

Fluctuations in recall across successive test trials*

THOMAS O. NELSON and COLIN M. MACLEOD

University of Washington, Seattle, Washington 98195

The purpose of the experiment was to determine whether item fluctuation (intertrial forgetting and spontaneous recovery) depends upon the time allowed for recall and/or the demand characteristics of the experimental situation (manipulated via instructions to S). One hundred and forty-four university undergraduates studied a 36-item list and then had two successive recall trials which were 9, 36, or 144 sec in duration. On the second recall trial, Ss were instructed to recall previously recalled items only ("old"), previously nonrecalled items only ("new"), or both previously recalled items and previously nonrecalled items ("both"). The results suggest two conclusions: (1) Although recall trial duration affects the overall amount of recall, it does not affect the amount of item fluctuation; and (2) an instruction to recall only "old" items or to recall only "new" items results in suppression of the nonrequested items without a corresponding facilitation for the requested items.

Brown (1923) first emphasized that the repeated testing of items does not always produce consistent recall from trial to trial. He noted that two kinds of fluctuations occur across test trials: (1) spontaneous recovery (i.e., recall on Test Trial $n + 1$ of an item not recalled on Test Trial n) and (2) intertrial forgetting (i.e., failure to recall on Trial $n + 1$ an item that was recalled on Trial n). More recently, Tulving (1967) has revived interest in this phenomenon by using a study-test-test free recall procedure and has reported a considerable amount of fluctuation in the recall of individual words over successive test trials.

Interestingly, the magnitude of the fluctuation effect is not stable across different experimental procedures. For example, Birnbaum and Eichner (1971) found a relatively large amount of fluctuation across test trials (in accord with Tulving), while the amount found by Rosner (1970) was relatively small. Also, Lewis (1971) has shown that, if the first test trial (T_1) immediately follows the study trial, a high estimate of intertrial forgetting is obtained at T_2 , due to the loss from T_1 to T_2 of the items in short-term memory (STM). If interest centers on fluctuations in recall from only long-term memory (LTM), then the experimental procedure must prevent recall from STM on T_1 , e.g., by interspersing number counting between study and test (Glanzer & Cunitz, 1966).

The present study employed the study-test-test procedure to examine two variables that could be responsible for item fluctuations in successive free recall from LTM. The first variable was the amount of time permitted for recall on the two test trials. Certainly, if T_1 is too short, S will not be able to output all of his stored items (Donaldson, 1971). As a result, some of the items that are stored but not recalled on T_1 may appear on T_2 , resulting in an artifactual increase in the amount of spontaneous recovery. If time for recall on T_2 is also

limited and if S has already recalled "new" items on T_2 , he may not have enough time to again recall all of those items output on T_1 . This, in turn, will result in an overestimate of intertrial forgetting. Consequently, the recall interval duration may influence estimates of item fluctuation, particularly in studies using relatively short recall intervals. For example, Tulving (1967) used 36-sec intervals for a 36-item list; the curves in Murdock and Okada (1970) suggest that the duration of these recall intervals is too short.

The second variable relates to the demand characteristics of the experimental situation (Orne, 1962). The amount of item fluctuation may depend upon what S believes is expected of him. For example, if S believes that he is supposed to demonstrate how many different items he has retained, he may use T_2 to search for new items not recalled on T_1 ; such an emphasis on outputting new items would almost certainly yield an increased amount of item fluctuation (both intertrial forgetting, particularly if the recall interval is short, and spontaneous recovery). As an example of this type of situation where instructions might increase the estimate of the item fluctuation effect, Tulving (1967, p. 178) told his Ss "not to worry if they could not recall as many words as they should, since they had a chance to improve their recall as they went on." Conversely, if S believes he is supposed to show consistency in recall, then he should use T_2 primarily to output old (T_1) items and the amount of item fluctuation should be relatively low. To examine this possibility, our Ss were instructed to recall on T_2 (a) only new items, (b) only old items, or (c) as many items as possible (both new and old). If this instructional set is effective, the amount of item fluctuation should be greatest for Ss who are told to recall new items only and smallest for Ss who are told to recall old items only.

Finally, an interaction of the recall interval variable with the instruction variable might be expected. Specifically, item fluctuations should decrease as the test interval is lengthened in the "new" and "both"

*This research was supported by United States Public Health Service Grant MH-21037 to the first author. Requests for reprints should be sent to the first author.

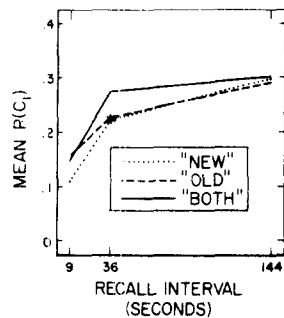


Fig. 1. Mean proportion of correct responses on the first test trial as a function of recall interval (9, 36, 144 sec) and instruction ("new," "old," "both"). (Asterisk at 36 sec represents performance for Group R in Tulving, 1967, Experiment 2.)

instructional conditions but not in the "old" condition. In the limit, very long test intervals might allow S to maximize his performance on both T_1 and T_2 , thereby attenuating the amount of item fluctuation.

To test the above predictions, we used a procedure similar to that of Tulving (1967), with the two major changes being the manipulation of (a) recall interval duration and (b) instructions.

METHOD

Stimuli and Apparatus

The stimuli were the 36 unrelated nouns comprising List A in Tulving (1967, Experiment 2)¹. These were presented on a Lafayette high-speed memory drum programmed by an external tape reader. The memory drum displayed the following, in the order indicated: two rows of asterisks, 36 nouns typed in uppercase, a three-digit number, "RECALL," and "RECALL AGAIN." The S's output was tape recorded.

Design and Subjects

A 3 by 3 factorial design was employed. The first between-Ss variable was the duration of the two test trials. Both T_1 and T_2 were of the same duration for a given S—9, 36, or 144 sec. The second between-Ss variable was the instruction for T_2 output. On T_2 , S was to recall (a) only those words recalled on T_1 ("old"), (b) only those words not recalled on T_1 ("new"), or (c) words recalled on T_1 and words not recalled on T_1 ("both"). Note that the "both" condition corresponds to Group R in Tulving (1967).

The Ss were 144 University of Washington undergraduates whose participation partially fulfilled a course requirement. Factorial combination of the two independent variables produced nine groups, 16 Ss in each. Four different random presentation orders of words were used equally often within each group.

Procedure

The instructions were read to S, who reviewed them aloud before beginning. The S was cautioned to closely attend to each word, because there would be only one presentation of the list. The two rows of asterisks (included as a signal that list presentation was about to begin) and the 36 nouns were presented at a 3-sec rate. Following the last list word, a three-digit number was presented from which S counted aloud, backward by threes, as rapidly as possible for 30 sec. This counting task insured that S was retrieving words only from

LTM during T_1 (Glanzer & Cunitz, 1966; Postman & Phillips, 1965). T_1 immediately followed the 30 sec of number counting. This recall interval lasted 9, 36, or 144 sec and was followed immediately by T_2 , where the to-be-recalled words were "old," "new," or "both."

RESULTS

Figure 1 displays the mean proportion of items correctly recalled on T_1 by each of the nine experimental groups. (The asterisked value represents Tulving's Group R, the reference group for our study.) A two-way analysis of variance showed that duration of the recall interval was a significant variable [$F(2,135) = 54.74$, $p < .001$]. Scheffé post hoc comparisons showed that the 144-sec group recalled more items than the 36-sec group, which in turn recalled more items than the 9-sec group (both comparisons, $p < .05$). This indicates that the levels of the recall interval variable were sufficiently spread to produce reliable differences in the number of items recalled. Neither the instruction variable [$F(2,135) = 2.40$] nor the interaction of Instruction by Recall Interval [$F(4,135) = 1.08$] was significant.

Figure 2 presents the intertrial forgetting data for each of the nine groups in terms of the mean proportion of items incorrect on T_2 that had been correct on T_1 , $P(I_2|C_1)$. The asterisked value is the comparable proportion, $P(I_3|C_2)$, for Tulving's Group R (1967, Experiment 2, Cycle 1). We chose Tulving's intertrial forgetting between T_2 and T_3 as a comparison value because his $P(I_2|C_1)$ overestimates intertrial forgetting from LTM; Tulving did not eliminate recall from STM on T_1 . A two-way analysis of variance showed that intertrial forgetting is affected by type of instruction [$F(2,135) = 782.48$, $p < .001$]. Scheffé post hoc comparisons showed that the "old" and "both" conditions do not differ significantly but, when taken together, they had significantly less forgetting than the

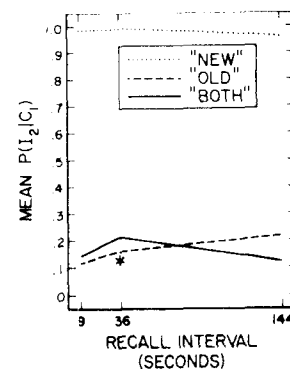


Fig. 2. Intertrial forgetting: Mean proportion of incorrect responses on the second test trial, given correct on the first test trial, as a function of recall interval (9, 36, 144 sec) and instruction ("new," "old," "both"). (Asterisk represents intertrial forgetting for Group R in Tulving, 1967, Experiment 2.)

"new" condition ($p < .05$). This demonstrated that Ss were able to comply with the instructions (cf. instructions to forget in Bjork, 1972). Neither recall interval [$F(2,135) = 1.43$] nor the interaction of Instruction by Recall Interval [$F(4,135) = 2.16$] was significant. The magnitude of intertrial forgetting for the "old" and "both" conditions is quite low [mean $P(I_2 | C_1) = .16$ for these conditions], corresponding rather well to Tulving's value (asterisk in Fig. 2).

Figure 3 presents the mean proportion of items correctly recalled on T_2 that were incorrect on T_1 , $P(C_2 | I_1)$, which represents the amount of spontaneous recovery in each of the nine groups. The asterisk indicates the corresponding proportion for Tulving's Group R, i.e., the proportion of items correct on T_3 that had been incorrect on T_2 (T_2 - T_3 gains were chosen to maintain parallelism with the intertrial forgetting analysis). A two-way analysis of variance showed that the type of instruction significantly affects the amount of spontaneous recovery [$F(2,135) = 13.62$, $p < .001$]. Scheffé post hoc comparisons revealed that this effect was due to the "new" and "both" conditions differing significantly from the "old" condition, again showing that Ss were able to comply with the instructions. We point out that, although significant, this effect is extremely small, with very little spontaneous recovery occurring at all (note the unusually small values on the ordinate of Fig. 3). Both the recall interval duration [$F(2,135) = 2.75$] and the interaction of Instruction by Recall Interval [$F(4,135) = 2.00$] were nonsignificant.

DISCUSSION

The duration of the recall interval, at least in the fairly wide range of 9 to 144 sec, does not seem to affect the amount of item fluctuation; this conclusion applies to both intertrial forgetting and spontaneous recovery. Therefore, these findings do not support the notion that the item fluctuation occurring in previous studies was the result of a too short recall interval.

The type of instruction and, hence, the demand characteristics of the experiment do influence the amount and kind of item fluctuation: Relative to a condition with standard ("both") instructions, "new" instructions increase intertrial forgetting without increasing spontaneous recovery; "old" instructions decrease spontaneous recovery without decreasing intertrial forgetting. Thus, an instruction to recall only "old" items or only "new" items produces suppression of the nonrequested items *without* a corresponding facilitation of the requested items. This implies that standard ("both") instructions serve to maximize spontaneous recovery of "new" items without any loss of "old" items; direct evidence of this comes from our finding that the "both" condition displayed just as much spontaneous recovery as the "new" condition and no more intertrial forgetting than the "old" condition.

Originally, we set out to determine whether item

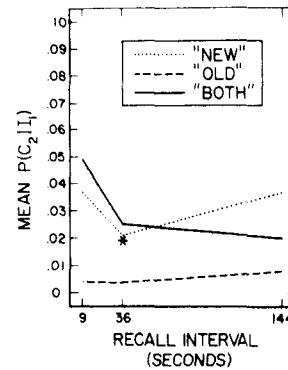


Fig. 3. Spontaneous recovery: Mean proportion of correct responses on the second test trial, given incorrect on the first test trial, as a function of recall interval (9, 36, 144 sec) and instruction ("new," "old," "both"). (Asterisk represents spontaneous recovery for Group R in Tulving, 1967, Experiment 2). Notice values on ordinate.

fluctuation is an artifactual result of (a) a too short recall interval and/or (b) demand characteristics due to instructions. Given the conditions of most free recall experiments (i.e., with the duration of the recall interval being between 9 and 144 sec and with standard "both" instructions), we conclude that the item fluctuation effect is *reliable* and not artifactual. We also emphasize, however, that the *magnitude* of the item fluctuation effect is not large. The 16% intertrial forgetting (for the "old" and "both" conditions shown in Fig. 2) does not seem high enough to warrant being classified as "a truly remarkable instability of recall of individual words [Tulving, 1967, p. 183]." Even less remarkable is the amount of spontaneous recovery (shown in Fig. 3), which is so small as to be nearly undetectable.

One explanation² for the present study's extremely small amount of spontaneous recovery [mean $P(C_2 | I_1) = .02$] concerns the conditionalizing term (I_1) of the conditional probability. I_1 may be comprised of two very different kinds of items: (a) items not present in LTM during T_1 and (b) items present in LTM but not recalled during T_1 . $P(C_2 | I_1)$ must necessarily decrease as the ratio of a items to b items increases. The number of a items may have been relatively large in our experiment because there was only one study trial on the 36 items. In a recent experiment (see Fig. 3 in Buschke, 1973) where the number of a items was minimized, the amount of spontaneous recovery was much greater [$P(C_{n+1} | I_n) = .26$]. Other indirect evidence supporting this notion can be seen in Tulving (1967, Table 1), where the mean number of items that spontaneously recovered on T_3 after being incorrect on T_1 and T_2 increased monotonically with study-test-test cycle (i.e., for Cycles 1-6 the respective means were .33, .81, 1.03, 1.28, 1.39; and 1.67). Future research should make a direct test of this notion that the likelihood of spontaneous recovery depends upon the amount (or strength) of items stored in LTM.

REFERENCES

- Birnbaum, I. M., & Eichner, J. T. Study versus test trials and long-term retention in free-recall learning. *Journal of Verbal Learning & Verbal Behavior*, 1971, 10, 516-521.
- Bjork, R. A. Theoretical implications of directed forgetting. In A. W. Melton and E. Martin (Eds.), *Coding processes in human memory*. New York: Wiley, 1972. Pp. 217-235.
- Brown, W. To what extent is memory measured by a single recall? *Journal of Experimental Psychology*, 1923, 6, 377-382.
- Buschke, H. Selective reminding for analysis of memory and learning. *Journal of Verbal Learning & Verbal Behavior*, 1973, 12, 543-550.
- Donaldson, W. Output effects in multitrial free recall. *Journal of Verbal Learning & Verbal Behavior*, 1971, 10, 577-585.
- Glanzer, M., & Cunitz, A. R. Two storage mechanisms in free recall. *Journal of Verbal Learning & Verbal Behavior*, 1966, 5, 351-360.
- Lewis, M. Q. Categorized lists and cued recall. *Journal of Experimental Psychology*, 1971, 87, 129-131.
- Murdock, B. B., Jr., & Okada, R. Interresponse times in single-trial free recall. *Journal of Experimental Psychology*, 1970, 86, 263-267.
- Orne, M. T. On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist*, 1962, 17, 776-783.
- Postman, L., & Phillips, L. W. Short-term temporal changes in free recall. *Quarterly Journal of Experimental Psychology*, 1965, 17, 132-138.
- Rosner, S. R. The effects of presentation and recall trials on organization in multitrial free recall. *Journal of Verbal Learning & Verbal Behavior*, 1970, 9, 69-74.
- Tulving, E. The effects of presentation and recall of material in free-recall learning. *Journal of Verbal Learning & Verbal Behavior*, 1967, 6, 175-184.

NOTES

1. We thank Endel Tulving for providing his experimental list.
2. This explanation evolved in a discussion with Herman Buschke.

(Received for publication March 12, 1974;
accepted April 6, 1974.)