

Are the costs of directed forgetting due to failures of sampling or recovery? Exploring the dynamics of recall in list-method directed forgetting

Gregory J. Spillers · Nash Unsworth

Published online: 17 November 2010
© Psychonomic Society, Inc. 2010

Abstract This study investigated the costs of directed forgetting within the framework of a search model. In such models, retrieval failure can occur at either the sampling or the recovery stage of recall. Multiple measures of performance were employed to answer two primary questions of interest: (1) Where does the locus of forgetting occur in the directed forgetting paradigm? and (2) What current theory of directed forgetting can best account for the pattern of data observed? Converging evidence from these measures suggested that the costs of directed forgetting are the result of sampling issues during retrieval. Further, these results were best explained by a contextual-change account that argues an instruction to forget creates a break in context causing individuals to sample more items at retrieval. It is argued that the retrieval deficits observed in list-method directed forgetting paradigms are due, in part, to the fact that individuals sample a greater number of items (that include both targets and intrusions) during retrieval in the forget-condition, thereby creating greater response competition compared with retrieval in the remember-condition.

Keywords Directed-forgetting · Random-search model · Context-change account

Introduction

Few people would argue that forgetting is not an often-troublesome feature of the human mind, especially in the wake of vanishing car keys or a missed anniversary. It turns out, however, that this often pernicious occurrence is not

without its merits. In fact, the act of forgetting has provided researchers with an interesting (if not paradoxical) means of investigating and understanding human memory more broadly. There is even some consensus that forgetting information is actually a necessary requisite for successful remembering (Bjork, 1978, 1989). A popular technique for studying forgetting phenomenon is to try and manipulate it intentionally within an individual through the use of the directed forgetting paradigm (Bjork, LaBerge, & LeGrand 1968; Muther, 1965).

In the typical list-method directed forgetting paradigm, participants are given two lists of words to remember for a later free recall test. However, after presentation of the first list, participants are instructed to either try to forget that list under the guise that they will not be tested on it or to keep remembering it. A second list is then presented followed by a distractor task. Critically, participants are tested on both lists regardless of instruction. The primary finding being that first list recall performance is significantly reduced for those individuals told to forget relative to those instructed to keep remembering (termed the *costs* of directed forgetting; for reviews, see Bjork, 1998; Johnson, 1994; Macleod, 1998).

The specific mechanism underlying the effectiveness of a forget instruction in the directed forgetting paradigm has been hotly debated; the three most prominent accounts of the phenomenon being selective rehearsal, retrieval inhibition, and the contextual-change hypothesis. In the selective rehearsal account, forgetting is argued to occur because the presentation of a forget-cue elicits differential rehearsal of to-be-forgotten (TBF) List 1 items. That is, when told to forget, individuals stop rehearsing List 1 items and begin instead devoting rehearsal solely to the to-be-remembered (TBR) List 2 items, thereby reducing their ability to recall List 1 items at test compared with participants in the

G. J. Spillers (✉) · N. Unsworth
Department of Psychology, University of Georgia,
Athens, GA 30602, USA
e-mail: gjspille@uga.edu

remember-cue condition who continue to rehearse both List 1 and List 2 items to the same extent (Sheard & Macleod, 2005). On the other hand, proponents of retrieval inhibition argue that upon presentation of a forget-cue, all items encoded under the designated TBF list are actively suppressed or inhibited, thereby making them inaccessible at recall (Bjork, 1989; Geiselman, Bjork, & Fishman 1983). And finally, the context-change hypothesis argues that the presentation of a forget-instruction after List 1 prompts participants to actively induce a change in their mental context and, therefore, encode List 2 items in a different contextual state. At test, it is assumed that participants fail to recall List 1 items because they are unable to reinstate the original encoding context (Sahakyan & Kelley, 2002).

Although a wealth of convincing evidence for and against each of the above accounts has amassed over the last several years, a great majority of the work done on directed forgetting has been plagued by a reliance on only a single measure of performance, namely, proportion correct. Only recently have researchers begun to investigate directed forgetting phenomenon from a broader perspective. Malmberg and colleagues (Lehman & Malmberg, 2009; Malmberg, Lehman, & Sahakyan 2006), in particular, have begun to account for directed forgetting effects using a search-model framework. Amongst other things, such a framework is appealing for its use in theory constraining and because it makes predictions concerning a host of dependent measures.

The most basic search-model framework is the random-search model (Bousfield, Sedgewick, & Cohen 1954; McGill, 1963; Rohrer, 1996; Rohrer & Wixted, 1994; Wixted & Rohrer, 1994). In the random-search model, it is assumed that a retrieval cue (composed of temporal-contextual information generated by the participant) delimits a mental search set comprised of both target and extraneous items. Following delimitation of the search set, two distinct stages of retrieval are argued to occur in order for items to actually be recalled. First, items must be *sampled*. Sampling of items is assumed to follow a relative-strength rule, wherein the probability of sampling an item (k) is dependent upon its strength divided by the sum of the strengths of all items in the search set (e.g., $s_k / \sum s_i$). The strength of an item is determined by its relation with the retrieval cue being used to search. Further, this sampling process is assumed to occur at a constant rate, one item at a time and with replacement, such that after a target item has been sampled and recalled, it still has an equal chance of being selected on the next sample. Once an item has been sampled, it must be *recovered* into consciousness in order to be recalled (Raaijmakers & Shiffrin, 1980; Rohrer, 1996). In some models, the process of recovery involves a fixed threshold that items must exceed in order to be recalled (Rohrer, 1996). The probability of an item

exceeding this threshold and being recovered is contingent upon its absolute-strength to the cue regardless of the other items in the search set. Although all studied items should eventually be sampled, only those items whose strength exceeds the threshold can actually be recalled.

As the retrieval process proceeds, the probability of recalling a new target item necessarily decreases because each sample is likely to be an already recalled target item or an extraneous item. This means that as the recall period progresses, the rate of recall will begin to slow (i.e., more items are output during earlier stages of the recall period than later stages). When cumulative recall is plotted as a function of time, the resulting curve has been found to be well described by the cumulative exponential:

$$F(t) = N(1 - e^{-\lambda t}), \quad (1)$$

where $F(t)$ represents the cumulative number of items recalled by time t , N represents asymptotic recall (the estimated number of items that could be recalled given infinite time), and λ represents the rate of approach to asymptote (the rate in which N items are recalled).

Within the framework of the random search model, researchers have argued that recall latency (λ) estimates the number of items in the search set and thus reflects sampling (i.e., relative-strength) whereas recall total (N) estimates the number of target items in the search set whose absolute-strength exceeds some threshold and thus reflects recovery (i.e., absolute-strength). Important evidence for this distinction comes from a number of studies that have manipulated aspects of free recall and found that some variables affect recall total (i.e., absolute-strength), but have no effect on recall latency (i.e., relative-strength) and vice versa. For instance, Rohrer and Wixted (1994) manipulated presentation duration and found that this manipulation affected the number of items recalled (N), but had no effect on recall latency (λ). Thus, the absolute-strength of each item was affected by presentation duration, but because all items were equally incremented in absolute-strength, relative-strength remained unchanged (see also Shiffrin, 1970). In another experiment, Wixted and Rohrer (1993) examined the build-up and release of proactive interference. The authors found that as PI accrued, estimates of both N and λ decreased, suggesting that the search set increased for subsequent trials using the same category. However, in the release from PI condition, estimates of N and λ increased slightly. These results suggested that as PI accrued, the search set became progressively larger because all prior category instances were based on the same retrieval cue. Therefore, relative-strength was reduced leading to a drop in probability of recall and an overall increase in recall latency. Under release conditions, the retrieval cue specified only the new category instances, and thus the search set excluded items from the previous trials.

These and other results suggest that the random search model can be a useful tool in interpreting recall performance under a variety of conditions. Using recall latency and recall total in the current investigation permits a fine-grained analysis of the directed forgetting paradigm. Specifically, employing both measures of performance allows one to examine whether the locus of forgetting in the paradigm is in the sampling or the recovery stage of retrieval. Such a determination is important as it provides a more thorough description of directed forgetting from which competing accounts of the phenomenon can be contrasted and tested against.

The present study

The question of where the costs of directed forgetting originate is a pertinent one; therefore, the present study sought to examine the list-method directed forgetting paradigm within the framework of the random-search model. Multiple measures of performance were used in order to investigate whether forgetting in the paradigm occurs because items are no longer sampled at retrieval because of reductions in relative-strength or because the items are degraded such that their absolute-strengths no longer exceed a threshold for recovery. Forgetting at either stage should exhibit a distinct pattern of results. Below we detail specific predictions concerning how an instruction to forget can affect the course of retrieval and provide various ways in which the three dominant theories of directed forgetting *may* possibly account for the results.

A reduction in relative-strength could manifest itself in a number of ways differing primarily in how participants' latency to recall is affected. One possibility is that an instruction to forget affects the size of the search set from which participants sample from. On the one hand, the search set of individuals in the forget-cue condition could be smaller such that when they try to recall the to-be-forgotten list, they focus on and sample only a portion of the items on the list. If this were true, we would expect to see reduced recall totals and *shorter* recall latencies compared with participants in the remember-cue condition. On the other hand, the search set could become larger such that when individuals in the forget-condition try to recall the to-be-forgotten list, they sample both the to-be-forgotten items *and* extraneous items from the intervening to-be-remembered list. This would result in reduced recall totals, increased intrusions, and *longer* recall latencies compared with remember-cue individuals. Notably, the contextual-change account of directed forgetting can be interpreted to explain both sampling possibilities: the change in context either prohibits participants in the forget-condition from using efficient enough cues to sample the entire to-be-

forgotten list or distorts their ability to use more specific cues, forcing them to include items from both the to-be-forgotten list and the succeeding to-be-remembered list in their search-set (see Mensink & Raaijmakers, 1988 for a discussion on the various ways contextual change can affect recall). The retrieval inhibition account is suited to explain how an individual's search set becomes smaller after a forget-instruction as well; that is, because the individual has inhibited the to-be-forgotten list, it follows that the items could be deleted from the search set and are, thereby, incapable of being sampled. At present, it is unclear how a selective-rehearsal account could explain forgetting because of a reduction in relative-strength.

In contrast, if directed forgetting were a problem of reductions in absolute-strength then one would expect to find recall totals affected by a forget-cue but *not* recall latencies. The contextual-change hypothesis could account for such results if one assumes that all items on the to-be-forgotten list become less associated with the retrieval context after a context change, thus affecting their absolute-strengths but not their relative-strengths (Mensink & Raaijmakers, 1988). By the same token, from an inhibitory account, it is possible to argue that all items on the to-be-forgotten list are weakened due to inhibition, leaving their relative-strength unchanged, but reducing their absolute-strength such that they can no longer exceed the recovery threshold. Finally, selective-rehearsal accounts may argue that given differential rehearsal of the to-be-forgotten list, items are less strong compared with items in the remember-cue condition that were allotted more extensive rehearsal. These different possibilities were examined in the current study via a collective examination of cumulative recall functions, proportion correct, recall latency, and intrusion errors.

Method

Participants

Fifty undergraduate students were recruited from the participant pool at the University of Georgia. Each participant was individually tested and was randomly assigned to either the remember-instruction condition ($N = 25$) or the forget-instruction condition ($N = 25$).

Design

The list-method of directed-forgetting paradigm was used with cue (remember vs. forget) manipulated between subjects. Both cue conditions consisted of two lists of 15 nouns each. Words were presented for 2 s each. The distractor task used required participants to order three-digit

numbers from largest to smallest (e.g. 538 would be correctly ordered as 853). The numbers were presented on the computer for 2 s each while participants recorded their responses on paper.

Procedure

Participants in both the remember- and forget-instruction conditions were given two lists of words to remember for a later test. Following the presentation of List 1, however, participants were given one of two instructions, either to keep remembering List 1 or to try and forget List 1. Participants in the remember-instruction condition were shown: “You have just been shown List 1. Now you will be shown List 2. You will then be tested on your memory for both lists.” Participants in the forget-instruction condition were shown: “Please try and forget the list that you just studied as you will not be tested on it. You will only be tested on the next list.” Following presentation of List 2, both groups completed the 30 s distractor task, followed by a separate 60 s recall test first for List 1 and then for List 2. Participants were required to type in their responses. For each recalled item the computer logged when in the recall period the item was output.

After completing the experimental session, participants in the forget-instruction condition completed a manipulation check questionnaire explicitly asking whether they attempted to forget when instructed to or if they intentionally tried to remember forget items.¹

Results

The results of the experiment are divided into three primary sections for analyses devoted to cumulative recall functions, accuracy, and recall latency associated with each list. Accuracy is further divided into three subsections: proportion correct, intrusion rates, and first-response accuracies. Recall latency is divided into two subsections: overall latency and first accurate response latency. Planned comparisons were conducted on List 1 performance in order to derive directed forgetting costs and on List 2 performance in order to derive directed forgetting benefits.

Cumulative recall functions

Overall results in the form of the cumulative number of correct items recalled as a function of time for each list are

¹ Compliance with the forget instruction was very high. No participants reported explicitly trying to remember the words they were asked to forget; however, three reported not doing anything in particular to forget. These participants were kept in the analyses.

shown in Fig. 1. For each function, correct responses were first placed into 60 1-s bins, and then the cumulative number of items recalled for each bin was computed. An examination of these cumulative recall functions suggests a number of important findings. First, as shown in Fig. 1a, participants in the remember-condition appear to have recalled more items correctly during the test of List 1 compared with participants in the forget-condition (i.e., higher asymptotic levels, N). Additionally, Fig. 1a also shows that participants in the forget-condition tended to reach asymptotic levels at a slower rate ($\lambda = 0.012$) during retrieval of List 1 than participants in the remember-condition ($\lambda = 0.033$). Finally, participants in the forget-condition appear to pause slightly before recall of correct items from List 1 whereas participants in the remember-condition do not show this tendency.

Careful examination of Fig. 1b suggests that individuals tended to recall List 2 items in a similar fashion regardless of their experimental condition. Indeed, participants in both conditions appear to recall a comparable number of items correctly and do so at a similar rate (forget-condition: $\lambda = 0.036$; remember-condition: $\lambda = 0.05$). The only noticeable difference appears at the beginning of the recall period, when participants in the forget-condition show a pronounced pause before outputting a correct response, akin to the one observed during recall of List 1 (see Fig. 1a). In

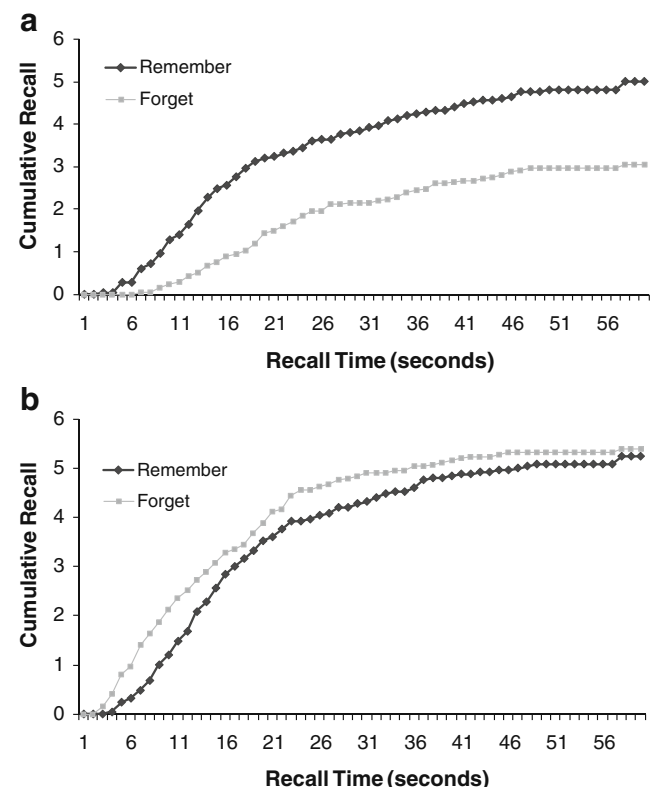


Fig. 1 Cumulative recall functions for List 1 (a) and List 2 (b) by cue condition. Functions are computed for correct responses only

the sections to follow are more fine-grained analyses confirming these initial observations.

Accuracy

Proportion correct As shown in Fig. 2, the classic directed forgetting costs were obtained in that participants in the forget-condition recalled significantly fewer List 1 items correctly than participants in the remember-condition, $F(1, 48) = 8.79$, $MSE = .18$, $p < .01$, $\eta_p^2 = .15$. Confirming the observations noted earlier in Fig. 1b, however, directed forgetting benefits were not found as the two conditions did not significantly differ in the number of items they correctly recalled from List 2 ($p > .65$).

Intrusion rates Intrusions, although notoriously rare, are nonetheless a useful indicator of how individuals are sampling items during retrieval. That is, they provide information about the types of cues (and the specificity of cues) participants are using to retrieve items. Therefore, in addition to examining proportion correct, intra-experimental intrusion errors (items presented in the experiment but belonging to other lists) output during recall of each list were tallied. As shown in Table 1, participants in the forget-condition recalled significantly more intrusions (items presented on List 2) during recall of List 1 compared with participants in the remember-condition, $F(1, 48) = 6.71$, $MSE = 21.78$, $p < 0.05$, $\eta_p^2 = .12$. These results suggest that the increase in intrusions during recall of List 1 is the result of participants using broad cues that encompass both Lists 1 and 2. The number of intrusions output during recall of List 2 (i.e., items originally presented on List 1) was also examined between the two conditions, however, this analysis failed to reach significance ($p > .87$).

First-response accuracies First-response accuracies give an indication of those responses that theoretically have the strongest overlap with the cues being used during retrieval.

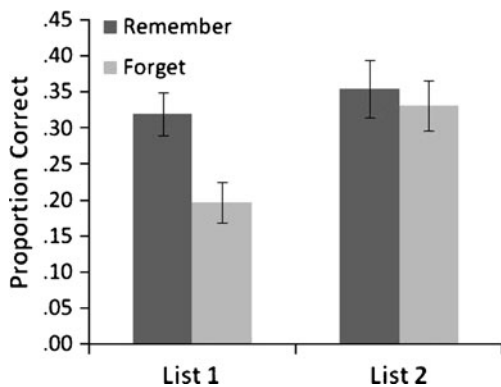


Fig. 2 Proportion correct for list as a function of condition. Error bars represent one standard error of the mean

Table 1 Mean number of intrusions for each list as a function of instruction condition

Measure	List 1		List 2	
	Remember	Forget	Remember	Forget
Intrusions	0.52 (0.17)	1.84 (0.47)	0.64 (0.18)	0.68 (0.17)

Numbers in parentheses represent one standard error of the mean

Given that the majority of intrusions output are from lists directly preceding the recall period, it would be interesting to see how often these items are output first. Therefore, we examined which type of response (correct vs. intrusion) participants started their recall with as a function of list. During the test of List 1, 92% of the participants in the remember-condition started their recall with a correct response compared with only 60% of the participants in the forget-condition. During the test of List 2, however, a similar proportion of participants in each condition began recall with a correct response, 88% in the remember-condition and 84% in the forget-condition. These results suggest that participants in the remember-condition are likely utilizing cues more specific to List 1 given that nearly all of the participants in that condition began their recall with List 1 items (while intrusions from List 2 were scarce). In contrast, participants told to forget List 1 appear plagued by intrusions from List 2 given that participants began with an error over a third of the time. Given the relative proximity of List 2 to the recall period in both conditions, it is not surprising that both groups of participants are able to begin recall with a correct response most of the time.

Recall latencies

Average recall latency Next, we analyzed average recall latency as a function of list. Here recall latency refers to the average time point in the recall period when a given response (both correct and intrusion error) was emitted. Thus, if responses were emitted 5, 10, and 15 s into the recall period, mean recall latency would be 10 s. As discussed, recall latency provides a theoretically sound means of measuring the search set size of participants. One would expect that if the forget-cue disrupts the sampling stage of recall, participants in the forget-condition would have significantly different recall latencies for List 1 compared with the participants in the remember-condition. If directed-forgetting is a recovery problem, however, latencies between the two conditions should not differ. As shown in Table 2, average recall latency during List 1 was much longer for participants in the forget-condition compared with those in the remember-condition, $F(1, 45) = 5.23$, $MSE = 2.69$, $p < .05$, $\eta_p^2 = .10$, providing further evidence for the notion that the forget-cue somehow

Table 2 Average and first accurate response latencies (in seconds) as a function of list

Measure	List 1		List 2	
	Remember	Forget	Remember	Forget
Recall latency	18.32 (1.50)	23.11 (1.44)	15.19 (1.24)	17.79 (1.41)
First accurate response latency	9.85 (1.43)	14.60 (1.29)	5.35 (0.65)	8.90 (1.18)

Numbers in parentheses represent one standard error of the mean

disrupts retrieval such that participants told to forget include more items in their search sets compared with those told to remember. Average recall latency during List 2 did not differ as a function of condition ($p > .17$).

First accurate response latency Given the high frequency in which participants in the forget-condition began their recall of List 1 with an intrusion, it may be informative to examine how long on average participants took to finally output a correct response. Therefore, we examined the average latency to the first correct response for each list. As shown in Table 2, participants in the forget-condition took much longer to output a correct response during recall of List 1 compared with participants in the remember-condition, $F(1, 36) = 5.29$, $MSE = 2.04$, $p < .05$, $\eta_p^2 = 0.12$. This result coupled with the first-response accuracies computed above suggest that participants in the forget-condition are likely having to sample and cycle through intrusions (presumably from List 2) as they begin their recall of List 1 items. Participants in the forget-condition took significantly longer to output a correct response compared with participants in the remember-condition during recall of List 2 as well, $F(1, 41) = 7.07$, $MSE = 1.35$, $p < .05$, $\eta_p^2 = .14$.

Discussion

This experiment examined the list-method of directed forgetting within the framework of the random-search model in order to determine the locus of forgetting in the paradigm. It was found that when participants were asked to recall a list of items they were instructed to forget earlier, they recalled significantly fewer items, made more intra-experimental intrusions, were more likely to begin recall with an error and recalled at a much slower rate compared to participants instructed to remember the same list of items. Collectively, these results reflect a number of important and interesting findings.

First and foremost, the current results strongly suggest that the forgetting observed in the list-method directed forgetting paradigm results primarily from problems in the sampling stage of retrieval. As discussed earlier, recall latency, the average time it takes to output items during recall, has been argued to be a reliable measure of search set size, such that the larger the search set a participant

searches through the longer their latency to recall items. Therefore, the prolonged mean recall latency values observed during the recall of List 1 by participants in the forget-condition coupled with the increase in intra-experimental intrusions observed are indicative of participants having to search through a larger search set compared with participants in the remember-condition. Such an increase in search set size lowers the probability of any one item being sampled and recalled. Therefore, the presence of a forget-cue affects the amount of competition between items which in turn affects their overall probability of being sampled (i.e., their relative-strength), rather than affecting an individual item's absolute-strength.

How does the size of the search set increase during recall in the forget-cue condition? It appears that the introduction of the forget-cue, through some mechanism to be discussed shortly, allows for the accrual of retroactive interference from items presented on List 2. That is, when tested in the forget-cue condition, participants are searching through a set of items composed of both targets from List 1 as well as intrusions from List 2, compared with the remember-cue condition in which they seem focused entirely on List 1. Evidence consistent with this notion comes from the intrusion analyses already discussed in which participants in the forget-condition output more intrusions (items from List 2) during recall of List 1 than those individuals in the remember-condition. Although not statistically significant, other researchers have also reported evidence of increased intrusions from the intervening list in forget-cue conditions (see Pastötter & Bäuml, 2007).

To make sense of these results, it is important to consider the assumptions of the random-search model described earlier. Recall that in these models, it is assumed that retrieval of episodic representations from memory is governed by the use of temporal-contextual cues available at the time of testing. The likelihood of retrieving a target representation depends on the relative overlap between the retrieval cues and the contextual elements of the item originally encoded. In reference to the current results, the contextual cues available during the recall period presumably overlap to a greater extent with the intervening list (List 2) given its proximity to the test compared to the to-be-forgotten list (List 1), thereby making inclusion of List 2 items in the search for forget items almost inevitable. By including these intruding items, the probability of sampling any to-be-forgotten item is substantially reduced due to increased

competition between representations in participants' search sets (the relative-strength rule). Further, it becomes increasingly more difficult to recall List 1 items the more List 2 items are sampled and recovered because of contextual overlap between items *within* that list. More complex instantiations of the random-search model like the Search of Associative Memory (SAM; Raaijmakers & Shiffrin, 1980) argue that once an item has been sampled it is used in conjunction with context information to aid in recall of further items; therefore, items that co-occur within a list are more likely to cue each other than items presented between lists. Evidence for this notion was observed in the current results where recall of List 1 often began with List 2 items being output first before any List 1 items were emitted. One can reasonably infer that only after several retrieval attempts using the recency context and recovering List 2 items was information retrieved that was sufficient to recall a target List 1 item. Once this item was output it was used to further probe for List 1 items, i.e., the observed recall of some List 1 items later in the recall period.

Although the pattern of results points to a reduction in relative-strength as the primary source of forgetting in the paradigm, the question of which existing theory of directed forgetting can best account for these results remains. A review of the predictions outlined in the introduction, however, leaves only one account that can clearly explain an increase in search set size after a forget-cue without any additional assumptions, namely, the contextual-change hypothesis. Specifically, the contextual-change hypothesis makes the unique prediction that the presence of a forget-cue creates a distinct break in context such that participants can no longer use cues that are diagnostic of only List 1 but overlap to a certain extent with List 2 as well. Note that this interpretation differs from the argument that participants simply cannot reinstate any aspect of List 1, which would predict floor levels of recall and shorter recall latencies. When there is no instruction to forget, as in the remember-cue condition, participants should have little trouble focusing directly on the list being tested, with little interference from the intervening list. One could speculate that the break in context caused by the forget-cue may serve to further exacerbate the effects of retroactive interference compared with a remember instruction. That is, a distinct change in context in the forget-condition acts as a catalyst for retroactive interference to accrue more rapidly compared to circumstances where context remains invariable. Recent work by Jang and Huber (2008) using the list-before-last paradigm provides supporting evidence for this notion.

In their paper, Jang and Huber used the list-before-last paradigm to examine the effect of retrieval on list-isolation and interference in free recall. Parallel in many respects to the design of list-method directed forgetting, the list-before-last paradigm requires participants to study two lists of

words before being instructed to recall the first of the two presented (the list before last). They argued that having a recall test in between encoding of the two lists would sufficiently isolate the lists contextually from each other such that at recall there would be little interference from the intervening list as participants attempted to recall items from the list before last. Interestingly, however, they found that intruding recall from the intervening list was still prominent despite evidence that the lists were contextually isolated. In fact, recall for lists not contextually isolated from each other was better overall and less affected by intervening list interference than lists that were contextually dissimilar, results commensurate with those found in the current study. The authors concluded that “there are two routes for recall of an item from the target list; either the target list context is properly reinstated, thus filtering out the intervening list, *or context is not reinstated but the current context is sufficiently similar to allow retrieval of target list items (as well as retrieval of intervening list items)*” (emphasis added, Jang & Huber, 2008, p. 116). Simply put, when individuals have contextually isolated two lists, as the contextual-change hypothesis assumes, unless they can retrieve (or reinstate) the target list context they will have to rely on the default recency context to conduct their search. The consequence of a change in context in directed forgetting experiments is that the target list is much harder to reinstate, therefore, use of the recency context is almost inevitable. By defaulting to the recency context, participants search with retrieval cues too inefficient to focus their search solely on the target list; therefore, they sample a large set of items that include a mix of both target and intruding items. In cases when there is no break in context (such as in the remember-cue condition), the context at test provides cues that allow participants to readily reinstate the target list, reducing their susceptibility to retroactive-interference and promoting superior recall performance.

Alternative explanations and limitations

An obvious critique of the current paper is that although the contextual-change account is consistent with the results this does not preclude explanations from the other theories. Indeed, more complex instantiations of inhibition and selective-rehearsal could possibly account for the results observed in the current study. However, if one were to use only the most basic of assumptions of these theories, as we have, they fall short of being able to predict the crux of the current results—that an instruction to forget causes an increase in the search-set size because of interference from the intervening list—compared with the contextual-change account. These shortcomings are likely tied to under-

specification of these theories to begin with and although post-hoc additions can serve to alter their explanatory power, these theories lose significant validity until they can make clear *a priori* predictions in the future about dependent variables beyond simply proportion correct. It is our hope that the current study will aid directed forgetting researchers in this regard by providing data across a number of variables that should serve to help constrain their theories.

Another possible objection to the ideas just discussed is that list-method directed forgetting could be rooted in both a reduction in relative-strength and absolute-strength. Indeed, the clever reader could easily figure that the two are not completely independent, and they would not be necessarily wrong. However, what is harder to refute is the fact that if the two conditions (remember vs. forget) did not differ in the size of the mental set of items they are theoretically searching through they would exhibit similar recall latencies. So although, within the forget-condition only, items could differ in absolute-strength, with the hypothesis that List 1 items are weaker than those on List 2, this would not affect their latency to be recalled but merely their probability of being recalled. With an eye towards parsimony and the present results being insufficient for any serious determination on the issue, further research is necessary to fully examine whether absolute-strength is actually altered by an instruction to forget. What cannot be mistaken is that relative-strength is affected during forget-conditions.

We would also be remiss not to discuss the fact that the intrusion data we have presented in the current paper differs from data reported by other researchers (cf. Lehman & Malmberg, 2009; Sahakyan & Delaney, 2010). Indeed, of the few studies that have examined intrusions in directed forgetting, many have reported null findings. While we cannot fully explain the significant effect of intrusions in the current study while in others they have been absent, we stand by our belief that the contextual-change account is not altogether inconsistent with the current data. The reason for the discrepancy, therefore, may be one of methodological differences between the current study and those reporting null effects of intrusions. Such differences could include (but are not limited to) longer presentation durations and the inclusion of lists prior to the target list to create proactive-interference (see Lehman & Malmberg, 2009; Sahakyan & Delaney, 2010). Presumably, the longer an individual has to rehearse an item, the more contextual information they encode, thus providing them at retrieval with a coherent contextual backdrop from which to make source-discrimination judgments about that item (Johnson, Hashtroudi, & Lindsay 1993). Therefore, one would expect fewer cross-list commission errors with increased presentation durations (but note this does not eliminate the

possibility that these items are still sampled). With the inclusion of prior lists, it is possible that you create a situation in which individuals, when attempting to retrieve the target list, use cues that overlap contextually with prior presented lists. Thus, intrusions and competition can arise from either the intervening list or lists presented prior to the target list depending upon the contextual-cues being used. Whatever the case may be, the very existence of conflicting data makes the question of intrusions an important one. For that reason, directed-forgetting researchers should continue the recent trend of examining intrusion errors so that an answer to this dilemma may no longer elude us.

One final point worth mentioning is that we have neglected to discuss the benefits of directed-forgetting in any great detail. This is simply because we failed to find any substantial evidence of benefits in the current study. Researchers have shown, however, that benefits are not a consistent finding in directed forgetting (Sahakyan, 2004) and further, are not directly tied to directed forgetting costs (Sahakyan & Delaney, 2005). Therefore, our present inability to find benefits should not substantially hinder our discussion of what drives costs.

Concluding remarks

The primary goal of the current study was to uncover the specific ways in which directed forgetting effects the process of retrieval. Converging evidence across multiple dependent measures suggests that the costs of directed forgetting are the result of reductions in relative-strength which in turn lead to sampling issues during retrieval. Specifically, the introduction of the forget-cue creates a break in context between the lists in the forget-condition. This break prevents participants from successfully reinstating the forget list at retrieval, forcing them to rely on the recency context to begin their search. Use of the recency context is inefficient for retrieving forget-list items but is also heavily matched to intervening list items. Therefore, participants search through a larger set of items (comprised of both targets and intrusions) reducing the probability of sampling and recalling target forget-list items. The implications of searching through a larger search set results in substantial changes in the dynamics of retrieval within the directed forgetting paradigm including a reduction in recall performance, an increase in intrusions, and an increase in recall latencies compared with the recall of participants in the remember-condition. Future work is necessary to fully understand the degree to which a forget-cue affects the contextual state at encoding and retrieval; nevertheless, directed forgetting researchers should continue to capitalize on the information that multiple performance measures can provide.

Acknowledgments We thank Gene Brewer, Miranda Hayworth, Ken Malmberg, Jeroen Raaijmakers, and an anonymous reviewer for comments on an earlier version of this article.

References

- Bjork, R. A. (1978). The updating of human memory. In G. H. Bower (Ed.), *The psychology of learning and motivation, Vol. 12* (pp. 235–259). New York: Academic Press.
- Bjork, R. A. (1989). Retrieval inhibition as an adaptive mechanism in human memory. In H. L. Roediger III & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honor of Endel Tulving* (pp. 309–330). Hillsdale: Erlbaum.
- Bjork, R. A. (1998). Intentional forgetting in perspective: Comments, conjectures, and some directed remembering. In J. M. Golding & C. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 453–481). Hillsdale: Erlbaum.
- Bjork, R. A., LaBerge, D., & Legrand, R. (1968). The modification of short-term memory through instructions to forget. *Psychonomic Science, 10*, 55–56.
- Bousfield, W. A., Sedgewick, C. H., & Cohen, B. W. (1954). Certain temporal characteristics of the recall of verbal associates. *American Journal of Psychology, 67*, 111–118.
- Geiselman, R. E., Bjork, R. A., & Fishman, D. L. (1983). Disrupted retrieval in directed forgetting: A link with posthypnotic amnesia. *Journal of Experimental Psychology: General, 112*, 58–72.
- Jang, Y., & Huber, D. E. (2008). Context retrieval and context change in free recall: Recalling from long-term drives list isolation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*, 112–127.
- Johnson, H. (1994). Processes of successful intentional forgetting. *Psychological Bulletin, 116*, 274–292.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3–28.
- Lehman, M., & Malmberg, K. J. (2009). A global theory of remembering and forgetting from multiple lists. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 35*, 970–988.
- MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1–57). Mahwah: Erlbaum.
- Malmberg, K.J., Lehman, M., & Sahakyan, L. (2006). *On the cost and benefit of taking it out of context: Modeling the inhibition associated with directed forgetting*. Proceedings of the 28th Meeting of the Cognitive Science Society, 549–554.
- McGill, W. J. (1963). Stochastic latency mechanism. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology, Vol. 1* (pp. 309–360). New York: Wiley.
- Mensink, G., & Raaijmakers, J. G. W. (1988). A model for interference and forgetting. *Psychological Review, 95*, 434–455.
- Muther, W. S. (1965). Erasure or partitioning in short-term memory. *Psychonomic Science, 3*, 429–430.
- Pastötter, B., & Bäuml, K.-H. (2007). The crucial role of postcue encoding in directed forgetting and context-dependent forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 33*, 977–982.
- Raaijmakers, J. G. W., & Shiffrin, R. M. (1980). SAM: A theory of probabilistic search in associative memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, pp. 207–262). New York: Academic.
- Rohrer, D. (1996). On the relative and absolute-strength of a memory trace. *Memory & Cognition, 24*, 188–201.
- Rohrer, D., & Wixted, J. T. (1994). An analysis of latency and interresponse time in free recall. *Memory & Cognition, 22*, 511–524.
- Sahakyan, L. (2004). Destructive effects of “forget” instructions. *Psychonomic Bulletin & Review, 11*, 555–5559.
- Sahakyan, L., & Delaney, P. F. (2005). Directed forgetting in incidental learning and recognition testing: Support for a two-factor account. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 31*, 789–801.
- Sahakyan, L., & Delaney, P. F. (2010). Item-specific encoding produces an additional benefit of directed forgetting: Evidence from intrusion errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 36*, 1346–1354.
- Sahakyan, L., & Kelley, C. M. (2002). A contextual change account of the directed forgetting effect. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 1064–1072.
- Sheard, E. D., & MacLeod, C. M. (2005). List method directed forgetting: Return of the selective rehearsal account. In N. Ohta, C. M. MacLeod, & B. Uttl (Eds.), *Dynamic cognitive processes* (pp. 219–248). Tokyo: Springer-Verlag.
- Shiffrin, R. M. (1970). Memory search. In D. A. Norman (Ed.), *Models of human memory* (pp. 375–447). New York: Academic Press.
- Wixted, J. T., & Rohrer, D. (1993). Proactive interference and the dynamics of retrieval. *Journal of Experimental Psychology: Learning, Memory, & Cognition, 19*, 1024–1039.
- Wixted, J. T., & Rohrer, D. (1994). Analyzing the dynamics of free recall: An integrative review of the empirical literature. *Psychonomic Bulletin & Review, 1*, 89–106.