Memory & Cognition
2010, 38 (4), 419-430
doi:10.3758/MC.38.4.419

Understanding the dynamics of correct and error responses in free recall: Evidence from externalized free recall

NASH UNSWORTH, GENE A. BREWER, AND GREGORY J. SPILLERS
University of Georgia, Athens, Georgia

The dynamics of correct and error responses in a variant of delayed free recall were examined in the present study. In the externalized free recall paradigm, participants were presented with lists of words and were instructed to subsequently recall not only the words that they could remember from the most recently presented list, but also any other words that came to mind during the recall period. Externalized free recall is useful for elucidating both sampling and postretrieval editing processes, thereby yielding more accurate estimates of the total number of error responses, which are typically sampled and subsequently edited during free recall. The results indicated that the participants generally sampled correct items early in the recall period and then transitioned to sampling more erroneous responses. Furthermore, the participants generally terminated their search after sampling too many errors. An examination of editing processes suggested that the participants were quite good at identifying errors, but this varied systematically on the basis of a number of factors. The results from the present study are framed in terms of generate–edit models of free recall.

A great deal of work has been done in which the dynamics of free recall have been examined in order to understand how individuals correctly recall items stored in their memory systems. Less work, however, has been done in which incorrect responses and the dynamics of correct and incorrect recalls have been examined. That is, although there is a burgeoning literature on false recalls in lists of semantically associated word lists (Deese, 1959; Roediger & McDermott, 1995), less work has been done in which the systematic effects of intrusion errors (items that were not presented on the current list) in standard free recall tasks have been examined. This is largely because these types of errors are generally quite rare, making analysis difficult. The aim of the present study was to examine the dynamics of correct and incorrect responses in free recall in order to better understand how individuals probe their memory systems and edit their responses in order to retrieve information from the recent past.

Generate–Edit Models and Context Retrieval

Many models of free recall can be classified as *generate–edit* models (Anderson & Bower, 1972; Kintsch, 1970; Metcalfe & Murdock, 1981; Watkins & Gardiner, 1979). For instance, in the search of associative memory model (SAM; Raaijmakers & Shiffrin, 1980) and its variants (Kimball, Smith, & Kahana, 2007; Mensink & Raaijmakers, 1988; Sirotin, Kimball, & Kahana, 2005), items are first generated (sampled) on the basis of the match between a general list cue and items stored in memory. Once an item has been generated, it is subjected to an editing and monitoring process that checks to see whether the generated item is correct. If the item is deemed correct, it is recalled. If the item is deemed incorrect, it is not recalled. Thus, recall performance is determined, in part, by two stages: a generation stage and an editing stage. Importantly, this type of model suggests that both correct and error responses are generated, but only correct responses (for the most part) are actually recalled because of a highly efficient editing process. This basic type of model naturally accounts for many recall findings and has been used to make novel predictions regarding recall performance.

In generate–edit models of this type, it is assumed that retrieval starts with a general context cue to probe the memory system. This cue is determined, in part, by the retrieval question and the retrieval plan that an individual develops. In standard free recall tasks, the retrieval question would be “What items were on the last presented list?” At the start of recall, it is assumed that individuals use this general list context cue to delimit possible items in memory into a search set composed of list items and possibly extralist items. Items are then generated (or sampled) on the basis of the match between information specified in the cue and information stored in the representations (encoding specificity; Tulving & Thomson, 1973). Importantly, less distinct contextual cues will tend to activate many items, leading to greater cue overload (Watkins, 1979). Thus, the probability of sampling correct items will

N. Unsworth, nunswor@uga.edu

be dependent not only on the match between the cues used and the stored items, but also on the extent to which the cues are distinct and isolate the target information relative to irrelevant information (Nairne, 2006).

On the first retrieval attempt, then, the first generated item should be the item that most strongly matches the general list context cue. That is, the first sampled item will be one that shares many features with the overall list context cue. Items that strongly match the list context cue will have a higher probability of being sampled first than items that weakly match the list context cue at the time of retrieval. After the item has been sampled, it is subjected to an editing and monitoring process that determines whether the item occurred on the last list. Theoretically, editing is based on source monitoring processes, whereby the item is judged as correct or incorrect on the basis of the overlap in contextual features between the sampled item and the list context (Johnson, Hashtroudi, & Lindsay, 1993). If there is a sufficient degree of overlap, the word is deemed correct and actually output (i.e., recalled). If there is a low degree of overlap, the word is considered incorrect and withheld (i.e., not recalled).

If the item successfully makes it past the editing stage, it is overtly recalled, and the search starts over for the next item. Unlike the first retrieval attempt, in these models, it is generally assumed that the next retrieval attempt relies not only on the general list cue, but also on the last retrieved item (or items). For instance, in the SAM model, it is assumed that the last retrieved item is used as a cue to sample the next item (Kimball et al., 2007; Raaijmakers & Shiffrin, 1980). Likewise, models such as the temporal context model (Howard & Kahana, 2002a) assume that the last retrieved item's associated context is also used as a cue. This means that the next item retrieved will be one that matches the overall list cue, shares features with the last retrieved item, and, importantly, shares contextual features with the last retrieved item. That is, recall transitions should be between items that are similar in terms of temporal-contextual features (Kahana, 1996) or in terms of semantic features (Howard & Kahana, 2002b).

Theoretically, the search process continues in a cyclical fashion in which items are continuously sampled on the basis of the overall list context cue, as well as the context from the most recently retrieved item. Throughout the search, items that are deemed correct will be recalled, and items deemed incorrect are covertly edited and withheld. This cyclical search process continues until a certain number of total failures to find new correct items is reached, at which point the individual terminates the search (Harbison, Dougherty, Davelaar, & Fayyad, 2009; Laming, 2009; Raaijmakers & Shiffrin, 1980).

The Present Study

Generate–edit models that rely on contextual retrieval as described above capture a number of important trends found for correct responses in free recall, including probability of first recall functions, serial position functions, contiguity and transition effects, interresponse times, and overall search termination decisions (Harbison et al., 2009; Howard & Kahana, 2002a; Raaijmakers & Shiffrin, 1980; Rohrer & Wixted, 1994; Sirotin et al., 2005; Unsworth, 2008). Furthermore, Sirotin et al. (2005; see also Kimball et al., 2007) demonstrated that an extended version of SAM can account for overall frequencies of intrusions in free recall and for transitions between some error responses (see Zaromb et al., 2006).

However, it is an open question as to whether this type of model can also explain the dynamics of both correct and error responses more fully. That is, these models typically account for errors that are actually produced, rather than accounting for errors that are generated (or sampled) but are withheld because of the editing process. Furthermore, given that overt errors are typically rare, they are usually not examined fully. Yet, despite this rarity, a number of systematic effects have been demonstrated. For instance, intrusions can be broken down into previous-list and extralist intrusions (PLIs and ELIs, respectively). PLIs represent words that were not presented on the current list but were presented on a previous list. ELIs represent words not presented on any of the lists but that tend to be phonologically or semantically related to one of the current target words (Craik, 1968). Previous research has demonstrated that both types of intrusion occur late in the recall period, with roughly 60% of both types of intrusion occurring at one of the last three output positions (Craik, 1968; Unsworth, 2008; see also Gardiner & Klee, 1976). Furthermore, prior work has suggested that when an intrusion is recalled, the next response also tends to be an intrusion (Zaromb et al., 2006). Finally, specific examination of PLIs has suggested that PLIs tend to come from the immediately preceding list and demonstrate a recency gradient (Murdock, 1974; Unsworth & Engle, 2007; Zaromb et al., 2006), and these PLIs tend to come predominantly from primacy and recency positions on the lists on which they were presented (Unsworth, 2008). Overall, it is clear that important information can be gleaned from an examination of error responses in free recall.

Our aim in the present study was to more fully examine correct and error responses in free recall. In particular, we wished to examine the generation of both correct and error responses in free recall in the absence of strong editing processes. That is, it is likely that many error responses are generated during retrieval, but because these error responses are subjected to an editing process, they are never overtly recalled, making inferences about them difficult. This lack of knowledge of the generation of error responses and how they interact with correct responses severely limits the predictive and explanatory power of many models of memory. For instance, as was noted above, intrusions (both PLIs and ELIs) tend to occur late in the recall period. But it is not known whether this is because intrusions are more likely to be generated late in the recall period or because the editing process is more likely to fail late in the recall period. It is entirely possible that intrusions are just as likely to be generated early in the recall period as late in the recall period, but intrusions tend to be found late in the recall period in standard free recall tasks because of changes in the editing process throughout the recall period. Thus, inferences regarding correct and error responses in standard free recall tasks are limited by a lack of information on generation and editing processes throughout the recall period.

To get around this limitation, we utilized an externalized free recall (EFR) task that has been used previously (Bousfield & Rosner, 1970; Kahana, Dolan, Sauder, & Wingfield, 2005; Roediger & Payne, 1985; Rosen & Engle, 1997; Unsworth & Brewer, in press). In EFR, participants are instructed to recall all of the words from the current list in a manner similar to that in standard free recall. In addition, the participants are instructed to recall any words that come to mind during the recall phase, even if they know that the word is not from the current list. Allowing the participants to recall all items that come to mind in the EFR task serves to minimize the editing process by making recall uninhibited (Bousfield & Rosner, 1970) and, thus, allows for an examination of the generation of correct and error responses.

Furthermore, in order to examine editing processes within EFR, Kahana et al. (2005) instructed participants to press a key immediately after any response that the participant knew was incorrect. Thus, in this version of EFR, participants are free to generate all items that come to mind (both correct and error responses) and can indicate whether they identify the item as a correct or an error response. This should allow for a more fine-grained examination of the editing of error responses than those of previous studies. For instance, we can ask what types of error responses are likely to pass the editing processes and what types of responses are likely to be caught by the editing process. Using the EFR task, we conducted a detailed examination of correct and error responses in order to better understand how individuals recall and edit information from memory.

METHOD

Participants and Design

The participants were 30 undergraduate students recruited from the subject pool at the University of Georgia. The participants were between the ages of 18 and 35 and received course credit for their participation. The participants performed two practice lists with letters and six lists of 10 words each. The words were 60 nouns selected from the Toronto Word Pool (Friendly, Franklin, Hoffman, & Rubin, 1982). The words were initially randomized and placed into the lists, and all of the participants received the exact same lists of words.

Procedure

The participants were tested individually. Items were presented visually for 1 sec each with a 1-sec blank screen in between the presentation of each word. After list presentation, the participants engaged in a 16-sec distractor task before recall: The participants saw eight 3-digit numbers for 2 sec each and were required to write the digits in descending order (e.g., Rohrer & Wixted, 1994; Unsworth, 2007). At recall, the participants saw three question marks in the middle of the screen. The participants were instructed not only to recall all of the items from the most recent list, but also to recall any other words that came to mind during the recall phase, even if they knew that the word was not presented on the most recent list. Furthermore, the participants were instructed that if they recalled a word that they knew was incorrect, they should press the space bar to indicate that the response was incorrect. The participants had 45 sec to recall as many of the words as possible in any order they wished. The participants typed their responses and pressed the "Enter" key after each response, clearing the screen. If the participants typed a word that they knew was incorrect (intrusions and repetitions), they were instructed to press the space bar before pressing the "Enter" key. Prior to the practice and test trials, the participants received a brief typing exercise (typing the words *one–ten*) to assess their typing efficiency.

RESULTS

The results are organized into two sections. In the first section, we examined the generation of correct and error responses. In the second section, we examined the editing process in terms of error responses as a function of whether they were correctly identified as errors or mistaken as correct responses.

Generation of Items

On average, the participants recalled 5.09 (*SE* = 0.22; range = 2.83–7.00) correct items per list, 1.57 (*SE* = 0.55; range = 0.00–11.67) PLIs per list, 3.53 (*SE* = 0.70; range = 0.00–17.17) ELIs per list, and 0.24 (*SE* = 0.05; range = 0.00–0.83) repetitions per list.[1] Thus, of the total number of errors, 66% were ELIs, 30% were PLIs, and 4% were repetitions. This overall distribution of errors is similar to the results in Keppel and Mallory (1969) over a variety of free recall instructions (i.e., normal, withhold guess, encourage guessing).

Next, we examined correct and error responses in more detail in order to better understand the dynamics of correct and error responses. Note that although some of these effects have been reported in previous studies, it was important to demonstrate them in the present study in order to get an overall picture of the dynamics of correct and error responses in the same data set. Furthermore, as was noted by Kahana et al. (2005), more research is needed to determine the extent to which similar effects (e.g., serial position and probability of first recall functions) are found in standard free recall and EFR.

First, we examined what items the participants started their recall with. Shown in Figure 1 are the probabilities of initiating recall with a correct response, a PLI, or an ELI. As can be seen, the participants were more likely to start recall with a correct item than with an intrusion (both

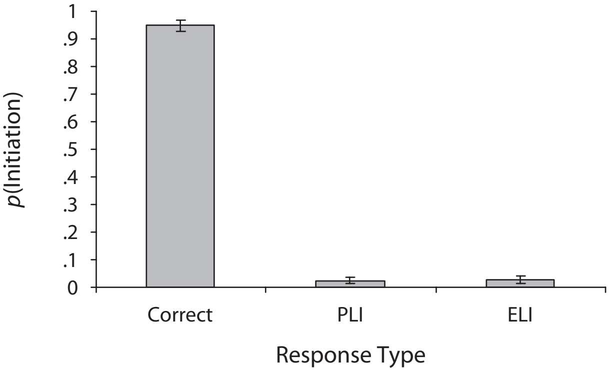

Figure 1. Probability of initiating recall as a function of response type. Error bars represent one standard error of the mean. PLI, previous-list intrusion; ELI, extralist intrusion.

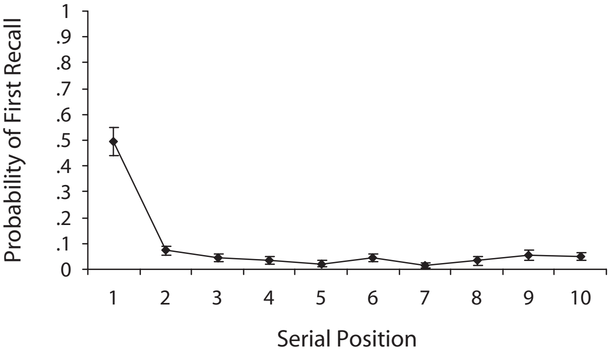

Figure 2. Probability of first recall for correct responses as a function of serial position. Error bars represent one standard error of the mean.

*t*s > 27, both *p*s < .01), and there was no difference in initiation probabilities for PLIs and ELIs [$t(29) = 0.21$, $p > .83$].

Given that the participants almost always initiated their recall with a correct item, we also examined the probability of first recall for correct items as a function of serial position. Shown in Figure 2 are the resulting functions. As can be seen, when the participants initiated their recall with a correct item, this item tended to be the first item presented [$F(9,261) = 36.55$, $MS_e = 0.02$, $p < .01$, $\eta_p^2 = .56$]. That is, 49% of the time, the participants started their recall with the first item presented. The probability of initiating recall with an item from any one of the other serial positions was much lower.

Overall, output functions for correct responses, PLIs, ELIs, and repetitions are shown in Figure 3A. As can be seen, the participants started out primarily with correct responses and recalled between five and six correct items in succession. The proportion of correct recalls started out high and then dropped substantially as a function of output position, such that very few correct items were recalled after Output Position 7. As is shown in Figure 4, these correct items came primarily from primacy portions of the serial position curve [$F(9,261) = 15.67$, $MS_e = 0.04$, $p < .01$, $\eta_p^2 = .35$]. That is, given that the task was a variant of delayed free recall, the probability of recalling correct items was greater for primacy items than for midlist or recency items (Glanzer & Cunitz, 1966).

After recalling correct items, the participants typically transitioned into recalling intrusions (both PLIs and ELIs). As can be seen in Figure 3A, the proportion of PLIs and ELIs increased from Output Position 1 to Output Positions 5 and 6. The proportion of both intrusion types tended to remain constant from Position 5 to Position 10 and then started to drop steadily thereafter. As is shown in Figure 5A, these PLIs tended to come predominantly from the immediately preceding list (Murdock, 1974; Unsworth & Engle, 2007; Zaromb et al., 2006) [$F(4,76) = 29.57$, $MS_e = 0.05$, $p < .01$, $\eta_p^2 = .61$]. Note that these numbers have been corrected for the total number of possible PLIs from each lag. Furthermore, as is shown in Figure 5B, these PLIs were predominantly from primacy and recency positions on the list that they were presented on (Unsworth, 2008), although a large number of PLIs also came from Position 5. It is not clear why so many PLIs came from Position 5, and this finding awaits further replication. In addition, 80% of PLIs were recalled correctly on their initial list but were then recalled incorrectly on subsequent lists. Thus, primacy items were more likely to be recalled correctly on the initial lists and then recalled incorrectly on subsequent lists. Examining those PLIs that were not initially correctly recalled suggested that 67% came from recency positions (Positions 8, 9, or 10) on their initial lists, and the remaining 33% came from Position 5 (again, it is unclear why so many PLIs came from Position 5). Thus, previously recalled PLIs tended to come from primacy positions, whereas PLIs not previously recalled tended to come from recency positions. Finally, PLIs also tended to be recalled in clusters of approximately 4.67 items, with 2.60 of these items coming from the same previous list.

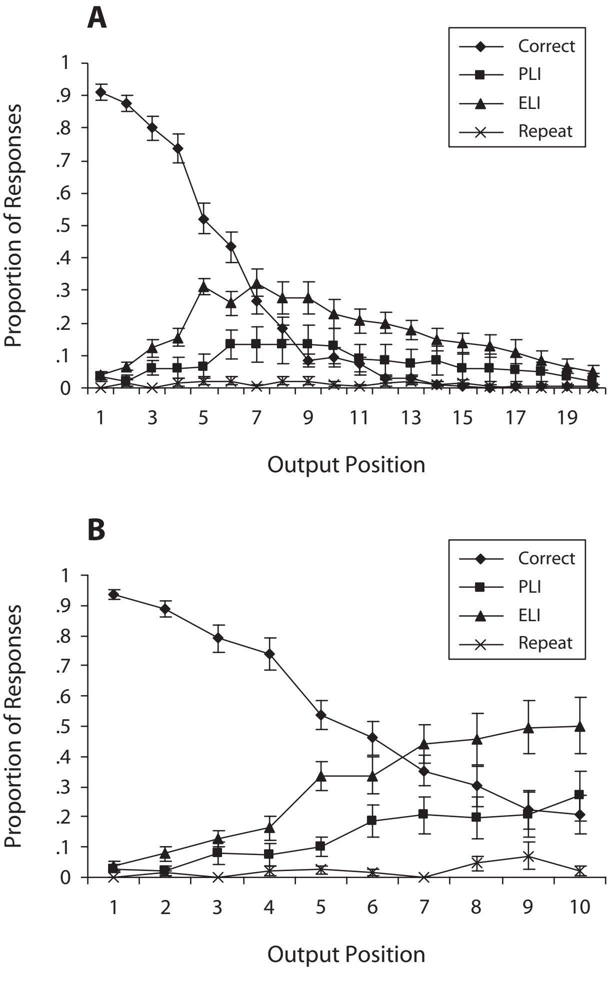

Figure 3. (A) Proportion of responses as a function of output position and response type. (B) Proportion of responses as a function of output position and response type, normalized such that the proportion of the four responses sums to 1.0 at each output position. Correct, correct responses; PLI, previous-list intrusions; ELI, extralist intrusions; Repeat, repetition errors. Error bars represent one standard error of the mean.

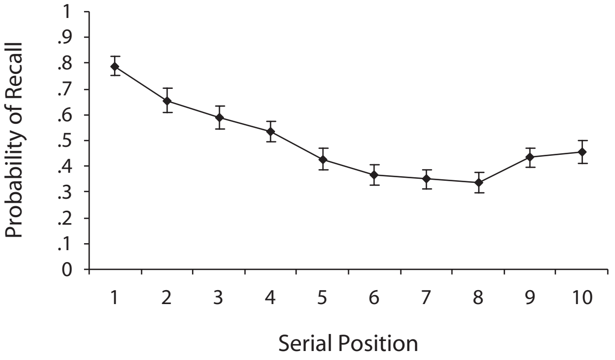

Figure 4. Probability of correct recall as a function of serial position. Error bars represent one standard error of the mean.

In terms of ELIs, the majority of responses from Output Positions 8–13 were ELIs. ELIs tended to be recalled in clusters of approximately 4.38 items, similar to the clusters seen for PLIs. Although no systematic attempt was made to determine the nature of these ELIs, a cursory inspection suggested that nearly all of the clustered ELIs were semantically related, consistent with prior work (Craik, 1968). For instance, 1 participant, after correctly recalling *rain*, then went on to output ELIs of *mud*, *fall*, *autumn*, *winter*, *summer*, and *spring*. Like the clusters of PLIs, which were associated via temporal-contextual relations (i.e., shared list membership), the clusters of ELIs tended to be related via semantic associations.

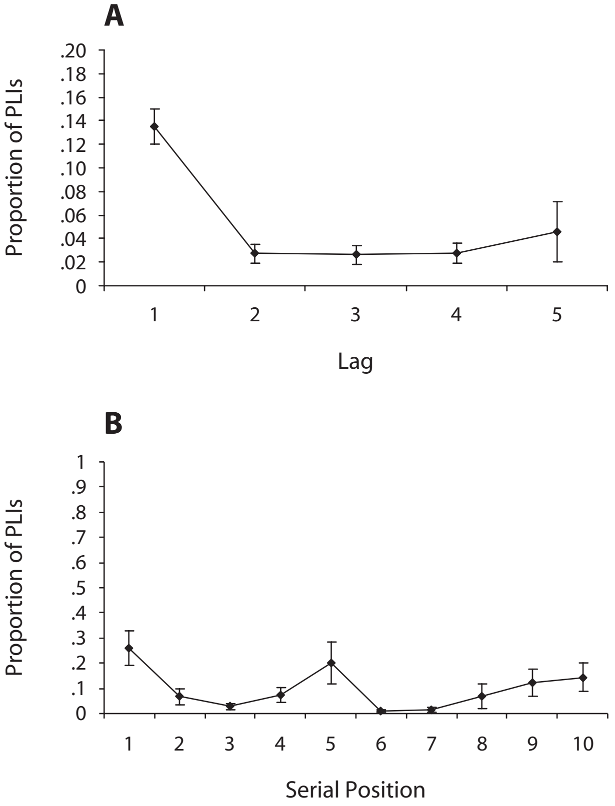

Figure 5. (A) Proportion of previous-list intrusions (PLIs) as a function of lag (list). (B) Proportion of PLIs as a function of their initial input serial position. Error bars represent one standard error of the mean.

Finally, an examination of repetition errors suggested very few systematic effects. First, although EFR was used, the overall frequency of repetitions was quite low (see also Kahana et al., 2005) and fairly similar to what is found in standard delayed free recall (see, e.g., Unsworth, 2008). Second, unlike correct recalls, PLIs, and ELIs, repetitions were fairly evenly distributed across output positions, as is shown in Figure 3A. Thus, unlike intrusion errors, the use of EFR in the present study did not reveal many systematic effects associated with repetition errors.

In order to get a better idea of the specificity of the generation process, we normalized the output functions shown in Figure 3A. Specifically, we examined only the first 10 output positions, given that roughly 10 items were emitted, and we computed the proportion of each response as a function of the total number of responses. Thus, for each output position, the proportion of the four different responses sum to 1.0. Shown in Figure 3B are the normalized output functions. As can be seen, like the overall output functions, correct responses were more likely to be generated early in the recall period, whereas errors were more likely to be recalled later in the recall period. Specifically, correct responses were more likely than intrusions at the first 5 output positions, but intrusions were more likely than correct responses at Positions 6–10 [$F(1,23) = 101.84$, $MS_e = 0.05$, $p < .01$, $\eta_p^2 = .82$]. Furthermore, despite overall low levels of repetitions, it can be seen that repetitions were more likely at Output Positions 8 and 9 than at earlier output positions.

Given that correct responses, PLIs, and ELIs tended to be recalled in clusters, we examined transition probabilities between all of the item types in order to better determine the relations among the different item types (see also Zaromb et al., 2006). Specifically, we computed transition probabilities for each possible transition for correct recalls, PLIs, ELIs, and repetitions individually. That is, we computed the probability of recalling a correct item followed by another correct item, as well as the probabilities of recalling a correct item followed by a PLI, followed by an ELI, or followed by a repetition. These transition probabilities were calculated separately for each response type. Specifically, for each participant, we computed the total number of transitions for each response type (i.e., total number of correct transitions, total number of PLI transitions, total number of ELI transitions, and total number of repetition transitions). Then we computed the total number of each type of transition (i.e., total number of correct-to-PLI transitions, total number of correct-to-ELI transitions, total number of correct-to-repetition transitions, etc.). The total of each type of transition was then divided by the overall number of possible transitions. For instance, the participants on average had 34 total correct transitions. Of these 34 total transitions, 23 were correct-to-correct transitions, 4 were correct-to-PLI transitions, 6 were correct-to-ELI transitions, and 1 was a correct-to-repetition transition. Therefore, the resulting transition

Table 1
Recall Transition Probabilities Between Correct Items, Previous-List Intrusions (PLIs), Extralist Intrusions (ELIs), and Repetitions for All Output Positions—Early Output Positions (1–5) and Late Output Positions (6–20)

Item	Correct Responses	PLI	ELI	Repetition Errors
	All Output Positions			
Correct	.68	.11	.17	.04
PLI	.07	.78	.09	.06
ELI	.12	.11	.72	.05
Repetition	.24	.22	.20	.34
	Early Output Positions			
Correct	.81	.04	.14	.01
PLI	.24	.50	.24	.02
ELI	.32	.11	.52	.05
Repetition	.63	.00	.25	.12
	Late Output Positions			
Correct	.60	.08	.27	.05
PLI	.03	.85	.09	.03
ELI	.06	.04	.89	.01
Repetition	.11	.21	.11	.57

probabilities would be .68 (23/34) for correct-to-correct transitions, .12 (4/34) for correct-to-PLI transitions, .18 (6/34) for correct-to-ELI transitions, and .02 (1/34) for correct-to-repetition transitions.

Shown in Table 1 are the resulting transition probabilities. As can be seen, when a participant recalled a correct item, the probability that the next item that they recalled would be another correct item was .68. The probability of recalling another item type was much smaller [$\chi^2(3) = 1{,}305.89, p < .01$]. Note that the comparisons in Table 1 are only meaningful within a row and not across rows, given that the transitions within a row were divided by the same baseline but those in different rows were divided by different baselines. Thus, Table 1 does not give an indication of the relative frequency of each response type. The transition probabilities within a row sum to 1.0.

Similar effects were found for PLIs and ELIs. Specifically, after recalling a PLI, the probability of recalling another PLI was .78, whereas the probability of recalling one of the other item types was much smaller [$\chi^2(3) = 377.67, p < .01$]. Note that the statistics are based on the raw frequencies. Furthermore, these PLIs tended to come from the same list. That is, 73% of the PLI–PLI transitions came from the same list, and this was significantly different from chance [$t(7) = 4.57, p < .01$]. Furthermore, given that most PLIs tended to come from the immediately preceding list (see Figure 5A), we examined these same-list PLI–PLI transitions as a function of list lag. These analyses suggested that the percentages of PLI–PLI transitions from the same list were roughly equivalent across lags (i.e., lag 1, 78%; lag 2, 64%; lag 3, 69%; lag 4, 71%; lag 5, 61%) [$F(4,16) = 1.63, MS_e = 0.02, p > .21$].

After recalling an ELI, the participants were more likely to recall another ELI (.72) than another response type [$\chi^2(3) = 938.78, p < .01$]. Thus, it seems clear that similar item types tended to be recalled in succession such that the item that was just recalled served as a cue for the next item. The only item type that did not necessarily follow this trend was repetitions. As can be seen in Table 1, the transition probabilities for repetitions were roughly equal [$\chi^2(3) = 4.80, p > .18$].

Given that that there were differences in the generation of correct responses and errors as a function of output position (see Figures 3A and 3B), we also examined the transition probabilities for early (Positions 1–5) and late (Position 6 and higher) positions separately. Shown in Table 1 are the resulting probabilities. As can be seen, and consistent with the results shown in Figures 3A and 3B, the transition probabilities tended to change as a function of output (all χ^2s > 10.80, ps $< .05$). Specifically, early in the output, the participants were more likely to transition to correct items, whereas later in the output participants were more likely to transition to errors. Again, this also changed as a function of whether the transition was of the same type (e.g., PLI–PLI) or was of a different type (PLI–ELI).

Our final set of analyses in this section was done to examine what responses the participants tended to stop their recall on. That is, we were interested in examining what types of responses tended to be the last response given. Therefore, we calculated termination probabilities for correct recalls, PLIs, ELIs, and repetitions. These termination probabilities were calculated by computing the frequency with which each response type occurred as the last response and then dividing each frequency by the total number of possible termination responses (i.e., dividing by 6, given that there were six lists). These analyses suggested that the participants' last response was a correct item 28% of the time and an error (PLI, ELI, or repetition) 72% of the time. Thus, the participants were far more likely to end with an error than a correct response [$t(29) = 4.72, p < .01$]. As is shown in Figure 6A, these termination probabilities can be further broken down for each of the response types. As can be seen, termination probabilities were highest for ELIs, followed by those for correct responses; those for PLIs and repetitions were roughly equal in their termination probabilities. All ts > 2.06, except for the comparison between PLIs and repetitions [$t(29) = 1.10, p > .28$].

We also examined the termination probabilities for each response type as a function of the total number of each response type (Kahana, Miller, & Weidemann, 2009). Specifically, we examined the frequency of stopping recall with a particular response (e.g., a correct response) divided by the total frequency of that response type (e.g., all correct recalls). Thus, in these analyses, the frequency of each response type occurring as the last response type was divided by the total number of that type of response, rather than by the total number of possible ending responses, as was done previously. This was done in order to determine whether a particular response was likely to occur as the last response type, corrected for the total number possible of that response type. For instance, repetitions were very rare, leading to an overall lower termination probability. However, it is possible that the majority of repetitions actually occur as the last response. Therefore, one needs to correct for the total number possible for each response type. Shown in Figure 6B are the resulting termination

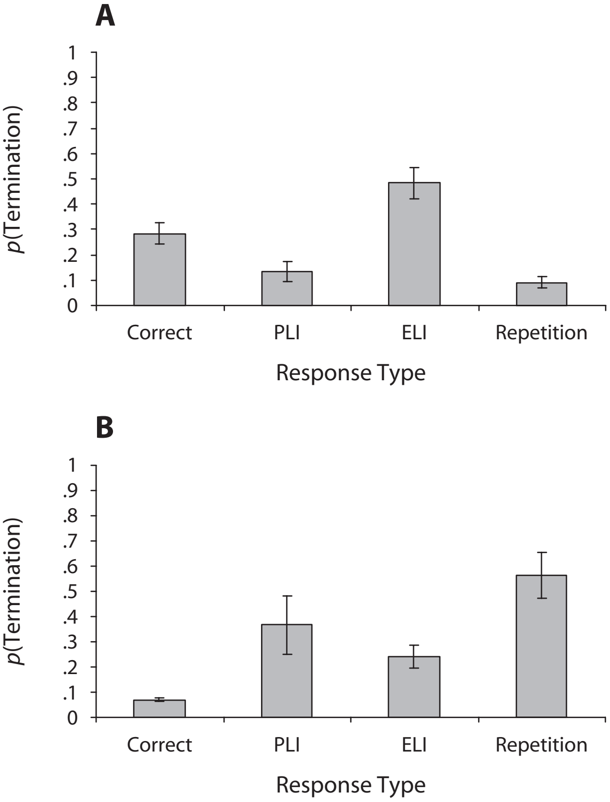

Figure 6. (A) Probability of terminating recall as a function of all of the response types combined. (B) Probability of terminating recall as a function of each response type individually. PLI, previous-list intrusion; ELI, extralist intrusion. Error bars represent one standard error of the mean.

probabilities. As can be seen, the probability of terminating recall in these analyses was greatest for repetitions, followed by PLIs, ELIs, and correct responses. That is, Figure 6B shows that 57% of repetitions occurred as the last response, whereas only 7% of correct recalls occurred as the last response. This suggests that the participants were very likely to terminate recall after an error, particularly if that error was a repetition (Kahana et al., 2009; Laming, 2009).

Identification of Items

Of the different response types, the participants correctly classified 98% (*SE* = 2) of their correct recalls as correct. The participants also correctly rejected 81% (*SE* = 9) of their PLIs, correctly rejected 68% (*SE* = 6) of their ELIs, and correctly rejected 47% (*SE* = 11) of their repetitions as incorrect. The participants were better at rejecting both types of intrusion than at rejecting repetitions (both *p*s $<$ .05), but there was no difference in the ability to reject the two types of intrusions ($p > .16$). Overall, the participants were fairly accurate in identifying correct and error responses, suggesting a fairly efficient editing/monitoring process. This efficiency is likely the main reason that intrusions are so rare in standard free recall tasks.

Next, we examined intrusion errors (PLIs and ELIs) in more detail to determine whether certain responses were more likely to make it past the editing process than others. For instance, the probability of rejecting PLIs that had been previously recalled was .86 (*SE* = .08), whereas the probability of rejecting PLIs that had not been previously recalled was only .50 (*SE* = .25). The rejection probability associated with previously recalled PLIs was significantly greater than chance [$t(19) = 4.42$, $p < .01$], whereas the probability of rejecting a PLI that was not previously recalled was not significantly different from chance ($p >$ .51). Note that given the relative scarcity of PLIs that were previously recalled, these values are associated with large *SE*s. Thus, one should be cautious in interpreting this result. Furthermore, although the participants only rarely started their recall with an intrusion (see Figure 1), the probability of correctly rejecting a PLI that was the first response was .50 (*SE* = .28). Likewise, the probability of rejecting an ELI that was the first response was only .25 (*SE* = .25). Neither of these rejection probabilities was significantly different from chance (both *p*s $>$ .54). Thus, when the participants started their recall with an error, this error was typically not recognized as being an error but, rather, was considered to be a correct item.

We also examined the probability of rejecting intrusions (both PLIs and ELIs) as a function of output position. Given the low numbers of intrusions early in the recall period (see Figures 3A and 3B), we computed the rejection probabilities for every two output positions separately (i.e., Positions 1 and 2 were collapsed, as were Positions 3 and 4, and so on). Shown in Figure 7A is the probability of rejecting PLIs as a function of output position. As can be seen, rejection probabilities were generally lower early in the recall period and increased thereafter. To test this, we examined the probability of rejecting a PLI as a function of output position in the first half versus in the second half (i.e., Positions 1–10 vs. Positions 11–20) [Wald $\chi^2(1) = 8.04$, $p < .01$]. In fact, all PLIs output at Positions 17 and 18 were correctly rejected. Similar results were obtained for ELIs. As is shown in Figure 7B, the probability of correctly rejecting an ELI was low early in the recall period and then increased steadily thereafter. Specifically, the probability of rejecting an ELI was lower during the first half of output than during the second half [Wald $\chi^2(1) = 111.50$, $p < .01$]. Collectively, these results suggest that the participants were much better at editing error responses late in the recall period (where most errors were generated) than early in the recall period (where only a few errors were generated). This suggests that errors output early are more likely to be confused for correct items, perhaps because of strong contextual, semantic, or phonological overlap with correct items.

Further examination of intrusions revealed a number of interesting findings. For instance, as is shown in Figure 5A, most PLIs were from the immediately preceding list. Are these PLIs from the most recent list the least likely to be correctly rejected? In order to examine this possibility, we computed rejection probabilities for PLIs as a function of list lag. Shown in Figure 8 are the resulting rejection probabilities. As can be seen, the probability of rejecting a PLI from the most recent list (lag 1) was quite low (.66), whereas the probability for rejecting

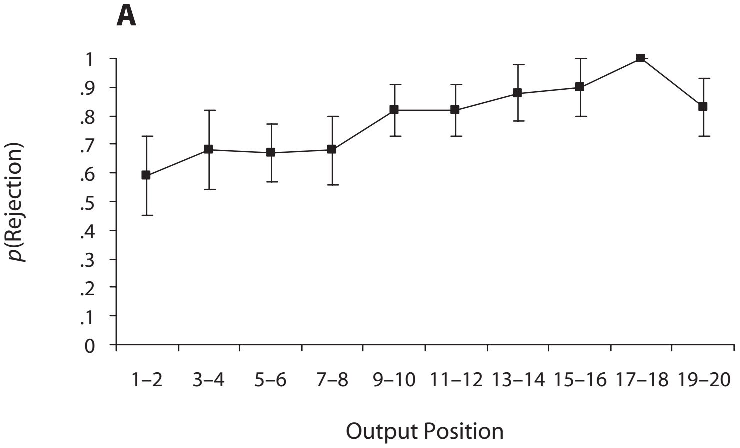

Figure 7. (A) Probability of rejecting a previous-list intrusion as a function of output position. (B) Probability of rejecting an extralist intrusion as a function of output position. Error bars represent one standard error of the mean.

PLIs from less recent lists was much better. To test this, we compared lag 1 rejection probabilities with the average rejection probabilities for all other lags. This analysis suggested that lag 1 rejection probabilities were significantly lower than rejection probabilities for the average of the other lags [$t(13) = 2.38$, $p < .05$]. An examination of serial positions of the PLIs suggested that the probability of rejecting a PLI from primacy positions was .92 ($SE = .08$), the probability of rejecting a PLI from midlist positions was .77 ($SE = .10$), and the probability of rejecting a PLI from recency positions was .88 ($SE = .10$).

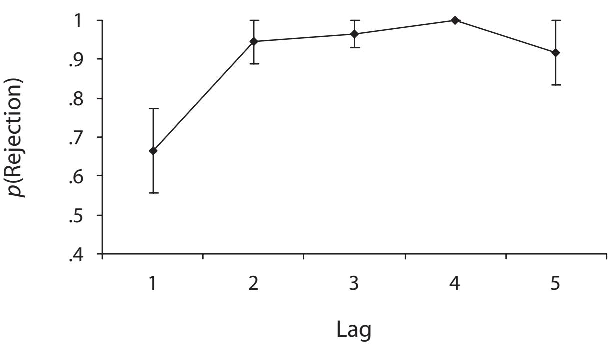

Figure 8. Probability of rejecting a previous-list intrusion as a function of list lag. Error bars represent one standard error of the mean.

Next, we examined the probability of rejecting intrusions (both PLIs and ELIs) as a function of what type of response preceded the intrusion. That is, are PLIs more likely to be rejected if preceded by another PLI or if preceded by a correct response? Shown in Table 2 are the resulting rejection probabilities for PLIs and ELIs as a function of the type of response that preceded them. As can be seen, when a correct item preceded a PLI, the rejection probability was .75. However, when one PLI preceded another PLI, the rejection probability was .92. Thus, the participants were generally better at rejecting a PLI when it was preceded by another PLI than when it was preceded by another response type [Wald $\chi^2(3) = 61.73$, $p < .01$]. This was true when compared with both correct responses and ELIs (both ps $< .01$), but the effect was not significant when compared with repetitions ($p > .59$). This is likely due to the relative scarcity of repetitions leading to large SEs. Furthermore, this did

Table 2
Probability of Rejecting an Intrusion (Previous List [PLI] or Extralist [ELI]) As a Function of the Response Type (Correct, PLI, ELI, or Repetition) That Preceded the Intrusion

Item	PLI	ELI
Correct	.75	.72
PLI	.92	.74
ELI	.69	.86
Repetition	.50	.44

not seem to differ as a function of whether the two PLIs came from the same list or from different lists (i.e., the probability of rejecting a PLI preceded by another PLI from the same list was .93; the probability of rejecting a PLI preceded by another PLI from a different list was .88). Likewise, when a correct item preceded an ELI, the rejection probability was .72. When one ELI preceded another ELI, however, the rejection probability was .86. Thus, the participants were generally better at rejecting an ELI when it was preceded by another ELI than when it was preceded by another response type [Wald $\chi^2(3) = 234.46$, $p < .01$]. This was true when comparing all response types (all *p*s $< .01$). Overall, it would seem that the participants were quite good at rejecting intrusions when they were preceded by the same response type but were poorer at rejecting intrusions preceded by another response type.[2]

Shown in Table 3 are the rejection probabilities, broken down as a function of whether the preceding response was itself associated with a correct identification (Id) or whether it was associated with an incorrect identification (Miss). As can be seen, when the preceding response was correctly identified, the participants were generally quite good at identifying the next item. However, when the preceding item was associated with an incorrect identification, the participants were near chance at identifying the next item. That is, when the preceding item was correctly identified, the participants were significantly above chance in identifying both PLIs and ELIs (all *p*s $<$ .01). However, when the preceding item was not correctly identified, the participants were at chance in correctly identifying both PLIs and ELIs (all *p*s $> .49$). It should be noted that given the relative scarcity of some of these transitions, one should be cautious in interpreting the results. That is, as was noted previously, the participants very rarely misclassified a correct item as incorrect. Thus, the transition between a misidentified correct item and a subsequent item rarely occurred, leading to large *SE*s for the associated rejection probabilities. Similar small absolute frequencies for the other transition types also likely increased the *SE*s for the resulting rejection probabilities, thus clouding their interpretation.

Table 3
Probability of Rejecting an Intrusion (Previous List [PLI] or Extralist [ELI]) As a Function of the Response Type (Correct, PLI, ELI, or Repetition) That Preceded the Intrusion and Whether That Preceding Response Was Associated With a Correct Identification (Id) or a Missed Identification (Miss)

Item	PLI	ELI
Correct		
Id	.75	.72
Miss	.50	.50
PLI		
Id	.93	.72
Miss	.67	.50
ELI		
Id	.75	.91
Miss	.50	.43
Repetition		
Id	.50	.50
Miss	.50	.60

Finally, we examined the probability of rejecting an error when it was the last response given. This should provide information regarding how well participants recognize that an error is incorrect late in the output sequence, which might influence their decision to terminate recall. Similar to the results shown in Figures 7A and 7B, the probability of rejecting a PLI when it was the last response was .83 ($SE = .11$), the probability of rejecting an ELI when it was the last response was .87 ($SE = .04$), and the probability of rejecting a repetition when it was the last response was .33 ($SE = .13$). Thus, when an intrusion was the last response given, the participants were quite accurate in rejecting the item and indicating that it was an error (i.e., rejection probabilities for both types of intrusion were significantly above chance; both *p*s $< .01$). In contrast, the participants were generally poor at rejecting repetitions overall (see above), and this trend continued even when repetitions were the last response given (i.e., rejection probabilities for repetitions when they were the last response given were not significantly different from chance; $p > .30$).

DISCUSSION

Our goal in the present study was to examine the dynamics of retrieval for correct and error responses in free recall. Specifically, we examined the extent to which generate–edit models that rely on contextual retrieval would be able to account for systematic effects associated with correct and error responses in EFR. According to models of this type, at the beginning of recall it is assumed that items are generated on the basis of the strength of the association between the item and the overall list context. Thus, correct items should be sampled first, given that they should share more contextual features with the current list context than would intrusions. This leads to an overwhelming probability of recalling correct items at the beginning of recall. Furthermore, given that the current task was a variant of delayed free recall, the first correct items tended to come from the primacy portion of the serial position curve. This could be due to increased rehearsals or attention at encoding or because these items are also the most recently rehearsed items (Tan & Ward, 2000).

After the first item is recalled, the overall list context, the previously recalled item, and its associated context are used to cue the next item (Howard & Kahana, 2002a; Raaijmakers & Shiffrin, 1980). Given that the first item recalled is usually correct, the next item should also be a correct item and should be an item that was presented in close temporal proximity to the previous item (Kahana, 1996). This means that participants should recall the first presented item first, and then recall should proceed in a forward direction, leading to strong primacy and reduced recency. This process of recalling items and using those items plus overall list context as cues repeats itself in a cyclical fashion, leading to a run of several correct responses (approximately four or five) and a high probability of recall transitions between correct items (i.e., .68).

At some point, the overall list context cue and the preceding items will no longer provide adequate cues exclusively for correct items. Rather, given that items are sampled on the basis of the overall strength of the cue–target relationship, at some point, weak correct items will start competing with errors that also share temporal or semantic features with the previously recalled items that serve as cues. That is, weak correct items and strong intrusions will compete for retrieval, leading to the retrieval of some intrusions. When this happens, the probability of sampling an intrusion should increase, and the probability of sampling a correct item should decrease. On the basis of the present data, it looks like this transition starts to occur around Output Position 6.

If a PLI is sampled, it is likely an item that was recalled correctly on the immediately preceding list, given that the current recall context should be fairly similar to the previous list recall context. This PLI will then be used as a cue for the next item, leading to an increased probability of sampling another PLI (around .78). Like correct recalls, this, in turn, leads to a cluster of PLIs (around four or five items). These PLIs likely share similar temporal contextual features, come from generally the same list (i.e., typically, the immediately preceding list), and come from primacy and recency portions of that list.

If an ELI is sampled, it is likely an item that shares semantic or phonological features with at least one of the previously recalled items that are acting as current cues. This ELI will then be used as a cue for the next item, leading to an increased transition probability for another ELI (i.e., .72). Thus, like correct recalls and PLIs, this leads to a cluster of ELIs (around four or five items). As was noted previously, these clusters seem to be strongly related on the basis of shared semantic features. For instance, 1 participant, after correctly recalling *career* and *skill* in succession, went on to recall *job*, *talent*, *résumé*, *salary*, *money*, and *happiness* in a cluster of ELIs. Clearly, *career* and *skill* acted as fairly powerful cues to generate related items from memory.

Finally, after long runs of generating incorrect items (PLIs, ELIs, and even repetitions), the participants terminated their recall. That is, after a certain number of total failures to generate another correct item, the participants decided to stop searching for new items (e.g., Harbison et al., 2009; Raaijmakers & Shiffrin, 1980). In fact, in the present study, the participants ended their recall on an error (PLI, ELI, or repetition) 72% of the time. Furthermore, examining each error type separately suggested that the majority of repetitions occurred at the last output position, suggesting that the participants had some sense that this item had already been recalled or that it generated items that had already been recalled. This latter result is consistent with those of Laming (2009), suggesting that recall termination is likely due to the fact that participants keep generating the same item (i.e., a repetition) over and over again. The present results extend this work by suggesting that it is not only repetitions that determine when recall will terminate, but also intrusions (see also Kahana et al., 2009). As was noted above, it is likely that the total number of failures (repetitions and intrusions) to generate a new correct item determines recall termination decisions (e.g., Harbison et al., 2009; Raaijmakers & Shiffrin, 1980).

The use of the EFR task in the present study has suggested a number of systematic findings about how participants generate correct and error responses in free recall. At the same time, this variant of EFR (Kahana et al., 2005) also allows for an examination of the editing processes insofar as participants are instructed to indicate those items that they know are incorrect. As was noted above, the use of this technique suggested that the majority of intrusions were correctly identified as incorrect. Thus, although many intrusions were generated during the recall phase, the participants were quite accurate in recognizing that these items were incorrect even though they shared both temporal-contextual (PLIs) and semantic (ELIs) features with the correct items. This suggests that the reason that intrusions are so rare in standard free recall tasks is not that these items are never generated (or sampled) but, rather, is that the monitoring component is quite effective at catching them before they are recalled.

Although the editing/monitoring process was generally effective in catching errors (particularly intrusions), its effectiveness varied as function of a number of factors. Specifically, intrusions output early in the recall period were associated with lower rejection probabilities than intrusions output late in the recall period. Furthermore, when an intrusion was preceded by another intrusion of the same type rejection, probabilities were generally high. However, when an intrusion was preceded by a correct item, rejection probabilities were lower. This suggests that intrusions generated early in the recall period are generated based on strong contextual as well as semantic and phonological overlap with one of the correct items on the list. For instance, an informal examination of ELIs output early in the recall period suggested that many of these were phonologically related to one of the target items on the list. That is, the participants recalled *agree* when the correct response was *degree*. This strong contextual and feature overlap leads to greater confusion and the mistaken belief that the item is in fact correct. Later in the recall period, when intrusions were generated on the basis of strong contextual and feature overlap with other intrusions (and weak overlap with correct items), the participants were much better at rejecting these items and identifying them as errors.

Similarly, the participants were poorer at rejecting PLIs from the immediately preceding list than at rejecting PLIs from lists presented further back. Again, this suggests that when there was strong contextual overlap between the current list and the intrusion being generated, the participants were more likely to confuse that intrusion for a correct response. When there was weak contextual overlap, the participants were much better at discriminating those errors as being incorrect. Thus, this suggests a strong coupling between items that are actually generated and subsequent editing of those items. Errors that were generated on the basis of shared features with correct items from the current list will be harder to discriminate than errors generated on the basis of shared features with other errors.

Finally, the rejection probabilities associated with errors that were the last response given suggested that when the last response was an intrusion, the participants were quite good at recognizing it as an error. Repetitions, however, were associated with much lower rejection probabilities. Despite this, these results suggest that the participants were generally accurate in determining the nature of their last response, and this knowledge may have influenced the participants' decision to terminate search. That is, at the end of the recall period, the participants generated many more errors than correct items (e.g., Figure 3B), and the participants were generally aware that these items were incorrect, which likely influenced their decision to stop searching for new correct items.

At this point, it would be remiss not to note that we are basing our interpretations of the data on generate–edit models of free recall, despite the fact that these models have encountered problems in the past. Specifically, the finding of recognition failure of recallable words (e.g., Tulving & Wiseman, 1975) has posed problems for generic versions of generate–edit models (see Watkins & Gardiner, 1979, for a review) given that any item that is recalled should also be correctly recognized. However, as was pointed out by Kintsch (1978), this is only problematic for generate–edit models if one assumes that the edit/recognition phase in recall is exactly the same as the processes used in a recognition task (i.e., the criterion is the same). If it is assumed that there are differences between editing/recognition in recall and recognition processes in a recognition task, generate–edit models can account for recognition failure of recallable words. Watkins and Gardiner (1979) countered Kintsch's (1978) point by noting that one of the main attractions of generate–edit models was that it should be possible to actually measure editing/recognition processes with a recognition task. Watkins and Gardiner went on to note that without an assessment of the editing/recognition process, generate–edit models lose much of their predictive power. However, the use of EFR, in which participants specifically indicate which responses they believe are correct, allows one to measure the editing process without assuming that it is exactly the same as those processes required in a recognition task.

Overall, the present results provide a fairly comprehensive examination of correct and error responses in free recall. These results are broadly consistent with a class of generate–edit models of free recall (Anderson & Bower, 1972; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann, & Usher, 2005; Kintsch, 1970; Metcalfe & Murdock, 1981; Raaijmakers & Shiffrin, 1980; Sederberg, Howard, & Kahana, 2008; Watkins & Gardiner, 1979). At the same time, these results have implications for generate–edit models, for other types of models of free recall, and for intrusions more broadly. In particular, although generate–edit models seem to provide an accurate verbal account of the pattern of correct and error responses demonstrated in the present study, it is not clear whether these models would be able to provide a more detailed quantitative account of the data. That is, the vast majority of quantitatively specified generate–edit models of free recall are primarily concerned with modeling correct responses and say little—if anything—about errors. Recently, however, two versions of the SAM model have been shown to account for more detailed patterns of errors (Kimball et al., 2007; Sirotin et al., 2005). For instance, Sirotin et al. demonstrated that their version of SAM, which incorporated both prior contextual and prior semantic information, could simulate the PLI recency effect, the overall low numbers of both PLIs and ELIs, and transitions between items. In addition, using a similar version of SAM, Kimball et al. showed that it could account not only for the overall frequency of PLIs and ELIs, but for the frequency of critical lure intrusions in the DRM paradigm (Deese, 1959; Roediger & McDermott, 1995), as well as for the fact that these critical lure intrusions tend to occur late in the output sequence. Thus, at least two generate–edit models of free recall seem capable of quantitatively explaining both correct and error responses. Future work is needed to determine whether these versions can account for all of the patterns of data demonstrated in the present study, as well as for the overall increase in intrusions that are seen in the EFR task relative to standard free recall conditions. Furthermore, although editing processes have been implemented in some models (e.g., Sirotin et al., 2005), more work is needed in order to examine editing processes, to determine the extent to which an editing process can successfully be implemented into existing models, and to account for the present pattern of data. Examining the dynamics of correct and error responses in free recall should provide a greater understanding of how individuals search and edit their memories in the absence of strong environmental cues.

AUTHOR NOTE

We thank Marc Howard and an anonymous reviewer for many valuable comments on a previous version of the manuscript. Correspondence concerning this article should be sent to N. Unsworth, Department of Psychology, University of Georgia, Athens, GA 30602-3013 (e-mail: nunswor@uga.edu).

REFERENCES

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review*, **79**, 97-123.

Bousfield, W. A., & Rosner, S. R. (1970). Free vs. uninhibited recall. *Psychonomic Science*, **20**, 75-76.

Craik, F. I. M. (1968). Types of error in free recall. *Psychonomic Science*, **10**, 353-354.

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.

Deese, J. (1959). On the prediction of occurrence of particular verbal intrusions in immediate recall. *Journal of Experimental Psychology*, **58**, 17-22.

Friendly, M., Franklin, P. E., Hoffman, D., & Rubin, D. C. (1982).

The Toronto Word Pool: Norms for imagery, concreteness, orthographic variables, and grammatical usage for 1,080 words. *Behavior Research Methods & Instrumentation*, **14**, 375-399.

Gardiner, J. M., & Klee, H. (1976). Memory for remembered events: An assessment of output monitoring in free recall. *Journal of Verbal Learning & Verbal Behavior*, **15**, 227-233.

Glanzer, M., & Cunitz, A. R. (1966). Two storage mechanisms in free recall. *Journal of Verbal Learning & Verbal Behavior*, **5**, 351-360.

Harbison, J. I., Dougherty, M. R., Davelaar, E. J., & Fayyad, B. (2009). On the lawfulness of the decision to terminate memory search. *Cognition*, **111**, 416-421.

Howard, M. W., & Kahana, M. J. (2002a). A distributed representation of temporal context. *Journal of Mathematical Psychology*, **46**, 269-299.

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

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, **114**, 3-28.

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

Kahana, M. J., Dolan, E. D., Sauder, C. L., & Wingfield, A. (2005). Intrusions in episodic recall: Age differences in editing of overt responses. *Journals of Gerontology*, **60B**, P92-P97.

Kahana, M. J., Miller, J. F., & Weidemann, C. T. (2009). *Recall termination in free recall*. Manuscript submitted for publication.

Keppel, G., & Mallory, W. A. (1969). Presentation rate and instructions to guess in free recall. *Journal of Experimental Psychology*, **79**, 269-275.

Kimball, D. R., Smith, T. A., & Kahana, M. J. (2007). The fSAM model of false recall. *Psychological Review*, **114**, 954-993.

Kintsch, W. (1970). *Learning, memory, and conceptual processes*. New York: Wiley.

Kintsch, W. (1978). More on recognition failure of recallable words: Implications for generation–recognition models. *Psychological Review*, **85**, 470-473.

Laming, D. (2009). Failure to recall. *Psychological Review*, **116**, 157-186.

Mensink, G.-J., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review*, **95**, 434-455.

Metcalfe, J., & Murdock, B. B. (1981). An encoding and retrieval model of single-trial free recall. *Journal of Verbal Learning & Verbal Behavior*, **20**, 161-189.

Murdock, B. B. (1974). *Human memory: Theory and data*. Potomac, MD: Erlbaum.

Nairne, J. S. (2006). Modeling distinctiveness: Implications for general memory theory. In R. R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 27-46). New York: Oxford University Press.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search of associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 14, pp. 207-262). New York: Academic Press.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **21**, 803-814.

Roediger, H. L., III, & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory & Cognition*, **13**, 1-7.

Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition*, **22**, 511-524.

Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General*, **126**, 211-227.

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

Sirotin, Y. B., Kimball, D. R., & Kahana, M. J. (2005). Going beyond a single list: Modeling the effects of prior experience on episodic free recall. *Psychonomic Bulletin & Review*, **12**, 787-805.

Tan, L., & Ward, G. (2000). A recency-based account of the primacy effect in free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 1589-1625.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, **80**, 352-373.

Tulving, E., & Wiseman, S. (1975). Relation between recognition and recognition failure of recallable words. *Bulletin of the Psychonomic Society*, **6**, 79-82.

Unsworth, N. (2007). Individual differences in working memory capacity and episodic retrieval: Examining the dynamics of delayed and continuous distractor free recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **33**, 1020-1034.

Unsworth, N. (2008). Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal-contextual search. *Journal of Memory & Language*, **59**, 223-236.

Unsworth, N., & Brewer, G. A. (in press). Variation in working memory capacity and intrusions: Differences in generation or editing? *European Journal of Cognitive Psychology*.

Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, **114**, 104-132.

Watkins, M. J. (1979). Engrams as cuegrams and forgetting as cue overload: A cueing approach to the structure of memory. In C. R. Puff (Ed.), *Memory organization and structure* (pp. 347-372). New York: Academic Press.

Watkins, M. J., & Gardiner, J. M. (1979). An appreciation of generate-recognize theory of recall. *Journal of Verbal Learning & Verbal Behavior*, **18**, 687-704.

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

NOTES

1. Given the large range of values for intrusion errors, one may wonder whether those participants who produced no PLIs also produced no ELIs and, thus, may have misunderstood the directions. However, an examination of those participants who produced none of one type of intrusion suggested that they always generated a large number of the other type of intrusion. That is, those participants who generated no PLIs generated 4.83 ELIs per list. Only 1 participant generated no ELIs; however, that participant generated 6.67 PLIs per list. Thus, all of the participants generated some intrusions, but the relative frequency of the different intrusions differed among participants (see also Unsworth & Brewer, in press).

2. We also examined this as a function of output (i.e., Positions 1–10 vs. Positions 11–20). Consistent with Figures 7A and 7B, all rejection probabilities increased from the first half to the second half. However, the increase in the rejection probabilities was virtually the same, regardless of the preceding item type.

(Manuscript received May 19, 2009;
revision accepted for publication November 3, 2009.)