flexibly access this positional representation and use it to cue item recall. In these models, items do not cue one another directly, only positions cue items. By successively cueing memory with increasing or decreasing positional codes, one can simulate either forward or backward recall. Whereas simple chaining models predict monotonic primacy and recency, simple positional models produce a symmetric U-shaped serial position curve. Both models can be extended with auxiliary assumptions to more closely match the empirical shape of the serial position curves (Lewandowsky & Murdock, 1989).

**Retrieved Context Theory.** Retrieved context theory (RCT, Kahana (2020)) provides an alternative framework to understand how participants may recall lists in either forward or backward order. Building upon earlier models that incorporate time-varying context into associative memory (Glenberg & Swanson, 1986; Mensink & Raaijmakers, 1988), Howard and Kahana (1999, 2002) championed the idea that remembering an item calls back its encoding context, which in turn serves as a retrieval cue for subsequent recalls. RCT proposes that items and context become reciprocally associated during study: context retrieves items, and items retrieve context. But because context is drifting through a high dimensional space, and because such a drift process is correlated with the passage of time, contextual retrieval enables the rememberer to jump back in time to an earlier contextual state. Howard and Kahana (2002) deviated from earlier approaches by positing that context evolves due to the nature of experience itself, rather than as a result of a random input to a stochastic process. They proposed a recursive relation in which the contexts retrieved by an item serve as the (additive) input to the evolution of a multidimensional context vector. This contextual retrieval process, in cueing subsequent recalls, generates temporal clustering. That is, retrieving the context associated with item $i$ triggers memories of items studied in similar contexts, such as items $i-1$ and $i+1$ (see Figures 3 and 7).

Because the current state of context is more similar to recent than remote contexts, the model naturally produces recency. Adding a primacy gradient (Sederberg, Howard, & Kahana, 2008; Polyn, Norman, & Kahana, 2009) can potentially allow the model to begin recalling at the start of the list, but only if the activation of the first item exceeds that of all other items. Reducing the reliance on semantic associations would allow the model to simulate forward serial recall by reducing the models' tendency to make non-adjacent transitions. Varying the primacy gradient would determine the balance between overall forward and backward recall performance under conditions of pre-cueing. Under conditions of post-cuing, it is hard to imagine a version of this model, with a single set of parameters, that could modulate its behavior between forward and backward recall.

Recent versions of RCT (Lohnas, Polyn, & Kahana, 2015; Healey & Kahana, 2016) borrowed response suppression ideas from theories of serial recall to avoid having the model revisit previously recalled items. RCT produces a forward-bias in the contiguity effect, favoring forward transitions. This forward bias arises because recalling an item, $\mathbf{x}_i$, recovers a mixture of the item's associated list context, $\mathbf{c}_i^{\text{list}}$, and its pre-experimental context(s), $\mathbf{c}_i^{\text{pre}}$. Whereas $\mathbf{c}_i^{\text{list}}$ will activate both its predecessors and successors, $\mathbf{x}_{i-1}$ and $\mathbf{x}_{i+1}$, $\mathbf{c}_i^{\text{pre}}$ will only activate items studied in later serial positions as it updated context after the encoding of item $i$ (see Howard & Kahana, 2002). The bias towards forward transitions would tend to produce fill-in transitions in backward recall and in-fill transitions in forward recall.

RCT shares elements with complex chaining models that allow for both remote associations and compound cueing (e.g., Chance & Kahana, 1997; Solway et al., 2012), and with models that explicitly encode temporal information into memory (e.g., Brown, Neath, & Chater, 2007). In RCT, the temporal information derives from the items themselves, and the past associations that they evoke. RCT struggles to explain how participants initiate their responses in forward serial recall. Chaining models face this same theoretical deficiency. The assumption of a primacy gradient offers a solution, but one lacking theoretical elegance.

**Control Processes.** Directional retrieval also calls for some type of control process that allows participants to direct the search of their memory. Although response suppression (e.g., Lewandowsky & Murdock, 1989; Page & Norris, 1998) prevents the model from continuously resampling already recalled items, participants can control their direction of retrieval even when probed to begin recall from the middle of the list (Kahana & Caplan, 2002; Caplan, 2005). Clearly, the direction of recall must be part of the cue. If one allows for list start markers and list end markers, including these in the cue-set could bias retrieval in the backward or forward direction (and would address the problem of initiation in forward recall). Yet, this solution to the cognitive control problem seems unsatisfying. Participants would appear to know, following recall of an item, whether that item came from an earlier or later position (a similar computation easily accomplishes this in most models). If participants can, in fact, steer their retrieval in a particular direction, models of recall should somehow capture this ability (e.g., Caplan, 2005).

## Predicted Forward and Backward Recall Dynamics

Here we discuss the qualitative predictions of the aforementioned models regarding the dynamics of forward and backward recall, and the effects of directional foreknowledge on performance. Chaining, positional coding, and retrieved-context models each provide a plausible account for people's ability to recall items in either the forward or backward direction. Consideration of recall dynamics, however, can help to explicate the strengths and weaknesses of each model.

Forward and backward recall differ in the cues available at the time of recall initiation: Whereas the last item remains active in memory at the moment of recall initiation, the first item must somehow be retrieved without an explicit cue. This asymmetry in cue availability challenges both chaining models and RCT, as these models lack a natural mechanism for retrieving the first list item. However, proponents have offered ad-hoc solutions to this issue, such as the presence of a start cue tagging the first item (Lewandowsky & Murdock, 1989), a primacy gradient enhancing retrievability of early list items (Sederberg et al., 2007), or a rehearsal process that selectively enhances memory for early list items (Atkinson & Shiffrin, 1968; Laming, 2006). In contrast, positional coding models assume that participants associate positional information with each item, thereby allowing recall initiation with either the start or end of the list. However, even positional models can predict superior backward initiation, as some theorists have hypothesized the loss of precision of positional associations over time (Glenberg & Swanson, 1986; Brown et al., 2007). In sum, each model can predict an initiation advantage for backward recall, consistent with the enhanced recency reported by Madigan (1971).

Having considered recall initiation, we now turn to the question of recall transitions. Here consider a participant who studied the sequence ABCDEFG and then recalled an initial sequence of correct responses: AB in forward recall, or GF in backward recall. When participants fail to make a correct transition, all three model families predict a greater tendency to transition to items from nearby positions and proximate lags to the just recalled item (e.g., skipping to item D in both forward and backward recall). Conditioning on a sequence of correct responses obviates any differences between positional distance and relative lag. This locality constraint (e.g., Henson, Norris, Page, & Baddeley, 1996) holds for both forward and backward recall.

Transitions after omissions offer a more interesting scenario for our models: Consider a participant who recalls ABD in forward recall. Participants could then make a backward 'fill-in' transition (recalling C), a forward 'in-fill' transition (recalling F), a repetition/intrusion error, or terminate recall. Considering the balance of the first two cases, positional-coding models predict a greater tendency for fill-in responses in forward recall, especially following omissions of more than one item (e.g. recalling ABED). In such cases, the retrieval cue for the item following the omission correctly matches the encoded position for items corresponding to fill-in transitions (e.g. position 4 for item D). This prediction reflects the advancement of the positional cue by one item on each retrieval, independent of the serial position of the just recalled item. RCT and chaining models, however, use the just recalled item as the cue for the next response and, as such, would predict a greater tendency to make forward in-fill responses.

The assumption of a primacy gradient, however, will tilt the prediction of each model towards greater fill-in tendencies, as primacy will favor early list items. But we would also reasonably expect that skipped items (omissions) will have lower-than-average encoding strength. As such, these items will likely lose in a retrieval competition with a non-recalled adjacent item, favoring forward in-fill transitions. Without formal model fits, we cannot infer a precise ratio of fill-in to in-fill transitions by considering forward recall alone.

Fortunately, the same analysis presented here should also apply to backward recall, allowing us to evaluate the relative balance of these transition types.

In backward recall, we can consider the case of a participant who recalls `GFD`, omitting item `E`. Howard and Kahana's (2002) implementation of context dynamics, inherited by subsequent variants of RCT, favors forward transitions. As such, RCT predicts that participants will tend to *fill in* item `E` in backward recall. RCT further predicts that this fill-in tendency will increase for early transitions due to the biasing effect of end-of-list context as a cue favoring recently encoded items. Chaining models with forward-biased associations similarly predict a greater proportion of fill-in to in-fill errors in backward recall, but without the further predicted biasing effect of recency. Although positional-coding models typically assume symmetric forward and backward associative gradients (e.g. Solway et al., 2012), they predict a tendency towards fill-in transitions in both backward and forward recall because the positional cue following omissions will generally lag behind the associative cue. Positional models' output-order predictions depend on assumptions regarding how positional uncertainty varies with serial position. In any of these models, participants' tendency to omit weakly-encoded items favors in-fill transitions in both backward and forward recall, but this effect should not depend on serial position. A primacy gradient would lead to even greater in-fill transitions to early list items.

Under the simplifying assumption that primacy gradients do not differ between forward and backward recall (as appropriate under conditions of directional post-cueing) clearer distinctions emerge among the models. Whereas positional coding models do not possess an inherent directional bias, RCT drives recall towards later serial positions, leading to relatively greater fill-in transitions in backward than in forward recall. Chaining models with forward-biased associations mimic RCT's predictions. Conditional on successful recall initiation, RCT predicts larger differences between forward and backward recall dynamics than positional coding theory. Further, RCT suggests that the combination of recency-sensitive cue strength, and forward-biased associations, should interfere with correct in-fill transitions in backward recall.

The foregoing discussion of retrieval dynamics failed to consider the possibility that participants could tune the parameters of their model, or even which model they use, based on their foreknowledge of recall direction. Knowing that one must recall a list in the forward vs. backward direction would likely alter one's encoding strategy. When expecting a forward recall test, for example, a participant might allocate fewer encoding resources to end-of-list items, reasonably assuming that they won't get to those items if they can not successfully initiate recall. Rehearsal processes would similarly be subject to cognitive control, and would likewise favor the encoding of early items. Thus, foreknowledge of forward recall could lead to a reduction, or even reversal, of the expected initiation advantage in backward recall. As foreknowledge of recall direction can impact any parameters governing the memory encoding process, we should consider the possibility that participants expecting a forward/backward recall test would increase the encoding strength of forward-going/backward-going associations. This would lead to a larger in-fill to fill-in ratio under directional pre-cueing.

## Experiments

To evaluate the qualitative predictions of chaining, positional coding and RCT models, and to provide a large dataset for future quantitative model evaluation, we conducted two extensive online experiments that recorded the dynamics of participants' responses in forward and backward recall of variable-length word lists. In a typical study of immediate serial recall, participants know that they will recall in forward order prior to the start of each encoding list. However, in studies comparing backward and forward recall, participants may either know the expected order of recall before the encoding period, or they may only learn of the direction of recall just before the retrieval phase. As participants' expectations regarding the direction of recall may influence the manner of recall (Murdock, 1962; Hintzman, 2015), Experiment 1 introduced a within-participant cueing manipulation. On half of the trials, we informed participants in advance of the direction of subsequent recall. On the other half of the trials, participants only learned of the recall direction immediately after studying the list. To the extent that pre-cueing participants 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 varied randomly across trials, with participants recalling in forward order on half of the trials and in backward order on the remaining half. As this variability in pre- vs. post-cueing across trials may have limited participants' ability to optimize their strategies for the direction of recall, we conducted a second experiment in which participants always knew the direction of expected recall. To help participants further optimize their strategy for recall direction, we organized each session into eight blocks of six consecutive trials – four blocks in the forward and four in the backward direction.

## Experiment 1

### Methods

**Participants.** 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, participants 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 participants (570 male, 425 female, 346 unreported) completed Experiment 1 for compensation of $7.50. One hundred and eighteen participants' data were impacted by technical issues, and an additional 698 participants met at least one of the three exclusionary criteria (see Appendix A for exclusionary criteria and analyses of excluded participants). The remaining 525 participants (280 male, 242 female, 3 unreported) contributed to our analyses, and ranged in age from 21.0 to 73.0 years, with a mean age of 37.78 years ($SD = 10.58$). Prior to participating, participants completed a consent form that was approved by the University of Pennsylvania Institutional Review Board.

**Procedure.** Each participant completed 48 study-test trials (list presentation & recall) in total, with 12 trials assigned to each of the four combinations of cueing condition (pre- vs. post-) and recall direction (forward vs. backward). Lists comprised common words, with variable list lengths manipulated both within and between participants such that each participant 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 two main objectives: (1) to reduce expectancy of list termination and any list length-specific encoding strategies by varying list length within participant, and (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.

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[1]. The word pool was open, meaning each trial used unique and unrepeated words.

After providing informed consent, and before beginning the experiment, we gauged participants' attention to our instructions with a simple task. Each participant was instructed to carefully read a description of memory prior to proceeding. A sentence

---

[1] We used the 300 dimensional word2vec model trained on the Google News dataset, as described in Mikolov, Chen, Corrado, and Dean (2013).

inserted within this description instructed participants to reply "never" to a subsequent survey question about their own memory quality, regardless of their subjective answer. Only participants who responded correctly advanced to the actual experiment.

After a brief test of their computer's audio output, participants 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). Participants were instructed to picture each word as it was presented and only focus on the most recently presented word. Following these instructions, participants performed a series of four practice trials. These trials introduced participants to our experimental manipulations (forward and backward recall, pre-cued and post-cued). All four practice trials were nine-word lists. Participants did not receive feedback about their performance at any point in the experiment (practice or test trials).

The experiment prompted participants 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 participants to type recalled words. Participants submitted responses by pressing the space bar, enter, comma, or semicolon keys. The text box was cleared after each word was submitted, preventing participants 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, participants participated in a math task to retain engagement, in which participants 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.

**Data and Code Availability.** The public repository, `http://memory.psych.upenn.edu` contains all experimental data and analysis code.

## Results

We begin our analyses by investigating the effects of direction, cueing, 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 participants 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 et al. (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 between-participant 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, participants exhibited the standard, primacy-dominated serial position effect (Figure 1a). 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). Given that primacy and recency often occur in tandem, the reader may be struck by the absence of recency in forward recall and the absence of primacy in backward recall. One typically finds strong recency in forward serial recall with short lists and auditory presentation. With longer lists of visually presented items, one sees far more modest recency effects unless participants recall the list by inputting responses into specific locations on a form, where the locations provide an additional spatial cue whose distinctiveness peaks at the start and end of the list (Ribback & Underwood, 1950). In studies involving longer lists and vocal recall, one typically does not observe significant recency in forward recall (e.g., Kahana, Mollison, & Addis, 2010; Kahana & Caplan, 2002).

When participants could anticipate the order of recall prior to encoding (pre-cued condition), they exhibited superior performance in both forward and backward recall (Figure 1a). This cueing effect asserts itself in early output positions, i.e. early serial positions in forward recall and late serial positions in backward recall. Accordingly, we conducted 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 and backward recall. For all subsequent statistical tests, we included participants who contributed at least one observation to each relevant condition. Tests of the data shown in Figure 1 revealed a significant main effect of cueing condition ($F(1, 524) = 26.30$, $MSE = 1.87$, $p < 0.001$), and a significant interaction between expected output position and cueing condition ($F(8, 4192) = 24.94$, $MSE = 0.66$, $p < 0.001$), confirming the above observations. A closer examination of Figure 1 indicates that the beneficial effects of cueing occurred at somewhat different output positions in forward and backward recall, as reflected in a significant three-way interaction of cueing condition, output position, and recall direction was significant ($F(1, 8, 4192) = 4.26$, $MSE = 0.11$, $p < 0.001$). Whereas pre-cueing benefits forward recall in early output positions, it appears to harm forward recall in later output positions, possibly because the recency effect 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 et al., 2003), we do not observe a significant main effect of recall direction ($F(1, 524) = 0.009$, $MSE = 0.002$, n.s.). Figure 1b shows the serial position curve calculated using the Solway et al. (2012) method of scoring, with results collapsed across cueing conditions. One 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 cueing condition) using four additional scoring techniques: Supplementary Figure 1b 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. Supplementary Figure 1c shows results using the free recall scoring method, in which a recall is scored as correct regardless of its order in the output sequence. Supplementary Figure 1d 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, Supplementary Figure 1e 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 ($F(8, 4192) = 21.12$, $MSE = 1.17$, $p < 0.001$) 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 et al., 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). Participants initiated recall correctly (with the first expected output position) more often in backward than in forward recall (see Figure 2, $F(1, 521) = 27.84$, $MSE = 4.14$, $p < 0.001$). We also found a significant interaction between cue timing and recall direction ($F(1, 521) = 17.23$, $MSE = 0.67$, $p < 0.001$). In forward recall, post-cueing reduced participants' tendency to initiate accurately ($t(521) = -5.21$, $p < 0.001$), whereas in backward recall, there was no effect of cueing on initiation accuracy ($t(521) < 1$, *n.s.*). Post-hoc tests revealed that the backward recall initiation advantage exists in both the pre-cued ($t(521) = 3.04$, $p < 0.01$) and post-cued conditions ($t(521) = 6.14$, $p < 0.001$). 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 the effects of recall direction and cueing on recall transitions.

In both free and serial recall, participants exhibit strong temporal clustering (Healey, Long, & Kahana, 2019) – 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[2], analyses of temporal clustering must condition on the availability of transitions to different lags. Figure 3a illustrates the temporal clustering effect separately for forward and backward recall and under pre- and post-cueing 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, participants were most likely to make a correct response (as defined by Solway et al. (2012) relative order scoring; Figure 3a). Although we failed to observe a reliable difference in overall forward and backward recall accuracy (see Figure 1b) the conditional probability of a $lag = +1$ transition exhibited a small but reliable forward recall advantage ($F(1, 479) = 17.41$, $MSE = 0.76$, $p < 0.001$) as well as an advantage of pre-cueing ($F(1, 479) = 19.21$, $MSE = 0.35$, $p < 0.001$).

Whereas $lag = +1$ signifies correct relative order transitions, $lag = -1$ transitions reflect participants' tendency to reverse direction and fill in a missing item. Statistical analyses of transitions of $lag = -1$ showed a main effect of recall direction ($F(1, 462) = 71.16$, $MSE = 3.07$, $p < 0.001$), with participants more likely to reverse the direction of recall in backward recall than in forward recall. Pre-cueing participants with the direction of recall also led to an increase in the probability of $lag = -1$ transitions ($F(1, 462) = 9.21$, $MSE = 0.20$, $p < 0.01$). These two factors, however, did not exhibit a reliable interaction ($F(1, 462) = 0.66$, $MSE = 0.01$, $n.s.$), as backward recall exhibited higher fill-in rates in both the pre- and post-cued conditions.

---

[2] Participants very rarely recall previously recalled items, averaging 0.079 repetition occurrences per list in Experiment 1 and 0.048 in Experiment 2. Therefore, we assume that these items do not play a significant role in the competition for retrieval.

As opposed to transitions of $lag = +1$, participants can only make transitions of $lag = -1$ after an omission. In addition, if a participant 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 participants committed the first error of $lag = +2$ in a recall sequence (Figure 3b), 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 of $lag = +2$ 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)). Restricting comparisons to those participants who contributed data in each condition limited the number of participants contributing to this very specific analysis. As this is a conditional analysis, only participants whose recall sequences presented an opportunity for the relevant fill-in and in-fill transitions following transitions of $lag = +2$ in each of the experiment's four conditions were included. Here, we see that fill-in transitions occurred more frequently in backward than in forward recall ($F(1, 69) = 4.56$, $MSE = 0.23$, $p < 0.05$). We did not observe a main effect of cue condition ($F(1, 69) = 2.05$, $MSE = 0.08$, $n.s.$), or an interaction between cue condition and recall direction ($F(1, 69) = 0.18$, $MSE = 0.01$, $n.s.$). 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, participants are more likely to make in-fill transitions (i.e. positive lags of any size) relative to backward ($t(232) = 2.16, p < 0.05$).

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 3a) separately at each output position in recall. Across nearly all output positions, participants are more likely to make fill-in transitions in backward than in forward recall (Figure 4). Consistent with the effect of recency, we found the most reliable difference in fill-in transitions at output position three.

Analyses of overall recall performance as a function of list position or output position (e.g., Figure 1) reflect not only successful initiation and transitions, but also whether recall terminates following a given transition. 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, participants may terminate recall rather than fill in skipped over items in forward recall. To test this hypothesis, we compared the probability of terminating recall immediately following the first-order error in forward and backward recall. No differences emerged as a function of recall direction ($t(361) = 1.35$, *n.s.*). This finding serves as evidence against the hypothesis that termination underlies the differential fill-in rates between forward and backward recall.

To extend the scope of our analyses on recall termination, we conducted a repeated measures ANOVA comparing stopping probability as a function of recall direction, output position, and relative proportion of correct vs. incorrect prior responses (i.e., correct > incorrect, or vice versa). Consistent with prior research (e.g., Miller, Weidemann, & Kahana, 2012), we observed significant main effects of both output position ($F(8, 8944) = 780.40$, $MSE = 65.86$, $p < 0.001$) and correctness ($F(1, 9046) = 27.68$, $MSE = 2.34$, $p < 0.001$) (participants tend to stop after errors and with increasing output position). A significant interaction between these factors indicates that participants tended to terminate recall earlier when they had made more incorrect responses ($F(1, 8, 8899) = 8.03$, $MSE = 0.76$, $p < 0.001$). Additionally, we observed a main effect of recall direction ($F(1, 8901) = 8.57$, $MSE = 0.72$, $p < 0.01$), indicating that participants terminated recall after fewer recall attempts in forward as compared to backward recall. A significant interaction of recall direction and output position ($F(1, 8, 8871) = 2.22$, $MSE = 0.19$, $p < 0.05$) supports this interpretation.

## Experiment 2

Results of Experiment 1 demonstrate three key differences between forward and backward recall: increased correct initiation in backward recall (especially in the post-cued condition), increased correct transitions in forward recall, and increased fill-in transitions in backward recall. If the presumed forward-recall advantage reported in some studies (e.g., Liu & Caplan, 2020) reflects specific encoding or rehearsal strategies, one would expect to find only modest differences between forward and backward recall under post-cueing conditions. Without knowledge of the recall direction in advance of encoding, participants may adopt a standard strategy that works well for both recall directions rather than optimizing their encoding or rehearsal to support forward recall.

By the above logic, however, the absence of notable asymmetries under pre-cueing conditions presents more of a puzzle. A critic may argue that the randomization of pre- and post-cued trials during each session hindered participants' ability to develop an effective strategy for forward recall. Suppose, for example, that forward recall was far easier than backward recall, but only when participants can "lock in" a forward-specific learning strategy. If participants cannot efficiently control their encoding strategy when faced with a situation where recall direction is post-cued on half of the trials, then pre-cueing participants to the direction of recall will not have a major impact on recall asymmetries.

Experiment 2 compares forward and backward recall under conditions designed to help participants differentiate their strategies for the direction of recall. Specifically, we eliminated the post-cued condition, thus ensuring that participants always knew the direction of recall prior to encoding. This should amplify any strategic factors that could take place during encoding to produce differences between forward and backward recall. To further encourage participants to use direction-specific strategies, we divided each session into six-trial blocks, each consisting of forward recall or backward recall trials. By combining pre-cueing and blocking, we sought to give participants the greatest opportunity to adopt direction-specific strategies.

## Methods

Experiment 2 replicated the methods of Experiment 1 with two exceptions: the elimination of the post-cued condition, and the introduction of recall direction blocks. Participants completed eight six-trial blocks in total, four in the forward direction and four in the backward direction (48 trials in total, 24 in each direction). The direction of recall changed between forward and backward with each consecutive block. Prior to the start of each new block, participants were required to acknowledge the new direction by pressing the "F" key for a block of forward recall or the "B" key for a block of backward recall. In addition, an arrow appeared prior to the start of each trial to remind participants of the recall direction. Experiment 2 was also deployed via Amazon Mechanical Turk (MTurk), and participants were required to be from the United States and possess at least a 95% approval rating on MTurk to qualify. Additionally, the task could not be accessed from mobile devices or tablets. A total of 1057 participants (607 male, 447 female, 3 unreported) completed Experiment 2 for compensation of $7.50. Sixty-nine participants' data were impacted by technical issues, and 436 participants met at least one of the three exclusionary criteria (see Appendix A for exclusionary criteria and analyses of excluded participants). The remaining 552 participants (284 male, 266 female, 2 unreported) contributed to our analyses, and ranged in age from 19.0 to 84.0 years, with a mean age of 36.48 years ($SD = 11.99$). Prior to participating, participants completed a consent form that was approved by the University of Pennsylvania Institutional Review Board.

## Results

Once again, we begin our analyses by investigating the effects of direction and serial position on correct recall. As in Experiment 1, we analyzed data using Solway et al. (2012) relative order scoring. The data from Experiment 2 replicate the primacy and recency-dominated serial position effects in forward and backward recall, respectively (Figure 5). To compare retrieval dynamics under conditions designed to help participants optimize their strategy according to recall direction, the remaining analyses again utilize expected output position rather than serial position. For all subsequent statistical tests, we included participants who contributed at least one observation to each relevant condition.

As in Experiment 1, we did not observe a main effect of recall direction on overall recall ($F(1, 551) = 1.42$, $MSE = 0.16$, $n.s.$). Supplementary Figure 1 applies the same analysis to four additional scoring methods, all demonstrating the near equivalence of forward and backward recall performance (a forward advantage did appear for one measure, conditional order scoring). Whereas overall performance (Solway et al. (2012) scoring) did not differ as a function of recall direction, we found a significant interaction between expected output position and recall direction ($F(8, 4408)) = 23.06$, $MSE = 0.50$, $p < 0.001$). In contrast to Experiment 1, the data demonstrate a forward recall advantage in early output positions, and a backward recall advantage in late output positions. These findings suggest a primacy-guided strategy that boosts performance at early output positions in forward recall and late output positions in backward recall.

We next proceeded to analyze recall dynamics. Figure 6 illustrates the probability of initiating recall as a function of expected output position. In contrast to Experiment 1, participants correctly initiated forward recall with higher frequency than backward recall ($t(549) = 3.69$, $p < 0.001$), supporting the idea that consistent foreknowledge of recall direction allowed participants to prioritize primacy items.

Whereas the effect of recall direction on initiation yielded different results between Experiments 1 and 2, participants in both experiments demonstrated similarities in recall transitions (Figure 7a). Among all possible transitions, participants were most likely to make a correct response. Participants demonstrated higher transition accuracy, however, in forward recall, as measured by the probability of $lag = +1$ transitions ($t(474) = 2.49$, $p < 0.05$). On the other hand, participants were more likely to reverse the direction of recall and fill in missing items in backward recall ($t(462) = 3.50$, $p < 0.001$). These results indicate that following a transition of $lag > 1$ (i.e. an error in the correct direction), the probability of reversing the direction of recall to fill in skipped items was higher in backward than in forward recall. Tests of transitions made after the first order error of $lag = +2$ did not demonstrate significance (Figure 7b, $t(241) = 1.19$, n.s.). This is likely due to the infrequent nature of these transitions in pre-cued block trials, decreasing the detectability of differences.

As in Experiment 1, the data on fill-in transitions suggest that recency guides transitions in forward and backward recall. In line with this theory of recency, 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 expected recall sequence. Figure 8 replicates the conditional response probability analyses of all transitions separately at each output position in recall. Consistent with the effect of recency, and with the results of Experiment 1, we found the most reliable difference in fill-in transitions at output position three.

As a final analysis of Experiment 2, we compared participants' propensity to terminate recall as a function of recall direction, output position, and relative proportion of correct vs. incorrect prior responses (i.e., correct > incorrect, or vice versa) using a repeated measures ANOVA. Replicating the results of Experiment 1, we observed significant main effects of both output position ($F(8, 8448) = 776.46$, $MSE = 59.91$, $p < 0.001$) and correctness ($F(1, 8533) = 39.54$, $MSE = 3.05$, $p < 0.001$) (participants tend to stop after errors and with increasing output position). In addition, the probability of termination varied reliably as a function of the interaction between these terms

($F(1, 8, 8425) = 19.53$, $MSE = 1.51$, $p < 0.001$), participants tending to terminate recall earlier in output with more incorrect responses. However, we did not observe any reliable differences as a function of recall direction ($F(1, 8396) = 2.93$, $MSE = 0.23$, $n.s.$), nor as a function of the interaction between recall direction and output position ($F(1, 8, 8383) = 1.99$, $MSE = 0.15$, $n.s.$).

Experiment 2 encouraged the development of direction-specific recall strategies by pre-cueing the direction of recall on all trials and blocking trials by recall direction. Participants exhibited superior initiation and transition accuracy in forward as compared to backward recall, with a higher probability of fill-in transitions for backward recall. Given these findings, one may question the equivalence in overall recall performance as a function of recall direction. Figure 9 revisits analyses of recall performance on nine-item lists across five previously-defined scoring methods, panel b illustrating data from Experiment 2. Only the conditional order scoring method showed a significant difference in performance as a function of recall direction (however, this difference did not survive FDR correction for multiple comparisons). This method is the only amongst the five that conditions on recall of the previously presented item in forward recall or the subsequently presented item in backward recall ($lag = +1$). While the Solway et al. (2012) method used throughout this paper considers $lag = +1$ transitions as correct recalls, it does not consider whether a $lag = +1$ transition was possible from the just-recalled item. For example, a participant could skip over the third item in a nine-item list during recall, in which case, it would be impossible to correctly recall the fourth item in forward recall or the second item in backward recall. By conditioning on the recall event, conditional order scoring reflects superior forward recall initiation and transition accuracy in Experiment 2, consistent with our observations of a forward recall advantage for +1 transitions.

## List Length Effects

To prevent participants from adopting list-length-specific strategies, we varied list length between six and 12 words in both Experiment 1 and Experiment 2. We also varied the range of list lengths across three participant groups who studied and recalled lists of six, nine and 12 items, seven, nine and 11 items, or eight, nine and 10 items. In Experiment 1, the list lengths within participants varied such that participants never studied two consecutive lists of the same length. In Experiment 2, two trials of each list length condition were randomly ordered within each block. 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 variance in the key features of the data across all tested list lengths.

As shown in Figure 10a and d, the near-equivalence of forward and backward recall performance appears consistently across all seven list lengths across both Experiments. To test for any effect of list length, as well as the list length condition that each participant was grouped into, we ran a 3 × 3 × 2 mixed-design ANOVA for each Experiment, with recall direction as a within-participant variable, relative list length as a within-participant 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-participant variable. In Experiment 1, we found an expected main effect of relative list length (recall probability declined with increasing list length ($F(2, 1044) = 407.60$, $MSE = 2.73$, $p < 0.001$)), but we did not find a main effect of recall direction ($F(1, 522) = 1.10$, $MSE = 0.03$, *n.s.*) or a main effect of list length grouping ($F(2, 522) = 1.37$, $MSE = 0.24$, *n.s.*). In Experiment 2, we also found the expected main effect of relative list length ($F(2, 1098) = 186.35$, $MSE = 0.79$, $p < 0.001$)), and no main effect of recall direction ($F(1, 549) = 1.06$, $MSE = 0.04$, *n.s.*) or list length grouping ($F(2, 549) = 0.60$, $MSE = 0.18$, *n.s.*). We did, however, find a significant interaction between relative list length and recall direction in experiment 1, as evident in the slightly higher performance in forward serial recall at the shortest list lengths ($F(2, 1044) = 22.41$, $MSE = 0.08$, $p < 0.001$). While the data trended towards this same result in Experiment 2, the effect was not significant ($F(2, 1098) = 2.79$, $MSE = 0.01$, $n.s.$).

Participants initiated recall more accurately in backward than in forward recall for nearly all list lengths in Experiment 1 (Figure 10b). Conducting the same ANOVA model as described above on recall initiation revealed a main effect of relative list length ($F(2, 1042) = 42.82$, $MSE = 0.83$, $p < 0.001$), a main effect of recall direction ($F(1, 521) = 23.32$, $MSE = 4.23$, $p < 0.001$) and a significant interaction between these variables ($F(2, 1042) = 33.88$, $MSE = 0.62$, $p < 0.001$). In addition, we found a significant three-way interaction between recall direction, relative list length, and list length grouping ($F(2, 4, 1042) = 3.36$, $MSE = 0.06$, $p < 0.01$). This interaction suggests that the effects depend on quantitative differences in list lengths, e.g., differences are pronouncedly more different between six, nine, and 12-item lists than between eight, nine, and 10-item lists (visual inspection of the plot suggests that the effect is likely to be approximately linear).

In contrast, Figure 10e shows that participants initiated recall more accurately in forward than in backward recall ($F(1, 545) = 15.02$, $MSE = 2.26$, $p < 0.001$). Experiment 2 also saw a main effect of relative list length on recall initiation, with higher initiation accuracy for the shortest lists in each condition ($F(2, 1090) = 14.08$, $MSE = 0.23$, $p < 0.001$). However, the interaction between recall direction and relative list length was not significant ($F(2, 1090) = 0.12$, $MSE = 0.00$, $n.s.$), indicating that the forward recall advantage did not differ as a function of list length. Experiment 2 also saw no reliable interaction between recall direction, relative list length, and list length grouping ($F(2, 4, 1090) = 0.46$. $MSE = 0.01$, $n.s.$).

Finally, we examined participants 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 (Farrell, Hurlstone, & Lewandowsky, 2013). Once again, we conducted an ANOVA to investigate the role of our three variables in predicting errors.

In Experiment 1, we found a main effect of relative list length ($F(2, 954) = 7.99$, $MSE = 0.11$, $p < 0.001$), as well as a main effect of recall direction ($F(1, 477) = 88.75$, $MSE = 4.02$, $p < 0.001$). In addition, we found an expected significant interaction between relative list length and list length grouping ($F(4, 954) = 8.42$, $MSE = 0.12$, $p < 0.001$), demonstrating an effect of the quantitative list length. Surprisingly, the degree of difference between forward and backward fill-in probability did not vary with list length ($F(2, 954) = 0.48$, $MSE = 0.01$, $n.s.$). Similarly, in Experiment 2 we found a significant main effect of relative list length ($F(2, 872) = 3.24$, $MSE = 0.04$, $p < 0.05$), as well as a main effect of recall direction ($F(1, 436) = 12.29$, $MSE = 0.84$, $p < 0.001$), and no interaction between the two ($F(2, 872) = 0.08$, $MSE = 0.00$, $n.s.$). We also did not observe an interaction between relative list length and list length grouping ($F(4, 872) = 0.24$, $MSE = 0.00$, $n.s.$), suggesting predictable recall conditions remove the effect of quantitative list length on fill-in transitions.

For each list length condition, we also evaluated serial position effects, recall initiation functions and temporal clustering effects (following the methods in Figures 1B, 2 and 3A, respectively). We report these functions, separately for Experiments 1 and 2, in *Supplementary Materials.*

## Initiation times and inter-response times

In addition to analyses of recall probability, response times can also shed light on latent retrieval dynamics, and aid in determining conditions under which participants have an "easier" time recalling, particularly when recall performance is similar between conditions. For example, when lists are very short, it is likely that a large subset of participants will perform perfectly in both forward and backward recall. However, the speed at which a participant recalls items could indicate whether they retrieved items from memory more quickly, and therefore more easily, in the forward or backward condition. Response times can also provide crucial evidence regarding the degree to which the usage of certain strategies, such as peeling off (Norris, Hall, & Gathercole, 2019), differs between conditions.

To effectively compare inter-response times for forward and backward recall, we elected to analyze lists of any length where participants initiated recall with the first four items in perfect serial order. This method removed instances of incorrect recall and maximized data inclusion compared to other common methods that limit analyses to perfectly recalled lists. Figure 11 shows initiation and inter-response times for Experiments 1 and 2. The data demonstrated a main effect of expected output position, with participants taking more time to initiate recall than they take to transition from word to word (Experiment 1: $F(3, 750) = 59.35$, $MSE = 59.23$, $p < 0.001$; Experiment 2: $F(3, 537) = 98.15$, $MSE = 108.53$, $p < 0.001$). The data from Experiment 1 also demonstrated a significant interaction between expected output position and recall direction ($F(3, 750) = 22.31$, $MSE = 13.38$, $p < 0.001$), participants tending to initiate recall more quickly in backward recall, then transition more quickly in forward recall. However, this interaction was not significant in Experiment 2 ($F(3, 537) = 2.49$, $MSE = 2.27$, n.s.) Neither experiment demonstrated a significant main effect of recall direction (Experiment 1: $F(1, 250) = 0.18$, $MSE = 0.19$, n.s.; Experiment 2: $F(1, 179) = 0.01$, $MSE = 0.01$, n.s.).

While we elected to conduct the above analyses across all list lengths combined, list length is a potential source of variability in these findings. Figure 12 provides visualizations of the difference between forward and backward initiation and inter-response times in Experiments 1 and 2 for each list length, again only including lists with perfect recall of the first four items. Linear mixed effects models of these data demonstrate a main effect of recall direction in both experiments, with forward recall requiring more time on average to initiate than backward recall (Experiment 1: $F(1, 620.36) = 20.36$, $MSE = 27.87$, $p < 0.001$; Experiment 2: $F(1, 550.34) = 12.13$, $MSE = 23.52$, $p < 0.001$). We did not observe a significant main effect of list length (Experiment 1:

$F(6, 750.95) = 1.25$, $MSE = 1.71$, *n.s.*; Experiment 2: $F(6, 643.95) = 0.90$, $MSE = 1.74$, *n.s.*) or an interaction between list length and recall direction (Experiment 1: $F(6, 620.36) = 1.31$, $MSE = 1.79$, *n.s.*; Experiment 2: $F(6, 550.34) = 0.86$, $MSE = 1.66$, *n.s.*) on initiation times. Mirroring the prior analyses, Experiment 1 saw slower average inter-response times for backward recall transitions compared to forward, whereas this did not appear in Experiment 2 (Experiment 1: $F(1, 620.85) = 26.08$, $MSE = 7.26$, $p < 0.001$; Experiment 2: $F(1, 551.51) = 3.75$, $MSE = 1.22$, *n.s.*). List length did not affect the average speed of transitions in either experiment (Experiment 1: $F(6, 727.03) = 0.97$, $MSE = 0.27$, *n.s.*; Experiment 2: $F(6, 628.21) = 0.53$, $MSE = 0.17$, *n.s.*), nor did list length interact with recall direction in Experiment 1 ($F(6, 620.85) = 0.93$, $MSE = 0.26$, *n.s.*). However, a significant interaction between list length and recall direction in Experiment 2 provides evidence that participants transitioned between recalls faster in forward recall at the shortest list lengths ($F(6, 551.51) = 2.64$, $MSE = 0.86$, $p < 0.05$).

## Discussion

The present study examined the dynamics of forward and backward serial recall in two online experiments that included within-participant manipulations of recall direction and list length. Experiment 1 also included a cueing manipulation: In the pre-cued condition, participants 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[3] 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 for most measures of aggregate performance (see Figure 9). The equivalence of overall performance appeared consistently across all list lengths tested, ranging from six to 12 items (see Figure 10a & d).

In studies of free recall, a decomposition of the recall process into initiation and transitions 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. In forward recall, successful initiation requires participants to somehow retrieve the first list item despite the intervening occurrence of numerous other items. In contrast, successful initiation of backward recall only requires participants to repeat the item they last studied. This difference is exemplified in the results of Experiment 1, where participants exhibited more

---

[3] Closed set refers to experimental designs where participants know, prior to the study period, the entire set of stimuli that will be presented on any list.

accurate initiation in backward recall (this effect was particularly evident in the post-cued condition, where participants could not anticipate the direction of recall during list encoding). The post-cueing of recall direction (on half of the trials) and the randomization of recall direction for both pre-cued and post-cued lists made it very difficult for participants in Experiment 1 to optimize their strategy for a given recall direction. Experiment 2 removed the post-cued trials and blocked forward and backward trials with the goal of allowing participants time to adjust their strategy for forward vs. backward recall from block to block. This change in method reversed the strong initiation advantage for backward recall. In Experiment 2, participants exhibited more accurate initiation in forward recall, despite the interference caused by the intervening list items.

Associative chaining, positional coding, and retrieved-context theories of recall can readily account for findings of superior initiation in backward recall. The reversal of the backward initiation advantage in Experiment 2 invokes the use of control processes to enhance the encoding of early list items when participants expect forward recall. Allowing for such control over encoding would enable these models to account for the recall initiation results reported across these experiments.

Analyses of recall transitions also provided valuable insights into the differences between forward and backward recall. When participants could make a transition in either the forward or the backward direction, they tended to make forward transitions in forward recall and backward transitions in backward recall; transitions opposite to the direction of recall (fill-in responses) occurred far less often than transitions in the direction of recall (in-fill responses). The reversal of the directional asymmetry between forward and backward recall stands in stark contrast to findings from diverse free recall experiments, which indicate a near universality of the forward asymmetry in recall transitions. Indeed, the high degree of similarity in the dynamics of forward and backward recall, independent of participants' foreknowledge of recall direction, suggests that encoding processes place modest directional constraints on serial recall and, therefore, retrieval processes largely govern the nature of transitions.

A closer look at recall transitions, however, revealed distinctions between forward and backward recall. In both experiments, participants committed significantly more fill-in responses in backward than in forward recall, with this effect evident across list lengths and independent of participants' foreknowledge of recall direction (see Figures 3, 7, and 10c & f). Under conditions that deprive participants of reliable foreknowledge of recall direction, theories of memory make distinctive predictions regarding the relative prevalence of fill-in vs. in-fill responses across forward and backward recall. The context-evolution process assumed by RCT creates a forward asymmetry in the learned associations among items. It thereby predicts a greater proportion of fill-in responses in backward compared to forward recall. Chaining theories that assume stronger forward associations than backward associations would make a similar prediction about the relative balance of fill-in to in-fill responses (Caplan, 2005; Solway et al., 2012). Positional coding theory, however, allows for bidirectional control of the positional markers used to cue items (Caplan, 2005). As such, this theory can produce forward and backward transitions with equal facility. However, in standard implementations of positional coding theory, positional cues increasingly lag behind prior-item cues with increasing output position (Solway et al., 2012). This suggests that participants' propensity to make fill-in errors should increase across successive recalls. Examination of Figures 4 and 8, however, fails to provide evidence for this prediction. Modifying positional coding theory to allow items to retrieve their positions would overcome this problem. Such a retrieved-positional-context model would closely resemble retrieved-temporal context-models.

Many memory models possess mechanisms that account for the primacy and recency effects seen robustly in a wide range of recall tasks. Context-based models produce recency effects due to the end-of-list context cue being more similar to the contexts associated with recently experienced items (Howard & Kahana, 1999). As mentioned in the introduction, many models posit that early list items enjoy greater attention and/or more frequent rehearsals, leading to a primacy effect (Atkinson & Shiffrin, 1968; Laming, 2006; Burgess & Hitch, 1999). Brown et al.'s SIMPLE model assumes logarithmic compression of temporal distances, with more recent items being temporally closer (and thus more similar) to a cue item than older items of the same lag. Recency- or primacy-generating mechanisms will impact participants' propensity to make fill-in vs. in-fill transitions in forward and backward recall. In backward recall, enhanced availability of recent items will tend to produce fill-in responses, as these items benefit from a recency advantage. In forward recall, enhanced availability of early list items will also produce fill-in responses. Unlike the recency effect, which fades across successive retrievals (Kahana, 2012), the primacy effect endures, leading to in-fill errors in later output positions in backward recall. Our finding of increased fill-in errors in backward relative to forward recall appeared most prominent during early output positions (see Figures 4 and 8). This suggests that perhaps recency-sensitive retrieval processes, rather than forward-biased retrieval, may offer the most parsimonious account for these effects. Accordingly, recency would have to exert a greater influence on backward recall than primacy exerts on forward recall.

It would appear that recency-sensitive retrieval processes could account for the differential fill-in and in-fill responses in forward and backward recall observed across list lengths and foreknowledge manipulations. Under this interpretation, the main difference between dynamics in forward and backward serial recall is the increased interference caused by recent items, potentially impeding correct transitions in backward recall. Accepting this conclusion leaves us in the difficult position of explaining why transitions in backward recall reverse the asymmetry effect that has been so thoroughly documented in free recall across widely varying experimental manipulations (Kahana, 1996; Healey et al., 2019) and in nearly every individual participant (Healey & Kahana, 2014). RCT, and many chaining models, assume that the learning process enforces a forward bias in associations. Yet, lag-CRP analysis of backward recall demonstrates a nearly complete reversal of this asymmetry. Such a finding appears highly problematic for both RCT and asymmetric chaining theories. Arguing that participants process context and associations differently when expecting backward recall does not save these models when the results appear robustly under pre-cuing conditions when participants have no foreknowledge of recall direction.

One may further ask whether a theory that predicts stronger forward than backward associations can even produce the backward-asymmetry lag-CRP curves seen in all of our experimental conditions. We can explain this negative asymmetry (i.e., the preponderance of in-fill errors in backward recall) by recognizing that variability in the strength of memory encoding and/or the quality of cues available at retrieval will lead to omissions. Because participants failed to recall these items, they will tend to in-fill in subsequent transitions in both forward and backward recall, as seen in the data. Knowing whether this explanation can rescue asymmetric retrieval models depends on carrying out quantitative model fits that extend beyond the scope of the present paper.

Our comparison of recall dynamics across output positions implicates recency in differentiating forward and backward recall. After making an early omission error in backward recall, recency serves as a source of negative interference, leading participants to reverse direction and thereby making it harder to reach the earlier, and less recent, list items (Figures 4 and 8). When participants cannot easily predict the direction of recall, the advantage of backward initiation over forward becomes more pronounced as list length increases (Figure 10b), however, the difference in making reversals (fill-in transitions) remains constant across list lengths (Figure 10c) and appears both under conditions of predictable (Experiment 2, Figure 10f) and unpredictable recall direction (Experiment 1, post-cued condition). 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 10a, d).

The convenience of recruiting participants through an online platform allowed us to efficiently gather data from $> 50,000$ trials of serial recall at a time when

COVID-restrictions greatly limited our ability to conduct in-person experiments. Although these large datasets afforded the power for our conditional output-order analyses, we nonetheless recognize an important limitation that should be addressed in future work. 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 participants 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. Another key difference in methodology is that we instructed participants to attempt to visualize each item and only to focus only on the just-presented items during encoding. These instructions likely discouraged sub-vocal rehearsal, and consequently minimized any differences between forward and backward recall that may have emerged from differential rehearsal strategies. Future work should compare the present results with those obtained using recall procedures that allow for the marking of omissions and also to those obtained under conditions that do not discourage rehearsal during encoding.

The present findings buttress recent calls for a unified theoretical analysis of serial and free recall (Grenfell-Essam, Ward, & Tan, 2017; Farrell, 2012; Logan & Cox, 2023). Whereas models of free recall have emphasized contiguity- and recency-mediated retrieval processes (Lohnas et al., 2015; Sederberg et al., 2008; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Farrell, 2012), several models of serial recall emphasize primacy mechanisms and lack a specific mechanism for supporting temporally-mediated associations (e.g., Burgess & Hitch, 1999, 2006; Page & Norris, 1998; Farrell & Lewandowsky, 2004). The present comparison of forward and backward recall implicates both contiguity and recency as potential explanatory factors in the analysis of serial order memory. Our data also indicate that under less predictable foreknowledge of recall direction, a recency advantage in backward recall initiation obscured a small but reliable forward advantage in making directionally-correct transitions. Under more predictable conditions of recall, participants developed direction-specific strategies that resulted in a more pronounced primacy effect in forward recall initiation. Our finding of directionally-accurate backward recall following omissions, and similar overall performance levels in forward and backward recall, even under pre-cueing conditions, challenges models that impose a forward asymmetry in associative memory. Rather, our findings suggest that participants can direct recall transitions in a desired direction, a feature omitted by many memory models, including most versions of RCT.

# References

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

Atkinson, R. C., & Shiffrin, R. M. (1968). Human memory: A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 2, p. 89-105). New York: Academic Press.

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

Botvinick, M. M., & Plaut, D. C. (2006). Short-term memory for serial order: A recurrent neural network model. *Psychological Review*, *113*(2), 201-233.

Brown, G. D. A., Neath, I., & Chater, N. (2007). A temporal ratio model of memory. *Psychological Review*, *114*(3), 539–576. doi: doi: 10.1037/0033-295X.114.3.539

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.

Caplan, J. B. (2005). Associative isolation: Unifying associative and list memory. *Journal of Mathematical Psychology*, *49*, 383–402.

Chance, F. S., & Kahana, M. J. (1997). Testing the role of associative interference and compound cues in sequence memory. In J. Bower (Ed.), *Computational neuroscience: Trends in research* (pp. 599–603). New York: Plenum Press.

Conrad, R. (1965). Order error in immediate recall of sequences. *Journal of Verbal Learning and Verbal Behavior*, *4*(3), 161–169.

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.

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. (2006, Nov). Mixed-list phonological similarity effects in delayed serial recall. *Journal of Memory and Language*, *55*(4), 587-600.

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., Hurlstone, M. J., & Lewandowsky, S. (2013). Sequential dependencies in recall of sequences: Filling in the blanks. *Memory & Cognition*, *41*, 938-952.

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

Glenberg, A. M., & Swanson, N. G. (1986). A temporal distinctiveness theory of recency and modality effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *12*(1), 3-15. doi: 10.1037/0278-7393.12.1.3

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

Guitard, D., & Saint-Aubin, J. (2022). Backward recall and foreknowledge of recall direction: a test of the encoding-retrieval matching hypothesis. *Memory*, *30*(8), 1057-1072. doi: 10.1080/09658211.2022.2079675

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

Healey, M. K., & Kahana, M. J. (2014). Is memory search governed by universal principles or idiosyncratic strategies? *Journal of Experimental Psychology: General*, *143*(2), 575–596. doi: 10.1037/a0033715

Healey, M. K., & Kahana, M. J. (2016). A four-component model of age-related memory change. *Psychological Review*, *123*(1), 23–69. doi: https://doi.org/10.1037/rev0000015

Healey, M. K., Long, N. M., & Kahana, M. J. (2019). Contiguity in episodic memory. *Psychonomic Bulletin & Review*, *26*(3), 699–720. doi: 10.3758/s13423-018-1537-3

Henson, R., Norris, D. G., Page, M. P. A., & Baddeley, A. D. (1996). Unchained memory: Error patterns rule out chaining models of immediate serial recall. *The Quarterly Journal of Experimental Psychology*, *49A*, 80-115.

Hertzog, C., Dixon, R. A., Hultsch, D. F., & MacDonald, S. W. S. (2003). Latent change models of adult cognition: are changes in processing speed and working memory associated with changes in episodic memory? *Psychology and Aging*, *18*(4), 755–769.

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

Hurlstone, M. J. (in press). Oxford handbook of human memory. In M. J. Kahana & A. D. Wagner (Eds.), (2nd ed., chap. Serial Recall). Oxford, U. K.: Oxford University Press.

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

Kahana, M. J. (2012). *Foundations of human memory.* New York, NY: Oxford University Press.

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). Laws of human memory. In M. J. Kahana & A. D. Wagner (Eds.), *Oxford handbook of human memory* (Vol. 1). 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.

Ladd, G. T., & Woodworth, R. S. (1911). *Elements of physiological psychology: A treatise of the activities and nature of the mind from the physical and experimental point of view.* New York, NY: Charles Scribner's Sons.

Laming, D. L. (2006). Predicting free recalls. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(5), 1146-1163.

Lewandowsky, S., & Farrell, S. (2008). Short-term memory: New data and a model. *The Psychology of Learning and Motivation*, *49*, 1-48. doi:

10.1016/S0079-7421(08)00001-7

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review*, *96*, 25-57. doi: 10.1037/0033-295X.96.1.25

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

Logan, G. D. (2021). Serial order in perception, memory, and action. *Psychological Review*, *128*(1), 1-44.

Logan, G. D., & Cox, G. E. (2023). Serial order depends on item-dependent and item-independent contexts. *Psychological Review*. doi: 10.1037/rev0000422

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

Mensink, G.-J. M., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, *95*, 434-55. doi: 10.1037/0033-295X.95.4.434

Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arxiv preprint arXiv:1301.3781*. doi: 10.48550/arXiv.1301.3781

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

Murdock, B. B. (1956). Backward learning in paired associates. *Journal of Experimental Psychology*, *51*, 213-215.

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). Similarity in a distributed memory model. *Journal of Mathematical Psychology*, *39*, 251-264.

Norris, D., Hall, J., & Gathercole, S. E. (2019). How do we perform backward serial recall? *Memory & Cognition*, *47*, 519-543. doi: https://doi.org/10.3758/s13421-018-0889-2

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.

Osth, A. F., & Hurlstone, M. J. (in press). Do item-dependent context representations underlie serial order in cognition? commentary on logan (2021). *Psychological Review*. doi: https://doi.org/10.1037/rev0000352

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

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

Ribback, A., & Underwood, B. J. (1950). An empirical explanation of the skewness of the bowed serial position curve. *Journal of Experimental Psychology*, *40*(3), 329-335. doi: https://doi.org/10.1037/h0054262

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

Sederberg, P. B., Schulze-Bonhage, A., Madsen, J. R., Bromfield, E. B., McCarthy, D. C., Brandt, A., ... Kahana, M. J. (2007). Hippocampal and neocortical gamma oscillations predict memory formation in humans. *Cerebral Cortex*, *17*(5), 1190–1196. doi: 10.1093/cercor/bhl030

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

Spurgeon, J., Ward, G., Matthews, W. J., & Farrell, S. (2015). Can the effects of temporal grouping explain the similarities and differences between free recall and serial recall? *Memory & cognition*, *43*(3), 469–488.

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.

Ward, G., Tan, L., & Grenfell-Essam, R. (2010, September). Examining the relationship between free recall and immediate serial recall: the effects of list length and output order. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *36*(5), 1207–1241.

## Appendix A

We excluded participants who completed the experiment based on three factors: we excluded participants who self-reported being a non-native English speaker, accounting for 64/698 exclusions in Experiment 1 and no exclusions in Experiment 2. We excluded participants who failed to recall any words, accounting for 30/698 exclusions in Experiment 1 and no exclusions in Experiment 2. Upon completing the experiment, we asked participants whether they took notes to boost their performance. Participants who answered "yes" were also excluded from the analysis, accounting for 300/698 exclusions in Experiment 1 and all 435 exclusions in Experiment 2. Early launches of Experiment 1 did not require an answer to this question, resulting in 226 participants for whom no answer was collected. Analyses of both Experiments only included participants who explicitly answered "no" to the aforementioned question, thereby excluding these 226 Experiment 1 participants for whom no answer could be determined. As participants who responded "yes" or did not answer this question represent 39.22% of completed data sets in Experiment 1 & 41.15% in Experiment 2, we report further analyses of these data below.

Figure 13 compares recall performance among participants who did and did not write notes in Experiments 1 and 2 using the Solway et al. (2012) relative order scoring method. Although one may expect that writing notes would boost overall performance, the trends for both experiments suggest the opposite, with self-reported note-writers displaying worse overall recall performance. This performance deficit among note-writers appeared in both forward and backward recall. One could explain these trends in several ways. Firstly, it is difficult to transcribe words within the experiments' short inter-stimulus interval (ISI). If this explanation applied to the majority of note writers, however, one would also expect to see near-perfect initiation in forward recall, regardless of cue direction. Alternatively, it is possible this question indirectly screened participants based on poor attentiveness. For example, several participants in this group chose one word from the first presented list and recalled it during every subsequent list, with little to no additional recalls. Participants in this group averaged 9.9 and 21.6 lists out of 48 with zero recall attempts in Experiments 1 and 2, respectively.

Regardless of data quality, analyzing these participants' data for the same phenomena observed in those who did not write notes is not productive. An experiment with explicit instructions to write each presented word would yield vastly different predictions for forward and backward recall. On the other hand, a coding oversight did not require an answer to this question in early launches of Experiment 1. We cannot be sure whether these individuals, or *no-answer* participants, followed the instructions of the Experiment, and as such, we excluded this group in our primary analyses. However, if every single no-answer participant did not write notes, excluding such a large amount of data could potentially skew phenomena reported in the main text. To address this concern, we conducted analyses of recall performance, initiation, transitions, and the consistency of these results across list lengths for no-answer participants.

Figure 14 separately analyzes the *no-notes* Experiment 1 participants analyzed in the main text and the no-answer participants for the major phenomena reported in this paper: consistent performance in forward and backward recall, superior backward recall initiation, higher fill-in transition rates ind backward recall, and the relative consistency of these findings across all tested list lengths. Analyses of recall probability as a function of recall direction, cueing condition, and expected output position in no-answer participants replicated findings in no-notes participants, demonstrating a main effect of cueing condition ($F(1, 225) = 10.82$, $MSE = 0.75$, $p < 0.01$), a significant interaction between expected output position and cueing condition ($F(8, 1800) = 14.04$, $MSE = 0.45$, $p < 0.001$), a significant interaction of expected output position and recall direction ($F(8, 1800) = 10.26$, $MSE = 0.65$, $p < 0.001$), a significant three-way interaction of cueing condition, output position, and recall direction ($F(1, 8, 1800) = 3.48$, $MSE = 0.10$, $p < 0.001$), and no main effect of recall direction ($F(1, 225) = 10.82$, $MSE = 0.09$, $n.s.$). List length also affected the probability of recall in no-answer participants

($F(2, 446) = 193.88$, $MSE = 1.44$, $p < 0.001$), this effect again more pronounced in forward compared to backward recall (($F(2, 446) = 10.91$, $MSE = 0.04$, $p < 0.001$), given the slightly higher forward recall performance at the shortest list lengths. No-answer participants similarly demonstrated no main effect of recall direction on recall probability across list lengths ($F(1, 223) = 0.61$, $MSE = 0.04$, *n.s.*).

Analyses of recall initiation in no-answer participants also replicated the results reported in no-notes participants. Participants exhibited superior initiation performance in backward compared to forward recall, across and within cueing conditions (Across: $t(225) = 3.77$, $p < 0.001$, Pre-cued: $t(225) = 2.65$, $p < 0.01$, Post-cued: $t(225) = 4.00$, $p < 0.001$.) Post-cueing reduced participants' tendency to initiate accurately in forward recall ($t(225) = 2.05$, $p < 0.05$), but not backward recall ($t(225) = 0.28$, *n.s.*). Similar to no-notes participants, no-answer participants demonstrated superior backward recall initiation across list lengths ($F(1, 223) = 17.40$, $MSE = 3.36$, $p < 0.001$). Increasing list length again affected the overall accuracy of recall initiation $F(2, 446) = 18.90$, $MSE = 0.30$, $p < 0.001$), with this effect being more pronounced in forward recall ($F(2, 446) = 32.47$, $MSE = 0.52$, $p < 0.001$).

Finally, post-hoc tests of recall transitions demonstrated similar trends in no-answer and no-notes participants. No-answer participants filled in items more often in backward compared to forward recall ($t(215) = 4.42$, $p < 0.001$), and showed a higher probability of correct ($lag = +1$) transitions in forward compared to backward recall ($t(220) = 3.02$, $p < 0.01$). Fill-in transitions were consistently more common in backward compared to forward recall across list lengths ($F(1, 207) = 25.31$, $MSE = 2.02$, $p < 0.001$). A main effect of list length appeared in no-answer participants that did not appear in no-notes participants ($F(2, 414) = 8.52$, $MSE = 0.15$, $p < 0.001$), however, the consistency of the effect of recall direction was not impacted by list length ($F(2, 414) = 0.64$, $MSE = 0.01$, *n.s.*). Taken together, these analyses indicate that exclusion of no-answer participants did not bias the major results of Experiment 1.

## 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.

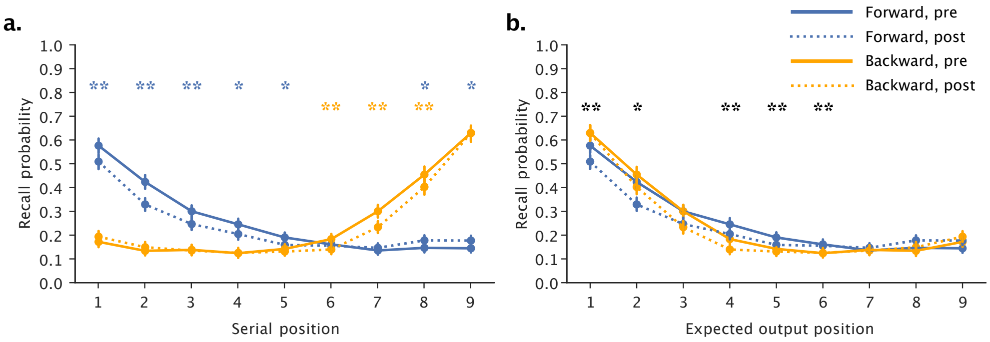

*Figure 1*. **Serial position effects on forward and backward recall (Exp. 1)**. Panel a illustrates the probability of correctly recalling an item as a function of its serial position, calculated using relative order scoring described in Solway et al. (2012). Blue lines indicate forward recall for pre- and post-cueing, orange indicates backward recall for pre- and post-cueing. Significance markers represent paired t-tests comparing pre- vs. post-cued conditions at each serial position separately for forward recall (blue) and backward recall (orange). $\star$ indicates FDR corrected $p < 0.05$, $\star\star$ indicates FDR corrected $p < 0.001$. Panel b illustrates the probability of correctly recalling an item as a function of its expected output position. Significance markers (black) denote comparisons between forward and backward recall collapsing across cueing condition. Error bars indicate 95% confidence intervals, calculated using 1000 iterations of bootstrapped resampling of the data.

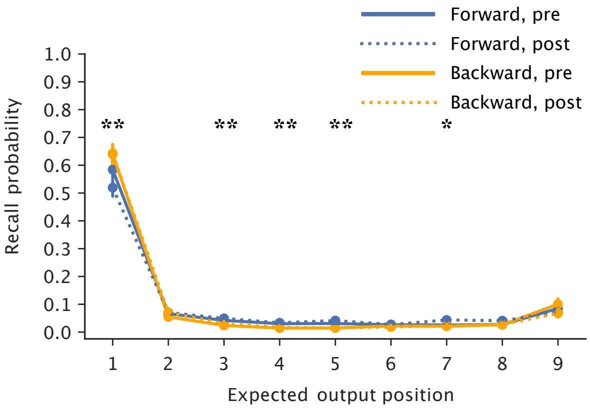

*Figure 2.* **Probability of first recall as a function of expected output position (Exp. 1).** The graph illustrates the probability that a participant initiated a recall sequence with an item at each expected output position. Expected output position one corresponds to correct initiation for both forward and backward recall. Significance markers represent results of paired t-tests on the probability of recall initiation in forward vs. backward recall at each expected output position. $\star$ 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 data resampling.

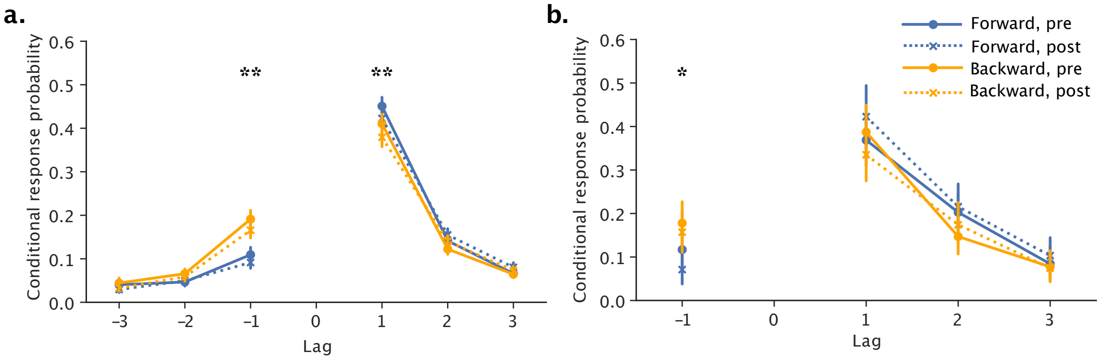

*Figure 3*. **Temporal clustering in serial recall (Exp. 1).** Probability of recalling an item as a function of its lag from the just-recalled item (the difference between its predecessor's expected output position and its expected output position). A lag of +1 marks a correct transition in both forward 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 represent results of paired t-tests on the probability of correct ($lag = +1$) and fill-in ($lag = -1$) transitions in forward vs. backward recall. $\star$ indicates $p < 0.05$, $\star\star$ indicates $p < 0.001$. Error bars indicate 95% confidence intervals, calculated using 1000 iterations of bootstrapped data resampling.

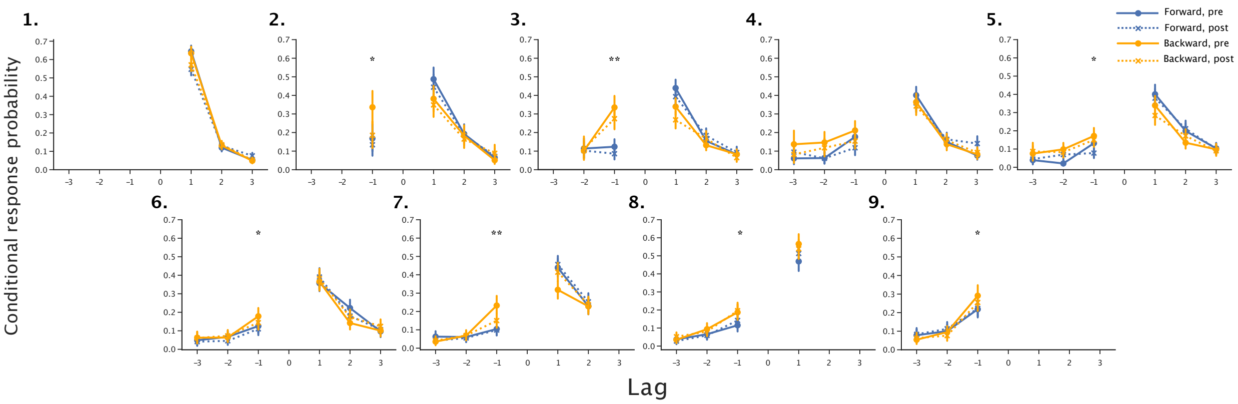

*Figure 4*. **Temporal clustering in serial recall for each expected output position (Exp. 1).** Probability of recalling an item as a function of its lag (distance, in items) from the just-recalled item at each expected output position. Each panel illustrates conditional response probabilities for each of the four experimental conditions: forward pre-cued, forward post-cued, backward pre-cued, and backward post-cued. Significance markers represent results of paired t-tests on the probability of fill-in ($lag = -1$) transitions in forward vs. backward recall at each expected output position. $\star$ 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 data resampling.

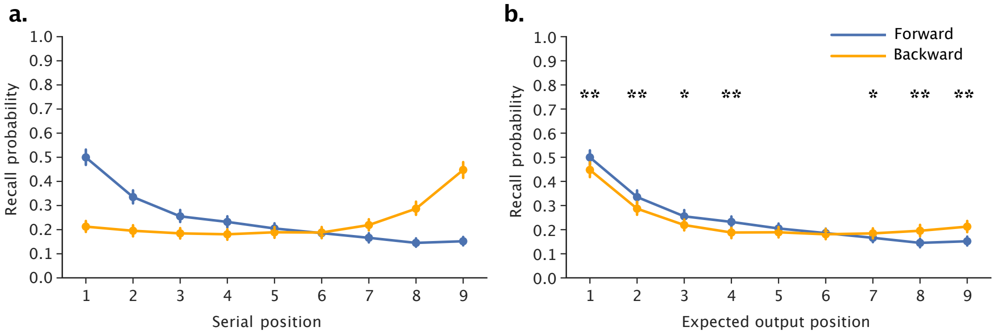

*Figure 5*. **Serial position effects on forward and backward recall (Exp. 2)**. Panel a illustrates the probability of correctly recalling an item as a function of its serial position, calculated using the relative order scoring method described in Solway et al. (2012). Blue indicates forward recall, orange indicates backward recall. Panel b 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 represent results of paired t-tests on recall probability in forward vs. backward recall at each expected output position. $\star$ 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 data resampling.

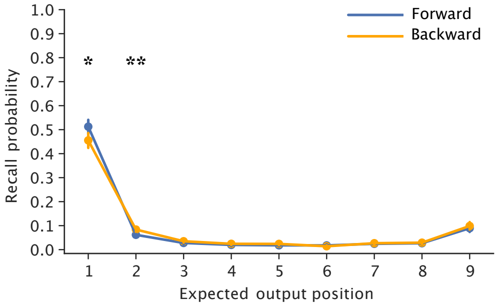

*Figure 6*. **Probability of first recall as a function of expected output position (Exp. 2).** The graph illustrates the probability that a participant initiated a recall sequence with an item at each expected output position. Expected output position one corresponds to correct initiation for both forward and backward recall. Significance markers represent results of paired t-tests on the probability of recall initiation in forward vs. backward recall at each expected output position. $\star$ 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 data resampling.

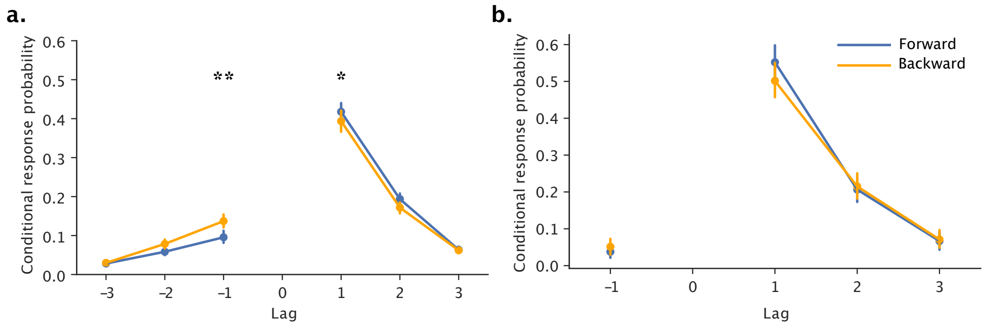

*Figure 7.* **Temporal clustering (Exp. 2).** Probability of recalling an item as a function of its lag from the just-recalled item (the difference between its predecessor's expected output position and its expected output position). Panel a compares conditional response probabilities for forward and backward recall. Panel b illustrates the same comparison as in a, but only for recalls following the first order error of $lag = +2$. Significance markers represent the results of paired $t$-tests on the probability of correct ($lag = +1$) and fill-in ($lag = -1$) transitions in forward vs. backward recall. $\star$ indicates $p < 0.05$, $\star\star$ indicates $p < 0.001$. Error bars indicate 95% confidence intervals, calculated using 1000 iterations of bootstrapped data resampling.

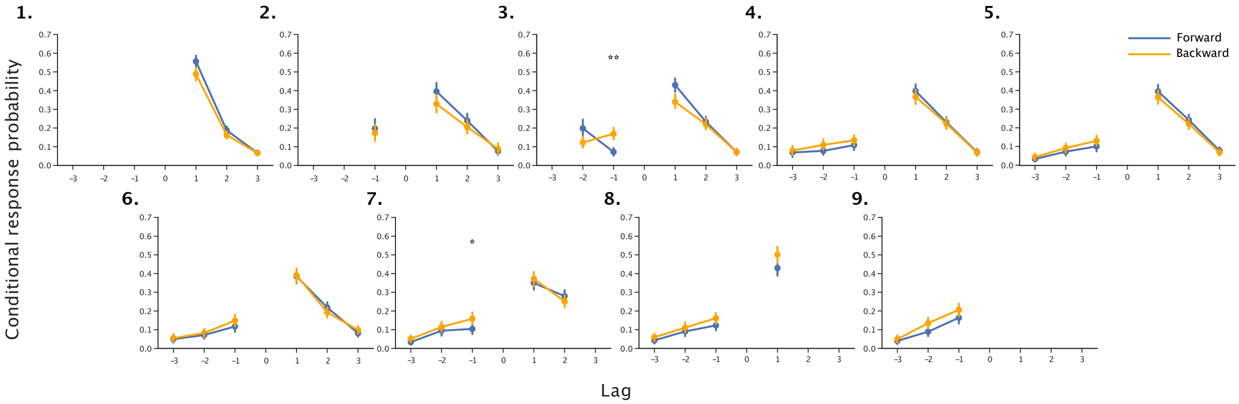

*Figure 8*. **Temporal clustering for each expected output position (Exp. 2).** Probability of recalling an item as a function of its lag (distance, in items) from the just-recalled item at each expected output position. Each panel illustrates conditional response probabilities for forward and backward recall. Significance markers represent results of paired t-tests on the probability of fill-in ($lag = -1$) transitions in forward vs. backward recall at each expected output position. $\star$ 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 data resampling.

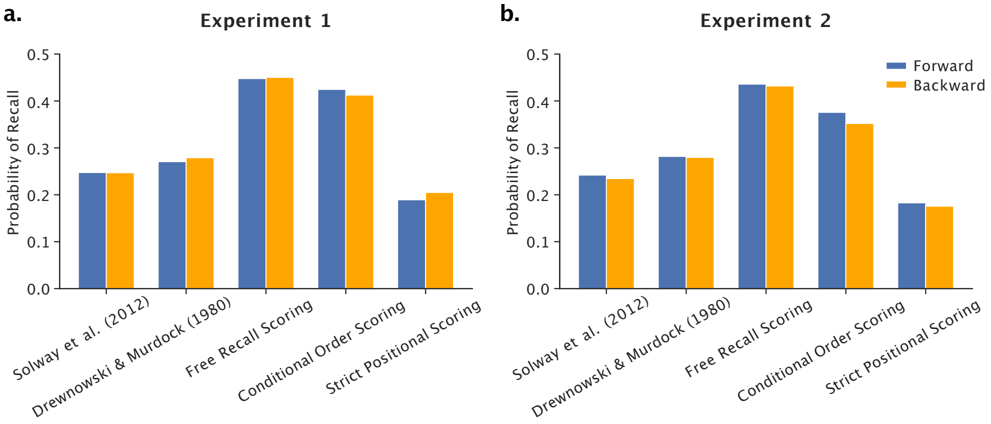

*Figure 9*. **Comparison of forward and backward recall performance under five scoring methods.** We measured recall performance using two relative order scoring methods (Solway et al., 2012; Drewnoswki & Murdock, 1980), free recall scoring, conditional order scoring, and strict serial position scoring, separately for Experiments 1 and 2 (panels a, b). Only conditional order scoring in Experiment 2 demonstrated a difference between overall forward and backward recall. $\star$ indicates $p < 0.05$, after conducting an FDR test for multiple comparisons.

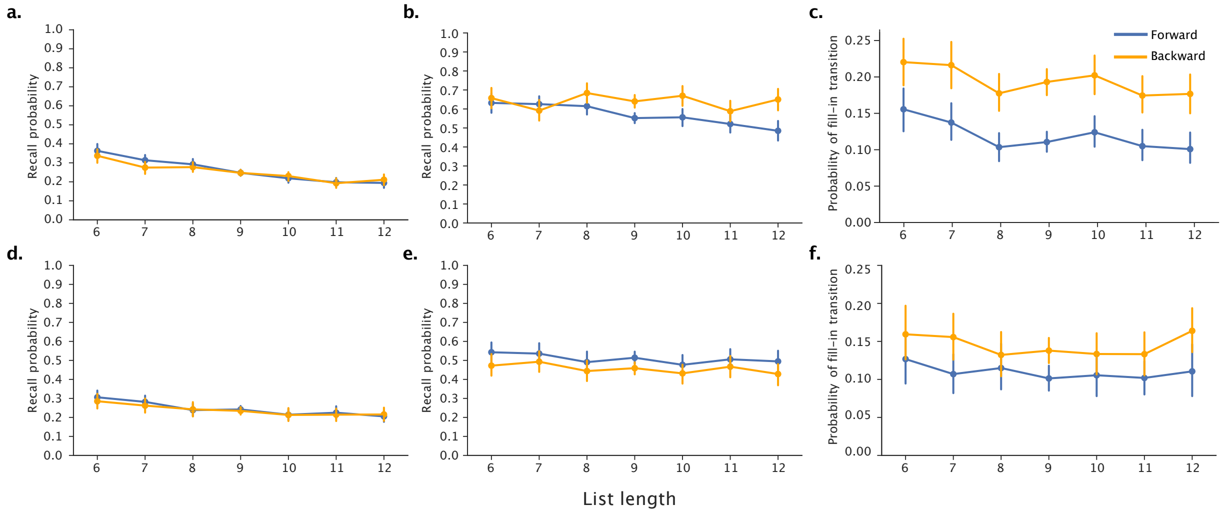

*Figure 10*. **List Length Effects.** Effect of list length condition on **a., d.** recall probability using Solway et al. (2012) scoring, **b., e.** correct recall initiation, and **c., f.** the conditional probability of committing a fill-in transition after the first order error in backward and forward recall. Error bars indicate 95% confidence intervals, calculated using 1000 iterations of bootstrapped data resampling.

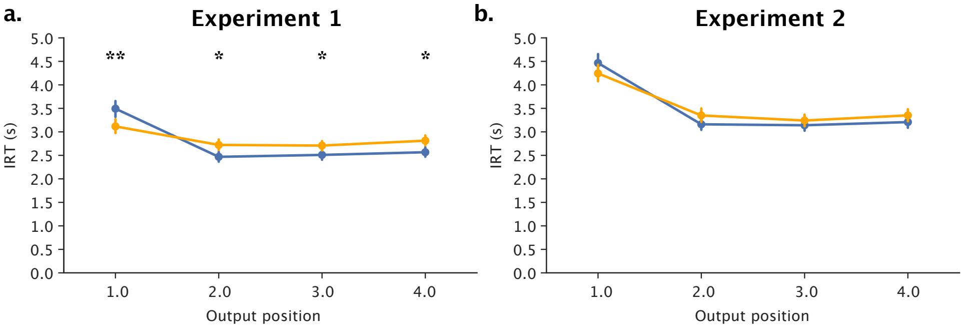

*Figure 11*. **Initiation and Inter-Response Times for Lists with Perfect Starts.** Panel a shows data from Experiment 1, panel b from Experiment 2. Initiation times were measured from the beginning of the recall period until submission of the first word; inter-response times were measured from the submission of the previous word to the submission of the current word. Significance markers indicate results of paired t-tests on inter-response times in forward vs. backward recall 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 95% confidence intervals, calculated using 1000 iterations of bootstrapped data resampling.

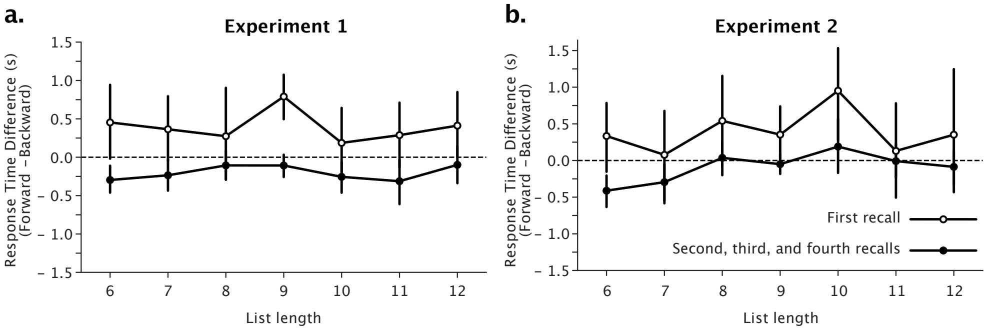

*Figure 12.* **Initiation and Inter-Response Times for Lists with Perfect Starts across All List Lengths.** Panel a shows data from Experiment 1, panel b from Experiment 2. Lines with open circles indicate the average difference in initiation times in forward minus backward recall for lists where participants perfectly recalled the first four items. Lines with filled circles indicate the average of the differences of inter-response times in forward minus backward recall across the second, third, and fourth recalls in lists where participants perfectly recalled the first four items. Error bars indicate 95% confidence intervals, calculated using 1000 iterations of bootstrapped data resampling.

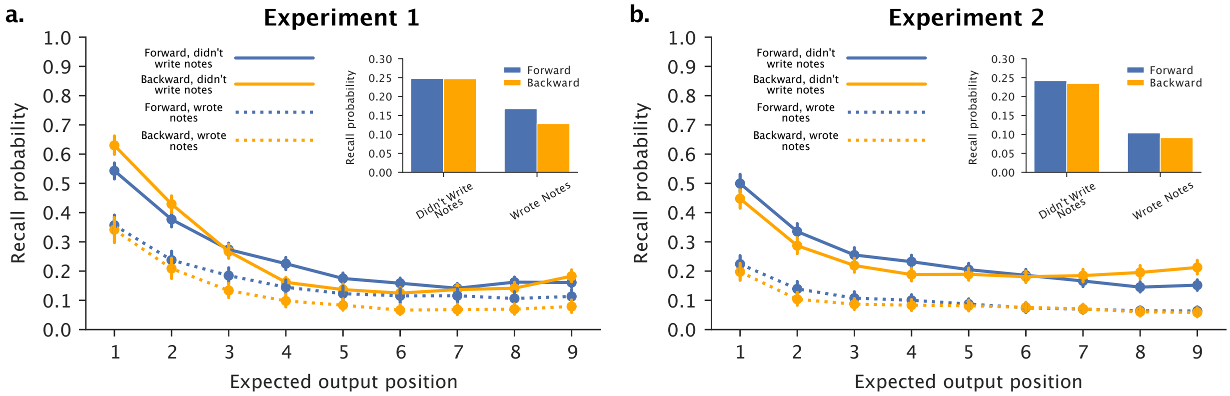

*Figure 13*. **Expected output position curves and recall rates of participants who did and did not write notes.** Panel a shows data from Experiment 1, panel b from Experiment 2. Each panel shows the probability of recall as a function of expected output position for forward and backward recall for participants who did not write notes (solid lines) and participants who did write notes (dotted lines), calculated using the Solway et al. (2012) relative order scoring method. Error bars indicate 95% confidence intervals, calculated using 1000 iterations of bootstrapped resampling of the data. Inlaid graphs demonstrate the overall probability of recall for each participant group according to this same relative order scoring method.

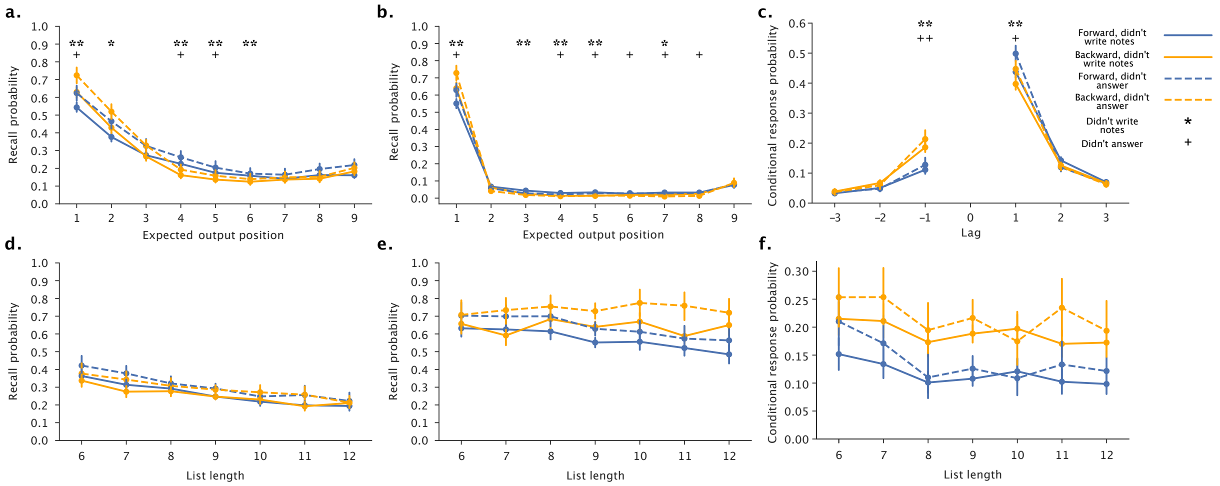

*Figure 14.* **Expected output position, probability of first recall, conditional response probability, and list length curves for participants who did not write notes in Experiment 1 and participants who did not answer the notes question.** Panel a replicates the analysis conducted in Figure 1, panel b replicates Figure 2, panel c replicates Figure 3, and panels d through f replicate Figure 10 panels a through c. ⋆ indicates $p < 0.05$, ⋆⋆ 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.