Psychonomic Bulletin & Review
1999, 6 (1), 123-129

The item and list methods of directed forgetting: Test differences and the role of demand characteristics

COLIN M. MACLEOD
University of Toronto, Scarborough, Ontario, Canada

In directed forgetting, the item method presents instructions to remember or to forget individual items; the list method presents a single mid-list instruction to forget the first half of the list. Initial free recall was better for remember (R) words than for forget (F) words under both methods. Offered 50¢ for each additional F word, subjects could recall almost no more items, eliminating a demand characteristics explanation. On a *yes/no* recognition test, only the item method showed directed forgetting. Retrospective instruction identification was good except for F words under the list method, where performance was at chance. There was no evidence of speed–accuracy tradeoff on the recognition or instruction identification tests. These results bring together the major findings concerning directed forgetting and support a method-based theoretical distinction.

There are facts that we need to forget, as is the case for a former address or the command keystrokes from a word processor no longer used. There are also facts that we want to forget, such as the death of a family member or just an embarrassing moment. How do we accomplish such forgetting and how can we understand it? Over the past 30 years, one technique has emerged as the primary way to examine "intentional forgetting" in the laboratory: *directed forgetting*. There is now a substantial literature on directed forgetting in human memory research (see Johnson, 1994; MacLeod, 1998, for reviews) and more broadly (see Golding & Long, 1998, for a review). The wide interdisciplinary impact of this work is thoroughly documented in a recent book edited by Golding and MacLeod (1998).

The present article is aimed at bringing together two crucial issues in directed forgetting in one place. The first goal is to compare the impact of the two dominant methods for conducting directed forgetting experiments on several prevalent tests of memory. The second goal is to test whether demand characteristics play any role in directed forgetting effects.

Cuing by the Item Method Versus the List Method

From the beginning of work on directed forgetting (e.g., Block, 1971; Muther, 1965), there have been two common procedures for presenting the instructional cues to subjects regarding which items to forget versus remember. Under the *item method*, the subject is given an explicit cue for each to-be-forgotten item (F word), such as "forget" or "FFFF." Usually, researchers also signal each to-be-remembered item (R word) with a complementary cue, such as "remember" or "RRRR." Typically, the cue comes after the relevant word, ensuring that the subject registered the word. Under the *list method*, only one cue is presented, ordinarily an instruction to forget all preceding items, given at the middle of the list.

Studies had been done using the two procedures more or less interchangeably until Basden, Basden, and Gargano (1993) made a crucial observation and confirmed it empirically. Until then, there had been a puzzle concerning why directed forgetting sometimes affected recognition (e.g., Davis & Okada, 1971; MacLeod, 1975) and sometimes did not (e.g., Block, 1971; Elmes, Adams, & Roediger, 1970). Basden et al. (1993) pointed out that the item method consistently produced a recognition effect whereas the list method consistently did not. They suggested that the item method fostered selective rehearsal favoring the R words whereas the list method promoted inhibition of the F words, as R. A. Bjork (1989) had conjectured and Johnson (1994) had then highlighted in her review. Both selective rehearsal and inhibition affected recall, but selective rehearsal had a more powerful effect, so that the remember–forget (R–F) difference in recall was smaller under the list method. Recognition, on the other hand, was sensitive only to selective rehearsal. The very act of re-presenting an item on a recognition test lifted or released the inhibition in the case of the list method (cf. Geiselman, Bjork, & Fishman, 1983).

On a *yes/no* recognition test, Basden et al. (1993, Experiments 1 and 2) observed an R–F difference only under the item method (R, .92; F, .77) and not under the list

This research was supported by Natural Sciences and Engineering Research Council of Canada Grant A7459. For her assistance in programming and in collecting the data, I am very grateful to Shelley Hodder. For helpful comments on earlier versions, I thank Barbara Basden, David Basden, David Elmes, and Jonathan Golding. Correspondence should be addressed to C. M. MacLeod, Division of Life Sciences, University of Toronto, Scarborough, ON M1C 1A4, Canada (e-mail: macleod@scar.utoronto.ca).

method (R, .90; F, .89). The disadvantage for F words under the item method seemed to disappear under the list method, although the possibility of a ceiling effect on the R words complicates interpretation. Most intriguingly, on a recall test following recognition, the cue effect so apparent for the item method (R, .64; F, .15) was absent for the list method (R, .57; F, .57). This is consistent with the claim that list method inhibition was released by the prior recognition test but that item method selective rehearsal continued to operate.

In the present study, the two methods will be directly compared on free recall, *yes/no* recognition, and instruction identification (or "tagging") measures of memory. The accuracy results should replicate those of Basden et al. (1993). In addition, the present study examines response time on the recognition and instruction tagging tests to determine whether accuracy differences could result from a tradeoff with latency. Although memory for instructional cue has been investigated (e.g., Geiselman et al., 1983), and latency has occasionally been used as a dependent measure (e.g.,Whetstone, Cross, & Whetstone, 1996), there are no preexisting data to address this tradeoff concern.

Is There a Motivational or Demand Characteristics Problem in Directed Forgetting?

Upon hearing about the finding of an R–F difference, some people will understandably be highly skeptical. Are the subjects *really* forgetting the F words? Perhaps they simply do not try as hard or search as long to recover the F words as they do for the R words. This would amount to a motivational explanation rather than a processing explanation. Indeed, subjects may actually recover F words but then simply not report having done so. This is the issue of demand characteristics (Orne, 1962), important in any psychological study, but especially in those manipulating instructions to the subject.

This potentially serious concern has been recognized from the earliest research on directed forgetting. Weiner (1968, p. 233) wrote that "Perhaps the most serious experimental problem . . . is differential withholding (suppression)." In the past, relatively little comfort has been provided that the standard R–F difference is truly a processing effect rather than the result of subjects' attempting to cooperate with the perceived desires of the experimenter. In fact, in the posthypnotic amnesia paradigm, Coe (1978) argued that response withholding did actually occur. Basden, Basden, Coe, Decker, and Crutcher (1994) took this a step further, comparing item method directed forgetting to posthypnotic amnesia instructions. Critically, on the second of two recognition tests, subjects were told that they could now remember all previously forgotten words. Performance for F words increased under the posthypnotic amnesia condition, implicating demand effects there, but no increase occurred in the directed forgetting condition, suggesting no demand effects there.

Where motivation has been explored, the manipulation has not been adequate to dispel concerns about withholding F words at the time of test. Gross, Barresi, and Smith (1970) manipulated payoff only during study, which does not address the question of motivation to succeed on R words versus F words at the time of test. Woodward and R. A. Bjork (1971) improved on this by manipulating payoff just prior to test, but the small amount of the payoff and the fact that it was equivalent for R words and F words also does not provide a suitable test of the motivation/demand characteristics account. The same failure to discriminate the R and F words is apparent in the study by R. A. Bjork and Woodward (1973), who manipulated incentive both prior to study and again just before a final recall test.

More recently, Geiselman, Rabow, Wachtel, and MacKinnon (1985, Experiment 4) examined this question, using the item method. On the recall test, subjects were given point values for the R and F words, 9 for R and 1 for F, 5 for each, or 9 for F and 1 for R. This payoff manipulation had absolutely no effect on recall: All conditions showed large and equivalent R–F differences. Perhaps a lack of explanation to the subjects about the purpose of the points manipulation undermined any potential motivational effect. Even had their points system been meaningful to the subjects, the Geiselman et al. (1985) manipulation still took place on the initial recall test.

It would seem best to allow subjects an initial recall, thereby determining the apparent R–F difference—"apparent" because it could hinge on withholding F words. Only then would a substantial payoff be offered targeting F words. If subjects had initially withheld F words, they should show the reverse pattern on this second recall test with enhanced motivation to recall only F words. That is what the present study sought to do. After an initial recall test, subjects were asked to try to recall additional F words, each now worth 50¢. Previously recalled R or F words could be recalled again, and new R words could be added, but these would not affect payoff. The focus would be on recovering F words, which should be straightforward if these words were available but simply being withheld from output.

METHOD

Subjects

Forty-eight undergraduate students from the University of Toronto at Scarborough participated, 24 under each method. All received bonus points toward their final grade in introductory psychology.

Apparatus

The experiment was controlled by an IBM-AT compatible microcomputer with a 14-in. color VGA monitor. The controlling program was written in QuickBasic 4.5, using the Graves and Bradley (1987, 1988) routines to achieve millisecond timing accuracy. Item displays were printed in white on a black background in 80-column mode. Response times were recorded as the time between the stim-

ulus screen onset and the subject's keypress response, which sent an interrupt to the computer.

Materials

The word pool consisted of 1,031 nouns of frequency greater than 9 per million taken from the Thorndike and Lorge (1944) norms. For each of the 24 subject files, the 1,031 words were randomized, after which 114 were selected. Twelve of the 114 words were then randomly chosen to form a practice list and 6 to be buffer items on the critical list. The remaining 96 words were randomly divided into two sets of 48, one set to be studied and one to be unstudied. Then the 24 R and 24 F instructions were randomly assigned to the 48 studied words, and the order of the study set was randomized. For both the recognition test and the instruction identification test, the order of the 96 test words (i.e., 48 studied and 48 unstudied) was independently randomized. Each of the 24 stimulus files thus contained the practice, study, recognition, and identification items and their condition identifiers. To maximize comparability, correspondingly numbered subjects under the two methods experienced the same stimulus file.

Procedure: Item Method

Study phase. Subjects were told that the task was to study a list of about 50 words for a later memory test. They were informed that each word would be followed by an instruction telling them whether that word would later be tested. If a word was followed by an "RRRRRR" instruction, they should try to remember it because it would be tested; if a word was followed by an "FFFFFF" instruction, they need not remember it as it would not be tested. The subjects were also informed that each word would be presented at the center of the screen for 1 sec (750-msec word plus 250-msec blank) followed by that word's instruction for 3 sec (2,500-msec instruction plus 500-msec blank). Finally, they were told that there was an initial 12-word practice list (6 R and 6 F) to help familiarize them with the task.

Following the practice study trials, the subject was asked to write down the six R words only, helping to reinforce the validity of the instructional manipulation. When the subject finished recalling the practice items, the experimental trials began. The first three words and the last three words were buffers included to reduce the effects of primacy and recency on recall; all of these buffers were R words and were not included in any of the analyses.

Each study trial began with a 500-msec blank screen followed by a warning string (i.e., ******) at the center of the screen for 1,000 msec. After another 500-msec blank screen, a word appeared on the screen for 750 msec, followed by a 250-msec blank screen. At this point, the remember (RRRRRR) or forget (FFFFFF) instruction was printed for 2,500 msec. The computer then proceeded to the next trial.

Recall test phase. Immediately following the study phase, the subjects were instructed to write down as many words as possible from the list just studied; contrary to initial instructions, they were encouraged to try just as hard to recover F words as R words. Once a subject finished recalling, the experimenter collected the first recall sheet and informed the subject that on a second recall test there would be a 50¢ reward for every new F word added (i.e., F words not recalled on the first test). However, if in doubt, the subject was to write down any possibly studied word that came to mind even though only new F words would be rewarded. When the subject finished the second recall, the experimenter immediately scored the number of additional F words recalled and paid the subject.

Recognition test phase. Following the second recall test, the instructions for the recognition test were presented. The subject was told to decide whether each test word was previously studied (regardless of the instruction) or new, and to press the "/" key for a studied word and the "z" key for an unstudied word, guessing if uncertain. Each trial began with a 500-msec blank screen; then the word appeared at the center of the screen until the subject responded. The computer scored the accuracy and recorded the latency of the trial, and then proceeded to the next trial.

Instruction identification ("tagging") test phase. Finally, subjects saw each studied word individually at the center of the screen, the task being to indicate which instruction (R or F) had followed that word during the study phase. Subjects were to press the "/" key for R words or the "z" key for F words. There were 48 test words, with the procedure conducted in the same way as for the recognition test.

Procedure: List Method

The 24 subject file item sequences under the list method were identical to those under the item method with one crucial change: Instead of individual-word R and F instructions, a single instruction to forget appeared after the first 24 words (and 3 buffer words), and a single instruction to remember appeared after the second 24 words (and 3 buffer words).

The subjects were informed that they would study individual words presented one at a time at the center of the screen for 1 sec each, and then would receive an instruction halfway through the list telling them whether they would later be tested on this first half of the list. They were further told that after this instruction, they would study the second half of the list and then receive an instruction at the end of the list telling them whether they would later be tested on this second half of the list. In fact, all subjects received the forget instruction for the first half of the list and the remember instruction for the second half of the list, following standard practice using the list procedure (see MacLeod, 1998, for a review). The practice list also had a mid-list forget instruction and an end-of-list remember instruction.

Presentation of the study phase was identical to that for the item method except that (1) there were now only two list-half instructions instead of item-by-item instructions, and (2) the first three buffer words were now necessarily F words. The instructions at the middle and the end of the list were each 2,500 msec long, followed by a 500-msec blank period. All three tests were carried out exactly as under the item method using the identical test sequences.

RESULTS AND DISCUSSION

Table 1 displays the R and F data for both recall tests under both methods. A 2 × 2 mixed analysis of variance (ANOVA) was conducted on each recall test, with the between-subjects factor being method and the within-subjects factor being instruction.

First Recall Test

On the first recall test, overall recall of R words (.295) was significantly better than that of F words (.112) [$F(1,46)$

Table 1
Recall as a Function of Instruction and Method: Mean Proportion of Words Correct on the First Recall Test and Mean Proportion of Words Correct on the Second Recall Test (Conditional on Non-Recall on the First Test)

	First Recall Test				Second Recall Test			
	Remember		Forget		Remember		Forget	
Method	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Item	.280	.025	.056	.010	.011	.004	.018	.006
List	.311	.018	.168	.022	.010	.006	.030	.011

= 104.20, MS_e = .008, $p < .001$]. As well, words were recalled significantly better under the list method (.240) than under the item method (.168) [$F(1,46) = 12.03$, $MS_e = .010$, $p < .001$]. These two reliable main effects must be interpreted, however, in terms of their significant interaction [$F(1,46) = 5.17$, $p < .05$]. Follow-up t tests indicated that recall of R words did not differ under the two methods [$t(46) = 1.03$, $p > .30$], but that performance on the F words was significantly better under the list method than under the item method [$t(46) = 4.69$, $p < .001$].[1]

The results of the first recall test clearly confirm that directed forgetting occurs under both methods (see MacLeod, 1998, for a review). But the data also show that the R–F difference in recall is larger for the item method than for the list method, owing entirely to the F words' being less well forgotten under the list method. This confirms the pattern observed by Basden et al. (1993). Apparently, the delay in receiving the instruction to forget in the list method undermines intentional forgetting. This is probably because the words that will eventually be designated as the F words under the list method have already received considerable processing before the instruction can be acted on. In contrast, under the item method, processing of each item can be held in abeyance until the instruction appears, shortly after the item, resulting in little processing of F items under the item method.

Second Recall Test

Because the second recall test emphasized previously unrecalled F items, the data reported in Table 1 for this test are conditionalized data: These recall scores represent the proportion of previously unrecalled items that were recalled on this second test—often called "spontaneous recoveries." The first feature to note is that there was extremely little additional recall of F words on this second test, despite a fairly strong incentive in the form of 50¢ per additional F word. This implies that subjects were doing very little withholding of items on the initial recall test and provides little support for the notion that the usual R–F difference is the result of withholding the F words in response to demand characteristics.

On this second test, there was no effect of method, nor did method interact with instruction (both Fs < 1). However, although the effect was not significant, new F words (.024) were recalled slightly better than new R words (.011) [$F(1,46) = 3.14$, $MS_e = .001$, $p = .08$]. This is consistent with a very slight tendency to withhold F words on the first test, although it must be borne in mind that subjects were only trying to recall additional F words, not additional R words. Alternatively, this difference may simply reflect that there were more F words left to be recalled on the second test. Regardless, it is a very small difference that was not statistically significant. Directed forgetting as measured by recall is not the result of withholding F words under either method.

Recognition Test

Table 2 displays both the accuracies and the latencies for the recognition test, separately for each instruction and method. The latency data are for correct responses only. The distractor data correspond to correct rejections, so the false alarm rates for the item and list methods would be .162 and .217, respectively. I chose to report correct rejections primarily because of the latencies: There were very few false alarms, which would mean that latency means for false alarms would be quite untrustworthy. Note that the distractor conditions are included in the table for completeness but were not included in the analyses, which were again 2 × 2 (instruction × method) mixed ANOVAs.

For the accuracy data, R words (.811) were recognized significantly better than F words (.655) overall [$F(1,46) = 41.96$, $MS_e = .014$, $p < .001$]. As well, words were recognized significantly better under the list method (.800) than under the item method (.666) [$F(1,46) = 23.60$, $MS_e = .018$, $p < .001$]. Once again, though, these two reliable main effects must be interpreted in terms of their significant interaction [$F(1,46) = 34.47$, $p < .001$]. Follow-up t tests indicated that recognition of R words did not differ under the two methods [$t(46) = 0.19$, $p > .80$], but that performance on the F words was significantly better under the list method than under the item method [$t(46) = 7.18$, $p < .001$].

As Basden et al. (1993) highlighted, recognition typically shows an R–F difference under the item method but not under the list method (see MacLeod, 1998, for a review). Table 2 clearly shows that the advantage for R words over F words was in fact restricted to the item

Table 2
Recognition as a Function of Instruction and Method: Mean Proportion Correct and Mean Latency Correct

	Proportion Correct						Latency Correct					
	Remember		Forget		Distractor		Remember		Forget		Distractor	
Method	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Item	.814	.021	.517	.029	.838	.023	1,093	71.41	1,191	95.53	1,150	63.70
List	.808	.028	.793	.026	.783	.026	1,098	66.65	1,074	62.32	1,209	49.68

Note—The values for distractors correspond to correct rejections.

method. Furthermore, the absence of an R–F difference under the list method derived entirely from the F words. Under the list method, then, recall showed less of an effect for F words and recognition showed no effect.

Turning to the latency data, none of the effects were reliable (Fs < 1.44). There was no evidence of any speed–accuracy tradeoff: The latency pattern mimicked the accuracy pattern. Thus, the lower accuracy for F words under the item method does not derive from less effort or time spent on retrieval of these words. When Whetstone et al. (1996) examined recognition latencies in the list condition only, they did find R words to be reliably faster than F words (perhaps because they used much shorter lists); this result is also inconsistent with a speed–accuracy tradeoff. Indeed, in their study, the same pattern occurred in recall, where latency was measured from the keystroke of the first letter of each word.

Instruction Identification ("Tagging") Test

Table 3 displays both the accuracy and the latency data for the instruction tagging test, separately for each instruction and method. Latencies are for correct responses only, and the analyses again were 2 × 2 mixed (instruction × method) ANOVAs. Chance performance on this task is .5, because there were only two possible responses and passing was not permitted.

Consider first the accuracy data. The originally assigned instructions were remembered significantly better for R words (.682) than for F words (.586) overall [$F(1,46) = 11.60, MS_e = .019, p < .001$]. As well, study instructions were remembered significantly better under the item method (.678) than under the list method (.590) [$F(1,46) = 10.68, MS_e = .017, p < .005$]. This time, the interaction was not significant [$F(1,46) = 1.55, p > .20$]. However, the standard pair of t tests revealed that although the advantage for R word tagging under the item method was not significant [$t(46) = 1.58, p = .12$], this method difference was significant for F word tagging [$t(46) = 2.82, p < .01$].

The instruction tagging data cannot be taken as definitive, in that this test followed two recall tests and one recognition test on the same items, certainly permitting contamination. However, they are suggestive. It would appear that instruction tagging is fairly good for R words under both methods and for F words under the item method, but that subjects perform essentially at chance for F words under the list method. It is worth noting that Geiselman et al. (1983) also found poorer instruction identification for F words than for R words under the list method. This may be because remembering the instruction for F words in the list method constitutes remembering whether the word occurred in the first half of the list, whereas the task is to remember individual word-instruction pairs in the case of the item method.

Table 3
Instruction Tagging as a Function of Instruction and Method: Mean Proportion Correct and Mean Latency Correct

	Proportion Correct				Latency Correct			
	Remember		Forget		Remember		Forget	
Method	*M*	*SE*	*M*	*SE*	*M*	*SE*	*M*	*SE*
Item	.708	.027	.648	.034	1,104	81.48	1,426	131.52
List	.655	.020	.525	.027	1,846	237.35	1,875	182.99

The latency data for instruction tagging are quite noisy and shed little light on the processing. Only the effect of method was significant, with overall responding slower under the list method (1,860 msec) than under the item method (1,265 msec) [$F(1,46) = 7.82, MS_e = 1{,}088{,}042, p < .01$]. This may again reflect the different strategies of trying to decide which half of the list an item was in for the list method versus trying to remember the individual word-instruction pairing for the item method. There was a trend toward a significant effect of instruction [$F(1,46) = 2.68, p = .11$], with faster classification of R words (1,475 msec) than of F words (1,651 msec). Although the interaction was not significant [$F(1,46) = 1.86, MS_e = 277{,}004, p = .18$], an R–F difference appears to have been present only for the item method. This observation is supported by t tests: The R–F difference was nonsignificant under the list method [$t(24) = 0.15, p > .80$], but significant under the item method [$t(24) = 3.42, p < .005$]. For both methods, slower responses were associated with less accurately labeled F words, inconsistent with a speed–accuracy tradeoff.

GENERAL DISCUSSION

The present study replicates and extends existing findings, in accord with its first goal. It is clear that an R–F difference emerges in initial free recall under both methods. However, the effect is reliably reduced under the list method, the difference stemming from better recall of the F words, not worse recall of the R words. This replicates the findings of Basden et al. (1993), the only previous direct comparison of the methods. In recognition, only the item method shows reliable directed forgetting; there is no R–F difference under the list method. Once again, this is because of the F words, which "rise" to the level of the R words under the list method. This finding also replicates that of Basden et al. (1993), alleviating possible concerns about ceiling-level performance on the R words in their study.

Going beyond Basden et al. (1993), the recognition data also show that the reason for the now familiar accuracy patterns does not lie in a speed–accuracy tradeoff. Subjects were actually slower on F words, as if they were trying harder but simply not succeeding in retrieving them. A more direct test of the effort–motivation–demand characteristics account of directed forgetting comes from the second recall test. Here it is clear that subjects were not withholding F words during initial re-

call. When motivated by a reward of 50¢ for each additional F word, subjects recalled almost no additional words. Clearly, the R–F difference reflects differential processing, not a strategic decision to suppress F words.

Finally, in her review, Johnson (1994) noted that few studies had examined discrimination of study cues. Under the list method, Geiselman et al. (1983) found that subjects could not discriminate R words from F words, nor could they use the F cue as the basis for retrieval (confirmed by Basden et al., 1993). Under the item method, MacLeod (1975, Experiment 2) showed that providing subjects with the original cues conferred no advantage over not providing these cues. When MacLeod (1975) specifically required subjects to retrieve study cues, their performance was reliably above chance.

The present results on memory for instructions are generally consistent with this prior work. Subjects under the item method did show fair discrimination of study cues. The same was true for the R words under the list method, but apparently not for the F words. In agreement with Johnson (1994), my strong impression is that more data will be required in order to define the role of cue retention in directed forgetting, particularly data not contaminated by other prior tests.

In comparing the item and list methods directly, the present data confirm and extend the key findings in the directed forgetting literature all in one place. Additionally, the familiar criticism of directed forgetting—that the R–F difference is the result of demand characteristics and reduced motivation to retrieve F words—is laid to rest both by the latency data and, more directly, by the data on the second recall test. With R. A. Bjork (1989), Basden et al. (1993), and Johnson (1994), I am in agreement that the two methods lead to reliably different patterns of data, with the recall effect stronger under the item method and the recognition effect present only under the item method. Both of these results hinge on poorer forgetting of F words under the list method.

The patterns reported here and in the past are entirely consistent with the hypothesis (Basden et al., 1993; R. A. Bjork, 1989) that the two methods produce their effects differently. The stronger effect under the item method is the consequence of on-line processing being suspended until the instruction appears, at which point processing is cancelled for F words and initiated for R words. Thus, as R. A. Bjork (1972) and Johnson (1994) have argued, F words are simply not as well encoded, and this will be evident on any retention test. This has come to be called the selective rehearsal account. In contrast, under the list method, much of the on-line processing has already occurred for words in the first half of the list by the time the forget instruction appears, so any resulting R–F difference is more likely to be the result of post-processing restructuring of the material in memory. Retrieval inhibition (see E. L. Bjork, R. A. Bjork, & Anderson, 1998) provides one such potential restructuring mechanism and is consistent with much of the existing literature. Taken together, these two accounts provide a coherent picture of directed forgetting.

REFERENCES

BASDEN, B. H., BASDEN, D. R., COE, W. C., DECKER, S., & CRUTCHER, K. (1994). Retrieval inhibition in directed forgetting and posthypnotic amnesia. *International Journal of Clinical & Experimental Hypnosis*, **42**, 184-203.

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, & Cognition*, **19**, 603-616.

BJORK, E. L., BJORK, R. A., & ANDERSON, M. C. (1998). Varieties of goal-directed forgetting. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 103-137). Mahwah, NJ: Erlbaum.

BJORK, R. A. (1972). Theoretical implication 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., & WOODWARD, A. E., JR. (1973). Directed forgetting of individual words in free recall. *Journal of Experimental Psychology*, **99**, 22-27.

BLOCK, R. A. (1971). Effects of instructions to forget in short-term memory. *Journal of Experimental Psychology*, **89**, 1-9.

COE, W. C. (1978). The credibility of posthypnotic amnesia: A contextualist's view. *International Journal of Clinical & Experimental Hypnosis*, **26**, 218-245.

DAVIS, J. C., & OKADA, R. (1971). Recognition and recall of positively forgotten items. *Journal of Experimental Psychology*, **89**, 181-186.

ELMES, D. G., ADAMS, C. A., III, & ROEDIGER, H. L., III (1970). Cued forgetting in short-term memory: Response selection. *Journal of Experimental Psychology*, **86**, 103-107.

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.

GEISELMAN, R. E., RABOW, V. E., WACHTEL, S. L., & MACKINNON, D. P. (1985). Strategy control in intentional forgetting. *Human Learning*, **4**, 169-178.

GOLDING, J. M., & LONG, D. L. (1998). There's more to intentional forgetting than directed forgetting: An integrative review. In J. M. Golding & C. M. MacLeod (Eds.), *Intentional forgetting: Interdisciplinary approaches* (pp. 59-102). Mahwah, NJ: Erlbaum.

GOLDING, J. M., & MACLEOD, C. M. (Eds.) (1998). *Intentional forgetting: Interdisciplinary approaches*. Mahwah, NJ: Erlbaum.

GRAVES, R., & BRADLEY, R. (1987). Millisecond interval timer and auditory reaction time programs for the IBM PC. *Behavior Research Methods, Instruments, & Computers*, **19**, 30-35.

GRAVES, R., & BRADLEY, R. (1988). More on millisecond timing and tachistoscope applications for the IBM PC. *Behavior Research Methods, Instruments, & Computers*, **20**, 408-412.

GROSS, A. E., BARRESI, J., & SMITH, E. E. (1970). Voluntary forgetting of a shared memory load. *Psychonomic Science*, **20**, 73-75.

JOHNSON, H. M. (1994). Processes of successful intentional forgetting. *Psychological Bulletin*, **116**, 274-292.

MACLEOD, C. M. (1975). Long-term recognition and recall following directed forgetting. *Journal of Experimental Psychology: Human Learning & Memory*, **1**, 271-279.

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

MUTHER, W. S. (1965). Erasure or partitioning in short-term memory. *Psychonomic Science*, **3**, 429-430.

ORNE, M. T. (1962). On the social psychology of the psychological experiment. *American Psychologist*, **17**, 776-783.

THORNDIKE, E. L., & LORGE, I. (1944). *The teacher's word book of 30,000 words*. New York: Columbia University, Teachers College Press.

WEINER, B. (1968). Motivated forgetting and the study of repression. *Journal of Personality*, **36**, 213-234.

WHETSTONE, T., CROSS, M. D., & WHETSTONE, L. M. (1996). Inhibition, contextual segregation, and subject strategies in list method directed forgetting. *Consciousness & Cognition*, **5**, 395-417.

WOODWARD, A. E., JR., & BJORK, R. A. (1971). Forgetting and remembering in free recall: Intentional and unintentional. *Journal of Experimental Psychology*, **89**, 109-116.

NOTE

1. There is one peculiar feature to the initial recall data. A subject could rehearse each R word for at least 3 sec under the item method (its instruction duration) but for only 1 sec under the list method (its presentation duration). Why, then, is recall of R words not greater under the item method? At least two possibilities come to mind: (1) Subjects may have rehearsed an R word for only about 1 sec regardless of opportunity, and (2) for R words, greater average recency under the list method may offset greater rehearsal opportunity under the item method. This matter warrants further investigation.

(Manuscript received November 24, 1997;
revision accepted for publication March 26, 1998.)