Journal of Experimental Psychology:
Learning, Memory, and Cognition
2015, Vol. 41, No. 5, 1447–1455

0278-7393/15/$12.00 http://dx.doi.org/10.1037/xlm0000113

The Fill-In Effect in Serial Recall Can Be Obscured by Omission Errors

Adam F. Osth and Simon Dennis
University of Newcastle, Australia

Henson (1996) provided a number of demonstrations of error patterns in serial recall that contradict chaining models. Chaining models predict that when participants erroneously recall an item too early, recall should proceed from the point of error. In contradiction to such a prediction, Henson found evidence for a fill-in effect: participants were much more likely to revisit an erroneously skipped item than to continue onward to later list items. However, recent reanalyses of serial recall data sets have found evidence for the opposite pattern in serial recall experiments that use open sets of items. We tested the hypothesis that open sets of items produce fill-in effects by comparing serial recall with an open set and a closed set, and when participants were allowed and prohibited from skipping over responses, and comparing serial recall with a reconstruction of order task. Fill-in effects were observed in all cases except when participants were not encouraged to skip over responses. Subsequent analyses indicated that when omission rates were equated, a fill-in effect was observed for all conditions when lists contained no omissions. These results suggest that high omission rates in open-set designs obscure a fill-in effect and further sound a cautionary note about interpreting cases in which recall continues in the forward direction after a skipped response.

Keywords: serial recall, fill-in effects, chaining models, conditional response probabilities

A central question in the study of memory for serial order concerns the qualitative nature of the representations that are used to encode and retrieve an ordered sequence of items. What is possibly the oldest class of theories that addresses this question is the class of *chaining theories*, which assume that only associations among the to-be-learned items are stored. Retrieval consists of retrieving the first item in the sequence and using that item as a cue for its successors; retrieval continues until the end of the sequence is reached.

Chaining models were no longer perceived as viable for serial order memory after the seminal dissertation of Henson (1996). Henson (1996) tested a critical prediction of chaining theories: If retrieved items are used as cues, then retrieval of an incorrect response should cause further errors. For instance, consider if the sequence *ABCDEF* was learned, and after retrieval of *A*, the participant erroneously recalls *D*; chaining models predict that *D* should be used as a cue and that the next response should be an item that is near *D*. However, this prediction has failed in a number of different experiments. In mixed lists of phonologically similar and dissimilar items, transpositions among the phonologically similar items cause no detriment on the dissimilar items (Baddeley, 1968; Farrell & Lewandowsky, 2003; Henson, Norris, Page, & Baddeley, 1996). In addition, when participants erroneously recall a late list item too early, recall is more likely to continue with its predecessor than its successor. That is, if a participant studies the list *ABCD* and erroneously skips from *A* to *C*, recall is more likely to continue with *B* than with *D*. Henson (1996) dubbed this the *fill-in effect*, because it is as if participants are filling in the missing response. The opposite error, in which participants continue forward after skipping an item, is dubbed an *in-fill error*. Henson (1996) found that fill-in errors dominated over in-fill errors by a 2:1 ratio, and this ratio was subsequently replicated by Surprenant, Kelley, Farley, and Neath (2005) using a *reconstruction of order* paradigm, in which participants are presented with the list items at retrieval and are only asked to indicate the order of the items.

Fill-in effects offer a considerable challenge to chaining models in which the associations among items are asymmetric in the forward direction (*A* gets associated to *B*, but *B* does not get associated back to *A*), because they predict a higher incidence of in-fill errors than fill-errors.[1] The fill-in effect has been argued to support the primacy model, in which order is stored as an exponentially decaying curve of activation strengths among list items (a primacy gradient; Page & Norris, 1998). At each stage of the retrieval process, noise is added to the strengths of the list items, and the item with the greatest strength is selected to be output. Referring to our earlier example, if *C* is erroneously output before *B* because of noise in the selection process, *B* is more likely to be selected than *D* because of its greater strength.[2]

[1] One should note that chaining models with symmetric associations are also possible (Lewandowsky & Murdock, 1989) and would predict equivalence between fill-in and in-fill errors.

[2] Positional models like the start-end model (SEM; Henson, 1998) are also able to predict fill-in effects using primacy-based mechanisms. In SEM, the absolute strength of the start marker exceeds the absolute strength of the end marker. Thus, when an item is recalled one position too early, such as the third item in Position 2, the cue for Position 3 matches the item in Position 2 more than the item in Position 4 because the higher absolute strength of the start marker biases retrieval toward the second item.

This article was published Online First April 20, 2015.

Adam F. Osth and Simon Dennis, School of Psychology, University of Newcastle.

Correspondence concerning this article should be addressed to Adam F. Osth, School of Psychology, Behavioral Sciences Building - W210, University of Newcastle, University Drive, Callaghan, NSW 2308 Australia. E-mail: adamosth@gmail.com

Despite the fill-in effects observed by Henson (1996) and Surprenant et al. (2005), recent evidence for an in-fill effect in immediate serial recall was observed by Solway, Murdock, and Kahana (2012), who conducted a reanalysis of several data sets and found that subsequent to a skip in the recall sequence, participants were much more likely to continue their recall in the forward direction than in the backward direction. Solway et al. argued that their results favored a compound chaining model with asymmetric associations and demonstrated a superior fit to the positional model of Burgess and Hitch (2006). However, subsequent to the publication of Solway et al.'s study, Farrell, Hurlstone, and Lewandowsky (2013) responded with a reanalysis of 21 data sets and found that the majority of these exhibited a fill-in effect.

An open question concerns why Solway et al. (2012) found evidence supporting chaining models whereas other experiments yielded evidence that was contradictory to chaining models (Farrell et al., 2013; Henson, 1996; Surprenant et al., 2005). Farrell et al. pointed to a number of differences between the experiments that they analyzed and the data sets analyzed by Solway et al. There were two major discussed discrepancies: (a) The experiments analyzed by Solway et al. used much longer list lengths (10–19 items in contrast to the lists of 5–7 items used by Farrell et al., 2013), and (b) in the experiments analyzed by Solway et al., participants were not given the option to mark their omissions if they were unable to recall an item. Omissions can be problematic for these analyses because they can essentially mask a fill-in effect. Consider if the sequence *ABCDEF* is stored in memory as *ACBDEF*, meaning that a fill-in effect will be produced if all items are successfully recalled. If a participant were to omit *B* (leading to recall of *AC_DEF*, where the underscore indicates a marked omission) or *C* (leading to recall of *ABD_EF*), neither case would contain a fill-in error. However, if the participant were to be unable to indicate the omission, recall would appear as either *ACDEF* or *ABDEF*, each of which contains an in-fill error. The data would thus appear to support chaining models despite the fact that the underlying process was not chaining.

To evaluate whether either of these possibilities was responsible for the discrepancy between the observed fill-in and in-fill effects, Farrell et al. (2013) conducted a reanalysis of the data of Grenfell-Essam and Ward (2012), who manipulated list length in immediate serial recall between four and 16 items and allowed for participants to indicate their omissions by leaving a blank on a lined response grid. Consistent with the list-length hypothesis, in-fill effects were observed for the longer lists of items: The effect was strongest for lists 15 items in length. However, an in-fill effect was also observed for lists that were five items in length, and there was neither a significant fill-in nor in-fill effect for lists four items in length. This was surprising given that the lengths of these lists were within the range that has demonstrated a fill-in effect and participants were also permitted to indicate their omissions. Such findings undermine the hypothesis that the differences between observed fill-in and in-fill effects can be solely attributed to differences in list length.

Open Versus Closed Sets of Stimuli

Another possibility that had not been discussed is that the data sets found to demonstrate a fill-in effect by Farrell et al. (2013) all used *closed sets* of stimuli, meaning that a small number of stimuli (typically digits or letters) were used and were sampled with replacement from trial to trial. The data sets of Solway et al. (2012), in contrast, used *open sets*, meaning that a large number of stimuli (typically words) were used and were sampled without replacement from trial to trial (i.e., each study list used novel stimuli). One of the motivations behind conducting studies that use closed sets is that memory for individual items quickly reaches ceiling, and only the order among the items has to be remembered, making it a relatively "pure" measure of order memory. A similar methodological approach concerns use of the reconstruction of order paradigm, in which participants are provided with items at retrieval and are only asked to indicate the order of the list items. A methodological concern with both closed-set serial recall and reconstruction of order paradigms is that participants are more easily able to guess the locations of the items when their order is not known. This can inflate the number of fill-in errors when participants guess on the final two positions when recalling a list. Consider the case in which the list *ABCDEF* is studied and a participant recalls *ABCD* but guesses on the remaining positions. The participant can either (a) recall *ABCDEF* or (b) recall *ABCDFE*, producing a fill-in effect. An in-fill effect cannot be produced in this circumstance.

Not only has a fill-in effect never been observed with an open-set serial recall procedure, other experiments using open-set serial recall have yielded results that are consistent with chaining models. Kahana, Mollison, and Addis (2010) found evidence for chaining with an open-set serial recall procedure using the *spin-list* paradigm, in which multitrial serial recall is compared between a baseline condition in which the same list is repeated from trial to trial and a spin-list condition in which the study list is rotated so that it begins at a different position on each trial. That is, a list such as *ABCDEFG* could be spun such that on Trial 2, participants would be presented with *DEFGABC*. A positional model predicts that performance should degrade from list to list, because each item is associated with a new position representation on each repetition. Chaining models, in contrast, predict improvement from list to list, because the linkages among the list items are mostly preserved. The data demonstrated that participants showed consistent improvement from trial to trial with the spin lists for all list-length conditions tested, supporting the chaining account.

In addition, McNicol (1978) conducted a set of experiments testing the positional and chaining accounts of serial recall by comparing different conditions of cross-list repetitions. In the positional repetition condition, items were repeated in the same position across lists. In the chaining condition, two items were repeated in adjacent positions on each list, but the position of the pair varied from list to list. Although there was considerable improvement in the positional repetition conditions over a baseline condition with no repetitions, there was either no evidence (Experiment 2) or limited evidence (Experiment 4) that repeating a pair of items in adjacent positions across lists improved performance. However, when an open set was used (Experiment 5), a large advantage for the chaining condition over the baseline condition emerged.

One possibility that the results of these studies highlight is that the use of open sets as stimuli may induce participants to rely on interitem associations and use chaining at retrieval instead of relying on positional representations. Consistent with this idea, the

Table 1
Differences Between the Four Conditions in the Experiment

Condition	*n*	Stimulus set	Response format	Blank responses encouraged
Open	100	Open	Serial recall	No
Blanks	101	Open	Serial recall	Yes
Closed	101	Closed	Serial recall	No
Reconstruction	95	Open	Reconstruction of order	No

experiments of Hitch, Fastame, and Flude (2005) used a condition very similar to the spin-list paradigm of Kahana et al. (2010) in which a list of items was rotated on every trial to preserve the pairwise associations among the items but used a closed set of stimuli instead of an open set. Contrary to the results of Kahana et al., no improvement was seen in this condition relative to a control condition in which novel lists were used on each trial. If open lists induce participants to rely on chaining, then a chaining model may provide a viable account of serial recall data using the open-set procedure.

We tested the hypothesis that open sets may be responsible for the fill-in effect by comparing serial recall for lists of six items under a number of different methodologies, each of which was manipulated between participants. Included were three conditions that used an open set: (a) one in which participants were not instructed or allowed to indicate omissions in the sequence (hereafter referred to as the *open* condition; this was similar to the experiments analyzed by Solway et al., 2012), (b) one in which participants were instructed to mark their omissions by typing the word "blank" (the *blanks* condition), and (c) one in which a reconstruction of order paradigm was used (the *reconstruction* condition). In addition, there was a condition that used a closed set, such that only six words were reused across all trials in the experiment (the *closed* condition). Because of the relative infrequency of fill-in and in-fill errors, we used a relatively large number of participants for each condition. Differences between all conditions are highlighted in Table 1. The open condition data set was previously reported and analyzed for intrusions in a separate article (Osth & Dennis, 2014). Although open-set and closed-set procedures have been directly compared in previous investigations, neither of these analyzed for the presence of a fill-in effect but, instead, investigated effects such as phonological similarity (Coltheart, 1993) and word frequency (Roodenrys & Quinlan, 2000).

Method

Participants

Totals of 100, 101, 101, and 95 participants were recruited for the open, blanks, closed, and reconstruction conditions, respectively. All participants were Ohio State University undergraduate students who participated in this experiment in exchange for credit in an introductory psychology course.

Materials

Participants studied words that were randomly selected from a word pool of 625 words from Google's word frequency counts, which is a collection of all Google search terms and how frequently they are searched.[3] Words ranged from four to seven letters in length and 30 to 250 counts per million in word frequency and comprised all parts of speech. In the open, blanks, and reconstruction conditions, six new words were sampled without replacement on every trial. In the closed condition, six words were sampled at the beginning of each experimental session to be used on all trials and were reshuffled before each trial began.

Procedure

Participants were instructed that their goal in the experiment was to recall all words on a list in the order in which they were presented. To familiarize them with the nature of the task, all participants began each session with four unscored practice lists of three, four, five, and six items to recall in order. Participants were given feedback on completion of each of the practice lists. If any errors were made, participants were reminded that they had to recall the items in the order in which they were presented. No feedback was given at any point in the experiment after the practice session was completed.

On completion of the practice session, participants were given 62 trials with lists of six items. During the study phase, participants were presented with each word for 1 s followed by a blank screen for 250 ms. Following completion of the study list, participants were presented with a recall prompt that was a series of three question marks (*???*) in the center of the screen. Participants were instructed to begin recalling the items on seeing the prompt by typing their responses on a keyboard and were given 20 s to recall the sequence. After the first key was pressed, the question marks disappeared and were replaced with the letters typed by the participant. Participants signaled completion of a word by hitting the ENTER key on the keyboard. On completion of a response, the response disappeared from view on the computer screen and was replaced by the question marks. Participants signaled completion of the recalled sequence by typing the word "done" and hitting the ENTER key. On completion of each trial, participants signaled readiness to begin the next trial by hitting the ENTER key. Participants were instructed to recall each list naturally and to not be too concerned with spelling. To ensure that participants did not get too fatigued with the task, halfway through the experiment, they were given a break in which they played a digital card game for 180 s.

[3] The Google word frequency counts used in this study can be found at http://mall.psy.ohio-state.edu/wiki/index.php/Google_Word_Frequency_Counts.

In the open, blanks, and reconstruction conditions, on each trial, a new six words were randomly sampled without replacement from the word pool. In the closed condition, a random selection of six words was sampled from the word pool and reused from trial to trial.

The recall phase across all conditions was identical with the exception of the reconstruction condition, in which participants were presented with all of the list words at the top of the screen. Words were presented in two rows of three, with all of the words randomly shuffled from their original presentation order. Participants were instructed that they would see the list words in a random order during the test phase and encouraged to use them as they recalled the list.

In the blanks condition, participants were additionally instructed that they should type the word "blank" when they could not recall a particular word. To motivate participants to do so, participants were told that the "blank" responses would act as placeholders to ensure that all words recalled after the "blank" response were scored correctly. Participants were additionally reminded to type "blank" in the practice phase during their feedback if they were not able to recall all of the list words.

Stimulus presentation and response collection was handled using the Python Experiment-Programming Library (Geller, Schleifer, Sederberg, Jacobs, & Kahana, 2007).

Results

All responses were spell-checked prior to analyses using the ratio function in the Levenshtein package in Python, which computes the similarity between two strings and returns a value between 0 and 1. If a participant entered a string that was not among the set of words used, it was substituted with a word among the word set if the similarity of the two strings exceeded .8 and the response was not semantically dissimilar from the set word. For instance, if the studied word was "house" but the participant typed "mouse," the similarity among the two strings is .8, but "mouse" would not be accepted because it is semantically dissimilar to the studied word.

Serial Position Effects

Serial position curves for each condition can be seen in Figure 1 under two different scoring procedures. Depicted on the left is the more traditional strict scoring procedure, in which items are scored as correct only if they are recalled in the same position they held at study. However, strict scoring can be quite punitive in cases in which omissions are high and participants are not permitted to mark their omissions. For instance, if a participant studies *ABCDE* and recalls *ABDE*, omitting *C* in the response, the outputs *D* and *E* are scored as incorrect even though they are in their correct relative positions. For this reason, we also include a relative scoring procedure, in which items are scored as correct if they occupy a position after the previously recalled item, a scoring procedure that has been used in several conditions using open sets (Drewnowski & Murdock, 1980; Golomb, Peelle, Addis, Kahana, & Wingfield, 2008; Kahana et al., 2010; Klein, Addis, & Kahana, 2005).

Significant effects of serial position were found in all conditions under both strict scoring—open, $F(5, 495) = 724.4$; blanks, $F(5, 500) = 415.6$; closed, $F(5, 500) = 241.2$; and reconstruction, $F(5, 470) = 225.6$ (all $ps < .001$)—and relative scoring—open, $F(5, 495) = 316.5$; blanks, $F(5, 500) = 284.6$; closed, $F(5, 500) = 34.09$; and reconstruction, $F(5, 470) = 55.8$ (all $ps < .001$). Even under relative scoring, a recency advantage for the last item over its preceding item emerged only in the closed, $t(100) = 6.22$, and reconstruction, $t(94) = 10.96$, conditions (both $ps < .001$). A negative recency effect emerged that manifested as worse performance for the last item relative to its preceding item in both the open, $t(99) = -6.43$, and blanks, $t(100) = -4.42$, conditions (both $ps < .001$). The negative recency effect for the open condition was previously reported in Osth and Dennis (2014). The discrepancy between the negative and positive recency effects is somewhat perplexing. Generally, recency effects are attenuated under visual presentation (Watkins, Watkins, & Crowder, 1974), but they are not usually eliminated. One possibility is that response suppression exerts a greater influence in closed sets and reconstruction of order tasks because of the smaller number of possible responses (e.g., Farrell & Lewandowsky, 2012). However, given that this investigation was not focused directly on recency effects, we hesitate to draw strong conclusions on the basis of the present data.

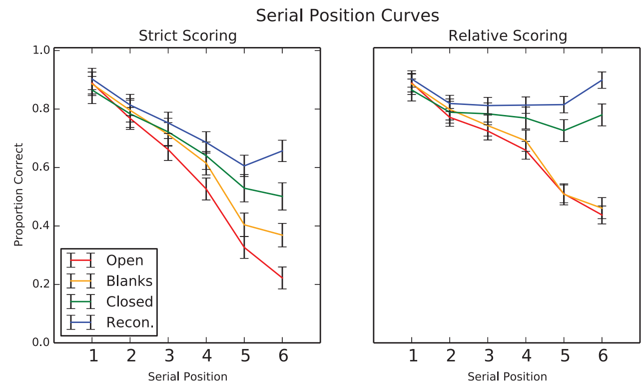

Figure 1. Serial position curves for all conditions, depicted under both strict and relative scoring. Error bars indicate 95% within-participant confidence intervals. See the online article for the color version of this figure.

Error Rates

Error types were analyzed as follows: An *omission* error involves an item not being recalled, a *transposition* error involves an item being recalled in the incorrect position, a *repetition* error involves an item being recalled more than once during the sequence, and a *prior-list intrusion* (PLI) involves an item from a prior list being recalled. Mean numbers of errors per list for omissions, transpositions, repetitions, and PLIs can be seen in Figure 2. Because the closed condition used the same stimuli on every trial, PLIs were undefined. One can see from the figure that the biggest differences among the conditions lay in the omission rates, which dominated over the other error types in the open and blanks conditions, both of which used open-set serial recall. Omission rates were not significantly different between the open and blanks conditions, $t(193.85) <$

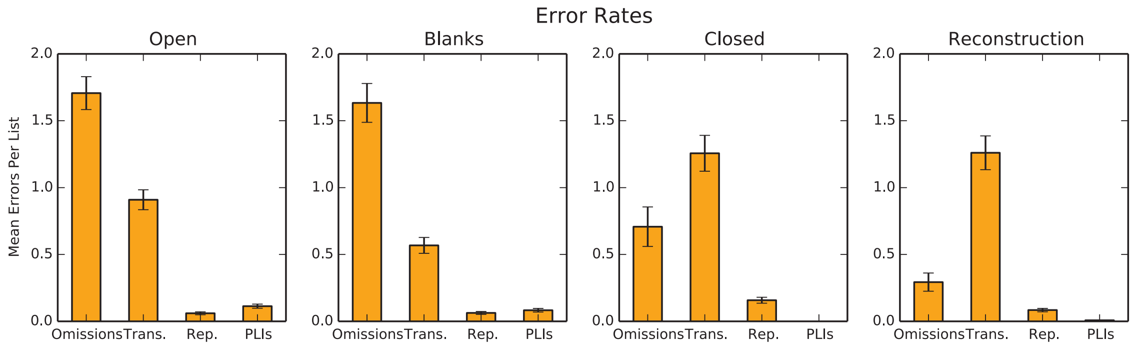

Figure 2. Mean errors per list in each condition, including omissions, transpositions (Trans.), repetitions (Rep.), and prior-list intrusions (PLIs). Error bars represent 95% confidence intervals. See the online article for the color version of this figure.

1.00,[4] although in the blanks condition, participants successfully gave a "blank" response to only 44% of their omissions. This estimate may be low in part because of a failure on the part of some participants to follow instructions, with some not indicating any of their omissions. Nonetheless, the mode of the distribution across participants was around 40%.

Omission rates were much lower in the closed condition relative to both the open, $t(192.85) = 10.26$, and blanks, $t(199.94) = 8.81$, conditions (both adjusted $ps < .001$). Unsurprisingly, omission rates were lower in the reconstruction condition than in the open, $t(153.70) = 19.89$; blanks, $t(141.21) = 16.48$; and closed, $t(140.03) = 5.03$, conditions (all adjusted $ps < .001$). (All p values were adjusted using the Bonferroni correction.) These analyses are quite similar to the results from the analysis of Coltheart (1993), who found that omissions in an open-set serial recall experiment were the dominant error, their rates greatly exceeding those found using a closed-set serial recall procedure.

Transposition rates also differed noticeably across conditions, with the lowest transposition rates observed in the open-set serial recall conditions. Transpositions in the closed condition exceeded those in both the open, $t(155.88) = 4.44$, and blanks, $t(137.56) = 9.29$, conditions (both adjusted $ps < .001$). Transposition rates in the reconstruction condition also exceeded those in both the open, $t(152.83) = 4.74$, and blanks, $t(133.78) = 9.81$, conditions (both adjusted $ps < .001$). Transposition rates did not significantly differ between the closed and reconstruction conditions, $t(193.83) < 1.00$.

Conditional Response Probabilities

Both Solway et al. (2012) and Farrell et al. (2013) presented the results of their fill-in analyses in the form of *lag conditional response probabilities* (lag CRPs), a technique that was pioneered for the analysis of free recall data by Kahana (1996). Lag CRP analysis focuses on analyzing recall transitions to get a measure of how participants are traversing a list being recalled. All recall transitions a participant makes are analyzed by measuring the relative distance between the serial positions of the list items, which is referred to as the *lag* between the items. For instance, if a participant were to study the list *ABCDE* and recall *ACBDE*, the recall transition *A* to *C* has a lag of $+2$, *C* to *B* has a lag of -1, *B* to *D* has a lag of $+2$, and *D* to *E* has a lag of $+1$. Lag CRPs are calculated by taking the frequency of transitions at each lag and dividing them by the number of opportunities to make the transitions at each lag. For instance, the Lag $+2$ CRP in the previous example is .667, because there were two $+2$ lag transitions but only three opportunities for it to occur (after *D* was recalled, a $+2$ transition was no longer possible). The Lag -1 CRP is 1.0, because although there was only one transition at Lag -1, there was also only one opportunity for it to occur.

Lag CRP calculations ignore all transitions to and from repetitions, intrusions, and "blank" responses. However, as stated in the introduction, it has been hypothesized that the allowance of "blank" as a response option should affect the relative proportions of fill-in and in-fill errors. For this reason, we performed an additional analysis on the blanks condition in which the "blank" options were removed. This meant that if a participant were, for example, to recall *ABCDE* as *A_CDE* (where the underscore indicates a "blank" response), the recall sequence would be scored as if it had been output as *ACDE*.

Lag CRPs in free recall show a strong forward asymmetry, in that participants tend to exhibit higher CRPs for positive than for negative lags (Bhatarah, Ward, & Tan, 2008; Howard & Kahana, 1999; Kahana, 1996). Lag CRPs in serial recall have generally revealed the same pattern (Bhatarah et al., 2008; Farrell et al., 2013; Golomb et al., 2008; Klein et al., 2005). However, lag CRPs by themselves do not give an appropriate measure of whether participants predominantly make fill-in or in-fill errors, because they consider all transitions among items in the list. Both Solway et al. (2012) and Farrell et al. (2013) quantified the relative predominance of fill-in errors by conditionalizing the lag CRPs on the first occurrence in which an item was recalled too early. We follow their example and report both the standard and conditionalized lag CRP curves.

Standard lag CRP curves. Lag CRP curves for all conditions can be seen in Figure 3. Because of the relatively short lists used, we restricted our analyses to lags between -3 and $+3$. Although recall transitions did occur at greater lags, they occurred with very low frequency, and several participants did not have any observations in these lags. From observation of the figure, one can see

[4] All t test degrees of freedom in the between-participants comparisons are corrected degrees of freedom from the Welch–Satterwaith equation.

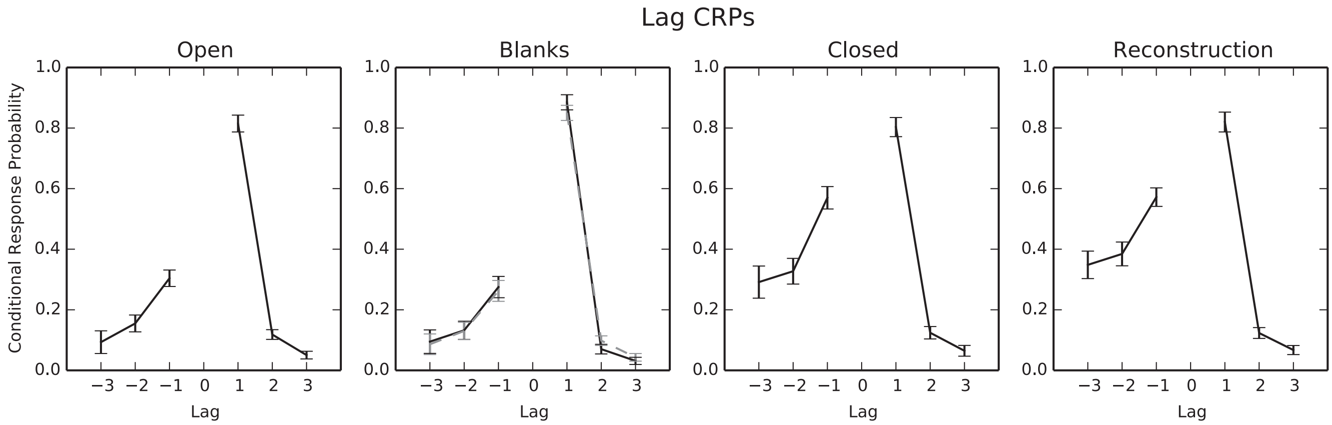

Figure 3. Lag conditional response probability (CRP) curves for all conditions, depicting the frequency of transitions at each transition lag divided by the number of opportunities for each transition lag to occur. For the blanks condition, the gray dashed lines indicate CRPs for when "blank" responses were removed. Error bars represent 95% within-participant confidence intervals.

strong forward asymmetry in all conditions. However, this forward asymmetry was only manifested in the comparison of the Lag 1 transitions. For lags of greater than 1, backward asymmetry was observed. This was the case in each condition: Lag +1 CRPs exceeded Lag −1 CRPs for the open, $t(99) = 32.11$; blanks, $t(100) = 31.85$; closed, $t(99) = 16.08$, and reconstruction, $t(94) = 18.01$, conditions (all adjusted $ps < .001$). Lag −2 CRPs exceeded Lag +2 CRPs, and Lag −3 CRPs exceeded Lag +3 CRPs. This effect appeared to be much stronger in the closed (Lag +2/−2, $t[99] = -11.45$; Lag +3/−3, $t[95] = -9.02$ [both adjusted $ps <$ .001]) and reconstruction conditions (Lag +2/−2, $t[94] = -13.64$; Lag+ 3/−3, $t[92] = 13.64$, [both adjusted $ps < .001$]) than in the open (Lag +2/−2, $t[99] = -2.82$; Lag +3/−3, $t[98] = -2.48$ [both adjusted $ps < .05$]) and blanks conditions (Lag +2/−2, $t[99] = -4.59$ [adjusted $p < .001$]; Lag +3/−3, $t[91] = 3.28$ [adjusted $p < .01$]). Although such a backward asymmetry was not observed in previous studies that analyzed lag CRPs in serial recall (Bhatarah et al., 2008; Golomb et al., 2008; Klein et al., 2005), we used considerably shorter lists and had a relatively low omission rate in comparison.

Conditionalized lag CRP curves: Fill-in and in-fill effects. To analyze for the presence of fill-in and in-fill effects, lag CRP curves were conditionalized to transitions that occurred after the first time an item was skipped during recall (i.e., a negative transposition). All responses prior to the skipped item were correct responses. For the blanks condition, trials on which the skipped item was preceded by a "blank" response (e.g., if a participant were to recall *A*, "blank," *C*) were not considered. Conditionalized lag CRPs depict how participants continue their recall transitions after they have recalled an item too early. For instance, if the list *ABCDE* is learned, and participants skip from *A* to *C* in their recall sequence, analysis focuses on where recall is most likely to continue after recalling *C*. Recall transitions with negative lags are fill-in effects, in that participants are filling in the missing responses (e.g., *B*). Recall transitions in the forward direction are in-fill errors, in that participants are continuing forward in their recall sequence from the point at which the error occurred.

An advantage of quantifying fill-in and in-fill errors in the form of lag CRPs is that the frequencies of fill-in and in-fill errors are conditioned on their availability. If a participant was recalling the *ABCDE* sequence and skipped from *A* to *D*, two fill-in errors are possible (*B* and *C*), whereas only one in-fill error is possible (*E*). Thus, conditionalized lag CRPs ensure that a higher incidence of fill-in errors is not merely a result of more opportunities for fill-in errors to be observed. To counteract the aforementioned confound that responses with guesses in the final output position will inflate the number of fill-in errors, analyses were restricted to output positions prior to the sixth output position.

Lag CRPs conditionalized on the first instance in which an item was skipped can be seen in Figure 4. Because several participants lacked errors for lags greater than 1 and less than −1, all analyses are restricted to the comparison between +1 and −1 lag CRPs. For the open condition, there was neither a fill-in nor an in-fill effect, $t(99) = .19$, $p = .84$. However, for the blanks condition, there was a fill-in effect, with the −1 lag CRPs significantly exceeding the +1 lag CRPs, $t(98) = 3.15$, $p < .01$, confirming the suggestion of Farrell et al. (2013) that allowing participants to indicate their omissions is critical to the observation of the fill-in effect. When the "blank" responses were removed, in contrast, a significant in-fill effect was present, $t(100) = 2.48$, $p < .05$. Both the closed, $t(97) = 9.49$, and reconstruction, $t(94) = 13.74$, conditions showed reliable fill-in effects (both $ps < .001$), exhibiting Lag −1 to Lag +1 CRP ratios of 2.26 and 2.71, respectively, which are similar to the 2:1 ratio observed previously (Henson, 1996; Surprenant et al., 2005).

Although the blanks condition exhibited a reliable fill-in effect when the "blank" responses were included in the analysis, the fill-in to in-fill ratio was lower (1.50) than was found in the closed condition and the reconstruction condition. One possible reason for this discrepancy is that the omission rates in the open and blanks conditions greatly exceeded those in the closed and reconstruction conditions. In addition, although participants were instructed to mark their omissions in the blanks condition, only 44% of their omissions were indicated. Thus, it remains possible that the unmarked omissions were artificially inflating the frequency of in-fill errors. To address this concern, we evaluated the fill-in (Lag −1) and in-fill (Lag +1) CRPs as a function of the number of omissions in each trial. Specifically, we separately evaluated the con-

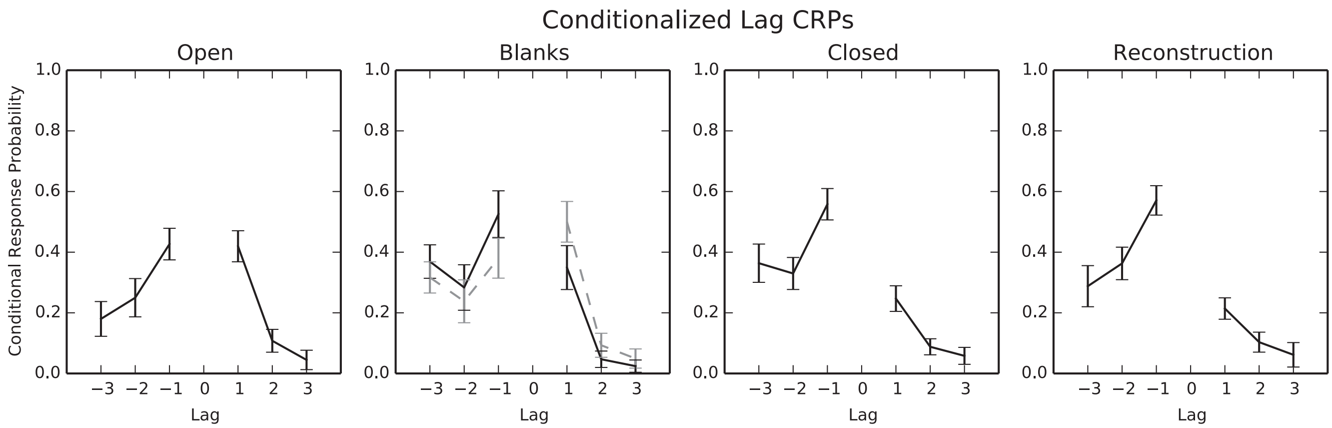

Figure 4. Conditionalized lag conditional response probability (CRP) curves, depicting all transitions after the first case in which an item was skipped. For the blanks condition, the gray dashed lines indicate CRPs for when "blank" responses were removed. Error bars represent 95% within-participant confidence intervals.

ditionalized CRPs for trials that contained zero omissions, one omission, two omissions, and three omissions. Also included was the percentage of trials at each omission frequency contributing to the conditionalized CRP analysis in Figure 4.

The results are depicted in Figure 5. There was a striking similarity between each condition: On trials with zero omissions, the fill-in CRPs greatly exceeded the in-fill CRPs. However, when trials with one or more omissions were considered, the fill-in and in-fill CRPs more closely resembled each other.[5] The critical difference between each condition lies in the different percentages of zero-omission trials contributing to the conditionalized CRP analysis. For the closed and reconstruction conditions, zero-omission trials were the most frequent, whereas for the open and blanks conditions, one- and two-omission trials were the most frequent, and zero-omission trials were relatively infrequent.

Some readers might contend that trials with zero omissions are inherently biased toward fill-in effects as the observation of zero omissions necessitates that participants follow the in-fill error with the skipped item somewhere in the recall sequence, which can be counted as a further fill-in error. However, this is not the case, because all analyses of fill-in and in-fill errors were restricted to responses following the first negative transposition in the recall sequence, such that any fill-in or in-fill errors that occurred following the initial fill-in or in-fill effects were not included in the analysis.

General Discussion

The fill-in effect was a major constraint on chaining models of serial recall. Although fill-in effects have been robustly observed in closed-set serial recall experiments (Farrell et al., 2013; Henson, 1996) and reconstruction of order tasks (Surprenant et al., 2005), there has not been an open-set serial recall experiment that has demonstrated a fill-in effect. To the contrary, Solway et al. (2012) found an in-fill effect in their analyses, and Farrell et al. found an in-fill effect in their reanalyses of the data of Grenfell-Essam and Ward (2012) for list lengths of five or more items. To evaluate whether this discrepancy was a result of the response format used, we conducted a serial recall experiment comparing open-set serial recall, open-set serial recall in which participants were encouraged to use blanks as responses, closed-set serial recall, and open-set reconstruction of order.

Consistent with prior findings, a robust fill-in effect was observed in the closed-set serial recall and open-set reconstruction of order conditions, yielding fill-in to in-fill ratios similar to those found in the previous literature (a 2:1 ratio). This means that after skipping an item, participants were extremely likely to revisit the item that was skipped. However, in open-set serial recall, neither a fill-in effect nor an in-fill effect was observed, but a robust fill-in effect was found in the condition in which participants were encouraged to skip over responses by typing "blank." This confirms Farrell et al.'s (2013) suggestion that the reanalyses conducted by Solway et al. (2012) may have found an in-fill effect because the selected data sets came from studies in which participants were not allowed to indicate their omissions. This result carries theoretical significance, because models that produce fill-in effects can produce a higher incidence of in-fill errors if the omission rate is high (for a demonstration of how a model with positional representations can produce an in-fill effect, see Farrell et al., 2013).

Although the fill-in effect was stronger in the closed-set serial recall and open-set reconstruction of order conditions, omission rates were quite low in those conditions, and they were much higher in the two open-set serial recall conditions (open and blanks). The hypothesis that a high omission rate can obscure a fill-in effect was supported by a separate evaluation of lists with different numbers of omissions. When trials with no omissions were considered, all conditions demonstrated a strong fill-in to in-fill ratio. However, as the number of omissions in a trial increased, the tendency for fill-ins decreased and that for in-fills increased. These results sound a cautionary note regarding ob-

[5] Some readers may note that for trials with one or more omissions, the blanks condition showed superior fill-in to in-fill effects, whereas the open condition did not. When "blank" responses are removed from the analysis, the figure looks nearly identical to the open condition (higher in-fill than fill-in for trials with one or more omissions).

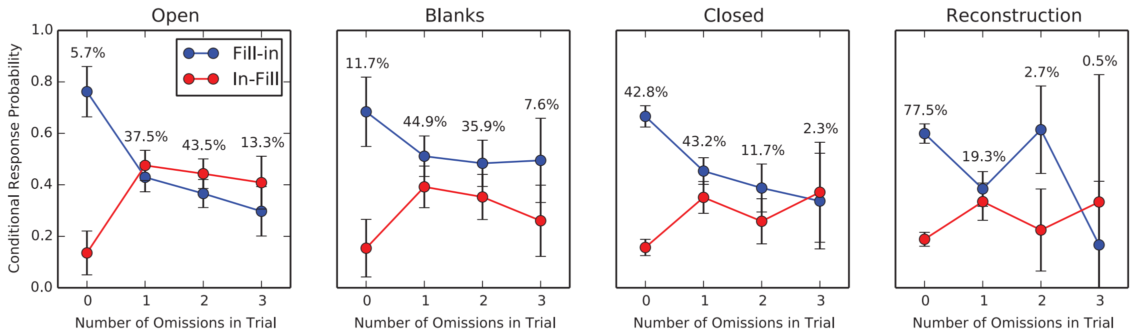

Figure 5. Conditionalized conditional response probabilities (CRPs) for the Lag −1 (fill-in) and Lag +1 (in-fill) transitions following the first case in which an item was skipped as a function of the number of omissions in the trial, with the percentage of all trials exhibiting that number of omissions depicted. Error bars represent 95% confidence intervals. See the online article for the color version of this figure.

served in-fill effects reported in the literature. In-fill effects have been reported with long lists, for which omission rates are quite high and trials with zero omissions are extremely rare. However, these in-fill effects may not be indicative of chaining per se. Although allowing participants to skip over responses can mediate this concern, a difficulty with such a procedure is that participants are not always aware of when they are making omissions, as evidenced by the fact that only 44% of omissions were marked in our condition that encouraged participants to indicate when they were skipping over responses.

The present results provide a strong challenge to chaining models of serial recall. Chaining models often use asymmetric associations in the forward direction, such that for an *A*–*B* association, *A* is associated to *B* to a greater extent than *B* is associated to *A*. Asymmetric forward associations predict a forward contiguity but, in addition, predict a higher preponderance of in-fill errors over fill-in errors. Chaining models can also possess symmetric associations, such that *A* is associated to *B* with the same strength as *B* is associated to *A* (e.g., Lewandowsky & Murdock, 1989). However, this model can at best predict an equal number of fill-in and in-fill errors without auxiliary assumptions. The present results undermine the possibility that open sets can produce a reliance on chaining. Instead, the results best support models in which items are stored using a primacy gradient (Farrell & Lewandowsky, 2002; Page & Norris, 1998), which make the a priori prediction of a fill-in effect.

References

Baddeley, A. (1968). How does acoustic similarity influence short-term memory? *Quarterly Journal of Experimental Psychology, 20,* 249–264.

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

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.

Coltheart, V. (1993). Effects of phonological similarity and concurrent irrelevant articulation on short-term memory recall of repeated and novel word lists. *Memory & Cognition, 21,* 539–545.

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

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. (2002). An endogenous distributed model of ordering in serial recall. *Psychonomic Bulletin and Review, 9,* 59–79.

Farrell, S., & Lewandowsky, S. (2003). Dissimilar items benefit from phonological similarity in serial recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29*.

Farrell, S., & Lewandowsky, S. (2012). Response suppression contributes to recency in serial recall. *Memory & Cognition, 40,* 1070–1080.

Geller, A. S., Schleifer, I. K., Sederberg, P. B., Jacobs, J., & Kahana, M. J. (2007). PyEPL: A cross-platform experiment-programming library. *Behavior Research Methods, 39,* 950–958.

Golomb, J. D., Peelle, J. E., Addis, K. M., Kahana, M. J., & Wingfield, A. (2008). Effects of adult aging on utilization of temporal and semantic associations during free and serial recall. *Memory & Cognition, 36,* 947–956.

Grenfell-Essam, R., & Ward, G. (2012). Examining the relationship between free recall and immediate serial recall: The role of list length, strategy use, and test expectancy. *Journal of Memory and Language, 67,* 106–148.

Henson, R. N. A. (1996). *Short-term memory for serial order*. Unpublished doctoral dissertation, University of Cambridge.

Henson, R. N. A. (1998). Short-term memory for serial order: The start-end model. *Cognitive Psychology, 36,* 73–137.

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

Hitch, G., Fastame, M., & Flude, B. (2005). How is the serial order of a verbal sequence coded? Some comparisons between models. *Memory, 13,* 247–258.

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

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

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

Klein, K. A., Addis, K. M., & Kahana, M. J. (2005). A comparative analysis of serial and free recall. *Memory & Cognition, 33,* 833–839.

Lewandowsky, S., & Murdock, B. B. (1989). Memory for serial order. *Psychological Review, 96,* 25–57.

McNicol, D. (1978). Initial codes for order in memory. *Quarterly Journal of Experimental Psychology, 30,* 681–691.

Osth, A. F., & Dennis, S. (2014). Prior list intrusions in serial recall are positional. *Manuscript submitted for publication*.

Page, M., & Norris, D. (1998). The primacy model: A new model of immediate serial recall. *Psychological Review, 105,* 761–781.

Roodenrys, S., & Quinlan, P. T. (2000). The effects of stimulus set size and word frequency on verbal serial recall. *Memory, 8,* 71–78.

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

Surprenant, A. M., Kelley, M. R., Farley, L. A., & Neath, I. (2005). Fill-in and infill errors in order memory. *Memory, 13,* 267–273.

Watkins, M. J., Watkins, O. C., & Crowder, R. G. (1974). The modality effect in free and serial recall as a function of phonological similarity. *Journal of Verbal Learning and Verbal Behavior, 13,* 430–447.

Received September 2, 2014
Revision received December 29, 2014
Accepted December 31, 2014 ■