Journal of Experimental Psychology:
Learning, Memory, and Cognition
2005, Vol. 31, No. 4, 789–801

0278-7393/05/$12.00 DOI: 10.1037/0278-7393.31.4.789

OBSERVATION

Directed Forgetting in Incidental Learning and Recognition Testing: Support for a Two-Factor Account

Lili Sahakyan and Peter F. Delaney
University of Florida

Instructing people to forget a list of items often leads to better recall of subsequently studied lists (known as the *benefits* of directed forgetting). The authors have proposed that changes in study strategy are a central cause of the benefits (L. Sahakyan & P. F. Delaney, 2003). The authors address 2 results from the literature that are inconsistent with their strategy-based explanation: (a) the presence of benefits under incidental learning conditions and (b) the absence of benefits in recognition testing. Experiment 1 showed that incidental learning attenuated the benefits compared with intentional learning, as expected if a change of study strategy causes the benefits. Experiment 2 demonstrated benefits using recognition testing, albeit only when longer lists were used. Memory for source in directed forgetting was also explored using multinomial modeling. Results are discussed in terms of a 2-factor account of directed forgetting.

Key words: directed forgetting, intentional forgetting, study strategies, multinomial modeling, inhibition

When people find out that they are talking to a memory specialist, they often ask about how to improve their memories and how to remember more effectively. The irony of this is that people should perhaps be interested in how to forget things better in order to improve memory, as it appears that forgetting often leads to better memory (R. A. Bjork, 1989).

One way to study the linkage between forgetting and learning is to ask people to forget something they have just studied—a procedure known as *directed forgetting*. A forget instruction can be delivered either after each item (known as the *item method*) or after a block of items (known as the *list method*). These two methodologies were shown to have different underlying mechanisms, with the item method reflecting encoding phenomena (better encoding of forget vs. remember items), and the list method most likely reflecting retrieval phenomena (e.g., Basden, Basden, & Gargano, 1993; R. A. Bjork, 1989). The current article focuses on the list method of directed forgetting because the mechanism underlying this phenomenon has created a variety of theoretical viewpoints as opposed to the mechanism supporting the item method. (For a more complete review of directed forgetting and related procedures such as specific intentional forgetting, see H. Johnson, 1994, or MacLeod, 1998.)

A typical list method directed forgetting study presents participants with two word lists to study. Between administration of the two lists, the experimenter instructs half of the participants to forget the first list and the remaining half of the participants to keep remembering the words. After studying the second list, participants are asked to recall all the items, including any items they were earlier instructed to forget (if applicable).

What is most commonly found from this procedure is that there are *costs* of directed forgetting—that is, the forget group participants recall fewer items from the first list compared with the remember group participants. Also, there are *benefits* of directed forgetting such that the forget group tends to recall more items from the second list than the corresponding remember group does. In the context of directed forgetting, the terms *costs* and *benefits* are used to describe List 1 (costs) and List 2 (benefits) recall performance of participants that receive forget instructions relative to participants that receive remember instructions (rather than vice versa) because prototypical list-learning studies imply a remember instruction. Most directed forgetting theories treat the costs and benefits as arising from a single underlying process, although recently we have proposed a two-factor account with different mechanisms for the costs and the benefits (Sahakyan & Delaney, 2003; Sahakyan, Delaney, & Kelley, 2004).

The selective rehearsal hypothesis was the first explanation provided for these phenomena. R. A. Bjork (1970, 1972) proposed that after hearing the forget instruction, participants stop rehearsing List 1 items and devote all their mnemonic and rehearsal abilities to List 2, whereas participants in the remember group continue rehearsing List 1 items during List 2 study. List 1 items therefore receive less cumulative rehearsal in the forget group,

---

Lili Sahakyan and Peter F. Delaney, Department of Psychology, University of Florida.

Peter F. Delaney is now at the Department of Psychology, University of North Carolina at Greensboro.

We are grateful to Xiangen Hu and William Batchelder for their help with the GPT.EXE software and multinomial models in general. Thanks also go to Kathryn Hughes, Kristin Kukharenko, Ryan Lobo, and Isabel Manzano for their help in data collection and scoring.

Correspondence concerning this article should be addressed to Lili Sahakyan, who is now at the Department of Psychology, University of North Carolina at Greensboro, 296 Eberhart Building, P.O. Box 26170, Greensboro, NC 27402-6170. E-mail: sahakyan@cas.usf.edu

leading to the costs of directed forgetting. As a result, List 1 items create less proactive interference on List 2 in the forget group, leading to the benefits of directed forgetting, which is often attributed to the escape from proactive interference (e.g., E. L. Bjork & Bjork, 1996; R. A. Bjork & Woodward, 1973).

The most compelling evidence challenging the selective rehearsal hypothesis was presented by Geiselman, Bjork, and Fishman (1983). In that study, participants were presented two types of items intermixed on each list. Half of the items were to be learned for a later memory test (intentional items), whereas half of the items were to be judged for pleasantness only (incidental items). Participants were specifically told not to memorize the incidental items and were told that two studies were being conducted simultaneously, one exploring how memorable certain words were, and the other exploring how pleasant the words were. The results revealed directed forgetting costs and benefits for both item types. Because the incidental items were not supposed to be memorized, they likely should not have been rehearsed. Consequently, ceasing rehearsal after List 1 in the forget condition should have impaired the recall of intentional items but not the incidental items. However, comparable directed forgetting effects were found for both items types, albeit at lower recall rates for incidental items. The findings of Geiselman et al. led to the retrieval inhibition account of directed forgetting (R. A. Bjork, 1989). According to this view, the forget instruction initiates a process that at the time of retrieval blocks or inhibits access to the first list of items.

The process by which retrieval inhibition takes place is not yet clear and requires further investigation. Various accounts differ considerably in their proposals for the underlying mechanisms supporting retrieval inhibition. For example, Sahakyan and Kelley (2002) proposed that upon receiving the forget cue, participants in the forget condition attempt to establish a new mental context to comply with the instruction. They encode the second list of items within that new mental context, which at the time of test mismatches the learning context for the first list. At the time of retrieval, the mismatch between the study and the test context leads to forgetting. Another retrieval inhibition mechanism was proposed by Basden, Basden, and Morales (2003), who suggested that the forget cue disrupts the retrieval strategy for the first list of items, whereas the remember cue mediates a retrieval strategy that favors both lists instead of just the last list.

In recent articles, we have argued that the directed forgetting phenomenon consists of two separate components having different underlying mechanisms: the costs (impaired List 1 recall) and the benefits (enhanced List 2 recall; Sahakyan & Delaney, 2003; Sahakyan et al., 2004). According to our view, the costs emerge from the context change mechanism proposed by Sahakyan and Kelley (2002). However, we proposed that the benefits emerge because the forget cue increases the likelihood that participants will adopt a better study strategy on List 2 compared with the remember cue (Sahakyan & Delaney, 2003). The evidence for the strategy change mechanism came from retrospective verbal reports, which we further validated by statistically and experimentally controlling study strategy. When participants were not free to choose an encoding strategy but rather were instructed to encode both lists using the same study strategy, they showed significant costs but no benefits (Sahakyan & Delaney, 2003). Our subsequent work suggested that strategy changes in the forget group are mediated by self-induced evaluation of the current study strategy (Sahakyan et al., 2004). For example, in one of the experiments, we asked the participants to recall List 1 after studying it but before they heard the midlist forget or remember instruction. We obtained no benefits on the final recall test (despite the significant costs) because the remember group participants performed as well as the forget group participants on List 2 (additional evidence for strategy changes with repeated testing can be found in Delaney & Knowles, 2005). In a subsequent experiment, before hearing the forget or the remember instruction, all participants were asked to predict how many words they would be likely to recall on a final recall test (Sahakyan et al., 2004). This experiment also revealed a similar pattern of findings—significant costs, but no benefits. We proposed that the forget group participants are more likely to spontaneously evaluate their performance and consequently change their study strategy, presumably because they perceive the first list as a practice list.

Several criticisms can be brought against the strategy change account of directed forgetting benefits. The influential, widely cited work by Geiselman et al. (1983) that was problematic for the selective rehearsal account poses similar complications for an account of benefits that invokes a purely encoding-based mechanism. Given that Geiselman and colleagues reported directed forgetting effects with incidental items, it is unclear why participants would adopt a better study strategy on List 2 for items they were not supposed to commit to memory.

Because Geiselman et al. (1983) strictly alternated intentional and incidental items on the same list, a pure assessment of directed forgetting mechanisms under incidental learning conditions could be complicated. Perhaps adjacent incidental and intentional items could have become integrated. For example, frequent switching between the two types of instructions (learning vs. judging pleasantness) may have increased the costs of fully complying with the instructions and thereby prompted some participants to integrate both types of items. Consequently, the List 2 benefits that emerged for incidental items in Geiselman et al.'s study could be due to inadvertent better encoding of those items because they were included with the intentional items.

A test of directed forgetting with incidental items would require replicating Geiselman et al.'s (1983) study using a between-subjects rather than a within-subjects approach to manipulate intentionality of learning because it would eliminate any potential interitem integration. A variation on Geiselman et al.'s study has been presented as a poster (Kimball & Metcalfe, 2001), but as of yet there are no published studies replicating the important findings that prompted the retrieval inhibition explanation of directed forgetting. Finding directed forgetting benefits on List 2 in the absence of intentions to learn the items would pose a difficulty for Sahakyan and Delaney's (2003) strategy-based account of benefits. This goal was addressed in Experiment 1.

A second criticism of the strategy change account of benefits can be traced to earlier research reporting the absence of directed forgetting effects in recognition testing—a finding also reported by Geiselman et al. (1983) as well as others (Basden & Basden, 1996; Basden et al., 1993; Gross, Barresi, & Smith, 1970; MacLeod, 1999; Schmitter-Edgecombe, Marks, Wright, & Ventura, 2004; Whetstone, Cross, & Whetstone, 1996; but see E. L. Bjork & Bjork, 2003). The absence of directed forgetting costs in recognition testing is not problematic for our two-factor model because we view the costs as a phenomenon arising from contextual change.

Most context effects are not detected with recognition tests (Smith & Vela, 2001; but see Murnane, Phelps, & Malmberg, 1999). However, if participants tend to change to a better encoding strategy following a forget instruction, then the List 2 benefits of directed forgetting should be observed in recognition testing as well. Prior research has consistently failed to detect the benefits (and the costs) with recognition tests using relatively short lists of 12–16 items per list (Basden & Basden, 1996; Basden et al., 1993; Geiselman et al., 1983; Gross et al., 1970; MacLeod, 1999; Schmitter-Edgecombe et al., 2004; Whetstone et al., 1996; but see E. L. Bjork & Bjork, 2003). We suspected that benefits in recognition studies were too small to detect with shorter lists; therefore, we systematically explored the role of list length in directed forgetting. This goal was addressed in Experiment 2.

Experiment 1

Experiment 1 manipulated intentionality of learning between subjects to test predictions of the strategy change hypothesis. Examining directed forgetting under conditions when no intention for learning is established should provide a useful test of this hypothesis because there would be no reason to change the learning strategy between the lists when items do not have to be committed to memory.

Method

Participants. The participants were 80 undergraduate students at the University of Florida who participated in partial fulfillment of course requirements. Twenty participants were randomly assigned to each of the four experimental conditions.

Materials. Thirty unrelated English nouns of medium frequency were drawn from the Kučera and Francis (1967) norms. Two lists of words were prepared with 15 words in each, which were followed either by the forget or remember cue, resulting in four possible combinations for counterbalancing.

Design. The study involved a between-subjects design with cue (forget vs. remember) and intentionality (intentional vs. incidental) as factors.

Procedure. The procedures for the intentional study condition were as follows. Before the experiment, the experimenter instructed participants to learn and memorize the words for a later memory test. They studied two lists of unrelated words containing 15 items per list presented one at a time at a rate of 5 s per word on the computer screens. Upon the completion of the first list, the experimenter interrupted the procedure to deliver the experimental instruction. Half of the participants were told that the list they had studied was only for practice to familiarize them with the task and that there was no need to remember those items. They were asked to try to forget the words (the forget condition). The remaining participants were told that the list they had studied included only the first half of the items and that they should attempt to remember them for a later memory test (the remember condition). All participants then studied List 2 in the same fashion; the study of List 2 was always followed by the remember instruction. The final recall test was preceded by a 90-s filled retention interval consisting of arithmetic problems. The recall test involved asking participants to recall as many items as they remembered from both lists in any order they wished during the allotted 4 min.

The participants in the incidental study condition were led to believe they were helping the experimenter select a group of words for a study that would be conducted later in the semester. At the beginning of the study session they were told the following:

> I would like you to help me design some study materials for two different experiments that I plan on conducting later in the semester. Before I actually begin the experiments I need to obtain the pleasantness ratings of some words. There are no published word norms for a college population, so I'd like you to help me find out which words are considered to be more pleasant by college students. I will present you the words at a rate of 5 s per word. For the sake of saving time, please do not write down the word. Only rate the word indicating its pleasantness on a scale from 1 to 5, where 1 is not pleasant at all, and 5 is very pleasant.

After rating the first list of words, participants in the forget condition were told the following:

> Thank you so much for helping out with these materials. We will use your responses to select a group of words for a later experiment. Could you please rate one more list of words for the second experiment? However, it is really important that you not be influenced by your prior ratings. Therefore, please make an effort not to think of those earlier words in order not to contaminate the judgments on these new items. Try to forget those words and pretend you did not rate anything.

The remaining half of the participants (remember condition) were given the same instructions except they were not told to forget or to not think of the words. They were asked to rate one more list of words in the same manner. Upon completion of rating the words for pleasantness, participants completed 90 s of arithmetic as a filler task. They were then given the same recall test as the participants in the intentional study condition. Because the memory test came as a surprise to the participants in the incidental learning condition, at the time of the test, they were told that in addition to obtaining pleasantness ratings for the words the experimenter was also interested in knowing how many words could be remembered. Following the recall test, all participants were fully debriefed and were given an option to withdraw their data after having been fully informed about the true purpose of the study.

Results and Discussion

In all subsequent analyses, the probability of Type I error was set at $\alpha = .05$. To analyze the costs of directed forgetting, we performed an analysis of variance on proportion correct List 1 recall with cue (forget vs. remember) and intentionality (intentional vs. incidental) as the factors. There was a significant main effect of cue, $F(1, 76) = 11.29$, $MSE = 0.025$, $\eta^2 = .13$, indicating that the forget group recalled fewer items from the first list (.27) than the remember group did (.39). There was also a significant main effect of intentionality, $F(1, 76) = 4.16$, $MSE = 0.025$, $\eta^2 = .05$, indicating that recall was higher under intentional conditions (.37) than under incidental conditions (.30). There was no interaction effect, $F(1, 76) = 1.40$, $MSE = 0.025$, $\eta^2 = .02$ (see Figure 1, top). To summarize, there was a cost of directed forgetting under both intentional learning and incidental learning conditions. To verify once again that there were significant costs in the incidental learning condition, we performed an independent t test, which confirmed that the forget group recalled significantly fewer List 1 items than the remember group did, $t(38) = 2.50$, $\eta^2 = .14$.

To analyze the directed forgetting benefits, we used the same factors to examine the proportion correct List 2 recall. There was a significant main effect of intentionality, $F(1, 76) = 6.81$, $MSE = 0.022$, $\eta^2 = .08$, confirming again that recall was higher in the intentional learning condition (.46) than in the incidental learning condition (.38). The main effect of cue did not reach significance, $F(1, 76) = 2.58$, $MSE = 0.022$, $\eta^2 = .03$. However, there was a significant interaction between the cue and intentionality, $F(1,$

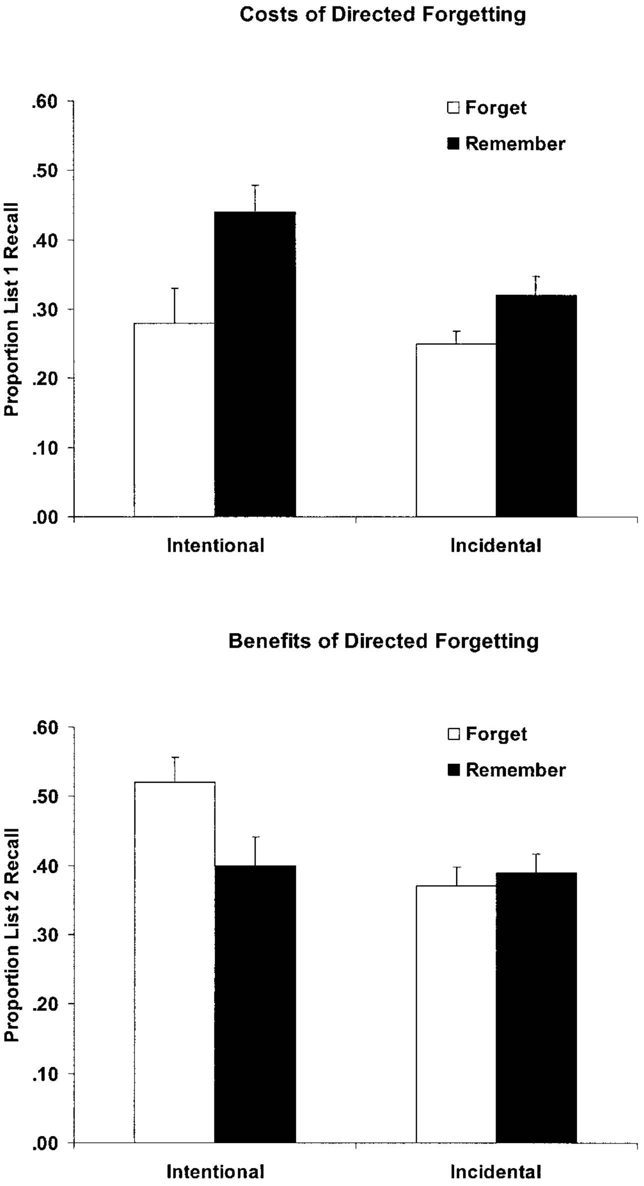

Figure 1. Proportion List 1 (top) and List 2 (bottom) recall by intentionality of learning and cue, Experiment 1.

76) = 4.87, *MSE* = 0.022, η^2 = .06. Follow-up analyses indicated that the forget group in the intentional learning condition recalled a greater proportion of List 2 items (.53) than the corresponding remember group (.40), revealing the standard directed forgetting benefits, $t(38) = 2.27$, η^2 = .12. However, there were no benefits in the incidental learning condition ($t < 1$; see Figure 1, bottom).

Given that the participants in the current study were allowed to recall the lists in any order, there is a possibility that the forget participants could start their recall from List 2 first, thus generating a greater output interference on List 1. To address this concern, we conducted the output order analyses by calculating an output position percentile according to the procedure developed by R. A. Bjork and Whitten (1974). Output position percentiles were calculated separately for each cue, group, and list condition (e.g., forget-incidental-List 1, remember-intentional-List 2, etc.). For a given participant, output position percentiles were computed as the average nominal output position divided by the total number of words recalled by that participant, where lower numbers indicate an earlier output order.

A Cue (forget vs. remember) × List (List 1 vs. List 2) mixed factorial analysis of variance on output percentile scores was conducted separately for incidental and intentional learning conditions. The results showed that in the incidental condition there was no main effect of cue ($F < 1$; .56 for forget and .57 for remember); there was also no main effect of list ($F < 1$; .58 for List 1 and .55 for List 2). The interaction was also not significant, $F(1, 38) = 1.24$, *MSE* = 0.039, η^2 = .03. Thus we found no evidence that the output order varied systematically across the cue conditions.

The results from the intentional condition indicated that there was a significant main effect of list, $F(1, 37) = 8.59$, *MSE* = 0.029, η^2 = .19, showing that List 1 was output before List 2 (.47 and .61, respectively). The main effect of the cue was not significant ($F < 1$), indicating that the forget and remember groups were similar in their output order preferences (.53 and .55, respectively). There was no significant interaction, $F(1, 37) = 1.22$, *MSE* = 0.029, η^2 = .03. Therefore, we found no evidence that output interference was responsible for the effects of cue on recall.

Overall, the results from the analyses of costs performed using a factorial design replicated the findings of Geiselman et al. (1983), who used a mixed design. We found that there was forgetting under both intentional and incidental learning conditions. This finding is predicted by the context change account of directed forgetting, which often derives its predictions from environmental context change studies. Because environmental context change studies show forgetting under both intentional study conditions (e.g., Smith & Vela, 2001) and incidental study conditions (Nelson & Goodmon, 2003), one would expect to detect significant directed forgetting costs both in intentional and incidental conditions. A failure to find costs with incidental material would have been problematic for the context change account.

The results from the analyses of List 2 (benefits), however, only partially replicated the findings of Geiselman and colleagues (1983). Although we obtained directed forgetting benefits for the intentional learning condition, there was no effect of the forget instruction in List 2 recall for the incidental learning condition. Geiselman et al. manipulated the intentionality of the items within subjects and found overall benefits both for incidental and intentional items without observing an interaction effect (although they indicated that the effect of the forget instruction was more pronounced for the intentional items than for the incidental items).

In the current study, intentionality was varied between subjects, and the List 2 benefits were significant only for the items that participants intended to learn for a later memory test. The presence of benefits in intentional but not incidental learning conditions is predicted by Sahakyan and Delaney's (2003) strategy change account, according to which participants in the forget condition often adopt a more efficient encoding study strategy when studying the second list. We have also previously shown that decisions to adopt better encoding strategy on the second list often can be mediated by participants' perceived degree of success with the current study strategy (Sahakyan et al., 2004). Because participants were not planning to learn the items in the incidental condition, they would be less likely to evaluate their study strategy and adopt a better encoding strategy on the second list. However, in Geiselman et al.'s (1983) study, some participants may have found it easier to encode all of the items using a better study strategy than to selectively apply the strategy to the intentional items.

Although the current results are consistent with the strategy change account and its predictions, we do not present any direct evidence that our results were due to strategy changes. Although it is clear that participants do change strategies in list-learning studies (see Delaney & Knowles, 2005; Sahakyan & Delaney, 2003) and do so more frequently in free recall studies following a forget instruction than following a remember instruction (Sahakyan & Delaney, 2003), the absence of List 2 benefits in incidental learning could potentially be explained by other mechanisms as well. For example, according to the selective rehearsal account (e.g., R. A. Bjork, 1970, 1972), the benefits in the forget group may emerge not because of study-strategy changes, but because participants drop List 1 items from rehearsal, thus freeing up more resources to rehearse List 2 items. In contrast, in the remember group, participants may keep rehearsing List 1 items along with List 2 items. Because there was no intention to learn the items in the current experiment, this explanation is consistent with the absence of List 2 benefits. Thus, the current findings can be explained by invoking a retrieval inhibition mechanism (to explain List 1 costs) and a selective rehearsal mechanism (to explain List 2 benefits).

An alternative way to construe the List 2 differences observed between the forget and remember groups in the intentional condition is to assume that there is a cost in the remember group rather than a benefit in the forget group. For example, in the intentional learning condition, the remember group may be retrieving List 1 items while studying List 2 items, producing retrieval-induced forgetting of List 2 items and hence poorer recall of List 2 items compared with the forget group. This can explain the absence of directed forgetting benefits in the incidental condition, which by definition lacks a remember instruction and consequently shows no retrieval-induced forgetting of List 2 items.[1]

Experiment 2

A second criticism of the strategy change account of benefits is the apparent absence of directed forgetting effects with recognition testing. If the benefits arise because of better encoding of List 2 by the forget group participants, then the benefits should be reflected both in recall and in recognition testing. However, benefits have not been observed in previous recognition studies.

In our previous work, we have shown that the participants in the forget group often adopt elaborative learning strategies when they study List 2 items (Sahakyan & Delaney, 2003). Elaborative encoding is likely to provide better organization of material and more efficient retrieval routes (e.g., Hunt & Einstein, 1981). However, the impact of organizational strategies is higher in recall than in recognition (e.g., Kintsch, 1970; Mandler, 1967). The effect of organization on recognition is more likely to emerge with longer lists because longer lists may lead to a greater list-length effect in the remember group compared with the forget group. This is because increasing the list length generally leads to impaired recollection (Cary & Reder, 2003; Yonelinas & Jacoby, 1994), but in the forget group better encoding of List 2 items may provide support for recollection and consequently lead to a smaller list-length effect compared with the remember group.

Another viewpoint is that more proactive interference on List 2 items is likely to accumulate in the remember group with longer rather than shorter lists because longer lists create more opportunities for interlist item similarity. Therefore, the benefits, which are often attributed to the escape from proactive interference by the forget group (e.g., E. L. Bjork & Bjork, 1996), are more likely to emerge under the conditions that permit sufficient accumulation of proactive interference in the remember group (i.e., longer lists).

Finally, longer lists can also impact the evaluation processes in the forget group, which were shown to mediate the strategy change decisions (Sahakyan et al., 2004). That is, the perceived degree of success with a particular study strategy is likely to differ between the longer and shorter lists because retrieving 5 items out of a 16-item list can be perceived as a less poor study strategy than retrieving 5 items out of a 36-item list. As the list length increases, participants' predictions about their future recollection performance are likely to decline, resulting in more pessimistic evaluations and perhaps more frequent strategy changes. Consequently, a study-strategy change may be more likely to occur with longer lists than shorter lists, maximizing the chances to detect the benefits of directed forgetting. In the current experiment, therefore, we used the standard list-method directed forgetting design, except that we varied the length of each study list from shorter lists (16 items) to longer lists (36 items).

The second goal of this experiment was to systematically explore source monitoring ability in directed forgetting using multinomial modeling analyses. In a typical source monitoring study, participants study items from various sources, where source may be visual versus auditory input, man's voice versus woman's voice, real versus imagined events, and so on. At the time of the recognition test, participants are asked to identify the source of origin of the detected items.

One could also apply source monitoring paradigms to understand participants' memory for list membership in directed forgetting (e.g., List 1 vs. List 2). Thus, in the remember condition, memory for list membership can be conceived of as a special case of typical source monitoring paradigms in which the two lists are the two sources. Forget instructions may affect list discrimination judgments. As we mentioned, previous research showed that the forget group is more likely to attribute List 1 items as having come from List 2 (e.g., Geiselman et al., 1983). These findings were based on the proportion of words that were correctly classified under the spaces designated for recall of List 1 and List 2. A recent study by E. L. Bjork and Bjork (2003) also suggested impaired source monitoring in forget group participants. In that study, participants studied a list of nonfamous names and then received either a forget or remember cue. After the cue, they studied a list of famous names. The final test included a fame-judgment task in which participants had to identify whether the presented name (some of which included previously studied items) was a famous or a nonfamous name. The results indicated that the forget group participants were more likely than the remember group participants to mistakenly judge the List 1 nonfamous names as being famous—in other words, they were more likely to confuse the list

---

[1] We thank Edgar Erdfelder for this intriguing idea.

membership of the items. It would be important to further explore whether the impaired source monitoring in the forget group reflects just a list discrimination problem or whether various response biases and guessing rates also play a role.

The analyses of traditional measures of source discrimination do not allow for an estimate of list discrimination ability that is independent of the overall performance levels, and multinomial modeling is recommended for assessing the impact of response biases, guessing rates, and measures of list discrimination that are independently estimated (e.g., Batchelder & Riefer, 1990; Riefer & Batchelder, 1988). Therefore, using the multinomial model could provide additional information about how people make list discrimination judgments and whether the forget cue affects such judgments. If the forget cue does affect list discrimination, is it because it affects the strength of List 1 items or because it affects various response biases?

Method

Participants. Participants were 88 University of Florida undergraduates who participated for course credit. None of the participants had previously participated in Experiment 1.

Materials. Two word sets of 72 medium-frequency nouns were chosen from the Kučera and Francis (1967) norms. Each word set served equally often as the study set and the distractor set on the recognition test. Each study set in turn was further divided into two equal halves, with each half serving equally often as the first and the second study list. Finally, the assignment of each word set to cue condition was also counterbalanced, resulting in eight possible conditions for counterbalancing. The recognition test consisted of studied items randomly intermixed with an equal number of foils.

Procedure. Participants studied two lists of equal length ranging from 16 to 36 items per list, with only the even list lengths being used (i.e., 16, 18, 20, . . . 36). The required number of words were randomly selected from the study set and presented in a random order. List length in statistical analyses was treated as a continuous variable rather than a discrete between-subjects variable. Eight participants were tested in each list-length condition to fulfill the counterbalancing requirements.

The procedures for this study resembled the procedures for intentional study condition of Experiment 1, the only exception being that the recall test was substituted by a recognition test. The recognition test consisted of intermixed foils and targets presented on the computer screen, and participants were asked to verbally indicate whether the item was “old” or “new.” If the participants judged an item as old, they then also had to indicate its source (e.g., List 1 or List 2).

Analyses. One can treat a directed forgetting experiment involving recognition testing as a source monitoring experiment in which the two sources correspond to List 1 and List 2, respectively. One way to represent outcomes from source monitoring studies is as a 3 × 3 source by response frequency matrix (cf. Batchelder & Riefer, 1990). An advantage of representing data in this way is that it allows the calculation of both traditional measures (such as hits and false alarms) as well as testing multinomial tree models, which are often recommended for the analysis of source monitoring experiments (e.g., Batchelder, Hu, & Riefer, 1994; Batchelder & Riefer, 1990; Bayen, Murnane, & Erdfelder, 1996). A generalized 3 × 3 frequency matrix for a given participant is given by

$$\begin{array}{ccc} & & \text{Response} \\ & & \begin{array}{ccc}\text{List 1} & \text{List 2} & \text{New}\end{array} \\ \text{Source} & \begin{array}{r}\text{List 1}\\ \text{List 2}\\ \text{New}\end{array} & \left[\begin{array}{ccc} Y_{11} & Y_{12} & Y_{13} \\ y_{21} & Y_{22} & Y_{23} \\ Y_{31} & Y_{32} & Y_{33} \end{array}\right] \begin{array}{c} Y_1 \\ Y_2 \\ Y_3 \end{array} \end{array}$$

where Y_{ij} is the frequency of responses of type j to items of type i. The marginal Y_1 and Y_2 values represent the list lengths of List 1 and List 2, respectively, and Y_3 is the number of foils presented (in our experiment, $Y_3 = Y_1 + Y_2$). For a group of participants, one can add the individuals’ frequency matrices together to get a 3 × 3 matrix for the group. Any of the traditional measures of recognition performance—such as hits and false alarms—can be obtained from the frequency matrix, both at the individual and group level.

Traditional measures of recognition. Traditional measures of recognition performance include hits and false alarms, which are sometimes combined into a single statistic that measures item detectability known as d'. In terms of the 3 × 3 frequency matrix, a participant’s hit rates for List 1 and List 2 are given as

$$\mathrm{HR}_1 = (Y_{11} + Y_{12})/\mathrm{Y}_1 \tag{2a}$$

and

$$\mathrm{HR}_2 = (Y_{21} + Y_{22})/Y_2. \tag{2b}$$

Total false alarms are given as

$$\mathrm{FA}_\mathrm{T} = (Y_{31} + Y_{32})/Y_3. \tag{3}$$

In addition, one can compute identification-of-origin scores for each list (I), which represent how well participants can discriminate the source of an item from that list (see Anderson, 1984; M. K. Johnson, Foley, & Leach, 1988; Voss, Vesonder, Post, & Ney, 1987). In terms of Equation 1, the identification-of-origin scores are derived as

$$I_1 = Y_{11}/(Y_{11} + Y_{12}) \tag{4a}$$

and

$$I_2 = Y_{22}/(Y_{21} + Y_{22}). \tag{4b}$$

Results and Discussion

We first analyze the costs and the benefits of directed forgetting with the traditional measures of recognition. We examine hits, false alarms, and source misattribution parameters. We then analyze the costs and the benefits using multinomial modeling technique. In all subsequent analyses, the probability of Type I error is set at $\alpha = .05$.

Traditional measures: Analyses of costs. The major difference between the current experiment and previous research is the systematic manipulation of list length. Because we believe that directed forgetting costs are caused by a contextual change mechanism, we did not predict effects of cue (remember vs. forget) on List 1 performance even with longer list lengths. Finding large directed forgetting costs in recognition would be problematic for the context change account because with words, context effects are typically small or nonexistent in recognition testing (cf. Smith & Vela, 2001; but see Murnane et al., 1999).

All subsequent analyses of traditional recognition measures made use of stepwise regression. We first entered the effect of cue, then the effect of list length, and we entered the interaction term (which we constructed by multiplying the list length and the cue) last to see whether it captured significant additional variance beyond the effects of cue and list length, following Jaccard and Turrisi’s (2003) recommendation.

Hit rates for List 1 as a function of cue and list length are shown as Figure 2A. The regression analysis indicated no significant effect of cue ($F < 1$; .77 for forget and .77 for remember). The effect of list length was significant, $F(1, 85) = 5.86$, $\Delta R^2 = .06$, reflecting a standard list length effect with performance decreasing

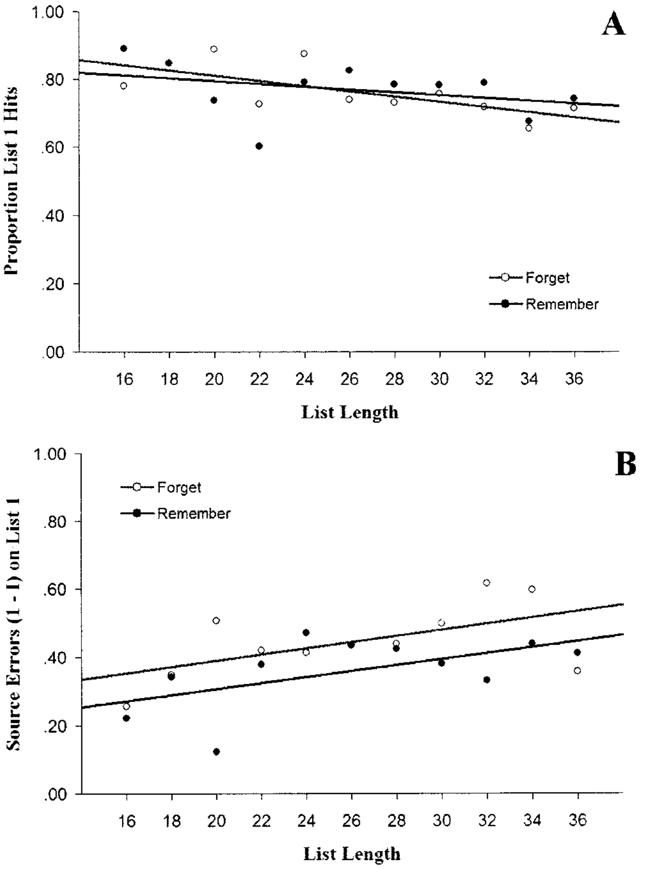

Figure 2. List Length × Cue interaction plots for List 1 hits (A) and List 1 source error statistic (1 – *I*) (B), Experiment 2. Points reflect means for each list length and cue condition (*n* = 4 per point).

as list length increases. The interaction was not significant ($F < 1$). The results indicate no effect of the forget instruction on the hit rate for List 1, consistent with earlier research.

The current data allow examination of whether the forget group participants were more likely than the remember participants to misattribute List 1 items as having come from List 2, as was shown in previous research (e.g., E. L. Bjork & Bjork, 2003; Geiselman et al., 1983). If so, it would tend to suggest that memory for source is also impaired by forget instructions in some way. The *I* statistic described earlier indicates the proportion of items that were attributed to correct source (List 1 or List 2) in addition to being correctly identified as old. Therefore, to examine source errors, we computed the statistic $1 - I$, which reflects the proportion of List 1 items mistakenly attributed to a wrong source. The source errors statistic $(1 - I)$ is shown as a function of list length and cue for List 1 in Figure 2B. Regression analysis showed that there was a significant effect of cue, $F(1, 86) = 4.84$, $R^2 = .05$. The forget group was overall more likely to misattribute the source of List 1 items to List 2 items (.44) than the remember group was (.36). In addition, there was a significant effect of list length, $F(1, 85) = 9.23$, $\Delta R^2 = .09$, reflecting more source errors as a function of list length. There was no interaction ($F < 1$). The results therefore suggest that the forget group would be more likely than the remember group to mistakenly attribute the source of items from List 1 to List 2, consistent with earlier research. However, these results demand caution in interpretation because some authors have suggested that *I* scores reflect the probability of a correct source identification that is conditional on the overall performance level (Batchelder & Riefer, 1990; M. K. Johnson et al., 1988; Riefer & Batchelder, 1988). *I* scores may also be influenced by other factors such as guessing rates and response biases. For more complete examination of source monitoring ability that is independent of the overall recognition rates and takes into account various guessing rates and biases, we later performed multinomial modeling in the forget and remember groups.

In summary, we found no evidence that there were costs on item recognition in directed forgetting, as evidenced by the absence of the cue effect on the hit rates. However, there was some evidence that the forget instruction affected memory for the source of List 1 items because forget participants were more likely to mistakenly attribute List 1 items as being from List 2.

Traditional measures: Analyses of benefits. As outlined in the introduction to the experiment, we expected to find significant directed forgetting benefits with longer list lengths. We analyzed List 2 performance as a function of cue and list length using the same stepwise regression analyses that we used for the costs.

List 2 hit rate is plotted as Figure 3A. There was a significant

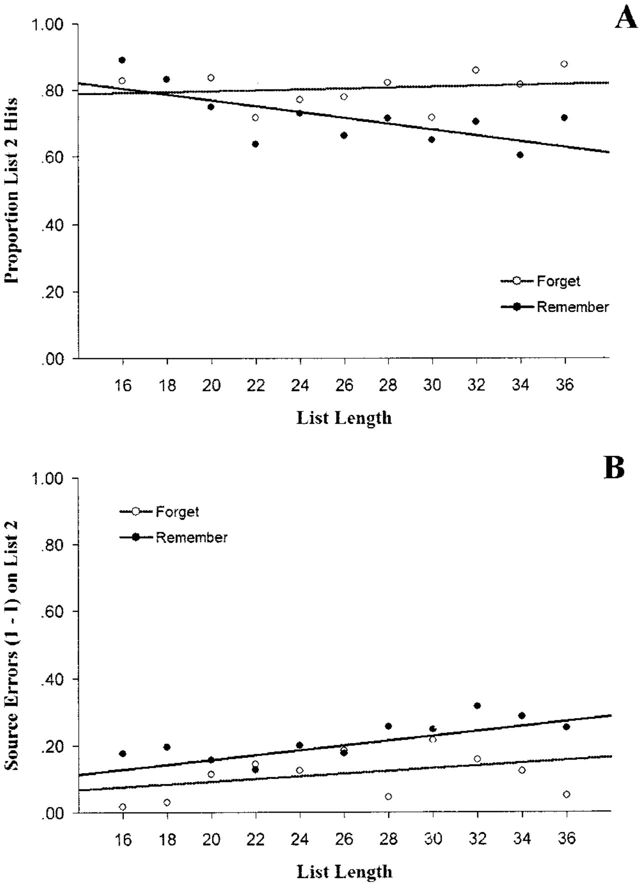

Figure 3. List Length × Cue interaction plots for List 2 hits (A) and List 2 source error statistic (1 – *I*) (B), Experiment 2. Points reflect means for each list length and cue condition (*n* = 4 per point).

effect of cue, $F(1, 86) = 7.88$, $R^2 = .08$, indicating better overall recognition in the forget group (.80) as compared with the remember group (.72). There was no effect of list length, $F(1, 85) = 2.32$, $\Delta R^2 = .02$. However, there was a significant interaction, $F(1, 84) = 4.37$, $\Delta R^2 = .04$. The interaction reflected reduced recognition rates in the remember group as a function of list length, $B = -.009$, significantly different from zero: $t(42) = 2.40$, but no effect of list length on recognition in the forget group, $B = .001$, which was not significantly different from zero: $t < 1$.

Taken together, the results suggest that benefits of directed forgetting on List 2 depend on the list length. Whereas the remember group suffered lower List 2 hit rates with increasing list length, the forget group maintained relatively stable List 2 performance as list length increased. Prior research has generally not found directed forgetting effects using recognition testing. We suspect this is because the studies used relatively short list lengths, and the current results suggest that the benefits of directed forgetting are a function of list length.

We also examined the source errors statistic ($1 - I$) for List 2, which is plotted as a function of cue and list length in Figure 3B. There was a significant effect of cue, $F(1, 86) = 18.37$, $R^2 = .18$, indicating that the remember group was more likely to mistakenly attribute List 2 items as having come from List 1 (.22) than the forget group was (.11). There was also a significant effect of list length, $F(1, 85) = 6.79$, $\Delta R^2 = .06$, indicating more source errors as a function of list length. There was no interaction ($F < 1$).

Traditional measures: Analyses of false alarms. False alarms rates as a function of cue and list length are shown as Figure 4. The main effect of cue was significant, $F(1, 86) = 9.46$, $R^2 = .10$, indicating that the forget group had on average fewer false alarms (.10) than the remember group did (.17). The main effect of list length was not significant, $F(1, 85) = 3.13$, $\Delta R^2 = .03$. However, there was a significant Cue × List Length interaction, $F(1, 84) = 5.18$, $\Delta R^2 = .18$, indicating that the effects of cue depended on list length. Specifically, false alarms increased with list length in the remember group, $B = .007$, reliably different from zero: $t(42) = 2.30$, but not in the forget group, $B = -.001$, not reliably different from zero: $t < 1$.

Summary of results from traditional measures of recognition. In summary, consistent with the previous research, there was no effect of the forget cue on recognition performance of List 1 regardless of the list length. However, contrary to the previous findings in the literature, the forget instruction produced better recognition performance on List 2 compared to the remember instruction, albeit only for longer lists. Longer lists produced higher false alarm rates in the remember group, but not in the forget group. As for the source discrimination judgments, the forget group was more likely to attribute List 1 items as having come from List 2, whereas the remember group was more likely to misattribute the source of List 2 items to List 1 items. The source discrimination errors tended to increase with longer list lengths in both the remember and forget conditions.

Multinomial modeling analyses. Multinomial models have become a popular tool for analyzing source monitoring data. Multinomial models propose that the likelihood of different behavioral outcomes can be mapped onto the likelihood of different underlying cognitive processes. They allow estimating independent parameters that correspond to the probability of measuring some processing capacity. These parameters can be used to conduct hypothesis testing across the experimental conditions.

Multinomial models are informative because they allow for independent estimates of source identification that do not rely on the overall recognition rates and take into account various biases and guessing rates (e.g., Batchelder et al., 1994; Batchelder & Riefer, 1990; Bayen et al., 1996; Riefer & Batchelder, 1988). Directed forgetting researchers have also noted that estimates of source discrimination that are not independent of overall recognition rates should be interpreted cautiously (see Basden & Basden,

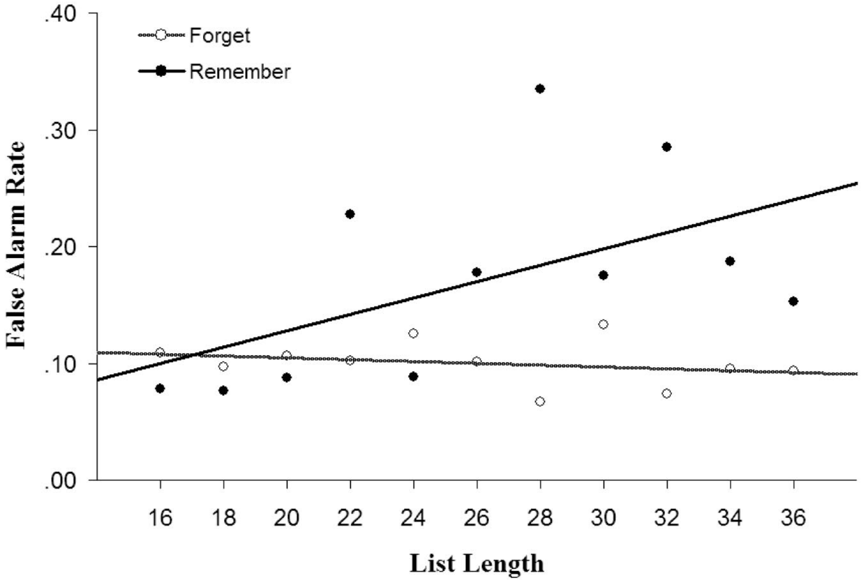

Figure 4. List Length × Cue interaction plot for false alarms, Experiment 2. Points reflect means for each list length and cue condition ($n = 4$ per point).

1996, 1998). For example, Basden and Basden (1996) found that forget and remember participants were equally good at identifying the serial positions of List 1 words, suggesting that at least within-list source memory was fairly intact. If participants lose access to information about list membership, then the forget participants ought to be less accurate in identifying serial positions as well.

We compared our shortest list lengths (16 to 22) against our longest list lengths (30 to 36) to understand how various parameters from multinomial modeling were affected by list length and cue. Selecting the four longest and four shortest groups was somewhat arbitrary, but it enabled us to maintain large enough samples at each list length group (32 per group) to draw reasonable conclusions while keeping the two groups sufficiently separated. Alternative grouping choices do not greatly affect the parameter estimates or their interpretation. The 3 × 3 response frequency matrices for these four experimental conditions are given in Table 1.

The general multinomial model for source monitoring from Batchelder and Riefer (1990) contains seven different free parameters, each of which is a probability that measures some processing capacity (and hence ranges from 0 to 1). The parameters D_1 and D_2 are the detection rates for List 1 and List 2, respectively, which give the probability of correctly detecting an item from that list as an "old" item. The parameters d_1 and d_2 are the source identification rates, which give the probability of correctly identifying the source of a correctly detected item (i.e., identifying whether an item came from List 1 or List 2). The parameter a is a guessing rate representing the probability that a correctly detected item whose source was not identified will be guessed as coming from List 1. The parameter g is the guessing rate representing the same probability except for undetected items. Finally, the parameter b is the bias, which represents the probability that an undetected study item or a new item will be guessed as being "old." Thus, a and g are guessing parameters relating to source, whereas b is a guessing parameter relating to detection.

Because the seven-parameter model is typically unidentifiable, several constraints may be imposed on its parameters to deal with this problem. The restriction that is most acceptable in these models is to assume that a equals g (rather than $d_1 = d_2$ or $D_1 = D_2$) because differences between the source-guessing parameters generally are less likely to be psychologically informative than other differences in the model. Batchelder and Riefer (1990) suggested conducting a series of tests to determine whether other parameters can be set equal in the model in order to select the most parsimonious, plausible model.

Table 1
3 × 3 Response Frequency Matrices Used in Multinomial Modeling of Experiment 2

	Short lists (16–22)			Long lists (30–36)		
Group	List 1	List 2	New	List 1	List 2	New
Forget						
List 1	151	95	58	178	197	153
List 2	19	224	61	54	379	95
New	28	35	545	38	66	952
Remember						
List 1	170	60	74	240	154	134
List 2	36	197	71	99	254	175
New	46	29	533	106	104	846

Following their recommendation, we first tested whether there were significant differences across the four experimental conditions in the detection rates for List 1 and List 2 using the chi-square test of independence. We found significant differences, $\chi^2(4, N = 3328) = 25.58, p < .01$. We therefore selected a multinomial model that assumed $D_1 \neq D_2$ (rather than one that assumed they were equal). The next step involved testing which of the two subsequent models would provide a better fit to the data—a model that assumes equal source identification parameters ($d_1 = d_2$) across the experimental conditions or a model that makes no such assumption (in other words, $d_1 \neq d_2$). These tests involved computing the log-likelihood ratio statistic G^2, where significant G^2 values indicate poorer fit to the data. The analyses revealed that the model that assumed equal source identification parameters across the experimental conditions provided a poor fit to the data, $G^2(4) = 44.23, p < .01$. This narrowed down the choice to the model that assumed $d_1 \neq d_2$ across the conditions. Accordingly we selected the six-parameter multinomial model, described by $D_1 \neq D_2$, $d_1 \neq d_2$, $a = g$, and b (see Figure 5). On the basis of this model, we estimated the parameters for the four experimental conditions (see Table 2).

Costs and benefits of directed forgetting: Multinomial model. Our first question was whether there were costs of directed forgetting in recognition, which correspond to changes in the D_1 parameter (item detectability of List 1). For short lists, there were no differences between the forget and remember groups in D_1, $G^2(1) = 2.91, p > .05$, indicating no costs. There were also no costs with longer lists, $G^2(1) = 0.02, p > .05$. Consistent with earlier research and the reported traditional analyses, the forget instruction had no effect on item detectability in recognition testing regardless of list length.

Our second question was whether there were benefits of directed forgetting in recognition with longer list lengths, as found by the traditional measures of recognition. This question can be answered by examining changes in the D_2 parameter (item detectability for List 2). For short lists, we observed no significant differences in D_2 parameter between the remember and forget groups, $G^2(1) = 1.25$, $p > .05$, indicating no benefits. However, for longer lists, the forget group had significantly better item detection for List 2 than did the remember group, $G^2(1) = 44.81, p < .01$. Furthermore, in the forget group, there was no effect of list length on D_2, $G^2(1) = 1.62, p > .05$, but in the remember group, longer lists resulted in lower List 2 item detectability, $G^2(1) = 16.64, p < .01$, suggesting a list-length effect. In summary, consistent with our traditional analyses, the benefits emerged only at longer list lengths, and they occurred because the remember group (but not the forget group) was less able to recognize List 2 items as a function of list length.

To examine false alarms, we looked at the differences in the bias parameter (b) because it indicates the propensity to identify the new items as old items (see Figure 5). We found that at shorter list lengths the bias parameter (b) did not differ between the forget and remember groups, $G^2(1) = 1.18, p > .05$. However, bias was reliably higher in the remember group than in the forget group at the longer list lengths, $G^2(1) = 42.75, p < .01$, indicating that on longer lists, remember participants were more likely than forget participants to guess that new items might have actually been "old"—in other words, to produce false alarms. As a result, the

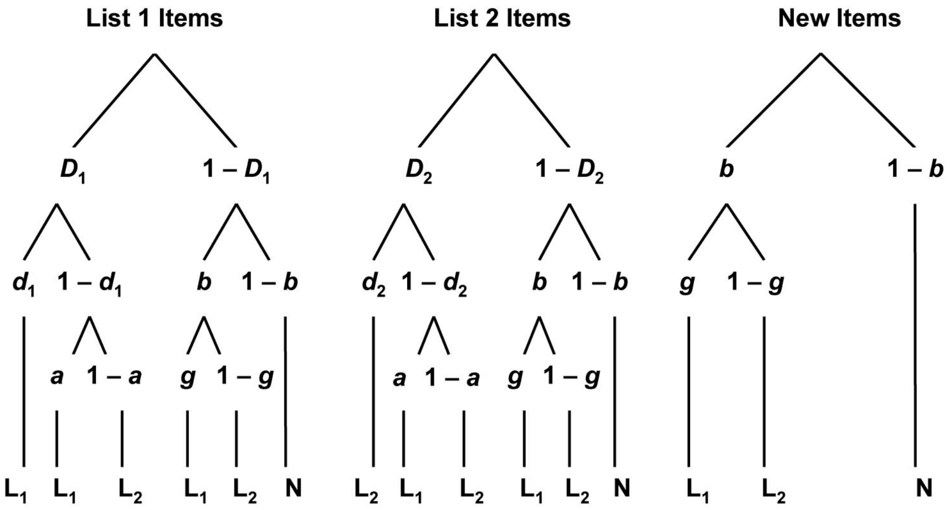

Figure 5. Multinomial model used in Experiment 2. D_1 = detection rate for List 1 (L_1); D_2 = detection rate for List 2 (L_2); d_1 = source identification rate for L_1; d_2 = source identification rate for L_2; a = guessing parameter representing the probability that a correctly deleted item with an unidentified source comes from L1; g = guessing parameter representing the probability that an undetected item comes from L_1; b = bias representing the probability of calling undetected study items or new items "old"; N = new item. Adapted from "Multinomial Processing Models of Source Monitoring" by W. H. Batchelder and D. M. Riefer, 1990, *Psychological Review, 97,* p. 551. Copyright 1990 by the American Psychological Association.

false alarm rate grows in the remember group with list length, while remaining relatively stable in the forget group (consistent with our traditional analyses).

Source discrimination measures: Multinomial model. Previous research has shown that forget group participants are worse at identifying the source of List 1 items than remember group participants are, whereas remember group participants are worse at identifying the source of List 2 items compared with the forget group (Geiselman et al., 1983). We observed a similar pattern in our traditional analyses of source errors scores ($1 - I$), which showed that the forget group was more likely than the remember group to mistakenly attribute the source of items from List 1 to List 2, whereas the remember group was more likely to mistake the source of List 2 items with List 1 compared with forget group.

To explore whether the multinomial modeling analyses concurred with the analyses on the traditional measures, we examined source identification parameters (d_1) and (d_2) between the forget and remember groups. We observed no differences in either source identification (d_1) or (d_2) between the forget and remember groups for either list length—for shorter lists, on d_1: $G^2(4) = 0.03$, $p > .05$; for longer lists, on d_1: $G^2(4) = 0.20$, $p > .05$; for shorter lists, on d_2: $G^2(4) = 0.10$, $p > .05$; and for longer lists, on d_2: $G^2(4) = 2.92$, $p > .05$. The results indicate that contrary to the findings of the traditional analyses, source discrimination for either List 1 or List 2 was equivalent in the forget and remember groups.

Although discrimination for either source was equivalent across the forget and remember groups, it is important to note that the source discrimination parameters between List 1 and List 2 were not equivalent. List 1 was consistently less well discriminated than List 2, regardless of list length and regardless of cue: For shorter lists in the forget condition, $G^2(1) = 16.46$, $p < .01$; for shorter lists in the remember condition, $G^2(1) = 10.81$, $p < .01$; for longer lists in the forget condition, $G^2(1) = 12.38$, $p < .01$; and for longer lists in the remember condition, $G^2(1) = 4.58$, $p < .05$.

If source discrimination is similar across the forget and remember groups, then what accounts for the observed differences in the traditional measures on source error rates? A possible explanation is that the source-guessing parameters differed across the two groups. The source-guessing parameter ($a = g$), which shows the probability that a correctly detected item whose source is not identified will be guessed as coming from List 1, was reliably larger in the remember group than in the forget group both at the shorter list lengths, $G^2(1) = 3.94$, $p < .05$, and at the longer list

Table 2
Parameter Estimates for Multinomial Model in Experiment 2

	Short lists (16–22)		Long lists (30–36)	
Parameter	Forget	Remember	Forget	Remember
D_1	.79	.72	.68	.68
D_2	.78	.73	.80	.59
d_1	.31	.34	.18	.23
d_2	.85	.78	.67	.51
$a = g$	.44	.61	.37	.50
b	.10	.12	.10	.20

Note. D_1 = detection rate for List 1; D_2 = detection rate for List 2; d_1 = source identification rate for List 1; d_2 = source identification rate for List 2; a = guessing parameter representing the probability that a correctly deleted item with an unidentified source comes from L_1; g = guessing parameter representing the probability that an undetected item comes from L_1; b = bias representing the probability of calling undetected study items or new items "old".

lengths, $G^2(1) = 5.50$, $p < .05$. This result indicates that participants in the remember group were more likely than participants in the forget group to guess that items with unidentified source were from List 1. This would lead to List 2 items being more often guessed as coming from List 1 in the remember group compared with the forget group.

Because the source-guessing parameter a indicates the probability of guessing the source of an identified item as being from List 1, conversely $1 - a$ will indicate the probability of guessing the source of an identified item as being from List 2. Because the a parameter was significantly smaller in the forget group, that would mean that the forget group was more likely to guess the source of items as being from List 2 (due to $1 - a$ being larger in the forget group compared with the remember group). This is consistent with the traditional analyses as well as earlier research showing that forget group often attributed the source of List 1 items to List 2.

Summary of multinomial model analyses versus traditional analyses. For the most part, the multinomial model analyses replicated the results of the traditional analyses. No directed forgetting costs were observed using either approach to data analysis, and the benefits emerged with longer lists in both approaches. The major difference between the approaches was in the results pertaining to the identification of source. Results from the traditional approach suggest that the forget group was better at identifying List 2 source than the remember group, but worse at identifying the List 1 source compared with the remember group. However, multinomial modeling results suggest that the source discrimination for either list was equivalent between the forget and remember groups. Because guessing rates favored List 2 in the forget group and List 1 in the remember group (albeit only with shorter lists), this led to poorer List 1 discrimination in the forget group compared with the remember group and poorer List 2 discrimination in the remember group compared with the forget group as evident in the analyses of source error scores.

Alternative explanations. Although we interpret the emergence of benefits with longer lists within the scope of the strategy change explanation, the results are consistent with alternative viewpoints as well. For example, if participants stop rehearsing List 1 items in the forget group and devote those resources to List 2 encoding (a selective rehearsal account), List 1 costs would emerge. The benefits could have emerged with longer lists because longer lists consumed more resources, and so dropping rehearsal of a long list might free up more resources than dropping rehearsal of a short list. However, the same selective rehearsal account has difficulty explaining why extra rehearsal benefited List 2 in the forget group, but the continued rehearsal of List 1 items during List 2 presentation did not benefit List 1 in the remember group, as was evident from the absence of costs. Accounts invoking a single mechanism seem insufficient to account for the entire directed forgetting effect. A combination of single-mechanism-based accounts (e.g., retrieval inhibition plus selective rehearsal) can better explain the findings. Perhaps recognition testing releases the costs by eliminating retrieval inhibition. In this fashion, costs could be observed with recall testing but released with recognition testing, but the benefits on List 2 due to extra selective rehearsal would be preserved.

Alternative accounts of the benefits exist that assume they emerge from an escape from the usual effects of previously studied items on newly encountered items. For example, E. L. Bjork and Bjork (1996) proposed that the forget group escapes from proactive interference that normally accumulates on List 2 in the remember group. It seems possible that in the remember group the proactive interference does not accumulate as fast with short lists in recognition performance. Therefore, the benefits cannot be observed under conditions of insufficient proactive interference accumulation. Alternatively, in the remember group people may be retrieving List 1 items during the study of List 2, resulting in retrieval-induced forgetting of the List 2 items (see Footnote 1). This mechanism would also result in benefits in the forget group because the forget instruction discourages retrieval of List 1 items. This latter account provides an intriguing alternative explanation of the results from both Experiments 1 and 2 that can be investigated in future research.

General Discussion

Our position on directed forgetting has been that two mechanisms are required to explain the benefits and the costs of directed forgetting. Most of the time, these mechanisms operate together, but there are cases when the effects of these mechanisms may be observed independently. Elsewhere, we have attributed the costs of directed forgetting to contextual change brought on by the forget cue (Sahakyan, 2004; Sahakyan et al., 2004; Sahakyan & Kelley, 2002). We have also argued that the benefits arise because participants often change their study strategy between the two lists in response to the forget instruction by adopting a more elaborate encoding of List 2. The more frequent choice of better encoding strategy by the forget group participants compared with the remember group gives rise to higher recall on List 2, leading to the benefits of directed forgetting (Sahakyan & Delaney, 2003; Sahakyan et al., 2004). There are several dissociations in the literature that report the costs without the benefits (Conway, Harries, Noyes, Racsma'ny, & Frankish, 2000; Sahakyan & Delaney, 2003; Sahakyan et al., 2004), or the benefits without the costs (Liu, 2001; Macrae, Bodenhausen, Milne, & Ford, 1997), lending further support to our two-factor account of directed forgetting.

Our article was mainly concerned with criticisms of the strategy change component of that model. Earlier research that found directed forgetting under incidental learning conditions (Geiselman et al., 1983) poses a challenge to the strategy change account of benefits because there would be no reason to suspect a change of study strategy between the lists when items do not have to be memorized. In Experiment 1, we repeated Geiselman et al.'s (1983) study, but we varied intentionality of learning between-subjects rather than within-subjects to avoid potential integration between the intentional and incidental items. Consistent with their findings, we found directed forgetting costs under both intentional and incidental learning conditions (a finding that could be expected on the basis of most current theories of directed forgetting). However, benefits were detected only under the intentional study condition—not in the incidental condition, as reported in earlier research. We propose that the findings are consistent with the strategy change account of directed forgetting benefits because there would be no obvious reason to change the study strategy in response to the forget cue when the items did not have to be studied in the first place. We suspect that the reason Geiselman et al. detected the benefits in the incidental condition was because the intentional and incidental study items became integrated with each other by virtue of being on the same study list. To summarize, we found significant costs, but no benefits, consistent with the two-factor account of directed forgetting.

Experiment 2 aimed to address another criticism of the strategy change account of benefits—specifically, the absence of directed forgetting benefits with recognition testing. Because we believe that the benefits arise from more frequent choice of better encoding strategy on List 2 in the forget group, we would expect the encoding differences to be reflected in recognition testing as well.

In this study, we systematically varied the length of the study lists to examine whether one could detect the benefits in recognition with longer lists as opposed to the shorter lists used in previous research. We found no directed forgetting costs regardless of the list length, but the benefits emerged as the list length increased. The absence of costs in recognition is consistent with previous research reporting similar findings, and it is also in accord with our two-factor view of directed forgetting in which we attribute the costs to the changes in mental context. Because recognition testing is often insensitive to contextual changes, we do not find the lack of directed forgetting costs to be problematic. In our view, the most important finding that emerged as a result of varying the list length is the presence of benefits in the forget group with longer lists. This finding is new and was not reported previously primarily because the majority of directed forgetting studies used shorter lists. To summarize, whereas in Experiment 1 we found significant costs but no benefits in incidental learning, in Experiment 2 we found no costs but significant benefits with longer lists. These results support our view that directed forgetting consists of two outcomes with different underlying mechanisms.

Experiment 2 also used multinomial modeling to examine source memory. The multinomial modeling generally confirmed the findings from the traditional measures, including hit rates and false alarm rates. However, the results differed on source memory. The traditional measures of source discrimination showed that the forget group often misattributed the source of List 1 items to List 2 items, whereas the opposite was true in the remember group. These traditional analyses were consistent with previous research by Geiselman et al. (1983), as well as recently reported findings of E. L. Bjork and Bjork (2003). The multinomial modeling analyses, on the other hand, showed that the source discrimination problems do not necessarily reflect a list discrimination problem. In fact, the list discrimination parameters were identical for both the forget and remember groups, although List 2 was always discriminated better than List 1 in both groups. What appears to give rise to problems with source discrimination is the significant difference in the guessing rates between the forget and remember groups. Whereas the forget group was more biased to guess the source of correctly detected items to be from List 2, the remember group is more inclined to guess the source to be from List 1 (at least for shorter lists). Because in our multinomial modeling analyses List 1 was always discriminated more poorly than List 2, that means there were more List 1 items with unidentified source that were subject to guessing than List 2 items. This can explain why source discrimination problems reported for the forget group are greater and more reliable than for the remember group (Geiselman et al., 1983).

Guessing biases may be relevant for explaining other recent findings, such as the recent E. L. Bjork and Bjork (2003) finding that there apparently are lingering effects of forget instructions in recognition testing. In their study, they instructed some people to forget a list of nonfamous names and others to keep remembering those names. Both groups then studied a list of famous names. Later, participants had to indicate whether a given name was famous. Participants in the forget group were more likely to falsely attribute fame to nonfamous names than participants in the remember group. The authors attributed the findings to be due partly to source discrimination problems. However, they also suggested that guessing biases could have also accounted for some of the observed effect. If forget group participants were unsure about the source of the item and had to guess, they would be more likely to guess an item was from List 2—that is, a famous name. Thus, they would make more false fame judgments than the remember group participants, who would be more biased to guess an item was from List 1. The multinomial model analyses confirmed this speculation and suggested that explanation based on guessing biases may be sufficient to explain their findings without having to invoke differences in source memory.

To sum up, the current studies provided further cases in which costs and the benefits did not occur simultaneously, suggesting that they may have different underlying mechanisms. The explanations that invoke a single process to explain both the costs and benefits are insufficient to account for the pattern of findings reported in these studies. Several combinations of processes may be sufficient to explain the pattern of results observed here, and future work is needed to address which particular combinations of mechanisms provide the best account of the phenomena.

References

Anderson, R. E. (1984). Did I do it or did I only imagine doing it? *Journal of Experimental Psychology: General, 113,* 594–613.

Basden, B. H., & Basden, D. R. (1996). Directed forgetting: Further comparisons of the item and list methods. *Memory, 4,* 633–653.

Basden, B. H., & Basden, D. R. (1998). Directed forgetting: A contrast of methods and interpretations. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting* (pp. 157–172). Hillsdale, NJ: Erlbaum.

Basden, B. H., Basden, D. R., & Gargano, G. J. (1993). Directed forgetting in implicit and explicit memory tests: A comparison of methods. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 603–616.

Basden, B. H., Basden, D. R., & Morales, E. (2003). The role of retrieval practice in directed forgetting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 389–397.

Batchelder, W. H., Hu, X., & Riefer, D. M. (1994). Analysis of a model for source monitoring. In G. H. Fischer & D. Laming (Eds.), *Contributions to mathematical psychology, psychometrics, and methodology* (pp. 51–65). New York: Springer-Verlag.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review, 97,* 548–564.

Bayen, U. J., Murnane, K., & Erdfelder, E. (1996). Source discrimination, item detection, and multinomial models of source monitoring. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 197–215.

Bjork, E. L., & Bjork, R. A. (1996). Continuing influences of to-be-forgotten information. *Consciousness & Cognition: An International Journal, 5*(1–2), 176–196.

Bjork, E. L., & Bjork, R. A. (2003). Intentional forgetting can increase, not decrease, residual influences of to-be-forgotten information. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 29,* 524–531.

Bjork, R. A. (1970). Positive forgetting: The noninterference of items intentionally forgotten. *Journal of Verbal Learning and Verbal Behavior, 9,* 255–268.

Bjork, R. A. (1972). Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory* (pp. 217–235). Washington, DC: Winston.

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 honour of Endel Tulving* (pp. 309–330). Hillsdale, NJ: Erlbaum.

Bjork, R. A., & Whitten, W. B. (1974). Recency-sensitive retrieval processes in long-term free recall. *Cognitive Psychology, 6,* 173–189.

Bjork, R. A., & Woodward, A. E. (1973). Directed forgetting of individual words in free recall. *Journal of Experimental Psychology, 99,* 22–27.

Cary, M., & Reder, L. M. (2003). A dual-process account of the list-length and strength-based mirror effects in recognition. *Journal of Memory and Language, 49,* 231–248.

Conway, M. A., Harries, K., Noyes, J., Racsma'ny, M., & Frankish, C. R. (2000). The disruption and dissolution of directed forgetting: Inhibitory control of memory. *Journal of Memory and Language, 43,* 409–430.

Delaney, P. F., & Knowles, M. E. (2005). Encoding strategy changes and spacing in free recall. *Journal of Memory and Language, 52,* 120–130.

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.

Gross, A. E., Barresi, J., & Smith, E. E. (1970). Voluntary forgetting of a shared memory load. *Psychonomic Science, 20,* 73–75.

Hunt, R. R., & Einstein, G. O. (1981). Relational and item-specific information in memory. *Journal of Verbal Learning and Verbal Behavior, 20,* 497–514.

Jaccard, J., & Turrisi, R. (2003). *Interaction effects in multiple regression.* Thousand Oaks, CA: Sage.

Johnson, H. (1994). Processes of successful intentional forgetting. *Psychological Bulletin, 116,* 274–292.

Johnson, M. K., Foley, M. A., & Leach, K. (1988). The consequences for memory of imagining in another person's voice. *Memory & Cognition, 16,* 337–342.

Kimball, D. R., & Metcalfe, J. (2001, November). *Directed forgetting: Evidence against a suppression mechanism.* Poster presented at the 42nd Annual Meeting of the Psychonomic Society, Orlando, FL.

Kintsch, W. (1970). Models for free recall and recognition. In D. A. Norman (Ed.), *Models of human memory.* New York: Academic Press.

Kučera, H., & Francis, W. N. (1967). *Computational analysis of present-day American English.* Providence, RI: Brown University Press.

Liu, X. (2001). On the dynamics of directed forgetting: Facilitation and interference in the updating of human memory. *Dissertation Abstracts International, 61*(11), 6159B.

MacLeod, C. M. (1998). Directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 1–57). Mahwah, NJ: Erlbaum.

MacLeod, C. M. (1999). The item and list methods of directed forgetting: Test differences and the role of demand characteristics. *Psychonomic Bulletin & Review, 6,* 123–129.

Macrae, C. N., Bodenhausen, G. V., Milne, A. B., & Ford, R. L. (1997). On regulation of recollection: The intentional forgetting of stereotypical memories. *Journal of Personality and Social Psychology, 72,* 709–719.

Mandler, G. (1967). Organization and memory. In K. W. Spence & J. T. Spence (Eds.), *The psychology of learning and motivation* (Vol. 1). New York: Academic Press.

Murnane, K., Phelps, M. P., & Malmberg, K. (1999). Context-dependent recognition memory: The ICE theory. *Journal of Experimental Psychology: General, 128,* 403–415.

Nelson, D. L., & Goodmon, L. B. (2003). Disrupting attention: The need for retrieval cues in working memory theories. *Memory & Cognition, 31,* 65–76.

Riefer, D. M., & Batchelder, W. H. (1988). *Statistical inference for multinomial processing tree models* (Tech. Rep. No. MBS 90–05). Irvine: University of California, Department of Mathematical Behavioral Sciences.

Sahakyan, L. (2004). The destructive effects of the "forget" instructions. *Psychonomic Bulletin & Review, 11,* 555–559.

Sahakyan, L., & Delaney, P. F. (2003). Can encoding differences explain the benefits of directed forgetting in the list-method paradigm? *Journal of Memory and Language, 48,* 195–201.

Sahakyan, L., Delaney, P. F., & Kelley, C. M. (2004). Self-evaluation as a moderating factor in strategy change in directed forgetting benefits. *Psychonomic Bulletin & Review, 11,* 131–136.

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.

Schmitter-Edgecombe, M., Marks, W., Wright, M. J., & Ventura, M. (2004). Retrieval inhibition in directed forgetting following severe closed-head injury. *Neuropsychology, 18,* 104–114.

Smith, S. S., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review, 8,* 203–220.

Voss, J. F., Vesonder, G. T., Post, T. A., & Ney, L. G. (1987). Was the item recalled and if so, by whom? *Journal of Memory and Language, 26,* 466–479.

Whetstone, T., Cross, M. D., & Whetstone, L. M. (1996). Inhibition, contextual segregation, and subject strategies in list method of directed forgetting. *Consciousness and Cognition, 5,* 395–417.

Yonelinas, A. P., & Jacoby, L. L. (1994). Dissociations of processes in recognition memory: Effects of interference and of response speed. *Canadian Journal of Experimental Psychology, 48,* 516–534.

Received June 14, 2004
Revision received January 18, 2005
Accepted January 21, 2005 ■