Journal of Experimental Psychology
1973, Vol. 99, No. 1, 22–27

DIRECTED FORGETTING OF INDIVIDUAL WORDS IN FREE RECALL [1]

ROBERT A. BJORK [2]
University of Michigan

ADDISON E. WOODWARD, JR.
Governors State University

This article reports an experiment designed to clarify the storage, rehearsal, and retrieval mechanisms that underlie the ability of *S*s to recall to-be-remembered words (R words) without intruding to-be-forgotten words (F words). The *S*s were presented seven 24-word lists, each consisting of a random mixture of 12 R words and 12 F words; the cues to remember or forget were presented subsequent to each word in turn. Six of the lists were followed by either an immediate test of R-word recall or a distractor activity, and one of the lists was followed by a prearranged signal to *S*s to recall both R words and F words from that list. In general, the results implicate rehearsal and organizational processes at input rather than suppression or selective retrieval processes during output as the mechanisms by which *S*s achieve the simultaneous recall and nonrecall of R words and F words, respectively.

Recently, Woodward and Bjork (1971) introduced a paradigm designed to study how *S*s remember some items in a free-recall list and forget others. The paradigm involves cuing *S*s to remember or to forget each item in the list in turn. To a remarkable extent, *S*s are able both to recall the to-be-remembered words (R words) and to avoid recalling the to-be-forgotten words (F words). From Woodward and Bjork's data, however, it is not possible to state conclusively whether the nonrecall of F words is attributable to their not being retrievable during recall or to their being actively suppressed.

The design of the Woodward and Bjork experiments included a delayed recall test at the end of the experiment during which *S*s were encouraged to recall any word they could remember independent of the initial cuing of the word when it was presented. The delayed recall of F words was very poor, related both to the delayed recall of R words and in absolute terms. This finding was interpreted as implying that the initial nonrecall of F words was not attributable to active suppression, in which case one might have expected a sizable recovery in their recall on the final test. One problem with this interpretation is that the immediate recall of R words might have contributed substantially to the likelihood that R words were again recalled on the delayed test. Thus, the very large difference between the final recall of R words and F words might be due to their immediate recall and nonrecall, respectively, rather than to their initial cuing.

Davis and Okada (1971) modified the Woodward and Bjork (1971) design to include an immediate test of the recall of F words. Without forewarning, they asked *S*s to recall all words, both R words and F words, from the last of a series of three 64-word lists. Davis and Okada found very poor immediate recall of F words; in fact, in terms of average number of F words recalled, there was an increase of less than one word over the number of F words intruded in the recall of any one of the preceding lists.

Davis and Okada's (1971) results support the notion that low F-item recall is not due to active suppression of F items; otherwise many more F items would have been recalled on their immediate recall test; nevertheless, problems exist with their procedure. It is possible that the unexpected instruction to recall everything following the presentation of the last list is a disruptive event. The instruction comes as

[1] This research was supported in part by National Science Foundation Research Participation in Psychology for College Teachers, 6T-6929, and in part by Grant OEG-5-71-0023 (509) from the U.S. Office of Education to the second author.

[2] Requests for reprints should be sent to Robert A. Bjork, Human Performance Center, 330 Packard Road, Ann Arbor, Michigan 48104.

a surprise, it takes time, and it violates what *S*s are told at the beginning of the experiment. Thus, such an instruction may disrupt *S*s' recall attempts and lead to impaired performance. That such a disruption may occur is supported by an examination of the results of a free-recall experiment by Bruce and Papay (1970). When they unexpectedly asked *S*s to recall all the words in a last list, that is, to recall both the F words preceding a forget signal in the list and the R words following the signal, the serial position curve they obtained shows a clear-cut disruption of the typical recency effect in the recall of the R words. Another possible problem with a surprise test of F-item recall following the last list of an experiment is that recall performance on the last list, due to practice effects, proactive interference, or whatever, may not be characteristic of recall performance on other lists during the experiment. (For a discussion of the various procedures of testing F items, see Bjork, 1972.)

Reitman, Malin, Bjork, and Higman (1973) devised a procedure that overcomes the objections that can be leveled at surprise tests such as the one used by Davis and Okada (1971), as well as the problems inherent in the kind of final test used by Woodward and Bjork (1971) to test *S*s' memory for all items presented during an experimental session. The procedure involves forewarning *S*s that one or more tests of F items will occur during the experiment, that such tests will be indicated by a special signal, that they will be infrequent, and that clearly the best strategy is to forget F items and to remember R items. There are several advantages of the Reitman et al. procedure. A test of F-item recall can be inserted at any point during the experimental session and more than one such test can be included. Such tests do not come as a surprise, and to the degree that performance on the normal trials is comparable to performance when there are no tests of F items, one has evidence that *S*s are consistently trying to forget F items and remember R items.

The present experiment was designed both to provide a better estimate of *S*s' memory for F words immediately following the presentation of a list and to assess the effect of the immediate recall of R words on the subsequent recall of those words at the end of the experiment. The *S*s were presented seven lists of 24 words. In all of the lists, *S*s were cued whether to remember or to forget each word in turn; 12 words were forget cued and 12 words were remember cued in an intermixed fashion. In order to assess *S*s' immediate memory for F words, one of the seven lists was followed by a prearranged signal to *S*s to recall all the words in the list independent of the cuing during the list. In order to assess the influence of immediate recall on the final recall at the end of the experiment, three of the lists were followed by a 30-sec. digit-shadowing task rather than a 30-sec. recall period. Thus, contrasting the final recall of words in lists followed by digit shadowing with the final recall of words in lists followed by an immediate recall test provides a measure of the extent to which final recall is facilitated by immediate recall.

Method

Subjects. The *S*s were 40 undergraduates at the University of Michigan. They were paid $1 plus any bonuses that accrued from the payoff system employed in the experiment.

Materials and apparatus. Every *S* viewed seven 24-word lists constructed of unrelated common four-letter nouns. The lists were shown on a high-speed (change time less than .05 sec.) memory drum. The words, the cues to remember or forget, the instructions to recall or get ready for the next trial, and the digits to be shadowed all appeared in the same window. The timing of advances of the memory drum was controlled by a high-speed paper-tape reader reading a prepunched tape.

Design. After each successive word in a list, a colored (red or green) dot appeared as a cue to *S* to forget or remember the item. For half of the *S*s a green dot meant remember and a red dot meant forget, and for the other half the meaning of the colored dots was reversed. Each word was shown for 2.3 sec., and each cue was shown for 1 sec.

Three of the seven lists were followed by a 30-sec. recall period, three by a 30-sec. period in which digits were shadowed, and one list was followed by a special recall period during which *S*s attempted to recall all list items independent of how they were cued in the list. The special recall list was always the fourth list seen. The remaining recall and

TABLE 1

IMMEDIATE AND FINAL RECALL PROBABILITIES

List and item type	List number							
	1	2	3	4[a]	5	6	7	Average[b]
Immediate recall								
IR lists								
R word	.642	.479	.554	.512	.517	.588	.567	.558
F word	.025	.021	.017	.050	.017	.004	.013	.016
Final recall								
IR lists								
R word	.170	.100	.134	.240	.188	.350	.500	.240
F word	.038	.050	.033	.035	.025	.025	.046	.036
NIR lists								
R word	.112	.100	.154	—	.176	.208	.300	.175
F word	.042	.017	.017	—	.046	.013	.021	.026

Note. The notation IR denotes lists followed by an immediate recall, and NIR denotes lists followed by digit shadowing rather than an immediate recall. Words were either to be remembered (R) or to be forgotten (F).

[a] The fourth list was always followed by a signal instructing *S*s to recall all list items, both R words and F words.

[b] Excluding List 4.

shadow lists were randomly arranged together. Suitable counterbalancing techniques insured that, across *S*s, all lists except the special list served in both the recall and shadow conditions. In addition, there were two different orders of presentation of the words in any one list. Finally, every quarter of each list contained three R words and three F words.

Procedure. The *S*s were tested individually. Every *S* was read a set of instructions and was shown two practice lists of 12 nonsense syllables. One practice list was followed by a recall period, the other by digit shadowing. After presentation of both practice lists, *S*s were informed that a payoff system would be in effect during the experiment: They would receive a 1¢ reward for each R word recalled during the immediate recall of a list, and they would lose 1¢ for each F word recalled (intruded). The *S*s were also told at this time of the special list; they were instructed that during the experiment a special list would occur, that they were to recall *all* words from that list, that the original payoff matrix would be suspended for the special list, and that they would receive 1¢ for any word recalled from that list whether or not the word was an R word or F word. The recall signal that designated the special list was the word "recall" highlighted in blue; all other recall signals were colored yellow. Finally, *S*s were urged not to anticipate the occurrence of the special list, and they were told that the best strategy was to try always to forget F items and remember R items.

Each of the seven lists was preceded by a 3-sec. ready signal. When digit shadowing followed a list, 13 eight-digit numbers appeared one by one at a 2.3-sec. rate. The *S*s wrote down words they remembered on response sheets, one sheet per list, and after *S*s had been presented all lists and had either shadowed digits or recalled list items, there was a phony debriefing period of 2 min. following which *S*s were asked, without having been forewarned, to recall all words they could remember from any of the lists they had seen. The *S*s were informed at the time of the final recall that they would receive a 1¢ bonus for any word they recalled, independent of the initial cuing of the word. When the *S*s could recall no more words, they were asked to circle any words among those they had recalled that they thought were F words.

RESULTS

Table 1 presents the immediate and final recall probabilities for R words and F words as a function of list type and list number. (IR denotes lists followed by an immediate recall; NIR denotes lists followed by digit shadowing.) In immediate recall, *S*s were proficient both at recalling R words (nearly 60%) and avoiding the recall of F words (less than 2%). In response to the special recall instruction immediately following the fourth list, *S*s were able to recall only about 5% of the F words in the list. Thus, even when *S*s were trying to recall F words immediately and were rewarded for doing so, the average *S* was able to recall less than one of the F words in the list. It appears

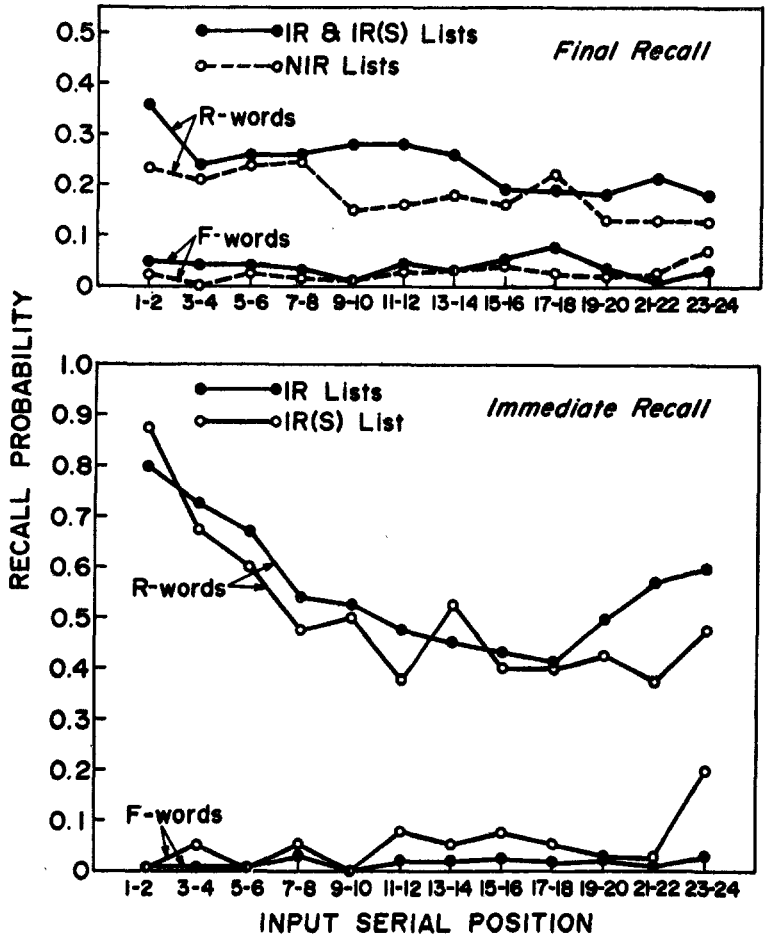

FIG. 1. Immediate and final recall probabilities as a function of serial position. (Lists: IR = list followed by an immediate recall; IR(S) = special list, followed by a signal to recall all words in the list; NIR = list followed by digit shadowing. Words: R = to be remembered; F = to be forgotten.)

that active suppression plays a negligible role in the immediate nonrecall of F words.

Several aspects of the final recall data merit comment. First, it is clear that the advantage of R words over F words in final recall is not attributable to their having a higher likelihood of being recalled immediately. The ratio of the final recall of R words to the final recall of F words is 6.6 (.240/.036) for IR lists and 6.7 (.175/.026) for NIR lists. Second, the final recall of both R words and F words from the special list corresponds exactly to the final recall of R words and F words from the other IR lists (.240 vs. .240, and .036 vs. .035). Finally, there is a list recency effect in the final recall of R words across the seven lists, and the anticipation of such an effect was one motivation for having the special list always be the middle list of the seven lists presented.

In Figure 1 the probability of recall of R words and F words is shown as a function of time of recall (immediate in the lower panel, final in the upper panel), serial position, and type of list. The notation IR(S) denotes the special list, that is, the list followed by a signal to recall all words in the list. The immediate recall of R words from IR and IR(S) lists is very similar

across serial positions, with the possible exception of Positions 21–24. The immediate recall of F words from IR and IR(S) lists is remarkably similar, especially in view of the fact that, from the *S*s' standpoint, F words recalled from IR lists are penalized intrusions and F words recalled from the IR(S) list are in response to an instruction to recall such words and are rewarded. Only for F words presented in Positions 23 and 24 is the recall of F words from the IR(S) list appreciable. The 20% likelihood of recalling F words presented in those two positions probably represents retrieval from short-term memory, and *S*s' efforts to retrieve F words from short-term store may account for their less efficient retrieval of R words presented at the end of the IR(S) list.

The final recall of R words and F words from IR and IR(S) lists was so indistinguishable that they are plotted together in Figure 1. Compared to the final recall of R words and F words from NIR lists, the final recall of IR and IR(S) lists is somewhat superior, but there appears to be no clear-cut interaction with serial position.

It is worth pointing out that the two procedural innovations in the present study, that is, preinstructing *S*s that there would be one immediate test of F-item recall and varying whether a list was followed by an immediate recall or digit shadowing, did not appear to influence the basic levels of R-word and F-word recall as established by Woodward and Bjork's (1971) initial results. For comparable lists, that is, those followed by an immediate attempt to recall R words, the immediate and final recall probabilities are .502 vs. .558 and .233 vs. .240 for R words, respectively, and .019 vs. .016 and .047 vs. .036 for F words, respectively.

DISCUSSION

Two questions prompted the current study, and the results answer those questions in an unambiguous way. First, the nonrecall of F words appears not to be explainable as arising from active suppression. The very poor performance on an attempted immediate recall of F words in our study combined with the same result in Davis and Okada's (1971) study rule out active suppression as an important factor in the nonrecall of F words. Second, the superiority of final R-word recall over final F-word recall appears not to be attributable to the greater frequency with which R words are recalled immediately. The final recall of R words and F words from lists that were not followed by an immediate recall reveals the same relative advantage of R words over F words as is shown in the final recall of words from lists that were followed by an immediate recall. These two results together argue that differences in immediate and final recall of R words and F words arise from differential processing during list presentation, rather than from differential editing of retrieval or output or from differential facilitation of delayed recall owing to differential frequency of immediate recall.

If not suppression, then what? The results of the present study taken together with the results obtained by Woodward and Bjork (1971) and by Davis and Okada (1971) argue convincingly against suppression as the mechanism that underlies the nonrecall of F words. It is worth noting that two different suppression mechanisms are ruled out: The *S*s might retrieve F words during tests of immediate recall, but reject them because they are tagged as F words, or they might suppress retrieval from the entire set of F words in memory. The latter kind of suppression might occur if F words were retrievable from memory, but were functionally segregated from R words in memory by virtue of the differential rehearsal and interassociation *S*s devote to R words. In any case, the negligible immediate recall of F words when *S*s are trying to recall them is evidence against any brand of output suppression.

A relatively simple and straightforward possibility is that F words do not exist in memory at the time of an immediate recall. Although the procedure forces *S*s to read every word as it is presented because the cue to remember or forget is not presented until after the offset of each word, F words may be lost from memory at a rapid rate characteristic of unrehearsed items. The very few F words *S*s are able to recall immediately may consist entirely of (*a*) words misencoded as R words, (*b*) words that have a strong idiosyncratic significance for a particular *S*, or (*c*) words still retrievable from short-term memory because they were presented in the last serial position or two in the list.

The possibility that F items do not exist in memory has been completely discredited in

other directed-forgetting paradigms for the very good reason that certain kinds of recognition and recall tests provide ample evidence that F items exist in memory (for a review of the evidence, see Bjork, 1972). In the present experimental context, however, which involves item-by-item cuing rather than the cuing of sets or blocks of items, part of the nonrecall of F words may be attributable to their simply not existing in memory. In the case where *Ss* are cued to forget items only after a block of items is presented, they are forced to rehearse and interassociate F items to a much greater extent than they are in the present procedure.

That the loss-from-memory explanation of the nonrecall of F items may be part, but not all, of the answer is suggested by some results from the Woodward and Bjork (1971) and Davis and Okada (1971) experiments. Davis and Okada included in their design a recognition test for R words and F words. They found the recognition of F words to be clearly inferior to the recognition of R words, a result that supports the possibility that F words are lost from memory to some extent, but they also found the recognition of F words to be very much higher than the false-alarm recognitions of words never presented, a result that clearly implies F words are not completely lost from memory. Woodward and Bjork found, using categorized lists, that the recall of a given F word was facilitated if R words from the same semantic category were presented in the list, but again, F-word performance was clearly inferior in all cases to R-word performance.

The most plausible explanation, in our opinion, is the following. During the presentation of a list, *Ss* rehearse and interassociate R words to the extent that the presentation sequence provides time to do so; they try to avoid any rehearsal or interassociation of F words. This differential treatment of R words and F words results in R words being grouped as a set in memory distinct from F words and words never presented. Whether a word exists in long-term memory, as measured by a recognition test, depends on its receiving a certain minimal amount of rehearsal: Some F words are rehearsed that much, some are not; nearly all R words are rehearsed more than the minimum necessary to support their recognition. Retrieval from long-term memory, as measured by a recall test, is almost entirely a function of interassociation. Thus, R words are retrievable to the extent that they are interrelated in memory; F words are typically not retrievable unless an experimental manipulation or idiosyncratic happenstance leads to interassociations or associations with R words.

REFERENCES

BJORK, R. A. Theoretical implications of directed forgetting. In A. W. Melton & E. Martin (Eds.), *Coding processes in human memory.* Washington, D.C.: Winston, 1972.

BRUCE, D., & PAPAY, J. P. Primacy effect in single-trial free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, **9**, 473–486.

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

REITMAN, W., MALIN, J. T., BJORK, R. A., & HIGMAN, B. Strategy control and directed forgetting. *Journal of Verbal Learning and Verbal Behavior*, 1973, in press.

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

(Received July 10, 1972)