The Role of Rehearsal in Short-Term Memory¹

Fergus I. M. Craik

University of Toronto

AND

MICHAEL J. WATKINS Yale University

Several widely accepted models of memory postulate that the adequacy of an item's registration in long-term storage is a positive function of its length of stay in the short-term store. However, when short-term storage times were measured, these times did not predict long-term recall or recognition. Two further experiments showed that neither the length of an item's stay in short-term storage nor the number of overt rehearsals it received was related to subsequent recall. It is concluded that the "maintenance" and "elaborative" aspects of rehearsal can be clearly separated, and that the duration of rehearsal is related to long-term memory and learning only in the latter case. Maintenance rehearsal does not lead to an improvement in memory performance.

Rehearsal must play a major part in any complete theory of memory, yet its function is still poorly understood. Waugh and Norman (1965) postulated that rehearsal serves the dual purpose of maintaining items in a shortterm store and transferring information about the items to a more permanent long-term store. This view was endorsed, in general, by Atkinson and Shiffrin (1968). They argued that the principal function of rehearsal was to maintain a small set of items in short-term store by repetition, but also that "any information in short-term store is transferred to long-term store to some degree throughout its stay in the short-term store" (p. 115).

The principle that an item's strength in long-term store is a direct function of its length of stay in the short-term store has been stated explicitly by Atkinson and Shiffrin (1968), Waugh (1970), and Norman and Rumelhart (1970). The notion has received good support from studies involving an overt rehearsal procedure (Rundus & Atkinson, 1970; Rundus, 1971). In these studies, the probability of an item's retrieval from longterm store varied directly with the number of rehearsals the item received. This finding also seems to provide a good explanation for the negative recency effect in free recall (Craik, 1970); that is, items at the end of a list are not rehearsed as often as prerecency items, and are thus poorly retrieved in a subsequent recall test. The adequacy of this explanation is examined in Experiment II of the present paper.

Initially, however, we were concerned with examining the notion that time in short-term store predicted later memory performance. The technique adopted involved measuring the time intervals between the presentation and recall of those items retrieved from shortterm storage in an immediate free recall test, and then relating these intervals to the probability of recalling the item from longterm storage in a second, delayed recall test. Although input–output intervals will accu-

¹ This research was started while the second author was at the North East London Polytechnic, England. Experiment II was supported by Grant A 8261 from the National Research Council of Canada to the first author.

rately measure short-term storage durations only if destructive readout is postulated, it seems reasonable to assume that these intervals will be at least highly correlated with shortterm storage times. The measurements were carried out in an experiment reported by Craik, Gardiner, and Watkins (1970, Experiment 3); the study will be described only briefly here. Twenty subjects were given ten 20-word lists for free recall. The words were presented auditorily at a 2-sec rate; recall was spoken and recorded on tape. Two minutes after recall of the last list, the subject was given an unexpected final recall test in which he wrote down as many words as possible from all ten lists. Finally, the subject was read a list of 200 words at a 3-sec rate. This list contained a randomly selected 100 words from the presentation list, plus 100 new words. The subject's task was to rate each word on a 5-point scale of recognition confidence, ranging from "certain old" (5) to "certain new" (1). It was assumed that the final recall and recognition tasks reflected retrieval from long-term storage only.

In immediate recall, a word was considered a short-term store item if no more than six words (either stimuli or responses) intervened between its presentation and recall (Tulving & Colotla, 1970). For each subject, the interval between the input and retrieval of each shortterm store item was measured with the aid of an event recorder and used as an index of time "in store." The mean short-term storage times for items which were retrieved and not retrieved in final recall, were 13.2 and 12.8 sec, respectively. While this difference is in the predicted direction, it is very small and not statistically reliable, p > .25. In the recognition analysis, mean durations in short-term store were 6.7, 5.7, 5.5, 5.9, and 4.2 sec for the confidence ratings 5 to 1, respectively. That is, there was some tendency for longer times in short-term storage to be associated with higher subsequent recognition ratings. However, a linear trend analysis revealed that the relationship was nonsignificant, F(1, 252) = 2.28, p > .10.

Thus, in the study under consideration at least, neither final recall nor recognition performance was significantly related to short-term storage time. The clear inference is that information was not in fact being transferred to long-term storage during the items' stay in the short-term store. Since this conclusion is directly opposed to the prediction from the Waugh and Norman (1965) and the Atkinson and Shiffrin (1968) models, further information was sought from an experiment in which short-term storage time was manipulated directly.

EXPERIMENT I

A paradigm was designed in which subjects were induced to hold single words in shortterm storage for varying lengths of time. The subjects were instructed to listen to a series of word lists, and to report after each list just the last word beginning with a particular letter. The subject was informed of the "critical letter" before list presentation, and he could therefore ignore words from the start of the list until the first critical word was presented. He then held that word in mind until a further critical word was presented at which time he dropped the first word and replaced it with the second. This procedure continued until the list ended, when he wrote down the latest critical word. Three rates of presentation were used. Thus the time for which a critical word was held in short-term storage varied both as a function of presentation rate, and of the number of noncritical words monitored between presentation and replacement (or report). For example, where "G" is the critical letter and the list begins DAUGHTER OIL RIFLE GARDEN GRAIN TABLE FOOTBALL ANCHOR GIRAFFE..., the subject would first hold GARDEN, replace it by GRAIN and then by GIRAFFE; if the remaining words were noncritical (that is, did not begin with G), GIRAFFE would be reported at the end of the list.

After presentation of all the lists, the sub-

jects were unexpectedly asked to recall as many words as possible from all lists. For each rate of presentation, the delayed recall of both the "replaced" and "reported" critical words was examined as a function of the number of noncritical presentations during which the words were held. Presumably if time in shortterm store predicts long-term store retrieval, then final recall performance should increase directly with the number of items monitored during the retention of the critical words, and inversely with presentation rate.

Method

Design. The experiment was carried out in three sessions with 18 subjects participating in each session. All 54 subjects heard the same 27 lists of 21 words. Within each session the 18 subjects were randomly allocated into three groups of six subjects, and each group was given a different set of critical words. That is, on any one list, the three groups received different critical initial letters. Three rates of presentation were used: slow, 1 word every 2 sec; medium, 1 word every sec; and fast, 1 word every half sec. Lists presented at one speed in the first session were presented at the other two speeds over the remaining sessions. In each session, nine lists were presented at each speed, with order of list presentation randomized separately for each session.

Two further within-subject variables were involved in the design. The first is *i*-value, which refers to the number of intervening (noncritical) words which were monitored during the retention of the critical words. There were nine *i* values: 0, 1, 2, 3, 4, 5, 6, 8, and 12. The second variable concerns the critical words, and whether they were "replaced" by further critical words during the list or whether they were "reported" as the last critical word in the list. Thus if a list contained two critical words, presented at serial positions 6 and 12, the first would be a "replace" word with an i value of 5, and the second a "report" word having an *i* value of 9. Construction of the lists ensured that all subjects received 81

critical words, equally distributed over *i* value and presentation rate. Hence the nine lists shown at a given rate included three critical words at each *i* value; of these two were "replace" words and one a "report" word. Order of *i* values was essentially random, so that when a critical word was presented, the subject knew neither its *i* value nor whether it would have to be replaced or reported. On the other hand, order of *i* value was not entirely random, in that an effort was made to avoid confounding serial position and *i* value of the "replace" words. In addition, the first three serial positions were not used. Apart from the 27 experimental lists, three practice lists were presented, one at each rate. All subjects were given a final free recall test following the 27 lists.

The words were one- and two-syllable concrete nouns. The lists were recorded on tape, and each list was preceded by a reference to the list number and the speed of presentation. A tone, presented at the same rate as the list words, signalled the end of the list. Separate randomizations and recordings were made for each session.

The subjects were 54 introductory psychology students from the North East London Polytechnic.

Procedure. The subjects in each session were randomly assigned to three groups of six. The subjects were informed that they formed the control group of a perception-memory experiment; whereas other subjects had a task with a substantial memory load, they merely had to keep track of the latest word beginning with a particular letter. Thus, they were to carefully monitor the lists and write down the last word in each list which began with the critical letter. It was stressed that their performance should be virtually perfect before any useful conclusions could be drawn from the other conditions. Before the presentation of each list, each subject turned over a card which contained the number of the list and the critical letter for that list (the critical letter changed from list to list). After the list ended, subjects wrote down on the card the last word beginning with the critical letter, and immediately placed the card into an envelope which had been provided for each subject. Lists were presented with an intertrial interval of 15 seconds.

Immediately following the last list, subjects engaged in a 1-min arithmetic task, during which time paper was distributed for the final recall test. A 10-min period was then allowed for the free recall of any of the words presented (both critical and noncritical).

Results

An average of 26.2 of a possible 27 "report" words were correctly identified and reported, with no subject making more than two errors. It was therefore assumed that the critical words were correctly perceived and retained over their respective intervals.

The recall data for the critical items were pooled over sessions and groups of subjects. The percentages of words recalled under each condition are shown in Table 1. An analysis of variance revealed significant effects of reporting, F(1, 36) = 24.10, p < .001, and presentation rate, F(2, 72) = 7.00, p < .005. The findings reflect the advantages of reported over replaced words, and of slower presentation. The only other significant effect was the second order interaction between rate, *i* value

and replaced-reported, the implications of which are not obvious. The most important finding for the present purpose is the nonsignificance of the *i* value variable, F(8, 36)= 1.14, p > .10. It could be argued that an overall test of the main effect is not the most sensitive method of assessing a potentially monotonic relationship between *i* value and recall. Accordingly, a linear test of trend was performed, with due allowance made for the fact that the *i* values represented were not evenly spaced. This also failed to achieve significance, F(1, 36) = 1.70, p > .10. For reported words, *i* value is necessarily confounded with serial position: Those reported words which had a small i value were at the end of the list. However, since the same pattern of results was obtained with replaced words (Table 1) and since i value was not confounded with serial position for replaced words, it is tentatively concluded that the confounding had no crucial effect on the recall of reported words.

Discussion

The most interesting finding of Experiment I was that the probability of recalling an item from long-term store remained. essentially independent of its i value. This result is clearly contrary to the idea that recall probability necessarily increases in direct proportion to the

TABLE 1

PERCENTAGE RECALL AS A FUNCTION OF EXPERIMENTAL CONDITION, *i* VALUE, AND PRESENTATION RATE

Condition	Presentation rate	<i>i</i> value									
		0	1	2	3	4	5	6	8	12	Mean
Replaced	Slow	12	13	22	10	21	19	19	18	19	17
	Medium	10	15	22	12	14	19	09	12	11	14
	Fast	14	07	- 11	06	06	14	09	16	15	11
	Mean	12	12	19	10	14	17	13	15	15	14
Reported	Slow	19	20	20	20	31	39	22	26	28	25
	Medium	20	22	19	19	31	26	20	28	20	23
	Fast	26	15	22	26	20	31	19	11	20	21
	Mean	22	19	20	22	28	32	20	22	23	23

total amount of time an item has been thought about or attended to (Waugh, 1970); a view which is implicit in most of the two-process models of short-term memory.

On the other hand, recall was inversely related to presentation rate. Apparently, the beneficial effect of slower input occurred during the interval before the next item was presented. It is speculatively suggested that the subject uses the interitem interval to process or analyse the last presented word, and that the level of analysis achieved increases with available time. The lack of an i value effect implies the use of a rehearsal process which was sufficient to retain a critical item in a highly accessible state (that is, in short-term storage) during the presentation of subsequent noncritical items, but did not have any long-term effects.

At first sight, the present results seem inconsistent with evidence indicating that retention varies directly with amount of rehearsal (Rundus, 1971; Rundus & Atkinson, 1970) and with repetition (Bjork, 1970; Waugh, 1963). It is suggested that this discrepancy can be reconciled by postulating two distinct modes of rehearsal-the one being a simple maintaining process, and the other an elaborative process (Craik & Lockhart, 1972). Time in short-term store will only predict later long-term store performance when the subject has used the time to encode the items elaboratively. Contrary to the models of Atkinson and Shiffrin (1968) and Waugh and Norman (1965), time in short-term store does not by itself lead to long-term retention. Experiment II reports further evidence on this issue.

EXPERIMENT II

Craik (1970) reported an experiment in which ten lists of words for immediate free recall were followed by a final free recall test, in which subjects attempted to recall all previously presented words. It was found that while the immediate recall data showed the typical large recency effect, the final recall data showed no such effect. In fact, terminal items were the least well recalled in the final recall test. Two possible explanations of this negative recency effect will be considered here. The first says simply that terminal items receive fewest rehearsals in immediate recall and are thus poorly registered in long-term store. This explanation obtains good support from the overt rehearsal experiments of Rundus (Rundus & Atkinson, 1970; Rundus, 1971), since he showed directly that the last items in a list are indeed rehearsed least often.

An alternative explanation of negative recency is that it is the type of rehearsal, not the amount of rehearsal, which is critical for later recall. Experiment I demonstrated that time in short-term store did not by itself enhance subsequent recall and it was suggested that the subject must rehearse in an associative or elaborative fashion for efficient later performance. It is possible that, if the subject knows the end of the list is near, he relies on highly effective, but transient, phonemic information for recall of terminal items. Earlier items have been encoded in a more semantic-associative fashion. Thus the last few items in the list are the best recalled in immediate recall but the least well recalled in final recall.

In the present experiment, subjects were induced to rehearse the last few items in the list at least as many times as they rehearsed the first few. If number of rehearsals is the critical factor, then negative recency in final recall should disappear and a slight positive recency effect, analogous to primacy, should appear. On the other hand, if type of rehearsal is critical, and the extra rehearsals are of a "maintaining" type, then these rehearsals should merely prolong the recency effect in immediate recall and should not lead to a strengthened trace. Thus if it is the quality, not the quantity, of rehearsal activity which is important, the recency positions in final recall should not be affected by the extra rehearsal period in immediate recall.

In summary, subjects were given several lists of words for free recall. They were

allowed to recall some lists at the end of presentation, but for other lists, a 20-sec unfilled delay was interpolated between presentation and recall. In all cases, subjects were informed that they must recall the last four words of each list, rehearsing where necessary. The subjects were asked to rehearse aloud and their spoken rehearsals were taperecorded and later counted. In this way it was hoped that subjects would rehearse terminal items a great deal in the delay condition. After all lists had been presented and recalled, the subjects were unexpectedly asked to recall all previous words in a final free recall test.

Method

Sixteen University of Toronto undergraduates were used as subjects. They were paid for their services. Each subject was tested individually. He was told that the experiment concerned the effect of rehearsal on immediate retention and that all his rehearsal must be spoken aloud. Each trial consisted of the visual presentation of a 12-word list at a 3-sec rate. On half of the trials, recall was immediate-signaled by a loud tap just after presentation of the last word. On the remaining trials, recall was delayed for 20 sec after the last word. The recall signal was again a loud tap, and in this case the subject was encouraged to rehearse aloud during the unfilled 20-sec interval. In all cases instructions were for free recall but with the additional strongly emphasized instruction that recall of the last four words was particularly important. These words were printed in block capitals to distinguish them from the first eight words in each list which were typed in lower case letters. By this means, the subject was encouraged to recall the last four words first of all and to rehearse them in particular, during the delay condition. Thus it was hoped that, in the delayed condition, subjects would rehearse the last four words as often as the first few.

The words used were common nouns. Twelve lists were presented—six under im-

mediate recall and six under delayed recall conditions. The order of immediate and delayed trials was randomized independently for each subject. Also, the subject did not know until the end of each list whether it was an immediate or a delayed trial. The subject was given 1 min for written recall on each trial. The 12 scored lists were preceded by three practice lists using letters as material since subjects required some practice and encouragement to keep them rehearsing aloud during presentation and during the delay interval. After the 12 lists had been presented and recalled, the subject was engaged in conversation for 2 min and was then asked to recall as many of the words as he could, in a final free recall test. No subject anticipated that he would be asked to recall the words a second time. Ten minutes were allowed for the final recall test.

Results

The results, shown in Figure 1, are clear cut. Even in the immediate condition the strong instructions regarding the last four words resulted in a boost in rehearsals for these words. Otherwise, the pattern of rehearsal is similar to that obtained by Rundus (1971). In the delayed condition, rehearsal of the first eight words was increased slightly, and rehearsal of the last four increased substantially by an average of six extra rehearsals per word. The initial free recall results were essentially identical for the two conditions, showing that subjects were able to maintain the words over the 20-sec interval. Of greater interest is the finding that the final recall scores for the immediate and delayed initial recall conditions were also identical. Despite the great increase in rehearsal for the last four words under the delayed condition, the mean final recall level for the last four words is less in the delayed condition (9.1%) than in the immediate condition (9.8%). Since the mean final recall score was less for the condition in which more rehearsal was demonstrated, no statistical test of the hypothesis that rehearsal leads

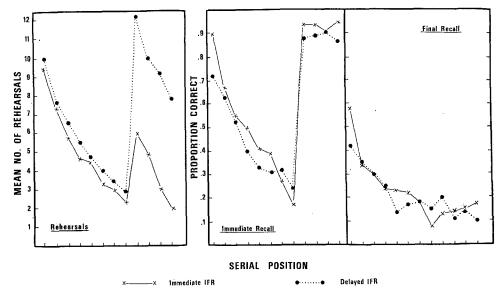


FIG. 1. Number of rehearsals, immediate free recall, and final free recall, with and without an unfilled 20-sec delay between presentation and immediate free recall (IFR).

inevitably to a stronger long-term trace was deemed necessary.

Discussion

Experiment II gives strong support to the position that time in short-term store does not by itself lead to good long-term retention. When subjects were forced to maintain items in short-term storage over a 20-sec interval, such maintenance rehearsal had no effect on later recall from the long-term store. It is important to note that in the present study subjects were probably encouraged to adopt a purely maintenance rehearsal strategy for the last four list items for the following reasons: (a) Since they were not expecting a final recall test there was no reason for them to indulge in complex semantic-associative coding. (b) They were very well aware that the list was nearly ended, since the last four words were printed in capital letters. (c) It is likely that phonemic encoding was encouraged by the necessity to rehearse aloud. To the extent that experimental procedures preclude these factors, some semantic encoding and some positive recency in final recall might be expected.

Although it is not easy to determine whether genuine negative recency effects have been obtained with list-length as short as 12 items, the present results clearly imply that negative recency is not due to fewer rehearsals of the last few items. In Experiment II, recency items were rehearsed at least as often as primacy items, yet the recall of recency items in final recall remained poor. It seems probable that it is the type of rehearsal rather than the amount of rehearsal which is critical for good long-term retention. If the last few items are encoded and maintained at the phonemic level only, final recall performance will be poor; performance should be enhanced if the subject encodes the words in a richer semantic-associative fashion.

GENERAL DISCUSSION

At first sight the results reported in this paper are surprising. They show that neither the prolonged maintenance of an item in short-term storage (Experiment I) nor a substantial increase in the number of overt rehearsals (Experiment II) increased final free recall performance. It is assumed that final recall performance depends only on the long-term store; thus the results imply that maintenance of an item in short-term storage does not necessarily lead to better registration in the long-term store. This finding is opposed to the conclusions of Waugh and Norman (1965), Atkinson and Shiffrin (1968), and Rundus (1971), since all these workers postulated that longer residence in short-term storage, or a greater number of rehearsals, led to better long-term performance.

On the other hand, there is evidence from a number of recent studies giving strong support to the notion that rehearsal will not enhance final recall performance when a simple maintenance type of rehearsal is sufficient to meet the immediate task requirements. Jacoby and Bartz (1972) showed that, relative to a silent delay, an interpolated filled delay reduced immediate recall of fiveword lists and reduced recall from the recency portion of 20-word lists, but that the filled delay condition yielded better recall performance in a final recall test. The results were interpreted as showing that subjects encoded the words semantically in order to survive the filled delays. Watkins and Watkins (1974) varied the list lengths of a series of free recall lists, and then in a subsequent recall test found negative recency only when subjects were informed of list length; it was also found that knowledge of list length gave rise to a larger recency effect in immediate free recall. Watkins and Watkins concluded that when subjects can identify during input which items are terminal ones, they will process them in a fashion which on the one hand enhances their immediate recall, but on the other hand gives rise to poor long-term retention. Jacoby (1973) has reported an experiment in which subjects recalled fiveword lists either immediately after presentation, or following 15 sec of overt rehearsal. Despite their greater amount of rehearsal, final recall performance for the second group was no higher than that of the "immediate recall" group. Meunier, Ritz, and Meunier

(1972) had subjects recall CVC trigrams in the Brown-Peterson paradigm. One group was given the orthodox counting backwards task as an interpolated activity while a second group was free to rehearse during the retention interval. Not surprisingly, the rehearsal group performed better in the immediate recall test, but the groups performed equally well in an unexpected final recall test. The authors concluded that rehearsal had maintained the items in short-term storage without increasing their strength in the long-term store. Glanzer and Meinzer (1967) showed that while silent and overt rehearsal gave rise to equal performance in the recency portion of free recall lists, overt rehearsal actually led to poorer recall from early list positions. Finally, Tulving (1966) demonstrated that word-list learning was no better for subjects who had previously read aloud the words in the list six times.

The present results, in conjunction with these past findings, demonstrate beyond all reasonable doubt that neither overt rehearsal not maintenance in short-term storage is by itself sufficient to enhance long-term memory performance. This conclusion opposes the models of Waugh and Norman (1965), Atkinson and Shiffrin (1968), and the "total time hypothesis" (see Cooper & Pantle, 1967). Obviously there are situations in which more rehearsal, longer residence in the shortterm store, a greater amount of learning time, and more repetitions do lead to better memory performance. Following Craik and Lockhart (1972), it is suggested that rehearsal can be usefully broken down into its "maintaining" function and its "elaborating" function. To the extent that the subject uses the rehearsal time to enrich and elaborate the memory trace, subsequent retention will be enhanced. If the time is used merely to maintain the trace in some simple form (a phonemic representation, for example), then further repetitions or a prolonged stay in the shortterm store will not lead to better learning and long-term retention.

More sophisticated models of human memory must be devised which show how the time devoted to learning interacts with the qualitative nature of the processing carried out during that time.

REFERENCES

- ATKINSON, R. C., & SHIFFRIN, R. M. Human memory. A proposed system and its control processes. In K. W. Spence & J. T. Spence (Eds.) *The psychology of learning and motivation*. New York: Academic Press, 1968. Vol. 2, pp. 89–195.
- BJORK, R. A. Repetition and rehearsal mechanisms in models for short-term memory. In D. A. Norman (Ed.) *Models of human memory*. New York: Academic Press, 1970. Pp. 307–330.
- COOPER, E. H., & PANTLE, A. J. The total-time hypothesis in verbal learning. *Psychological Bulletin*, 1967, 68, 221–234.
- CRAIK, F. I. M. The fate of primary memory items in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 143–148.
- CRAIK, F. I. M., GARDINER, J. M., & WATKINS, M. J. Further evidence for a negative recency effect in free recall. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 554–560.
- CRAIK, F. I. M., & LOCKHART, R. S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, 11, 671–684.
- GLANZER, M., & MEINZER, A. The effects of intralist activity on free recall. Journal of Verbal Learning and Verbal Behavior, 1967, 6, 928–935.
- JACOBY, L. L. Encoding processes, rehearsal, and recall requirements. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 302–310.

- JACOBY, L. L., & BARTZ, W. H. Rehearsal and transfer to LTM. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 561–565.
- MEUNIER, G. F., RITZ, D., & MEUNIER, J. A. Rehearsal of individual items in short-term memory. *Journal of Experimental Psychology*, 1972, 95, 465-467.
- Norman, D. A., & RUMELHART, D. E. A system for perception and memory. D. A. Norman (Ed.) *Models of human memory*. New York: Academic Press, 1970. Pp. 21–64.
- RUNDUS, D. Analysis of rehearsal processes in free recall. *Journal of Experimental Psychology*, 1971, 89, 63–77.
- RUNDUS, D., & ATKINSON, R. C. Rehearsal processes in free recall: A procedure for direct observation. *Journal of Verbal Learning and Verbal Behavior*, 1970, 9, 99–105.
- TULVING, E. Subjective organization and effects of repetition in multitrial free-recall learning. *Journal of Verbal Learning and Verbal Behavior*, 1966, 5, 193–197.
- TULVING, E., & COLOTLA, V. Free recall of trilingual lists. *Cognitive Psychology*, 1970, 1, 86–98.
- WAUGH, N. C. Immediate memory as a function of repetition. Journal of Verbal Learning and Verbal Behavior, 1963, 2, 107–112.
- WAUGH, N. C. On the effective duration of a repeated word. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 587–595.
- WAUGH, N. C., & NORMAN, D. A. Primary memory. Psychological Review, 1965, 72, 89–104.
- WATKINS, M. J., & WATKINS, O. C. Processing of recency items for free recall. *Journal of Experimental Psychology*, 1974, (in press).

(Received May 1, 1973)