Journal of Experimental Psychology:
Learning, Memory, and Cognition
1989, Vol. 15, No. 1, 13-21

0278-7393/89/$00.75

Directed Forgetting Affects Both Direct and Indirect Tests of Memory

Colin M. MacLeod
University of Toronto, Scarborough Campus, Scarborough, Ontario, Canada

In two experiments, subjects were instructed either to remember or to forget each word. Following study, two tests were given, one a direct test of memory requiring conscious recollection of the study list and the other an indirect test that could be performed without awareness of the study list. In Experiment 1, subjects recognized more *remember* than *forget* words (direct test) and completed more *remember* than *forget* fragments (indirect test) on both immediate and 1-week delayed tests. In Experiment 2, subjects showed superior recall (direct test) and greater repetition priming in lexical decision (indirect test) for remember than for forget words. The consistent directed forgetting effect on both types of tests is in accord with the idea that forget items are inhibited at the time of retrieval and that retrieval manipulations, unlike elaboration manipulations at encoding, affect direct and indirect tests in similar ways.

In their review chapter, Johnson and Hasher (1987) identified the relations among different memory tasks as a central problem for memory researchers to address. One distinction they particularly emphasized was that between direct and indirect measures of memory. In the words of Johnson and Hasher (1987), "*Direct* memory tasks (free recall, cued recall, recognition) require conscious expressions of remembering; *indirect* memory tasks (e.g., perceptual identification, homophone spelling, word completion, skill learning) do not" (p. 641). Johnson and Hasher went on to survey the considerable research effort that had been directed at this distinction.

In his recent review article, Schacter (1987) clearly defined a distinction between implicit and explicit tests of memory: "Implicit memory is revealed when previous experiences facilitate performance on a task that does not require conscious or intentional recollection of those experiences; explicit memory is revealed when performance on a task requires conscious recollection of previous experiences" (p. 501). As his review convincingly demonstrated, this dichotomy is attracting an increased amount of research attention.

Most recently, Richardson-Klavehn and Bjork (1988) focused their review chapter almost exclusively on this distinction. They opted for the direct–indirect terminology, as have others (Yaniv & Meyer, 1987), arguing that these terms are more theoretically neutral than the alternatives and that they avoid confusing tests with theoretical entities underlying test performance. Their distinction is operational in that subjects are told that events of the past are being queried on direct tests, whereas no reference is made to the past on indirect tests. (Note that the same operational distinction could be applied to the implicit–explicit dichotomy.) In terms of the Johnson and Hasher (1987) definition, Richardson-Klavehn and Bjork (1988) put the weight on what is "expressed," rather than on what is "conscious."

The hallmark direct tests of memory are recall and recognition, in which the rememberer must knowingly retrieve some specified episodes from the past. The most intensively investigated indirect tests of memory are those that rely on repetition priming, including lexical decision, perceptual identification, and word-fragment completion (cf. Schacter, 1987, p. 507). Repetition priming takes place when a prior occurrence of a stimulus, usually a single word, facilitates subsequent processing of the same stimulus when it appears again, without any requirement that the rememberer be aware of the repetition. Thus, having seen the word COCONUT recently (as opposed to not having seen it) speeds the decision about whether COCONUT is a word (cf. Scarborough, Gerard, & Cortese, 1979), enhances the probability of identifying a degraded version of the word COCONUT (Jacoby, 1983), and increases the likelihood that the fragment _O_O_UT will be successfully completed with the word COCONUT (Tulving, Schacter, & Stark, 1982).

One of the most compelling findings has been the apparent dissociation of the two types of memory tests in the face of certain manipulations. As one illustration, variation in degree of elaborative processing has large and consistent effects on recall and recognition but leaves repetition priming virtually unaffected (e.g., Graf, Mandler, & Haden, 1982; Jacoby & Dallas, 1981; Schacter & Graf, 1986). As a second illustration, changes in modality of presentation or other surface features of the stimuli strongly affect repetition priming but typically do not affect recall or recognition (Graf, Shimamura, & Squire, 1985; Jacoby & Dallas, 1981; Roediger & Blaxton, 1987). Such differences suggest a sort of "double dissociation" between direct and indirect tests of memory.

Yet Schacter (1987) and Richardson-Klavehn and Bjork (1988) also listed a number of manipulations to which the two types of tests respond similarly. As one illustration, Jacoby (1983) showed that increasing the proportion of words drawn from a previously studied list helped both perceptual identification and recognition. As a second illustration, several studies (e.g., McKoon & Ratcliff, 1979; Schacter & Graf,

This research was supported by Grant A7459 from the Natural Sciences and Engineering Research Council of Canada.

For their assistance in collecting the data, I thank Douglas Martin (Experiment 1) and Samuel Bruser (Experiment 2). I also thank Peter Graf, Robert Greene, Reed Hunt, Henry Roediger, and an anonymous reviewer for their helpful comments on a previous draft.

Correspondence concerning this article should be addressed to Colin M. MacLeod, Division of Life Sciences, University of Toronto, Scarborough Campus, Scarborough, Ontario, Canada M1C 1A4.

1986) have demonstrated that testing a target word in the context of a cue word with which it was studied improves performance on both direct and indirect tests. For a variety of reasons, similarities such as these have received less attention than have differences between the two types of tests.

There is an interesting feature to these illustrations. All three reviews build a very strong case that encoding manipulations affect the two types of memory tests differently. Thus, experimentally varying degree of elaboration or surface features at the time of study produces apparent dissociations. However, less has been said about what links the cases in which the two types of tests respond in the same way. One potentially viable hypothesis is that retrieval manipulations may affect the two classes of memory tests similarly. Restoration of study context, whether at the level of the single item or of the list, is one example of a retrieval effect evident on both types of tests.

Putting forth a retrieval hypothesis may seem an odd thing to do in that the difference between direct and indirect tests is necessarily one having to do with retrieval. After all, the two types of tests are distinguished by what the subject is told at the time of the test. Yet information must be retrieved from memory for any test, and selecting which information to recover and how to go about recovering it are driven in part by the subject's interpretation of the test demands. There is a large body of evidence suggesting that direct and indirect tests focus on different aspects of information in memory. What is being suggested here is that the choice of retrieval scheme may be less affected by whether the test is direct or indirect.

The research to be reported was partly motivated by this retrieval hypothesis. The technique adopted was the directed forgetting procedure (for reviews, see Bjork, 1972; Epstein, 1972). Although there have been a number of variations of this procedure published, one of the most common involves following the presentation of each item in a study list with an instruction to remember or to forget that item (e.g., Bjork & Woodward, 1973; Horton & Petruk, 1980; MacLeod, 1975; Wetzel, 1975). Prior to study, subjects are told that they will only be tested on the to-be-remembered words. At test, when asked to remember both the *remember* and the *forget* words, subjects show markedly better memory for the remember words. The effects of forget instructions are well established in both recall and recognition (e.g., Davis & Okada, 1971; MacLeod, 1975; Wetzel, 1975), but to date no research has examined the impact of directed forgetting on any indirect test of memory.

Theoretically, three mechanisms have been put forth to explain the directed forgetting effect. One emphasizes operations at encoding, suggesting that selective rehearsal favors the remember words, so that only the remember words are elaboratively processed (e.g., Bjork, 1972; Woodward, Bjork, & Jongeward, 1973). A second mechanism emphasizes storage and retrieval, suggesting the formation of distinct sets of remember and forget items in memory, with selective search of the higher priority remember set (e.g., Epstein, 1972). Under the third, more recent, view, a repression-like process at the time of retrieval prevents forget items from being recovered. The result is retrieval inhibition of the forget items (e.g., Geiselman, Bjork, & Fishman, 1983).

The selective search idea was intended primarily to apply to the situation where the subject is trying not to recall forget items. Yet most directed forgetting experiments require subjects to recall (or recognize) both remember and forget items and still display a quite robust directed forgetting effect. The selective encoding and the retrieval inhibition accounts have been put forward to explain why performance on remember items is better than that on forget items in such a situation.

Under the encoding account, if subjects are told to remember an item, elaborative processing is instituted; if subjects are told to forget an item, processing ceases (e.g., MacLeod, 1975; Wetzel & Hunt, 1977; Woodward et al., 1973). As stated by Wetzel (1975), the cue "directs either discontinued or more elaborative processing" (p. 556). Under the retrieval account, if the instruction is *remember*, further processing helps to facilitate later retrieval; if the instruction is *forget*, the item is repressed, obstructing retrieval at the time of test (e.g., Geiselman & Bagheri, 1985; Geiselman et al., 1983; Weiner & Reed, 1969). As Geiselman and Bagheri (1985) put it, "A command to forget initiates a process that serves to inhibit or block access routes to the target items" (p. 62).

In bringing together directed forgetting and the indirect–direct test distinction, the following predictions can be derived. Consider first the encoding view. Given the numerous findings that elaboration affects direct but not indirect tests of memory, then the encoding view holds that directed forgetting should not affect indirect test performance. Now consider the retrieval view. As pointed out, there is emerging evidence (e.g., Jacoby, 1983; Schacter & Graf, 1986) that retrieval manipulations affect both types of tests similarly. If directed forgetting leads to retrieval inhibition of forget items, the logical prediction from the retrieval view is that performance on forget items should be worse than that on remember items for both indirect and direct tests.

Testing these predictions was the major goal of the two experiments reported here. To maximize convergence, each experiment examined a different combination of direct and indirect tests: recognition and fragment completion in Experiment 1 and recall and lexical decision in Experiment 2. The common element in the two experiments was the directed forgetting manipulation applied to the items at the time of study.

Experiment 1

The first experiment followed as closely as possible the design used by Tulving et al. (1982), but with the addition of the directed forgetting manipulation. Subjects studied a long list of low-frequency words, half of which were followed by an instruction to forget and half by an instruction to remember that word. Then subjects were given two immediate tests: a word-recognition test and a word-fragment completion test for half of the studied remember and forget words. One week later, subjects returned for recognition and fragment tests on the other half of the studied words. To discourage any explicit recollection strategy on the indirect test, the fragment-completion test was represented to subjects as unrelated to the major recognition experiment.

Tulving et al. (1982) found substantial priming for studied, as compared with unstudied, words on the fragment-completion test. They also reported two findings suggesting that direct and indirect test performances were dissociated. First, there was a large decline across retention interval in recognition but not in priming on the fragment-completion test. Second, when recognition preceded fragment completion, performances on the two tests were stochastically independent. One motivation behind the present study was to determine how the addition of the directed forgetting manipulation would affect the pattern over the two types of tests.

If directed forgetting affects the degree of elaborative processing during study, there should be a consistent directed forgetting effect in recognition at both intervals, as previously demonstrated (MacLeod, 1975, Experiment 2). More novel, priming in fragment completion should be unaffected by the manipulation at either retention interval. That is, there should be consistent and equivalent priming for both remember and forget words, because both sets of words have been activated previously and elaborative processing is not critical to performance on indirect tests.

If directed forgetting affects item retrievability, perhaps by inhibiting forget items at the time of test, again it is reasonable to expect remember words to be better recognized than forget words. However, because it appears that retrieval manipulations affect direct and indirect tests similarly, the prediction is that remember words should also show more priming than forget words on the indirect fragment-completion test. Thus, although not testable with respect to the direct test results because they make the same prediction, the encoding and retrieval accounts do make a differential prediction with respect to the indirect test results.

Method

Subjects. The participants were 32 undergraduate students at the Scarborough Campus of the University of Toronto. They were paid $10 each for their participation in the two sessions of the experiment and were tested individually.

Materials. The 192 critical words were those used by Tulving et al. (1982) and contained in the appendix to their article. These were divided into eight sets of 24 words each for counterbalancing purposes. Each set was chosen so that word-initial consonants were roughly equivalent across sets. A further 6 common words comprised the buffer set, whose purpose is described below.

Procedure. The list presentation phase was carried out under control of an Apple IIe microcomputer. First, four of the eight 24-word subsets of items were selected for study. Two of these subsets were designated as the forget words, one subset to appear on the immediate test and the other on the delayed test. The other two subsets were designated as the remember words, one assigned to each test. The remaining four subsets were not presented for study and were reserved for use as distractors on the two tests. Over the 32 subjects, each 24-word subset appeared equally often (a) as remember versus forget, (b) as studied set versus distractor set, and (c) on the immediate test versus the delayed test.

Prior to list presentation, subjects read the following set of instructions on the screen.

> This is a memory experiment. You will be seeing a long list of unusual words, one word at a time. Each word will appear for one second and will be followed by an instruction lasting three seconds. You will be instructed either to remember or to forget each word. Because the list is very long—100 words—and you will only be tested on the half you are told to remember, it is a good idea to follow the instructions. Try to remember the "remember" words for the test that will follow the list. Here is how the word by word instructions work. After each word has been shown for one second, either "RRRR" or "FFFF" will appear at the center of the screen for three seconds. If "RRRR" appears, try to remember the word you just saw—it will be on the test. If "FFFF" appears, you need not remember that word—it will not be on the test. The instructions are there to help you select the words to learn from the long and difficult list.

After reading these procedural instructions, the subject called the experimenter, who summarized the instructions, answered any questions, and initiated the study phase. A 1-s blank preceded a 1-s display of "*****" at the center of the screen, and then list presentation began. Details of list presentation were as follows. List words and instructions were presented in upper case letters at the center of the screen, with instructions presented in inverse field. There were actually 102 words, the first 3 and last 3 being buffer words. These were always the same 6 words for every subject, and all of them were assigned remember instructions to reduce the impact of serial position effects on the 96 critical words. The set of critical words was presented in a different random order for each subject.

Once the list had been presented, the immediate test phase began. The order of administration of the two tests was counterbalanced, with 16 of the subjects receiving each test order. The recognition test consisted of 96 words, 24 studied remember words, 24 studied forget words, and 48 unstudied words. Subjects were asked to circle any words they remembered seeing in the study phase, regardless of the instruction those words had been assigned during study. They were given as long as they wished to complete this test. The studied and unstudied words appeared in the same random order on the test sheet for all subjects.

Following (or preceding) the recognition test, subjects did a fragment-completion test. It is by now clear that the manner in which an indirect test of memory is presented to the subject can be crucial (cf. Graf & Mandler, 1984). Consequently, the fragment test was characterized as a problem-solving task that was being administered to assist another professor in his research and that was unrelated to the memory experiment. The test contained 96 fragments of words and used the same letter-omission scheme as Tulving et al. (1982). These were exactly the same 96 words that appeared on the corresponding recognition test, but in fragmentary form and in a different random order. Before the test, three samples were given—_ R _ A _, FR _ _ M _ N _, and C _ _ LIO _ _—and their solutions were provided—BREAD, FRAGMENT, and CALLIOPE. Subjects were warned that the items were quite hard and were encouraged to do their best and not to give up. They were given as much time as they desired to complete the test.

At the end of this session, subjects were scheduled for a second session 1 week later. They were told that it was a different experiment and that the aim was to correlate performances in the two studies. They were reminded by telephone a day or two prior to their scheduled return, and all subjects returned for the second session. In this session, the remaining halves of the originally studied items were tested. The recognition and fragment-completion tests were administered to each subject in the same order in the second session as in the first, and under the same general procedure and instructions. The set of words was completely different from those on the tests in the first session. Subjects were debriefed following the second of their two tests in this second session.

Results

Recognition. Table 1 presents the data from the recognition tests. The false alarm rates on both immediate and delayed tests were all around 10%, regardless of test order. A 2 × 2 analysis of variance (ANOVA), with factors of testing order (fragment completion first or recognition first) and retention interval (immediate vs. delayed) revealed no significant differences in false alarm rates. The only tendency was for false alarm rate to rise slightly with retention interval, as revealed by the main effect of retention interval, $F(1, 30) = 2.78$, $MS_e = .004$, $p > .10$. Both the effects of order of testing and the interaction were nonsignificant, both $Fs < 1.18$.

The data of principal concern were the hit rates—the proportions of correctly identified studied words. A 2 × 2 × 2 ANOVA, with factors of retention interval (immediate versus delayed), order of testing (recognition first versus fragment completion first), and item instruction ("remember" versus "forget"), was conducted. The main effect of order of testing was nonsignificant, and order of testing did not significantly interact with either of the other variables. All of the *F*s involving order of testing were less than unity except for the nonsignificant interaction with retention interval, $F(1, 30) = 2.90$, $MS_e = .029$, $p > .10$, suggesting a tendency for hit rate in recognition to be higher on the immediate test if recognition was tested first, as might be expected.

Of primary interest were the effects of item instruction and retention interval. Both main effects were significant: for instruction, $F(1, 30) = 101.28$, $MS_e = .010$, $p < .001$, and for retention interval, $F(1, 30) = 97.44$, $MS_e = .029$, $p < .001$. There was also a marginally significant interaction of retention interval with instruction, $F(1, 30) = 4.03$, $MS_e = .008$, $.05 < p < .10$, indicating that the directed forgetting effect was somewhat larger on the immediate test than on the delayed test.

The 15% difference on the delayed test was smaller than the 22% effect on the immediate test, but both were quite large and significant. Indeed, the interaction may well be artificial, in the sense that performance on forget items after the delay was approaching the performance floor. What is important is that a substantial directed forgetting effect appeared on both the immediate and the delayed tests, replicating previous studies (e.g., MacLeod, 1975, Experiment 2).

Fragment completion. Table 2 presents the data from the fragment-completion tests. A 2 × 2 ANOVA showed that the base rate of completing fragments of unstudied words did not differ across retention intervals, $F(1, 30) = 1.66$, $MS_e = .003$; nor did order of testing interact with retention interval, $F < 1$. However, the base rate did depend strongly on which test was given first, $F(1, 30) = 33.59$, $MS_e = .014$, $p < .001$. Because the unstudied fragments were also the recognition distractors, considerably more unstudied fragments were completed if the fragment test followed the recognition test (.28) than if the fragment test was given first (.11). This is test-based priming. The different base rates should be kept in mind in the discussion of priming that follows.

Of central concern was the pattern for previously studied words. A 2 × 2 × 2 ANOVA like that performed on the recognition data demonstrated a significant main effect of order of testing, $F(1, 30) = 25.59$, $MS_e = .048$, $p < .001$, and a significant interaction of order of testing with retention interval, $F(1, 30) = 9.57$, $MS_e = .006$, $p < .01$. The other two interactions involving order of testing were both nonsignificant, both $Fs < 1$. For studied items as for unstudied (base-rate) items, fragment-completion performance was better when the recognition test came first (.42) than when the fragment completion test came first (.23). Still, priming was of roughly the same magnitude in both cases (for recognition first, a .14 difference between studied and unstudied; for fragment completion first, a .12 difference). The interaction indicated that, if recognition was the first test, there was no decline in fragment completion over retention interval (.43 to .42); if fragment completion was the first test, performance did decline over retention interval (.28 to .18). Put simply, an immediately preceding recognition test is very beneficial to a fragment-completion test that follows.

Examination of the effects of instruction and retention interval confirmed two significant main effects without interaction. For retention interval, $F(1, 30) = 11.23$, $p < .01$, performance was better on the immediate test (.35) than on the delayed test (.30). The implication of this finding with respect to those of Tulving et al. (1982) will be considered in more detail shortly. Most critical, the main effect of instruction was significant, $F(1, 30) = 9.03$, $MS_e = .009$, $p < .01$. There was more priming for remember words (.35) than for forget words (.30). Notice that this was true whether the fragment-completion test was administered first or second, assurance that the directed forgetting effect was not simply a consequence of contamination from a preceding recognition test. Instruction and retention interval did not interact, $F < 1$.

The key finding is that there was a directed forgetting effect on the fragment-completion test at both retention intervals and for both testing orders. The probability of correctly completing a studied word was greater for remember words; that is, there was more priming for remember words than for forget words. Although not a large effect in terms of difference scores, it was reliable and in keeping with the pattern observed in recognition.

Table 1
Experiment 1: Proportion of Words Correct on the Immediate and Delayed Recognition Tests

Retention interval	Types of words: Remember	Forget	False alarms
Recognition test first			
Immediate	.76	.55	.08
Delayed	.39	.22	.10
Fragment completion test first			
Immediate	.69	.46	.11
Delayed	.39	.26	.15

Discussion

The recognition data were thoroughly consistent with previous findings (MacLeod, 1975, Experiment 2), both in terms

Table 2
Experiment 1: Proportion of Words Correct on the Immediate and Delayed Fragment-Completion Tests

	Types of words		
Retention interval	Remember	Forget	Unprimed
	Recognition test first		
Immediate	.46	.40	.26
Delayed	.44	.41	.30
	Fragment completion test first		
Immediate	.31	.24	.10
Delayed	.20	.16	.11

of the magnitude of the directed forgetting effect and its persistence over retention interval. Clearly, there was a large effect on this direct memory test. What is critical for present purposes is that directed forgetting also affected the indirect memory test, fragment completion. At both retention intervals, studied items showed higher completion rates than did unstudied items, replicating the priming reported by Tulving et al. (1982). More to the point, there was also differential priming favoring remember over forget words.

It is useful to examine how the present results relate to those of Tulving et al. (1982). On the surface, the significant effect of retention interval on priming in fragment completion appears to be a failure to replicate. Recall that Tulving et al. observed no loss on their fragment-completion test; coupled with the large forgetting evident on the recognition test, they took this as evidence for a dissociation between the tests. But there is a difference between the two studies that may be crucial. Tulving et al. tested after 1 hr and after 1 week. In the present experiment, tests were administered immediately (within 5 min of study) and after 1 week. Recent work by Sloman, Hayman, Ohta, Law, and Tulving (1988) demonstrated a rapid drop in priming in the first 5 min, in addition to more forgetting over a week than had been reported previously. Thus, there is forgetting for primed fragment completion in this interval, and it is accentuated when the first test closely follows the study episode.

Tulving et al. (1982) reported a second result that they claimed was also in line with a dissociation of performance on the two tests. Specifically, when fragment completion followed recognition (but not when test order was reversed), performance on the two tests was stochastically independent. Similar analyses were carried out on the present data, producing values for both test orders equivalent to those that Tulving et al. obtained when fragment completion preceded recognition. Thus, there was no support for the stochastic independence argument.

In sum, the present results seem most consistent with a retrieval-based explanation of directed forgetting. Under the encoding view (e.g., MacLeod, 1975; Wetzel, 1975; Wetzel & Hunt, 1977; Woodward et al., 1973), remember words, but not forget words, are elaboratively processed. We know, though, that elaboration effects should appear only on direct tests (Graf & Mandler, 1984; Roediger & Blaxton, 1987). Yet directed forgetting affected both recognition and fragment completion in the same way here (although possibly not to the same extent). This is difficult to handle under the encoding view but is consistent with the retrieval view that inhibition of forget items at the time of retrieval will influence both direct and indirect tests. Although provocative, this is only a single finding that requires convergence.

Experiment 2

The major purpose of Experiment 2 was to extend the findings of Experiment 1 to different direct and indirect tests, providing converging evidence. The large effect of directed forgetting on free recall is widely replicated (e.g., Bjork, 1972; Epstein, 1972), so recall would seem to be a suitable benchmark direct test of memory. The main aim here was to choose a rather different indirect test from that used in Experiment 1. The choice was repetition priming in lexical decision (cf. Scarborough et al., 1979). The logic is similar to that of priming in fragment completion—a previous study of a word facilitates the decision on a later test about whether that same item is a word or a nonword.

When a variable affects performance on an indirect test, an immediate concern is that subjects may have adopted an explicit retrieval strategy. Indeed, this argument has been used to explain occasional instances of elaboration effects on purportedly indirect tests, which normally do not show such effects (e.g., Squire, Shimamura, & Graf, 1987). Unfortunately, there is at present no way to know when to invoke this account. A major advantage of using repetition priming in lexical decision is that this task has been identified as indirect (Richardson-Klavehn & Bjork, 1988; Schacter, 1987). Furthermore, the dependent variable in the task is response time, which would seem to make it more difficult for an explicit retrieval strategy to contaminate this task. This was one reason for selecting lexical decision as the indirect test in Experiment 2. Of course, lexical decision is also procedurally quite different from fragment completion, increasing the generalizability of the results.

There is a large body of evidence showing that varying the extent of elaborative processing markedly affects recall (e.g., Craik & Tulving, 1975). Recently, investigators have begun to examine whether varying type of processing also affects repetition priming in lexical decision. The basic message seems to be that it does not (e.g., Smith, MacLeod, Bain, & Hoppe, 1987), although the evidence is somewhat limited. This fits the pattern: Direct tests are affected but indirect tests are unaffected by manipulating elaboration. Thus, directed forgetting should affect recall but not repetition priming in lexical decision if the encoding view is correct.

Under the retrieval view of directed forgetting, both recall and repetition priming should be affected by directed forgetting. Forget items will be repressed at the time of test, thereby inhibiting retrieval regardless of the type of test. If the pattern of Experiment 1 generalizes, the results from these two different pairs of direct and indirect tests should correspond, with all four showing a directed forgetting effect.

Method

Subjects. Thirty-three students in residence at the Scarborough Campus of the University of Toronto volunteered to take part in the study. They were randomly assigned to the eight counterbalancing conditions in roughly equal numbers and were tested individually.

Materials. The stimuli consisted of the 80 words and 80 nonwords listed in the Appendix to this article, plus six buffer words and 15 practice items. All of the words were relatively common nouns. A single letter in each of a further set of common nouns was changed to create a pronouncable nonword (e.g., the word VICTIM became the nonword VISTIM). Every stimulus was 4–7 letters in length. As the Appendix shows, the words were divided into two subsets of 40 items. Half of the subjects studied Set A prior to the recall and lexical decision tests; the other half of the subjects studied Set B. Each of these 40-word sets was further divided into two subsets of 20 (the two columns), with one subset remember and one subset forget during study. Item instructions were also counterbalanced across subsets.

Procedure. The study phase was controlled by an Apple II+ microcomputer. The subject saw a list of 46 words, each with an instruction to remember or forget immediately after the word. The first 3 and the last 3 words all had remember instructions and were identical for every subject. These were buffer items to reduce serial position effects. The critical 40 words were evenly split between remember and forget instructions, with the order of items randomized anew for every subject.

Words were presented at the center of the monitor screen in uppercase letters. The timing sequence consisted of a 250-ms blank, the word for 2 s, another 250-ms blank, and the instruction for 3 s. The instruction, also centered on the screen, was either **RRRRRR** or **FFFFFF**, signifying remember and forget, respectively. Subjects were told that the list was long and that they would only be tested on the remember items, so that it was to their benefit to follow the instructions.

Following study, every subject did both a free-recall test and a lexical-decision task. Half of the subjects did free recall first; the other half did lexical decision first. Those subjects who had lexical decision first were told that this was a separate experiment unrelated to the list they had just learned and was included to fill the time between list study and the recall test. Those who had lexical decision second were told simply that it was a separate, unrelated experiment. Debriefing followed the second test.

In the free-recall test, subjects were given a blank sheet of paper and asked to write down as many of the 46 studied words as they could remember. They were encouraged not to worry about the instruction each word had been assigned, but to recall as many words as possible—both remember and forget—guessing when unsure. They were given as much time as they wanted.

The lexical-decision task followed quite standard methodology. There were 15 practice trials, made up of seven words and eight nonwords not included in the experimental sets. A trial began with a 250-ms warning of ****, followed by a 250-ms blank prior to the item. Each item was presented in uppercase letters at the center of the screen until 500 ms after the subject pressed a key to indicate the chosen response. There was a 250-ms blank period before the next warning stimulus.

Following practice, any questions were answered and the experimental trials began. The 160 experimental trials were made up of 20 studied forget words, 20 studied remember words, 40 unstudied words, and 80 nonwords. The procedure for experimental trials was identical to that for practice trials, except that item offset coincided with the subject's keypress response, and the intertrial interval was 1 s. The words and nonwords were presented in a different random order to every subject. Subjects were given the option of taking a break for as long as they wished halfway through the trials. They were encouraged to respond as rapidly as possible while avoiding errors.

Results

Free recall. Table 3 summarizes the free-recall data. A 2 × 2 ANOVA examined the effects of order of testing (free recall first or lexical decision first) and item instruction (remember versus forget). Neither the main effect of order of testing nor the interaction of order with instruction was significant, both Fs < 1. However, the effect of item instruction was highly significant. The proportion of correctly recalled remember items (.35) was significantly greater than the proportion of correctly recalled forget items (.05), $F(1, 31) = 82.50$, $MS_e = .017$, $p < .001$. This replicates the widely observed effect of directed forgetting on recall. Of the 328 response words, 55 (17%) were extralist intrusions. Intrusion rate was higher after lexical decision (38, 12%) than before it (17, 5%), as might be expected.

Lexical decision. Table 4 presents the lexical-decision data, both mean latencies and error proportions. Although shown in the table, the nonword data were not included in the analyses because they were not pertinent to the main questions. Thus, two ANOVAs were carried out, one on errors and one on latencies. These were 2 × 3 analyses, combining the three types of words (remember, forget, and new) with the two orders of testing (lexical decision first versus recall first).

For errors, the order of testing had no influence, $F < 1$. However, the main effect of type of word was significant. The probability of incorrectly identifying a word as a nonword was greater for words that had not been studied (new—.07) than for words that had been studied (remember and forget—.03 and .04, respectively), $F(2, 62) = 10.08$, $MS_e = .001$, $p < .001$. The interaction of word type with testing order also reached significance, $F(2, 62) = 3.14$, $p < .05$, but this was a slight effect and, as Table 4 shows, there was no obvious pattern.

The data of primary concern were the latency data. The main effect of testing order and the interaction of testing order with word type were both nonsignificant, both Fs < 1.29. Thus, unlike in Experiment 1 in which testing order had an impact, it exerted no influence at all in Experiment 2. Presumably, this is because a prior recall test does not expose all of the items to the subject, unlike a prior recognition test.

Of central interest, there was a significant effect of word type on latency, $F(2, 62) = 16.33$, $MS_e = 551.355$, $p < .001$. Planned comparisons demonstrated that, taken together, the studied words (remember—513 ms and forget—525 ms) were responded to more quickly than the unstudied words (new—545 ms), $F(1, 62) = 26.97$, $p < .001$. This replicates the phenomenon of repetition priming in lexical decision (Scarborough et al., 1979). More novel, however, was the significantly greater priming of remember words than of forget

Table 3
Experiment 2: Proportion of Remember and Forget Words Correct on the Free-Recall Test

Testing order	Types of words: Remember	Forget
Recall test first	.37	.04
Lexical decision first	.33	.06

Table 4
Experiment 2: Mean Latencies (in Milliseconds) and Error Probabilities in Lexical Decision

Testing order	Types of words			
	Remember	Forget	New word	Nonword
Recall test first				
Latency	503	510	530	620
Error	.03	.05	.06	.09
Lexical decision first				
Latency	523	540	560	638
Error	.03	.03	.08	.07

words, $F(1, 62) = 4.31$, $p < .05$. Although small in absolute terms, the 12-ms advantage for remember words was quite reliable. Twenty-six of the 33 subjects (79%) showed the effect ($p < .01$ by a sign test).

Discussion

The pattern in Experiment 2 conceptually replicated that in Experiment 1. Again, both the direct and the indirect tests were affected in the same way by the directed forgetting manipulation. Although one might be tempted to say that the effect on the direct test was greater than that on the indirect test, the absence of a common scale makes such arguments difficult to evaluate. The critical observation is that both indirect tests have shown a reliable directed forgetting effect, qualitatively similar to that observed on both direct tests. Furthermore, the quite rapid response times of Experiment 2 make it unlikely that subjects adopted an explicit recollection strategy to assist in lexical decision. Presumably, this would have slowed them down; instead, the data look like those of a standard repetition priming study (e.g., Scarborough et al., 1979).

In summary, the results again conflict with the prediction of the encoding view and support the prediction of the retrieval view of directed forgetting. If directed forgetting is an elaboration manipulation, then it should affect only direct tests, as do other elaboration variables. But directed forgetting affects both types of tests. Such an outcome is more consistent with what is known about the effect of retrieval manipulations on direct and indirect tests. Thus, it may be best to think of directed forgetting as having its influence at the time of retrieval rather than at the time of study.

General Discussion

In the early days of research on directed forgetting (e.g., Bjork, 1972), the prevalent explanation relied on differential rehearsal: The instruction drove a decision about whether to process an item elaboratively. If the instruction was to remember, elaboration was worthwhile; if the instruction was to forget, further processing was suspended (MacLeod, 1975; Wetzel, 1975; Woodward et al., 1973). Wetzel and Hunt (1977) described the effect of the instructional cue: "The suggested mechanism is a pendent form of processing that maintains an item in a rote fashion until a remember or forget decision leads to either discontinued or more elaborative processing" (p. 244).

Recent work on the distinction between direct and indirect memory tests (e.g., Graf et al., 1982; Jacoby & Dallas, 1981; Schacter & Graf, 1986) strongly suggests that manipulation of elaborative processing has no effect on indirect test performance, despite its well-documented influence on direct test performance. Putting together these ideas about encoding, it follows that if directed forgetting leads to differential elaboration at encoding, the effect of directed forgetting should be seen on direct tests but not on indirect tests. Instead, two experiments showed that both types of test were affected in the same way by directed forgetting.

On this basis alone, then, the present results argue against a selective rehearsal account of directed forgetting. The argument is strengthened by Greene's (1986) recent findings contrasting cued recall as a direct test and word-stem completion as an indirect test. A particular virtue of his study was that the same stems were used as the recall cues and as the to-be-completed stems, assuring comparability. What he showed was that both intentionality and rehearsal duration affected cued recall, but neither affected stem completion. This is converging evidence that rehearsal manipulations during encoding affect direct but not indirect measures of memory.

A recent alternative interpretation of directed forgetting provides a way to make sense of the present results. Geiselman and his colleagues (Geiselman & Bagheri, 1985; Geiselman et al., 1983) have argued that items followed by an instruction to forget become inhibited. Consequently, on a later test, forget items are more difficult to retrieve than are remember items. The argument being offered here is that although direct and indirect tests are affected differently by elaboration, an encoding manipulation, they may be affected similarly by inhibition, a retrieval manipulation.

Support for this line of reasoning can be marshalled from both domains. First, consider the directed forgetting literature. In a quite direct attack, Geiselman and Bagheri (1985) began by showing the standard advantage for remember over forget items on an initial test. They then presented all of the same items as to-be-remembered on a second study trial, and performance on the formerly forget items improved much more than that of the formerly remember items. On the basis of several control studies, they argued that this greater improvement of formerly forget items was due to release from retrieval inhibition. Thus, there is evidence that the directed forgetting procedure causes retrieval inhibition.

Now consider the studies of retrieval manipulations in contrasting direct versus indirect test performances. As mentioned in the introduction, at least some manipulations that do affect both types of test can be viewed as retrieval manipulations. Thus, manipulations of list context (e.g., Jacoby, 1983) and of item context (McKoon & Ratcliff, 1979; Schacter & Graf, 1986) affected both direct and indirect measures of memory in the same way. It is certainly the case that more work needs to be done on this topic, but the existing evidence is consistent with a retrieval-based interpretation. If directed forgetting is seen as a retrieval manipulation, then the overall pattern of results makes sense.

There is one complication that must be considered before accepting this retrieval interpretation. It may simply be that when subjects notice items from the study list on what is intended to be an indirect test, they adopt an explicit retrieval strategy that functionally transforms the indirect test into a direct test. Then it would not be very surprising that a manipulation would affect the two tests in the same way. This is always potentially a problem, and it is a difficult one to guard against, although it is perhaps worth noting that the present results rely on rejection of the null hypothesis, whereas the argument that a test is unaffected relies on accepting the null hypothesis. That the directed forgetting effect held up on the indirect tests even after a long retention interval and even in a test requiring a speeded response is comforting with respect to the present experiments. Indeed, the fact that there were two sets of words, remember and forget, both of which showed priming but to differing extents, also stretches the credibility of this possible criticism. The explicit retrieval strategy would have to be a very complicated one. Nevertheless, it remains a possibility and a topic worthy of further study.

In summary, the two experiments reported here showed that directed forgetting affected both direct and indirect tests of memory. This generalized to two different tests of each type and to lists of different length and word frequency. The result seems somewhat counterintuitive, but this may be because most of us think of directed forgetting as a manipulation of elaboration during encoding. If, instead, directed forgetting is seen as having its effect through inhibition at the time of retrieval, the results are readily interpreted. The findings suggest that future research might profitably be directed to the encoding–retrieval distinction in an effort to understand better the similarities and differences between direct and indirect memory tests.

References

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 & Sons.

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

Craik, F. I. M., & Tulving, E. (1975). Depth of processing and the retention of words in episodic memory. *Journal of Experimental Psychology: General, 104*, 268–294.

Davis, J. C., & Okada, R. (1971). Recognition and recall of positively forgotten items. *Journal of Experimental Psychology, 89*, 181–186.

Epstein, W. (1972). Mechanisms of directed forgetting. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 6, pp. 147–191). New York: Academic Press.

Geiselman, R. E., & Bagheri, B. (1985). Repetition effects in directed forgetting: Evidence for retrieval inhibition. *Memory & Cognition, 13*, 57–62.

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.

Graf, P., & Mandler, G. (1984). Activation makes words more accessible, but not necessarily more retrievable. *Journal of Verbal Learning and Verbal Behavior, 23*, 553–568.

Graf, P., Mandler, G., & Haden, P. (1982). Simulating amnesic symptoms in normal subjects. *Science, 218*, 1243–1244.

Graf, P., Shimamura, A. P., & Squire, L. R. (1985). Priming across modalities and priming across category levels: Extending the domain of preserved function in amnesia. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11*, 386–396.

Greene, R. L. (1986). Word stems as cues in recall and completion tasks. *Quarterly Journal of Experimental Psychology, 38A*, 663–673.

Horton, K. D., & Petruk, R. (1980). Set differentiation and depth of processing in the directed forgetting paradigm. *Journal of Experimental Psychology: Human Learning and Memory, 6*, 599–610.

Jacoby, L. L. (1983). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9*, 21–38.

Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110*, 306–340.

Johnson, M. K., & Hasher, L. (1987). Human learning and memory. *Annual Review of Psychology, 38*, 631–668.

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

McKoon, G., & Ratcliff, R. (1979). Priming in episodic and semantic memory. *Journal of Verbal Learning and Verbal Behavior, 18*, 463–480.

Richardson-Klavehn, A., & Bjork, R. A. (1988). Measures of memory. *Annual Review of Psychology, 39*, 475–543.

Roediger, H. L., III, & Blaxton, T. A. (1987). Retrieval modes produce dissociations in memory for surface information. In D. S. Gorfein & R. R. Hoffman (Eds.), *Memory and cognitive processes: The Ebbinghaus centennial conference* (pp. 349–379). Hillsdale, NJ: Erlbaum.

Scarborough, D. L., Gerard, L., & Cortese, C. (1979). Accessing lexical memory: The transfer of word repetition effects across task and modality. *Memory & Cognition, 7*, 3–12.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13*, 501–518.

Schacter, D. L., & Graf, P. (1986). Effects of elaborative processing on implicit and explicit memory for new associations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 432–444.

Sloman, S.A., Hayman, C. A. G., Ohta, N., Law, J., & Tulving, E. (1988). Forgetting in primed fragment completion. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*, 223–239.

Smith, M. C., MacLeod, C. M., Bain, J. D., & Hoppe, R. B. (1987). *Constraints on episodic priming: General versus specific associative facilitation.* Manuscript submitted for publication.

Squire, L. R., Shimamura, A. P., & Graf, P. (1987). Strength and duration of priming effects in normal subjects and amnesic patients. *Neuropsychologia, 25*, 195–210.

Tulving, E., Schacter, D. L., & Stark, H. A. (1982). Priming effects in word-fragment completion are independent of recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 8*, 336–342.

Weiner, B., & Reed, H. (1969). Effects of the instructional sets to remember and to forget on short-term retention: Studies of rehearsal control and retrieval inhibition (repression). *Journal of Experimental Psychology, 79*, 226–232.

Wetzel, C. D. (1975). Effect of orienting tasks and cue timing on the free recall of remember- and forget-cued words. *Journal of Experimental Psychology: Human Learning and Memory, 1*, 556–566.

Wetzel, C. D., & Hunt, R. E. (1977). Cue delay and the role of rehearsal in directed forgetting. *Journal of Experimental Psychol-*

ogy: Human Learning and Memory, 3, 233–245.
Woodward, A. E., Jr., Bjork, R. A., & Jongeward, R. H. (1973). Recall and recognition as a function of primary rehearsal. *Journal of Verbal Learning and Verbal Behavior, 12,* 608–617.
Yaniv, I., & Meyer, D. E. (1987). Activation and metacognition of inaccessible stored information: Potential bases for incubation effects in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 187–205.

Appendix
Word and Nonword Stimuli From Experiment 2

Word sets				Nonwords			
Set A		Set B					
ulcer	belt	bread	iron	hoise	palty	inselt	ronk
market	yarn	umpire	pants	naig	percil	isem	keader
oven	king	moon	wheat	weavon	tubnel	hemon	letner
jury	army	yolk	prison	alser	artow	masket	obour
apple	nose	organ	rodeo	winler	laby	ralace	pafer
glass	fence	juice	heart	mody	lutter	parmy	poek
window	mule	author	speech	cafin	cardy	priton	rebuge
lamp	duck	kettle	nickel	nattle	celtar	remarb	himal
elbow	quart	goat	ruler	cimy	poffee	javing	keason
drum	infant	angle	lawn	comlade	cogtest	fesies	fingle
barn	water	niece	train	dorner	coston	hoftly	jugar
pole	pipe	wood	horse	dehart	saster	sumply	hicket
hair	rail	flag	table	readow	genber	miny	tolay
needle	steam	master	coat	getal	jethod	undel	ubit
rabbit	ladder	door	volume	doney	pother	ulper	bessel
tool	ticket	leaf	crown	murdel	objent	vistim	vipit
cloud	hand	edge	vapour	elfow	engane	waber	weason
violin	cigar	quilt	storm	ement	forent	yelkow	boark
snake	vessel	doctor	forest	sarden	bammer	dirl	fleth
fish	glove	bell	grape	hoteg	bortle	golp	jamner

Received July 20, 1987
Revision received March 31, 1988
Accepted March 31, 1988 ■