Acta Psychologica 138 (2011) 19–29

Contents lists available at ScienceDirect

Acta Psychologica

journal homepage: www.elsevier.com/ locate/actpsy

Factors that influence search termination decisions in free recall: An examination of response type and confidence

Nash Unsworth *, Gene A. Brewer, Gregory J. Spillers

Department of Psychology, University of Georgia, United States

ARTICLE INFO

Article history:
Received 22 June 2010
Received in revised form 29 April 2011
Accepted 3 May 2011
Available online 28 May 2011

PsychINFO classification:
2343

ABSTRACT

In three experiments search termination decisions were examined as a function of response type (correct vs. incorrect) and confidence. It was found that the time between the last retrieved item and the decision to terminate search (exit latency) was related to the type of response and confidence in the last item retrieved. Participants were willing to search longer when the last retrieved item was a correct item vs. an incorrect item and when the confidence was high in the last retrieved item. It was also found that the number of errors retrieved during the recall period was related to search termination decisions such that the more errors retrieved, the more likely participants were to terminate the search. Finally, it was found that knowledge of overall search set size influenced the time needed to search for items, but did not influence search termination decisions.

In recall tasks participants are required to generate items from memory either in the presence of a particular cue (cued recall) or without any explicit cues (free recall). Although much previous work has been done examining how participants initiate recall (e.g., Howard & Kahana, 1999; Laming, 1999) and transition between successive recalled items (e.g., Howard & Kahana, 2002; Kahana, 1996), relatively less work has been done examining recall termination (e.g., Dougherty & Harbison, 2007; Winograd, 1970). In the current study we examined recall termination as a function of type of response (correct or errors) participants ended on as well as their confidence in those responses in hopes of better elucidating the mechanisms that underlie recall termination decisions in free recall.

1. Search termination decisions in free recall

Once a search of memory has been initiated, whether it be for an answer to a specific question or for a list of recently presented words, at some point that search will have to be terminated. What exactly dictates the decision to terminate memory search is an important issue for nearly all models of retrieval. Intuitively, one may think that the search is terminated once the correct answer (or answers) is generated. However, there are likely many situations in which the correct answer or the set of correct answers are not immediately generated, and thus there is likely some rule for deciding when to terminate the search. Despite the fact that search termination rules are explicit in many models of memory (e.g., Anderson, Bothell, Lebiere & Matessa, 1998; Davelaar, Goshen-Gottstein, Ashkenazi, Haarmann & Usher, 2005; Harbison, Dougherty, Davelaar & Fayyad, 2009; Laming, 2009; Metcalfe & Murdock, 1981; Raaijmakers & Shiffrin, 1980; Sederberg, Howard & Kahana, 2008; see Shiffrin 1970a for a number of possible termination rules), there is little evidence for when and why participants actually terminate their search.

Part of the reason for a lack of evidence on search termination decisions (especially in episodic memory paradigms) was that it was difficult to determine and measure when participants had decided to terminate their search. Without a metric of search termination it is difficult to determine what variables may affect search termination decisions. Recently, Dougherty, Harbison and colleagues (Dougherty & Harbison, 2007; Harbison et al., 2009) developed a means of measuring when during retrieval participants decided to terminate the search. Specifically, Dougherty and Harbison (2007) modified a traditional free recall paradigm such that participants, rather than the experimenter, determined how much time they were allowed during the recall period. That is, in many free recall tasks participants are given a fixed amount of time (e.g., 45–60 s) to recall the most recently presented list of items. In the paradigm utilized by Dougherty and Harbison, participants are given an unlimited amount of time to recall, and the recall period ends once the participant decides that they are done recalling items (usually via key press). Importantly, this method allows the researcher to measure when the participant has decided to terminate the search. Thus, although other researchers have allowed participants to decide when to terminate search (e.g., Bower, Clark, Lesgold & Winzenz, 1969; Klein, Addiss & Kahana, 2005; Shiffrin, 1970b), the novel aspect of this paradigm is that there is now a means of measuring when during the recall period the participant has decided to terminate the search (see also Winograd, 1970). In

* Corresponding author at: Department of Psychology, University of Georgia, Athens, GA 30602-3013, United States. Tel.: +995 706 542 8346.
E-mail address: nunswor@uga.edu (N. Unsworth).

0001-6918/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.actpsy.2011.05.001

particular, Dougherty and Harbison suggested that three recall latency variables could be determined via this paradigm. These include: (1) the time-to-last response which is the time from the onset of the recall period to when the last response is given; (2) exit latency, which is the time between the last response and the search termination decision; and (3) the total time spent searching which is simply the sum of the time-to-last response and exit latency.

A hypothetical recall period is shown in Fig. 1 depicting these three latency variables. As shown in the example, the time-to-last response is simply the amount of time that has elapsed from the onset of the recall period to when the last response (e.g., pen) is given by the participant. Exit latency, however, is the time from the last response to when the participant terminates recall (here by pressing the slash key). Finally, total time is the total amount of time from the onset of recall to when the participant terminates recall. Thus, this paradigm provides an effective means of measuring aspects of search termination decisions.

Empirically, Dougherty and Harbison (2007) demonstrated that exit latency was correlated with an individual's decisiveness such that individuals who were more decisive had shorter average exit latencies than individuals who were less decisive. Furthermore, Dougherty and Harbison found that exit latency changed as a function of the number of items retrieved such that exit latency decreased as more items were retrieved. Thus, if a participant recalled only two items they had a fairly long exit latency compared to an individual who recalled seven items. In subsequent work Harbison et al. (2009) replicated and extended these results. Specifically, Harbison et al. found that a measure of decisiveness was related to an individual's average exit latency and found that exit latency dropped as a function of the number of items retrieved. In this study Harbison et al. manipulated both word frequency and list-length in separate experiments to determine if these variables affected the temporal measures. Harbison et al. found that both word frequency and list-length affected total time and time-to-last response such that participants took longer to recall in pure word frequency lists than mixed word frequency lists and that participants took longer to recall as list-length increased (see also Rohrer & Wixted, 1994). In terms of exit latency, word frequency did not seem to affect participants' termination decisions, whereas list-length had a marginal effect on exit latency. Specifically, as list-length increased there was a small (i.e., approximately 1 s) increase in exit latency. Although this effect did not quite reach conventional levels of significance, it suggests that as list-length increased participants may have spent more time searching before they decided to terminate search.

In order to better examine search termination decisions, Harbison et al. (2009; see also Harbison, Davelaar & Dougherty, 2008 examined the ability of four different stopping rules to account for the results of decreasing exit latency and increasing total time as function of number of items retrieved. These rules included: 1) the total amount of time spent retrieving, 2) the time since the last retrieved item, 3) last inter-response time (IRT), and 4) the total number of retrieval failures. Each rule was implemented within a simplified version of the search of associative memory (SAM) model (Raaijmakers & Shiffrin, 1980). Of these four rules, Harbison et al. found that the total number of retrieval failures seemed to account for data best. Thus, this work suggests that one possible rule used to determine when to terminate search is the total number of retrieval failures that have occurred during the recall period.

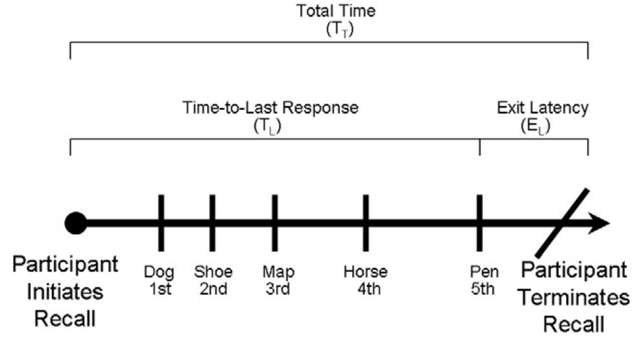

Fig. 1. Hypothetical recall period depicting time-to-last response, exit latency, and total time. See text for details. Adapted from Harbison et al. (2009).

Not only are retrieval failures a likely important component of search termination decisions, but so too are retrieval errors. Several studies have found that errors tend to occur relatively late in the recall period (e.g., Craik, 1968; Gardiner & Klee, 1976; Kahana, Miller & Weidemann, 2010; Roediger & McDermott, 1995; Unsworth, 2008; Unsworth, Brewer & Spillers, 2010), and more specifically, many of these errors tend to occur at the last output position (e.g., Kahana et al., 2010; Unsworth, 2008; Unsworth et al., 2010). For instance, in a standard delayed free recall task, Unsworth (2008) found that roughly 30% of intrusions (both previous list, and extra-list intrusions) tend to occur at the last output position, and over 60% of intrusions occur at one of the last three output positions. Furthermore, in a large scale reanalysis of many free recall datasets, Kahana et al. (2010) found that the probability of terminating recall increased as a function of output position and found that the probability of terminating recall was much higher for intrusions (particularly previous list intrusions) than for corrects. Thus, this work suggests that in standard delayed free recall tasks participants are quite likely to end their recall on an error response.

In order to examine these effects more thoroughly, we (Unsworth et al., 2010) utilized an externalized free recall paradigm in which participants were instructed to recall everything that came to mind during the recall period (i.e., uninhibited recall) including responses that they knew were errors. Furthermore, participants were instructed to press the spacebar for each response that they knew was incorrect (e.g., Kahana et al., 2005). Thus, this paradigm allows for an examination of errors that would not normally be recalled because they would be edited out and it allows for an examination of the editing process by comparing errors that were correctly rejected vs. those that were not rejected. With this paradigm we found that participants ended on an error over 70% of the time. Thus, this suggests that in standard delayed free recall participants are likely generating a number of errors towards the end of recall, but are correctly identifying those items as errors and are not outputting them. With uninhibited recall, however, we see that participants are far more likely to end on an error than a correct response suggesting that the type of response that participants end on may be an important component of search termination decisions. For instance, recently one of us (Unsworth, 2007) has specifically simulated cumulative recall functions with the use of a stopping rule that assumes individuals terminate their recall after a number of successive samples of not recalling anything new (i.e., both retrieval failures and retrieval errors). The use of a stopping rule that combined retrieval failures and retrieval errors successfully accounted for the overall shape of the cumulative recall functions as well as individual differences in working memory capacity in the cumulative recall functions. Thus, search termination decisions are likely not only influenced by the number of retrieval failures (i.e., sampling non-recoverable items), but also by the number of retrieval errors (i.e., sampling intrusions).

Given that people tend to terminate recall with an error response, a natural question is whether participants know that the response is incorrect. That is, are participants terminating their recall with an error because they know that the response is incorrect and thus, decide that it is no longer fruitful to continue searching for correct items? In essence this is a question of how meta-cognitive monitoring and meta-cognitive control interact (Dunlosky & Metcalfe, 2009; Nelson & Narens, 1990). Thus, in free recall it is possible that participants monitor their confidence in recalled items and use that confidence to decide whether to continue searching or whether to terminate search. Recalled items that are judged as correct, and hence have high confidence, may be used by the participant to decide that the search is going well and that it would be fruitful to continue searching for other items. However, recalled items that are judged as incorrect, and hence have low confidence, may be used by the participant to decide that the search is no longer generating correct items, and thus it should be terminated. Indeed, in our prior work (Unsworth et al., 2010) we found that participants were very accurate at classifying items as correct or incorrect. In particular, we found that participants correctly classified over 80% of the intrusions that were their last recalled item. Thus, it is possible that participants relied on their correct monitoring processes of already recalled items and this information was used to inform search termination decisions.

2. The present study

In the present study we examined search termination decisions in the form of exit latencies as a function of type of response (correct vs. intrusion) that participants ended on as well as their confidence in those responses. In particular, we were interested in determining whether exit latencies would be different when the last recalled item was a correct vs. an intrusion and whether exit latencies would change as a function of confidence in the recalled item. For response type, it is possible that exit latencies will be shorter for intrusions than for correct items based on the notion that participants know that the intrusions are incorrect and thus decide that it is not worth it to continue searching. Alternatively, it is possible that exit latencies will be shorter for correct items than for intrusions to the extent that once an intrusion is recalled participants might decide that they need to continue searching in order to not end on an error, and thus try and recall a few more corrects. That is, participants may feel like there are still correct items to be recalled, but need to change their search strategy to find them. This could lead to an increase in exit latency to the extent that extra time is needed to utilize a new search strategy to locate potential responses.

Similarly, for confidence it is possible that exit latencies will be shorter for low confidence responses compared to high confidence responses to the extent that participants use their uncertainty in the recalled item (or items) to decide that further search attempts will not lead to another correct response. Thus, as confidence decreases, so should exit latencies. Alternatively, it is possible that as confidence decreases, exit latencies will actually increase to the extent that participants use their uncertainty in the recalled items to decide that more items might be recalled with more effort, or with a change in the search strategy. In both cases, it is assumed that search termination decisions are determined, in part, by the confidence in the last retrieved item (or items). Thus, participants are relying on their monitoring processes (here confidence) to control their memory search (i.e., the decision when to terminate the search). It should be noted that these hypotheses are not meant to be necessarily specific. Rather they are merely general ideas about what might occur after examining response type and confidence and their relation with search termination decisions. Thus, the current study is not meant to be an explicit test of theoretical hypotheses, but rather is meant to provide more information on search termination decisions given the relative lack of research on this topic.

In order to examine these issues, we relied on the open-ended recall period paradigm of Dougherty, Harbison, and colleagues (Dougherty & Harbison, 2007; Harbison et al., 2009) in which participants press a key (here the slash key) to decide when they have terminated the search and want to move onto the next trial. In three experiments participants performed delayed free recall with this paradigm. In each experiment total time, time-to-last response, and exit latencies were measured for each individual. In addition, following each recalled item participants provided a confidence judgment (from 1 to 5) on the just recalled item. Thus, in all three experiments response type (correct vs. intrusion), recall latencies (total time, time-to-last, and exit latency), and confidence estimates were examined.

2.1. Experiment 1

In Experiment 1, we examined search termination decisions in a standard delayed free recall task. Participants performed multiple trials of delayed free recall in the open-ended recall paradigm. After the recall of each item participants provided a confidence judgment. When participants decided they were done recalling items they were instructed to press the slash key to end the recall period and move onto the next trial.

3. Method

3.1. Participants and design

Participants were 29 undergraduate students recruited from the subject-pool at the University of Georgia. Participants were between the ages of 18 and 35 and received course credit for their participation. Each participant was tested individually in a laboratory session lasting approximately 30 min. Participants performed two practice lists with letters and 10 lists of 10 words. Words were three- to six-letter medium frequency nouns (average frequency 62.47) nouns selected from the Toronto word pool (Friendly, Franklin, Hoffman & Rubin, 1982).

3.2. Procedure

Items were presented alone for 1 s each. After list presentation, participants engaged in a 16 s distractor task before recall: participants saw 8 three-digit numbers appear for 2 s each, and were required to write the digits in descending order (e.g., Rohrer & Wixted, 1994; Unsworth, 2008). At recall participants saw three question marks appear in the middle of the screen. Participants had as long as they needed to recall as many of the words as possible in any order they wished from the current trial. Participants typed their responses and after typing each response they were instructed to provide a confidence judgment for that response by pressing one of the 5 numbers on the numeric keypad to indicate their confidence with 1 equaling low confidence, a 5 equaling high confidence, and with 2–4 indicating mid-range levels of confidence.[1] Prior to the task participants were instructed to provide a confidence rating for each item and were given instructions on how to use the confidence scale. After typing the word and the associated confidence value, participants pressed Enter to clear the screen and type the next word. When participants had decided they were done recalling the words from the current list, they were instructed to press the slash key to end the recall period for that trial and move onto the next trial. Exit latency was measured as the time between pressing Enter on the prior word and the pressing of the slash key.

[1] Note, although verbal responses might be seen by some to be a purer measure of recall latency, some of our prior work has suggested that typed responses result in similar patterns of results in terms of recall latency and inter-response times (e.g., Unsworth, 2008). Furthermore, we would argue that typed responses are equally as ecologically valid as verbal responses given the widespread use of computers in many situations that require recall. Although clearly future work is needed to determine if there are any major differences between verbal and typed response in terms of various recall latency variables.

Table 1
Total number of items recalled, overall confidence in those items, probability of terminating recall, and termination confidence as a function of response type.

Response type	Total	Overall confidence	*p*(termination)	Termination confidence
Correct	47.97 (2.61)	4.75 (.04)	.78 (.04)	4.45 (.10)
PLI	3.30 (.57)	3.17 (.23)	.08 (.02)	2.73 (.33)
ELI	4.24 (1.14)	2.92 (.29)	.11 (.04)	2.69 (.37)

Note: Numbers in parentheses reflect one standard error of the mean.

4. Results

4.1. Accuracy

First, we examined correct and intrusion responses in terms of total numbers, overall levels of confidence, termination probabilities, and confidence for the last (termination) response. Shown in Table 1 are descriptive statistics for correct responses, previous list intrusions (PLIs), and extra-list intrusions (ELIs). As can be seen participants recalled more correct items than either type of intrusion (both p's<.01), but there was no difference in the number of PLIs and ELIs recalled (p>.84). In terms of overall confidence estimates, correct items were associated with higher confidence than either type of intrusion (both p's<.01) and there was no difference in confidence for the intrusions (p>.40). Furthermore, participants tended to terminate their recall following a correct item 78% of the time, whereas they tended to terminate recall following an intrusion only 19% of the time. Note, the other 3% of the time, participants terminated their recall following a repetition. Finally, an examination confidence of those responses that participants terminated their recall on suggested that participants were more confident in correct responses than intrusions (both p's<.01) and there was no difference in confidence of the intrusions (p>.90).

5. Latency

For all latency analyses we relied on linear mixed models. This type of analysis is a generalized case of the repeated measures analysis of variance that is slightly more powerful given that it utilizes all data points even for individuals who might have sporadically missing values. Overall, exit latency was 9487 ms (SE=1124), time-to-last response was 22,757 ms (SE=1077), and total time was 32,180 ms (SE=1417). Note the time-to-last response and total-time were a little longer than what prior research has found (Dougherty & Harbison, 2007; Harbison et al., 2009). This is likely due to the fact that participants had to make confidence judgments for each response, which likely shifted the overall latency distributions. Furthermore, consistent with prior research, exit latency dropped as a function of number of items recalled, $F(8, 105)=6.37$, $p<.01$. Specifically, as shown in Fig. 2, exit latency was longest when only one item was recalled and then dropped systematically as more items were recalled. Similar to prior work (e.g., Dougherty & Harbison, 2007; Harbison et al., 2009) we also computed the rank order gamma correlation between numbers retrieved and exit latency for each participant. The mean gamma correlation was −.38, p<.01, and all but 4 participants demonstrated this negative trend.

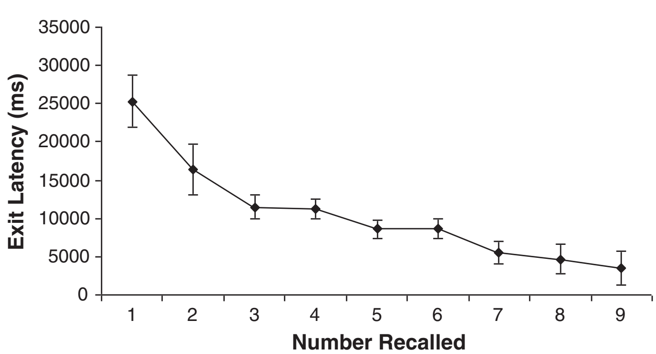

Fig. 2. Exit latency (in ms) as a function of number of items recalled. Note, error bars reflect one standard error of the mean.

Next, we examined whether exit latency would change as a function of response type (i.e., correct vs. intrusion). Therefore, exit latency was computed separately for correct responses, PLIs, and ELIs. Consistent with the notion that participants may be willing to search longer following a correct response, exit latency associated with correct responses was longer (M=10,420, SE=1092) than exit latency associated with either PLIs (M=5617, SE=1699), or ELIs (M=5778, SE=1699), both p's<.05. Exit latencies for PLIs and ELIs did not differ, p>.94. Thus, when the last retrieved item was a correct item, participants spent more time searching for additional items than when the last retrieved item was an intrusion.

In order to examine our other question of interest we examined exit latency as a function of confidence. That is, we computed exit latency for each of the five levels of confidence. As can be seen in Fig. 3a, exit latency was generally longer for high confidence (i.e., confidence=5) responses than for responses with lower values of confidence. To examine this, we tested exit latency associated with the highest level of confidence vs. exit latency for all other confidence levels. This was done because most of the responses were associated with a confidence of 5, and thus we combined the other four levels of confidence in order to increase power. We collapsed the other four levels of confidence in order to increase power given that so few responses were associated with low levels of confidence. This analysis suggested that exit latency was indeed longer for high confidence responses (M=10,420, SE=1092) than for all other confidence levels (M=7007, SE=1210), $F(1, 47)=4.38$, p<.05. This was also examined for each participant by computing the within-subject gamma correlations between confidence and exit latency. The mean within-subject gamma correlation was −.22, p<.05, and all but seven participants demonstrated this negative trend. Furthermore, in order

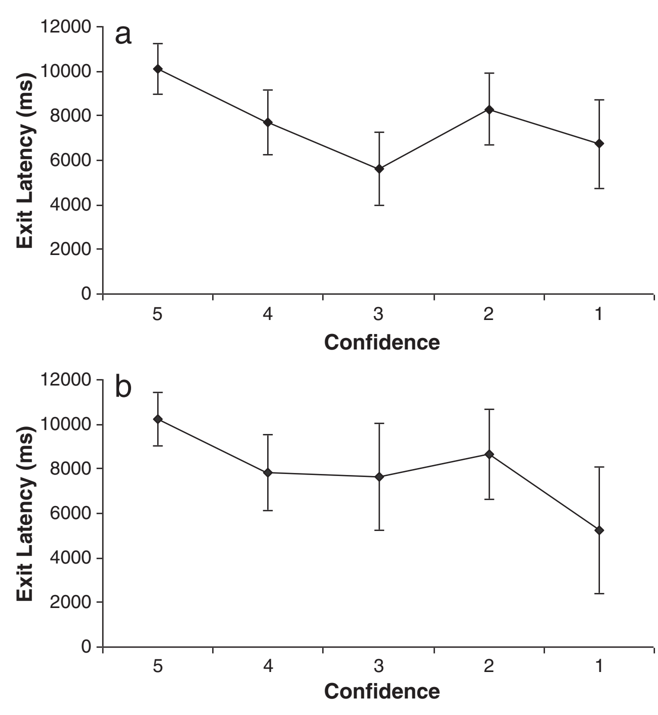

Fig. 3. (a) Exit latency (in ms) as a function of confidence for the last recalled item. (b) Exit latency (in ms) as a function of confidence for the last recalled item only if the last item was a correct response. Note, error bars reflect one standard error of the mean.

to ensure that this correlation was not influenced by the number of retrieved items which was also shown to be related to exit latencies, we computed the mean within-subjects gamma correlation between confidence and number retrieved. The correlation was −.02, $p>.81$. Thus, confidence and number retrieved were not related and seemed to make independent contributions to exit latency. Of course, it is also possible that how these two variables are used to determine exit latency might be systematically related.

Given that exit latency differed as a function of response type, we examined the relation between exit latency and confidence only for correct items. That is, the previous analysis could have been due to low confidence responses associated with intrusions. Therefore, we examined exit latency as a function of confidence for correct responses only. Shown in Fig. 3b are the results. Consistent with prior analysis, exit latency for high confidence responses was generally longer than exit latency for lower confidence responses. That is, exit latency for high confidence responses was significantly longer ($M=10{,}206$, $SE=1093$) than exit latency for the other levels of confidence ($M=6558$, $SE=1293$), $F(1,\ 46)=4.64$, $p<.05$. Thus, even when only examining correct responses, exit latency differed as a function of confidence. Exit latency was longer when participants were sure that the previously recalled item was a correct item.

5.1. Discussion

The results from Experiment 1 suggested that search termination decisions were related to the type of response that participants terminated their recall on as well as their confidence in that response. Specifically, participants mainly terminated their recall following a correct item rather than an intrusion, but when participants did terminate their recall with an intrusion they were generally aware that the item was incorrect. That is, confidence estimates for the last item given were much lower for intrusions than for correct items. Furthermore, when participants terminated their recall following a correct item, exit latencies were much longer than when the last retrieved item was an intrusion. In conjunction with the confidence results, this suggests that participants accurately monitored their responses and when a correct item was the last item retrieved, they were more likely to decide to continue searching. In contrast, when an intrusion was the last retrieved item, participants were more likely to decide to terminate the search as it was likely that no new correct items would be recalled. Additionally, participants relied on their confidence estimates to decide how long to continue searching even when the last retrieved item was correct. That is, when the last retrieved item was correct and participants were very sure that the item was correct, they spent much longer searching before termination than when the last item retrieved was correct and but they were unsure if it was correct. Overall, these results suggest that search termination decisions are influenced, in part, by monitoring processes that gauge how correct the last item retrieved was. When the information is consistent with the last item being correct, participants decide to continue searching. When the information is consistent with the last retrieved item being incorrect, participants decide to terminate the search.

5.2. Experiment 2

The purpose of Experiment 2 was to replicate and extend the findings from Experiment 1. In particular, given that so few intrusions were actually recalled in Experiment 1, more evidence is needed to determine the robustness of these effects. Additionally, Experiment 2 was conducted to better examine errors that are generated but are not recalled (i.e., are withheld), and how these errors can influence search termination decisions. Presumably in standard recall tasks participants generate many more errors than are actually produced, but these errors are edited out prior to recall. Thus, it is possible that it is these withheld errors that are generated in between the last recalled item and the decision to terminate which affects exit latency. By instructing participants to recall these errors, we should be able to not only increase the total number of errors emitted, but also should be able to better examine how these errors influence search termination decisions. In particular, we should be able to better examine when participants are likely to end on an error and examine how many errors (both successive and total number of errors) are emitted prior to termination. In order to examine these issues participants performed the same delayed free recall task as Experiment 1, but participants were further instructed to recall any words that came to mind during the recall phase even if participants knew that the word was not from the current list. Allowing participants to recall all items that come to mind in the task serves to minimize the editing process by making recall uninhibited (Bousfield & Rosner, 1970; Kahana et al., 2005; Roediger & Payne, 1985; Unsworth et al., 2010). Similar to Experiment 1, participants were also required to provide a confidence judgment for all responses (both corrects and errors). Thus, with this method we should be able to better examine how intrusion errors and judgments of confidence are related to exit latencies and, hence search termination decisions.

6. Method

6.1. Participants and design

Participants were 25 undergraduate students recruited from the subject-pool at the University of Georgia. Participants were between the ages of 18 and 35 and received course credit for their participation. Each participant was tested individually in a laboratory session lasting approximately 30 min. Participants performed two practice lists with letters and 10 lists of 10 words. Words were three- to six-letter medium frequency nouns (average frequency 62.47) nouns selected from the Toronto word pool (Friendly et al., 1982).

7. Procedure

Items were presented alone for 1 s each. After list presentation, participants engaged in a 16 s distractor task before recall: Participants saw 8 three-digit numbers appear for 2 s each, and were required to write the digits in descending order (e.g., Rohrer & Wixted, 1994; Unsworth, 2008). At recall participants saw three question marks appear in the middle of the screen. Participants were instructed to not only recall all of the items from the most recent list as they could, but to also 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. Participants had as long as they needed to recall as many of the words as possible in any order they wished from the current trial. Participants typed their responses and after typing each response they were instructed to provide a confidence judgment for that response by pressing one of 5 numbers on the numeric keypad to indicate their confidence with 1 equaling low confidence (i.e., know for sure that the word was not presented on the current list), a 5 equaling high confidence (i.e., know for sure that the word was presented on the current list), and with 2–4 indicating mid-range levels of confidence. Given that recall was uninhibited, participants were instructed to recall both correct and incorrect items and were instructed to provide a confidence estimate for each item. If they were positive that the item was correct they were instructed to give the item a 5. If they were positive that the item was incorrect they were instructed to give the item a 1. After typing the word and the associated confidence value, participants pressed Enter to clear the screen and type the next word. When participants had decided they were done recalling the words from the current list, they were instructed to press the slash key to end the recall period for that trial and move onto the next trial. Exit latency was measured as the time between pressing Enter on the prior word and the pressing of the slash key.

Table 2
Total number of items recalled, overall confidence in those items, probability of terminating recall, and termination confidence as a function of response type.

Response type	Total	Overall confidence	*p*(termination)	Termination confidence
Correct	41.84 (1.86)	4.83 (.05)	.50 (.05)	4.68 (.11)
PLI	6.61 (.57)	2.76 (.21)	.17 (.03)	2.38 (.36)
ELI	9.28 (1.14)	2.44 (.18)	.28 (.05)	2.12 (.26)

Note: Numbers in parentheses reflect one standard error of the mean.

8. Results

8.1. Accuracy

First, we examined correct and intrusion responses in terms of total numbers, overall levels of confidence, termination probabilities, and confidence for the last (termination) response. Shown in Table 2 are descriptive statistics for correct responses, previous list intrusions (PLIs), and extra-list intrusions (ELIs). As can be seen participants recalled more correct items than either type of intrusion (both p's$<$.01), but there was no difference in the number of PLIs and ELIs recalled ($p>$.13). In terms of overall confidence estimates, correct items were associated with higher confidence than either type of intrusion (both p's$<$.01) and there was no difference in confidence for the intrusions ($p>$.16). Furthermore, participants tended to terminate their recall following a correct item 50% of the time, whereas they tended to terminate recall following an intrusion 45% of the time. Note, the other 5% of the time, participants terminated their recall following a repetition. On average, participants terminated recall following 1.75 ($SE=$.17) successive intrusions. An examination of the distribution of successive intrusions, suggested that participants terminated recall following one error 45% of the time, terminated recall following two successive intrusions 31% of the time, terminated recall following three successive intrusions 12% of the time, and terminated recall following four or more successive intrusions 12% of the time. Finally, an examination of confidence for those responses that participants terminated their recall on suggested that participants were more confident in correct responses than intrusions (both p's$<$.01) and there was no difference in confidence of the intrusions ($p>$.46).

9. Latency

Overall, exit latency was 8521 ms ($SE=814$), time-to-last response was 23,854 ms ($SE=1095$), and total time was 32,375 ms ($SE=1258$). Consistent with Experiment 1, exit latency tended to drop as a function of number of items recalled, $F(8, 92)=5.72$, $p<$.01. Specifically, as shown in Fig. 4, exit latency was longest when only a few items were recalled and then dropped systematically as more items were recalled. Similar to Experiment 1 we computed the rank order gamma correlation between the number retrieved and exit latency for each participant. The mean gamma correlation was −.38, $p<$.01, and all but 6 participants demonstrated this negative trend.

Next, we examined whether exit latency would change as a function of response type (i.e., correct vs. intrusion). Exit latency was computed separately for correct responses, PLIs, and ELIs. Exit latency associated with correct responses was longer ($M=9160$, $SE=780$) than exit latency associated with either PLIs ($M=6216$, $SE=946$), or ELIs ($M=6692$, $SE=919$), both p's$<$.05. Exit latencies for PLIs and ELIs did not differ, $p>$.72.

Similar to Experiment 1, exit latency was examined as a function of confidence. As can be seen in Fig. 5a, exit latency was generally longer for high confidence (i.e., confidence = 5) responses than for responses with lower values of confidence. To examine this, we tested exit latency for associated with the highest level of confidence vs. exit latency for all other confidence levels. This was done because most of the responses were associated with a confidence of 5, and thus we combined the other four levels of confidence in order to increase power. This analysis suggested that exit latency was indeed longer for high confidence responses ($M=10{,}192$, $SE=1004$) than for all other confidence levels ($M=7338$, $SE=1026$), $F(1, 43)=4.01$, $p<$.05. Similar to Experiment 1 we computed the within-subject gamma correlations between confidence and exit latency. The mean within-subject gamma correlation was −.26, $p<$.01, and all but 7 participants demonstrated this negative trend. We also computed the mean within-subjects gamma correlation between confidence and number retrieved. The correlation was −.003, $p>$.97.

Similar to Experiment 1, we also examined exit latency as a function of confidence for correct responses only. Shown in Fig. 5b are the results. Consistent with Experiment 1, exit latency for high confidence responses was longer ($M=10{,}124$, $SE=1051$) than exit latency for the other levels of confidence ($M=6174$, $SE=1680$). This effect was marginally significant, $F(1, 30)=4.98$, $p<$.06. Exit latency

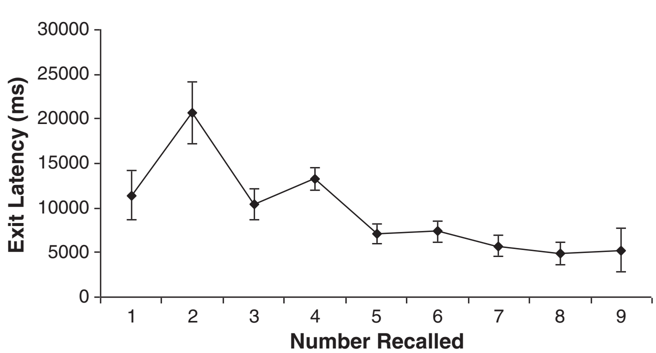

Fig. 4. Exit latency (in ms) as a function of number of items recalled. Note, error bars reflect one standard error of the mean.

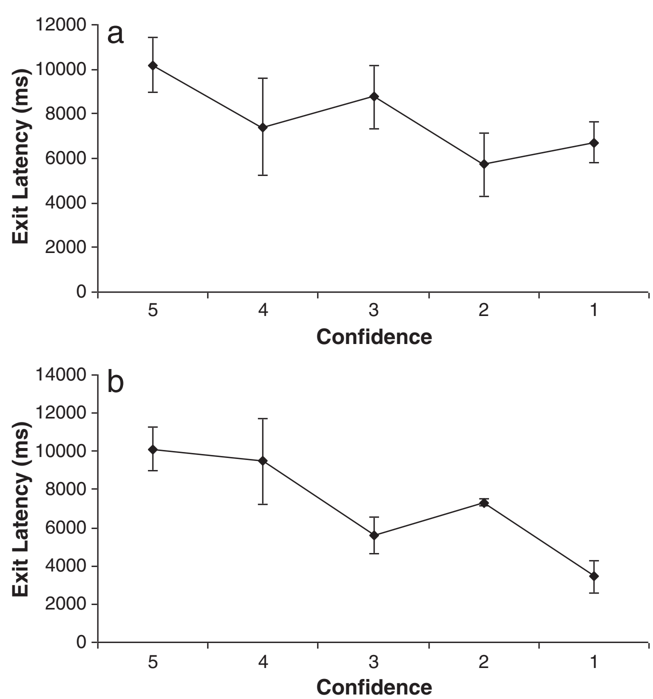

Fig. 5. (a) Exit latency (in ms) as a function of confidence for the last recalled item. (b) Exit latency (in ms) as a function of confidence for the last recalled item only if the last item was a correct response. Note, error bars reflect one standard error of the mean.

was longer when participants were sure that the previously recalled item was a correct item.

For our final set of analyses we examined exit latency as a function of the number of errors recalled prior to terminating recall. In particular, we examined exit latency as a function of the number of successive errors prior to search termination as well as a function of the total number of errors recalled during the entire recall period. In our first analysis we computed exit latency separately for when participants terminated recall following 1 error, 2 errors, 3 errors, or 4 or more errors. Note, here we combined PLIs and ELIs. As can be seen in Fig. 6a, exit latency was longest following a single error and then dropped systematically as the number of consecutive errors increased, $F(3, 38) = 2.91$, $p < .05$. Thus, exit latency was related, in part, to the number of successive errors that were emitted. When there was only one error immediately prior to termination, participants searched much longer before terminating the search compared to when several errors were recalled prior to termination. We also examined this for each individual subject by computing within-subject gamma correlations between number of successive errors and exit latency. The mean gamma correlation was −.23, $p < .06$, and all but 5 participants demonstrated this negative trend.

We also examined exit latency as a function of the total number of errors (both PLIs and ELIs) recalled prior to termination. On average, participants recalled approximately 1.54 ($SE = .21$) intrusions per list. However, this varied with a range of 0–7. Therefore, we computed exit latency separately for lists where no errors were recalled, when 1 total error was recalled, when 2 total errors were recalled, when 3 total errors were recalled, and when 4 or more total errors were recalled. As can be seen in Fig. 6b, exit latency was longest when participants recalled no errors, and then exit latency tended to drop as more total errors were recalled during the entire recall period, $F(4, 73) = 5.99$, $p < .01$. Thus, exit latency was related to the total number of errors a participant had recalled prior to terminating recall. We also examined this for each individual subject by computing within-subject gamma correlations between number of total errors and exit latency. The mean gamma correlation was −.37, $p < .01$, and all but 5 participants demonstrated this negative trend.

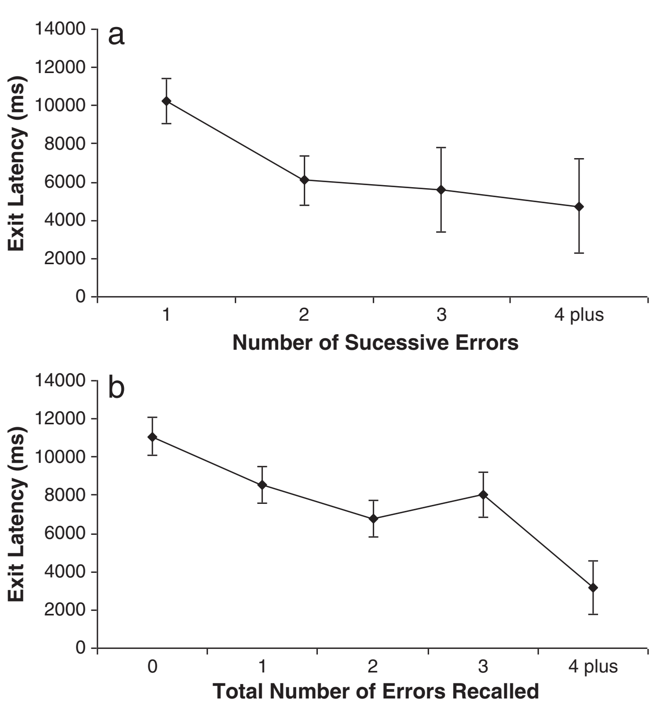

Fig. 6. (a) Exit latency (in ms) as a function of the number of successive errors recalled prior to search termination. (b) Exit latency (in ms) as a function of the number of total errors recalled prior to search termination. Note, error bars reflect one standard error of the mean.

9.1. Discussion

The results from Experiment 2 were consistent with the results from Experiment 1 suggesting that response type and confidence was related to search termination decisions. Specifically, participants tended to end their recall on a correct 50% of the time and on an intrusion 45% of the time. Thus, by making recall uninhibited, we see that participants are more likely to end on an intrusion than in standard delayed free recall. Consistent with Experiment 1, participants were more confident in correct responses than intrusions (as they should be) and this occurred both overall and for the last response. Importantly, replicating Experiment 1, the results suggested that exit latencies were longer following a correct item than an intrusion. Furthermore, exit latency dropped as a function of number of items retrieved, and dropped as a function of confidence with longer exit latencies being associated with higher levels of confidence. Thus, when the last retrieved item was correct and participants were highly confident that the item was correct, they spent much longer searching before termination than when the last item was incorrect and they were confident it was incorrect. Finally, the results from Experiment 2 suggested that the number of successive errors recalled prior to termination as well as the total number of errors recalled throughout the recall period was related to search termination decisions. Specifically, exit latencies were shorter when participants recalled many successive intrusions (as well as many intrusions overall) than when participants recalled few successive intrusions (and few intrusions overall). The more search errors that have accumulated during the recall period; the more likely participants are to terminate their search. Thus, these results provide support for the notion that the number (either successive or total) of search errors that have accumulated during the recall period is related to search termination decisions.

9.2. Experiment 3

The purpose of our final experiment was to replicate the major findings of the prior two experiments as well as examine other factors that can influence search termination decisions. In particular, we were interested in how knowledge of search set size might influence decisions to terminate the search. In perhaps the earliest study examining search termination decisions, Winograd (1970) had participants recall all of the states that began with either “M” or “N” as they could. One group was specifically told that there were 8 states that began with the desired letter (both M and N). The other group was not told how many states began with the desired letter. Thus, one group had knowledge of the size of the search set, whereas the other did not. Participants were told that when they felt like they couldn't recall any more names they should tell the experimenter. Winograd found that participants who have knowledge of the search set size searched longer than participants who did not explicitly know how big the search set size were. Thus, knowledge of search set size seemed to influence search termination decisions. Experiment 3 was conducted to determine if knowledge of search set size would similarly influence search termination in more standard free recall tasks. Furthermore, Experiment 3 was conducted to better determine the nature of the effect Winograd reported. In particular, Winograd only measured the time from recall onset to search termination, and thus only obtained a measure of total time to recall and did not break this down into time-to-last response and exit latency. It is possible that knowledge of search set size affects the overall time participants are willing to search (time-to-last response) or the time between the last recalled item and the decision to terminate (exit latency). To examine this, participants performed the same recall task as Experiment 1, but participants were informed that there were 10 items on each list. Additionally, onscreen participants were shown how many items they had recalled up to that point during recall period. That is, a number was shown for each word recalled. If participants had recalled 6 words so far, a 6 was presented at the bottom of the screen. As with the previous experiments, participants gave a confidence judgment after each response and pressed the slash key when they were done recalling items for that list and wanted to move onto the next trial.

10. Method

10.1. Participants and design

Participants were 28 undergraduate students recruited from the subject-pool at the University of Georgia. Participants were between the ages of 18 and 35 and received course credit for their participation. Each participant was tested individually in a laboratory session lasting approximately 30 min. Participants performed two practice lists with letters and 10 lists of 10 words. Words were three- to six-letter medium frequency nouns (average frequency 62.47) nouns selected from the Toronto word pool (Friendly et al., 1982).

10.2. Procedure

Items were presented alone for 1 s each. After list presentation, participants engaged in a 16 s distractor task before recall: Participants saw 8 three-digit numbers appear for 2 s each, and were required to write the digits in descending order (e.g., Rohrer & Wixted, 1994; Unsworth, 2008). At recall participants saw three question marks appear in the middle of the screen. Participants had as long as they needed to recall as many of the words as possible in any order they wished from the current trial. Participants typed their responses and after typing each response they were instructed to provide a confidence judgment for that response by pressing one of 5 numbers on the numeric keypad to indicate their confidence with 1 equaling low confidence, a 5 equaling high confidence, and with 2–4 indicating mid-range levels of confidence. After typing the word and the associated confidence value, participants pressed Enter to clear the screen and type the next word. When participants had decided they were done recalling the words from the current list, they were instructed to press the slash key to end the recall period for that trial and move onto the next trial. Prior to beginning the experiment, participants were informed that they would be recalling individual lists of 10 words each. Participants were further informed that during recall they would know how many items they had recalled up to that point.

11. Results

11.1. Accuracy

First, we examined correct and intrusion responses in terms of total numbers, overall levels of confidence, termination probabilities, and confidence for the last response. Descriptive statistics for correct responses, previous list intrusions (PLIs), and extra-list intrusions (ELIs) are shown in Table 3. As can be seen participants recalled more correct items than either type of intrusion (both p's$<.01$), but there was no difference in the number of PLIs and ELIs recalled ($p>.37$). In terms of overall confidence estimates, correct items were associated with either higher levels of confidence than either type of intrusion (both p's$<.01$) and there was no difference in confidence for the intrusions ($p>.18$). Furthermore, participants tended to terminate their recall following a correct item 76% of the time, whereas they tended to terminate recall following an intrusion only 23% of the time. Note, the other 1% of the time, participants terminated their recall following a repetition. Finally, an examination of confidence for those responses that participants terminated their recall on suggested that participants were more confident in correct responses than intrusions (both p's$<.01$) and there was no difference in confidence of the intrusions ($p>.23$).

Table 3
Total number of items recalled, overall confidence in those items, probability of terminating recall, and termination confidence as a function of response type.

Response type	Total	Overall confidence	p(termination)	Termination confidence
Correct	49.39 (2.21)	4.70 (.05)	.76 (.04)	4.37 (.12)
PLI	2.89 (.57)	2.96 (.32)	.08 (.02)	2.53 (.29)
ELI	3.81 (.53)	2.57 (.20)	.15 (.03)	2.15 (.24)

Note: Numbers in parentheses reflect one standard error of the mean.

11.2. Latency

Overall, exit latency was 9666 ms ($SE = 1188$), time-to-last response was 27,172 ms ($SE = 1053$), and total time was 36,807 ms ($SE = 1790$). Consistent with Experiments 1 and 2, exit latency dropped as a function of number of items recalled, $F(8, 122) = 2.46$, $p<.05$. As shown in Fig. 7, exit latency was longest when only one item was recalled and then dropped systematically as more items were recalled. Similar to the prior experiments we computed the rank order gamma correlation between numbers retrieved and exit latency for each participant. The mean gamma correlation was −.37, $p<.01$, and all but 5 participants demonstrated this negative trend.

Next, exit latency was examined as a function of response type. Consistent with the prior experiments, exit latency associated with correct responses was longer ($M = 10{,}935$, $SE = 1110$) than exit latency associated with either PLIs ($M = 5975$, $SE = 1516$), or ELIs ($M = 5328$, $SE = 1424$), both p's$<.05$. Exit latencies for PLIs and ELIs did not differ, $p>.76$.

We also examined exit latency as a function of confidence. As can be seen in Fig. 8a, exit latency was generally longer for high confidence responses than for responses with lower values of confidence. To examine this, we tested exit latency associated with the highest level of confidence vs. exit latency for all other confidence levels. This was done because most of the responses were associated with a confidence of 5, and thus we combined the other four levels of confidence in order to increase power. This analysis suggested that exit latency was indeed longer for high confidence responses ($M = 11{,}322$, $SE = 1253$) than for all other confidence levels ($M = 6574$, $SE = 1277$), $F(1, 51) = 7.04$, $p<.05$. Similar to the prior experiments we computed the within-subject gamma correlations between confidence and exit latency. The mean within-subject gamma correlation was −.26, $p<.01$, and all but 8 participants demonstrated this negative trend. We also computed the mean

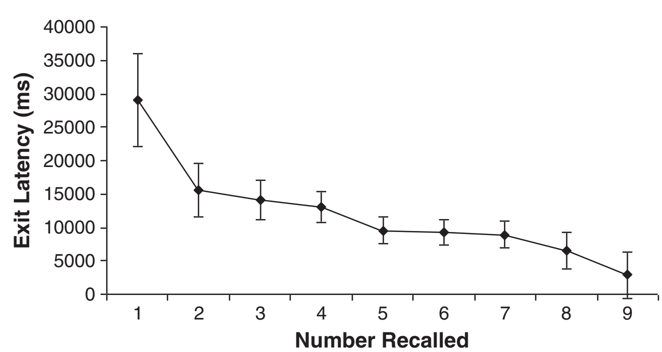

Fig. 7. Exit latency (in ms) as a function of number of items recalled. Note, error bars reflect one standard error of the mean.

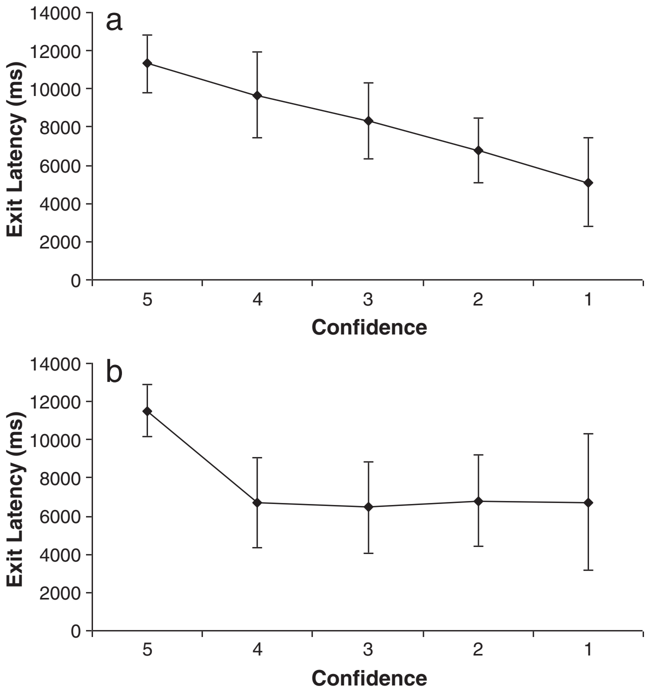

Fig. 8. (a) Exit latency (in ms) as a function of confidence for the last recalled item. (b) Exit latency (in ms) as a function of confidence for the last recalled item only if the last item was correct response. Note, error bars reflect one standard error of the mean.

within-subjects gamma correlation between confidence and number retrieved. The correlation was −.07, $p>.36$.

Next we examined exit latency as a function of confidence for correct responses only. As shown in Fig. 8b, and consistent with the prior experiments, exit latency for high confidence responses was longer ($M=11{,}525$, $SE=1416$) than exit latency for the other levels of confidence ($M=6874$, $SE=1688$), $F(1, 44)=4.46$, $p<.05$.

11.3. Cross-experimental analyses

In order to examine our primary question of interest, namely how knowledge of search set size would affect search termination decisions, we compared the results from Experiment 1 with the results from Experiment 3. That is, the only difference between the two experiments was whether participants were informed of the size of the search set (Experiment 3) or not (Experiment 1). First we examined the accuracy variables. There were no significant differences between the experiments in terms of accuracy, intrusions, or confidence (all p's>.31). Next, we compared the three latency variables (total time, time-to-last response, and exit latency) for the two experiments. Consistent with Winograd (1970) participant's total time was longer when they were informed of the size of the search set ($M=36{,}807$, $SE=1790$) compared to when they were not informed of the size of the search set ($M=32{,}180$, $SE=1417$), $t(55)=2.03$, $p<.05$. Breaking total time down into time-to-last response and exit latency suggested that only time-to-last response was influenced by knowledge of search set size. Specifically, time-to-last response was longer when participants were informed of the size of the search set ($M=27{,}172$, $SE=1053$) compared to when they were not informed of the size of the search set ($M=22{,}757$, $SE=1077$), $t(55)=2.93$, $p<.01$. However, there was no difference between the informed group in Experiment 3 ($M=9666$, $SE=1188$) and the uniformed group from Experiment 1 ($M=9487$, $SE=1124$) in terms of exit latency, $t(55)=.11$, $p>.91$. Thus, knowledge of search set size seems to influence time-to-last response latencies, but has no effect on exit latencies.

12. Discussion

The results from Experiment 3 replicated the basic effects from Experiments 1 and 2, suggesting that response type and confidence influence search termination decisions. Additionally, following up on the work of Winograd (1970) the results from Experiment 3 suggested that when participants have knowledge of search set size they are willing to search longer for the desired information. Thus, although Winograd examined retrieval from semantic memory on which participants had sufficient knowledge of the domain, the current results suggest that the same basic effects can be found in episodic memory tasks for items that were just presented. Importantly, the current results suggest that knowledge of search set size affects how long participants are willing to search throughout the recall period (time-to-last response) rather than affecting how long participants are willing to continue searching after the last retrieved item (exit latency). Thus, it seems that the type of response that participants end on, as well as their confidence in that response, is related to search termination decisions (exit latency), whereas knowledge of how many items one is searching for affects how long participants are willing to search throughout the entire recall period (time-to-last response).

13. General discussion

In the three experiments we examined search termination decisions in free recall. Specifically, we examined what types of responses participants ended their recall on, their confidence in those responses, and how these variables might influence exit latencies in free recall. We found that in standard delayed free recall (Experiments 1 and 3), participants tended to end their recall with correct response rather with an error. With uninhibited recall (Experiment 2) participants still primarily ended recall with a correct response. However, the tendency to end on an intrusion error increased suggesting that many times participants generate intrusions at the end of recall, but correctly recognize them as errors and do not output them. Indeed, in all experiments participants were quite good at identifying correct and error responses based on their confidence judgments.

In terms of exit latencies, we replicated prior work suggesting that exit latencies decrease as a function of the number of total items retrieved (e.g., Dougherty & Harbison, 2007; Harbison et al., 2009). Furthermore, we found that exit latencies decreased as a function of confidence in the last response given and this occurred not only for all responses, but for correct responses as well. We also found that exit latencies varied as a function of the type of response participants ended on as well as their confidence in those last responses. Specifically, we found that exit latencies were much longer when participants' last response was a correct item compared to when it was an intrusion. In line with these results, with uninhibited recall we found that exit latency varies as a function of the total number of intrusions recalled as well as a function of the number of successive intrusions recalled. The more intrusions that were recalled (either in total or successively) the shorter exit latencies were. Finally, we found that knowledge of overall search set size influenced participants' overall search time, but did not influence their exit latencies.

The current results provide important evidence that item type (correct vs. intrusion) and confidence in retrieved items influence search termination decisions. Although a number of well developed and quantitatively explicit models have been proposed to account for free recall performance and many of these models have explicit stopping rules (e.g., Anderson et al., 1998; Davelaar et al., 2005; Metcalfe & Murdock, 1981; Raaijmakers & Shiffrin, 1980; Sederberg et al., 2008); none of these models explicitly account for the possible influence of confidence assessments on search termination decisions. For instance, although the Search of Associative Memory (SAM; Raaijmakers & Shiffrin, 1980) theory specifically acknowledges the role of monitoring and evaluation processes, the quantitatively explicit version of the SAM model does not, as of yet, have a monitoring process that gauges the confidence of items, and then uses this information to influence search termination decisions (although see Sirotin et al., 2005 for a possible implementation). Furthermore, prior work has specifically examined stopping rules that rely on retrieval failures, but say little of how retrieval errors might influence search termination decisions. Future modeling work is needed to better examine the interplay of retrieval failures and retrieval errors in stopping decisions and how confidence may influence these decisions. Hopefully as more work is done examining search termination decisions in free recall, additional processes can be implemented in such models to give a better overall account of the dynamics of retrieval.

13.1. Limitations and future directions

One limitation of the current work is that it is not truly experimental in the sense that no real manipulations were conducted. Rather, the results of the current study are correlational demonstrating that exit latencies are related to response type and confidence in the last response given. Thus, causality cannot be inferred in terms of whether confidence in the last item retrieved compels individuals to terminate their search, however, the current study does indicate that the type of response and confidence in the last retrieved item is related to exit latencies. It is entirely likely that type of response, confidence, number of retrieved items, number of retrieval errors, number of non-recoverable items that are sampled, as well as other variables are taken together to inform search decisions by participants. Furthermore, given that the current study is correlational in nature, it is also possible that some other unmeasured variable is related to search termination decisions. Clearly future work is needed to see what experimental manipulations affect exit latencies overall and what experimental manipulations affect confidence which in turn might have an effect on exit latencies.

Another potential limitation of the current work is that the number of intrusions reported in each study was quite low. This was especially true for Experiments 1 and 3. Thus, some may question whether the exit latency data associated with intrusions is in fact stable given the overall low number of intrusions. However, the fact that the same pattern of results was found in three different experiments suggests that the results are quite stable despite the low number of intrusions. Furthermore, other work examining intrusions has suggested that despite the rarity of intrusions, intrusions tend to be quite systematic (e.g., Unsworth & Brewer, 2010). Thus, although the number of intrusions was low, which is consistent with many previous studies, the fact that the same pattern of results was found in all three experiments suggest that the results are, in fact, reliable. However, future work could further investigate these issues by utilizing procedures that are known to increase the number of intrusions in order to better examine the relation between intrusions and exit latency.

Finally, another important limitation of the current work is the fact that results could be biased by demand characteristics of the task. Specifically, having participants provide confidence responses after each recalled item might force participants to actually notice the confidence in the recalled items which then influences search termination decisions. It is possible that participants do not normally rely on confidence in deciding when to terminate search and that the relations found in the current study only arise when confidence judgments are required. Future work is needed to better examine if different relations arise when participants are instructed to provide confidence judgments and when they are not.

13.2. Conclusions

Collectively, the current results suggest that the type of response participants' end on as well as their confidence in the last response is related to exit latencies and, hence search termination decisions. These results suggest that an examination of response type and confidence can be fruitful in understanding search termination decisions and the dynamics of free recall more broadly. The current work combined with prior work on search termination (Dougherty & Harbison, 2007; Harbison et al., 2009; Kahana et al., 2010) should provide us with a better understanding of how and when individuals determine when to terminate memory search. However, as Winograd (1970) noted, the analysis of what variables are likely to influence search termination decisions has only just begun in earnest.

References

Anderson, J. R., Bothell, D., Lebiere, C., & Matessa, M. (1998). An integrated theory of list memory. *Journal of Memory and Language*, *38*, 341–380.

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

Bower, G. H., Clark, M. C., Lesgold, A. M., & Winzenz, D. (1969). Hierachical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior*, *8*, 323–343.

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.

Dougherty, M. R., & Harbison, J. I. (2007). Motivated to retrieve: How often are you willing to go back to the well when the well is dry. *Journal of Experimental Psychology: Learning, Memory & Cognition*, *33*, 1108–1117.

Dunlosky, J., & Metcalfe, J. (2009). *Metacognition.* Los Angeles: Sage.

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 and Instruments*, *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 and Verbal Behavior*, *15*, 227–233.

Harbison, J. I., Dougherty, M. R., Davelaar, E. J., & Fayyad, B. (2009). On the lawfulness of the decision to terminate memory search. *Cognition*, *111*, 146–421.

Harbison, J. I., Davelaar, E., & Dougherty, M. R. (2008). Stopping rules and memory search termination decisions. *Proceedings of the Cognitive Science Society Meeting*.

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.

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

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 differnces in editing of overt responses. *Journal of Gerontology: Psychological Sciences*, *60B*, 92–97.

Kahana, M. J., Miller, J., & Weidemann (2010). *Recall termination in free recall.* Manuscript submitted for publication.

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

Laming, D. (1999). Testing the idea of distinct storage mechanisms in memory. *International Journal of Psychology*, *34*, 419–426.

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

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

Nelson, T. O., & Narens, L. (1990). Metamemory: A theoretical framework and new findings. In G. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 125–173). San Diego: Academic Press.

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

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

Roediger, H. L., & 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*.

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.

Shiffrin, R. M. (1970a). Memory search. In D. A. Norman (Ed.), *Models of Human Memory* (pp. 375–447). New York: Academic Press.

Shiffrin, R. M. (1970b). Forgetting: Trace erosion or retrieval failure? *Science*, *168*, 1601–1603.

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, and 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 and Language*, *59*, 223–236.

Unsworth, N., & Brewer, G. A. (2010). Individual differences in false recall: A latent variable analysis. *Journal of Memory & Language*, *62*, 19–34.

Unsworth, N., Brewer, G. A., & Spillers, G. J. (2010). Understanding the dynamics of correct and error responses in free recall: Evidence from externalized free recall. *Memory & Cognition*, *38*, 419–430.

Winograd, E. (1970). Effect of knowledge of set size on search termination in long-term memory. *Psychonomic Science*, *20*, 225.