Research Report

DISPARATE EFFECTS OF REPEATED TESTING: Reconciling Ballard's (1913) and Bartlett's (1932) Results

Mark A. Wheeler and Henry L. Roediger, III
Rice University

Abstract—*Bartlett (1932) gave subjects a prose passage and showed how recall dropped when they were tested repeatedly. Ballard (1913), using poetry, and Erdelyi and Becker (1974), using pictures, reported improvements in performance (or hypermnesia) over repeated testing. We investigated two likely factors leading to the discrepant results: the type of material and the interval between tests. The primary cause of the differing outcomes is the interval between tests. In general, when the intervals between successive tests are short, improvement occurs between tests. When these intervals are long, forgetting occurs. The type of material used plays little role: Hypermnesia in recall of prose (even "The War of the Ghosts") occurred with short intervals between tests. We also report a striking confirmation of the power of tests to enhance memory: Repeated tests shortly after study greatly improved recall a week later.*

In the course of daily affairs, we are frequently asked or required to recollect the same events from memory. This repeated recollection may occur in educational contexts, as when people are tested on material during the course of a semester and later receive a cumulative final exam. Similarly, eyewitnesses to crimes may be queried repeatedly about what they observed by police, by friends, by lawyers, and then eventually in court. More commonly, we all repeatedly retell stories of favorite or notable events in our lives. Despite the ubiquity of such experiences, the strategy of repeatedly testing memory has not been the dominant method of studying how memories change over time. Rather, researchers have generally employed between-subjects comparisons, in which independent groups learn some material and then are tested at various times after the original learning to measure the course of forgetting. This design has the advantage of delayed tests being uncontaminated by recollections on previous tests, as would occur if a within-subjects design were used. However, the "contaminating" influences of previous tests can be of central interest in their own right, with the eyewitness testimony situation representing a prime example (Loftus, 1979). Therefore, the repeated testing of memory for the same events (without intervening study opportunities) also represents a valid, if less used, approach to changes in retention over time.

Several experimental paradigms and research traditions have been used to examine effects of repeated testing on memory. For present purposes, we confine interest to those cases in which the tests were given without overt cues (free recall). The most famous example of this technique is that of Bartlett (1932), who gave college students the Indian folktale "The War of the Ghosts" and then tested them repeatedly. He presented no aggregate data to support his conclusions, but rather provided sample recall protocols from his subjects. He reported that their performance became increasingly poor over time; they forgot the story but reconstructed plausible accounts that were skewed to the more typical schema of a fairy tale, a style that was presumably more familiar to his students. Despite the fact that Bartlett (1932, pp. 3–4) criticized the pioneering work of Ebbinghaus (1885/1964), his conclusions about the course of forgetting were broadly consistent with Ebbinghaus's prior work. Of course, his theoretical emphasis was markedly different.

Curiously, a second tradition of repeated testing research, begun by Ballard (1913), leads to the opposite conclusion from Bartlett's better publicized work. Ballard gave schoolchildren poetry and tested their memories both soon after learning and then again at various periods up to 1 week later. He reported that children frequently recalled lines of poetry on later tests that were not recalled on earlier ones, a phenomenon he termed *reminiscence*. Sometimes the gain in recall between tests outweighed forgetting, so overall performance improved. The basic phenomena were replicated (Brown, 1923; Williams, 1926) and became known in textbooks as the Ballard-Williams reminiscence phenomenon (e.g., Osgood, 1953, pp. 564–566).[1] Bartlett (1932) never cited these articles in *Remembering*, and never noted the discrepancy between these findings and his own.

The topic of reminiscence fell into disrepute in the 1940s and 1950s, when researchers failed to obtain absolute gains in performance across repeated testing under some conditions (Buxton, 1943).[2] The topic has enjoyed a resurgence of interest in the past 20 years, owing largely to the work of Matthew Erdelyi, who developed standard experimental settings that reliably produce the phenomenon. Erdelyi and Becker (1974) had subjects study either a large set of pictures or a set of words (the names of the pictures). Subjects received three tests without recall cues of any sort. Each test lasted 7 min, with short breaks

Address correspondence to the authors at the Department of Psychology, Rice University, Houston, TX 77251-1892.

1. Although Williams (1926) did replicate Ballard's (1913) reminiscence phenomenon under highly similar conditions, he actually failed to obtain the effect in most conditions.

2. Payne (1987) noted that reminiscence was redefined by researchers after Ballard as net improvements in recall over tests. They often failed to find such effects and concluded that the phenomenon was unreliable. Yet when defined as Ballard (1913, pp. 17–18) did originally—"remembering one or more [items] that were not remembered in a [prior] test"—the phenomenon is regularly observed in most studies. Curiously, Bartlett (1932, pp. 67, 72) provided only two brief mentions of subjects correctly recalling on a later test a fact that was not recalled previously.

occurring between tests. Subjects recalled an increasing number of pictures across the three tests, but overall recall of words was stable. Erdelyi and Becker labeled the improvement in recall across tests *hypermnesia* to distinguish the phenomenon from amnesia, or forgetting, over time. Others have reported hypermnesia even for verbal materials (Klein, Loftus, Kihlstrom, & Aseron, 1989; Payne & Roediger, 1987). Roediger and Thorpe (1978) obtained hypermnesia for verbal stimuli and distinguished between reminiscence and hypermnesia, in accord with the definitions of Ballard (1913) and Erdelyi and Becker (1974). Reminiscence refers to recall on a later test of items that could not be recalled on an earlier test. Virtually all studies of repeated free recall have shown this phenomenon. However, reminiscence is often neutralized or outweighed by intertest forgetting so that there is no overall net improvement in recall across tests, especially with verbal materials and short amounts of recall time (Tulving, 1967). Nonetheless, net improvements in recall across tests, or hypermnesia, can be reliably obtained in many situations, particularly those in which overall levels of performance are fairly high (Payne, 1987).

Various theoretical approaches for explaining hypermnesia have been explored, involving, for example, an imaginal encoding format (Erdelyi, 1984) and the retrieval dynamics of cumulative recall curves (Roediger & Challis, 1989; Payne, 1986, 1987). Raaijmakers and Shiffrin (1980, 1981) proposed a variety of mechanisms for exploring hypermnesia. The impetus for the present experiments, however, was not to explain reminiscence or hypermnesia. Rather, we sought to understand how two research techniques that appear so similar on the surface could lead to such different results and interpretations about how memory changes over time.

In the tradition of work begun by Bartlett (1932), subjects learned prose and were typically tested once about 15 min later, and then again either days or weeks later. In the research tradition initiated by Ballard (1913) and rejuvenated by Erdelyi and Becker (1974), subjects were more typically required to study a list of pictures or words and were tested repeatedly over short intervals of time during the experimental session. Two obvious differences between the traditions that might explain the discrepancy in outcomes are the type of material (prose or lists) and the amount of time between tests (short or long). We investigated both sets of factors in the current experiments. It is not obvious from prior work which should be more important. Ballard reported reminiscence with passages of poetry that may be considered connected discourse, like prose. He also obtained the effect with a 24-hr interval between successive tests. Similarly, Erdelyi and Kleinbard (1978) obtained hypermnesia for lists of pictures over a week after presentation by testing subjects three times a day. Finally, Roediger, Payne, Gillespie, and Lean (1982, Experiment 3) found hypermnesia with well-organized categorical materials retrieved from permanent memory (U.S. presidents, sports, and birds).

In Experiment 1, subjects studied a series of pictures that were embedded in a story read orally. They then took a series of up to six tests (depending on the condition) shortly after the study phase and after a week. We attempted to replicate both the findings of Bartlett (1932)—sharp forgetting between successive tests—and those of Erdelyi and Becker (1974) and Ballard (1913)—improvements over tests. In Experiment 2, we attempted to obtain hypermnesia in recall of a prose passage when tests were repeatedly given over short delays.

EXPERIMENT 1

All subjects studied 60 pictures, either in the context of a story (thereby simulating to some extent schematic processing as in Bartlett's work) or with their names (similar to the list-learning situation employed by Erdelyi and Becker). Subjects then took either zero, one, or three tests immediately. We expected to replicate the typical hypermnesia phenomenon in the three-test condition, at least under the listlike presentation conditions. All subjects returned a week later and took three more tests under identical conditions. We expected to see intertest forgetting over this long delay, at least in the conditions involving presentations of pictures in the story context. We were also interested in the subsidiary question of how recall of the pictures after a week would be affected by the number of tests taken initially.

Method

Subjects

Subjects were 120 Rice University undergraduate students participating either for a course requirement or for extra credit in a psychology course. They were tested in small groups, ranging in size from 3 to 9.

Materials

The same 60 slides were shown to all subjects in the same sequence. The slides represented pictures that could be named easily. Half of the subjects listened to a children's story as they watched the slides. The story was written so that each pictured object was being mentioned in the story at the time that slide was presented. The other half of the subjects, not given the story, listened to the name of each picture as it was being viewed.

Design

Half of the subjects experienced the presentation while hearing only the name of each picture being read from a cassette player (the Pictures + Names condition). For the remaining subjects, the slides were accompanied by a children's story, which was read from the cassette player during presentation (the Pictures + Story condition). Subjects differed in the number of recall tests they received on the 1st day of testing: none, one, or three. All subjects returned after a week and took three additional recall tests.

Therefore, there were six between-subjects conditions, with two types of presentation (Pictures + Names and Pictures + Story) and three different possible schedules of recall tests (zero, one, or three tests on Day 1). Twenty subjects were tested in each condition. In addition, there was the repeating measure of memory test for each subject, providing for a mixed-factorial design.

Procedure

All subjects were originally told that they would be watching a series of picture slides, about which they would "later be asked some questions." Subjects in the Pictures + Names condition were told that they would also hear the name of each picture from a cassette

player as the picture was viewed, so that no ambiguity would exist as to the identity of each object. The other subjects (in the Pictures + Story condition) were told that they would hear a children's story via a tape recorder and that the pictures would be shown when they arose in the story. There were 60 slides presented at a rate of 7 s each, with each slide immediately following the preceding one (with 0.75 s change time between slides).

Immediately following presentation of the slides, all subjects spent 5 min recalling as many U.S. presidents as they could. Then subjects were given a questionnaire on which they guessed how many slides they had seen, how long each slide had appeared, and the total length of the entire presentation. They were also asked to recall the instructions they had received before viewing the slides. Subjects were given 3 min to complete the questionnaire. Following the questionnaire, one third of the subjects in both the Pictures + Names and Pictures + Story conditions were dismissed, and they were reminded to return for the next week's session. The purpose of the questionnaire was to create a plausible cover story for the subjects who were not tested immediately, so they would be less likely to expect a test on the material on their return.

All other subjects were informed that they would be given a recall test for the slides using a procedure called forced recall. The experimenter passed out test sheets, with lines numbered 1 to 60, and told subjects that they would have 7 min to recall the names of the pictures. If they were unable to remember all 60 objects, then they were to fill in the remaining spaces with their best guesses. The experimenter stressed that all 60 spaces on the test sheet must be filled in so subjects would pace themselves to complete the memory test by the end of the 7-min period. (Erdelyi and Becker used forced recall in their original studies of hypermnesia, because this procedure holds response criteria constant across tests.) All subjects who took at least one recall test in the first experimental session began their first test 11.5 min after the end of the slide presentation.

Subjects then began their memory tests. If all 60 spaces had not been completed after the 7 min, the subjects were instructed to fill in the remaining spaces as quickly as possible with guesses. After this first test, subjects were either excused (in the one-test condition) or required to take two more tests using the same forced-recall procedure (in the three-test condition), with 1-min breaks between tests.

One week later, subjects returned to the same room at the same time and received three more forced-recall tests for the pictures under identical forced-recall conditions. Again 1 min elapsed between tests, and subjects were informed that the third test would be their last. Afterwards, subjects were debriefed and thanked.

Results and Discussion

The mean numbers of words recalled on all tests in the Pictures + Names and Pictures + Story conditions are shown in Table 1 as a function of the number of tests given in the initial testing session. These data can be used to answer four basic questions of interest.

First, did hypermnesia occur on the initial tests, and was it confined to only the list-learning situation (the Pictures + Names condition)? The answer appears in the Group 3-3 data on the left-hand side of Table 1. Hypermnesia (the improvement in recall across successive tests) did occur on the initial tests and occurred for both study conditions. The improvement over tests was greater in the Pictures + Story condition (3.8 items) than in the Pictures + Names condition (1.8 items). Also, the story schema seemed effective in that it improved overall recall for the pictures. These impressions from Table 1 were confirmed by a 2 (Pictures + Names or Pictures + Story) × 3 (Test 1, 2, or 3) analysis of variance (ANOVA), which indicated a main effect both for study condition, $F(1, 38) = 7.62$, $MS_e = 208.85$, $p < .01$, and for test number, $F(2, 76) = 17.20$, $MS_e = 4.32$, $p < .001$. The interaction between presentation context and test number fell short of significance, $F(2, 76) = 2.66$, $MS_e = 4.32$, $p < .08$.

The second question of interest is whether forgetting occurred when 1 week (rather than a few minutes) intervened between successive tests. The answer is clearly yes: There was reliable forgetting in each of the four relevant conditions across the 7-day interval, as measured from the last test taken on the 1st day to the first test taken a week later (smallest $t[39] = 3.88$). (The relevant contrast for Groups 3-3 is between performance on the third initial test and on the first delayed test; for Groups 1-3, the contrast is between their sole initial test and their first delayed test.) There was also a significant interaction between the number of prior tests and the amount of forgetting, $F(1, 78) = 13.01$, $MS_e = 7.62$, $p < .001$. Subjects who had taken three initial tests forgot less over 1 week (a mean of 3.9 items) than those who had taken only one test initially (7.0 items).

Table 1. *Mean number of pictures recalled as a function of presentation context and testing schedule in Experiment 1*

	Initial tests				Delayed tests			
Group	T1	T2	T3	T3 − T1	T1	T2	T3	T3 − T1
				Pictures + Names				
3-3	26.6	27.2	28.4	1.8*	25.2	26.3	26.0	0.8
1-3	25.7				20.2	21.7	23.0	2.8*
0-3					16.7	17.5	17.5	0.8
				Pictures + Story				
3-3	32.7	35.0	36.4	3.8*	31.8	33.0	33.4	1.6*
1-3	31.8				23.3	25.0	25.6	2.3*
0-3					17.4	17.2	18.4	1.0

Note. All groups took three tests in the delayed session. Group 3-3 received three tests in the initial session, Group 1-3 received one test in the initial session, and Group 0-3 took no memory tests until the delayed session.
* These conditions demonstrated reliable hypermnesia across the three consecutive tests as indicated by a simple analysis of variance.

The third question of interest is whether the number of initial tests, besides affecting forgetting over a week, also affected the total level of recall on the delayed tests. Once again, the answer is yes. As can be seen from the right side of Table 1, the number of initial tests had a powerful effect on delayed recall. For example, in the Pictures + Story condition, the mean number of words recalled on the first delayed test was 17.4 if no initial tests had been given, 23.3 if one prior test had been given, and 31.8 if subjects had been tested three times initially. In addition, recall was greater on the delayed tests in the Pictures + Story condition than in the Pictures + Name condition (at least, as long as one initial test had been given). These conclusions were confirmed by a 2 (presentation context) × 3 (initial testing schedule) × 3 (number of test: 1, 2, or 3) ANOVA, which showed a significant main effect for the number of initial tests, $F(2, 114) = 33.79$, $MS_e = 124.02$, $p < .08$, with the number of initial tests tending to affect recall more powerfully in the Pictures + Story condition.

Fourth, did hypermnesia occur on the delayed tests? In general, the answer seems to be yes, because recall on the third delayed test was greater than that on the first delayed test in all six conditions, and the main effect for the number of delayed tests (1, 2, or 3) in the abovementioned ANOVA was significant, $F(2, 114) = 14.35$, $MS_e = 5.03$, $p < .001$. However, it is also clear that the effect is not particularly robust, especially when no initial tests were given. The data from subjects in each of the six conditions were subjected to simple ANOVAs for number of delayed tests, and only three were significant (see Table 1). Few researchers have examined hypermnesia on delayed tests. Apparently, it occurs (at least when the delayed tests occur in rapid succession), but the magnitude seems variable, for poorly understood reasons.

The results provide straightforward conclusions: The primary factor causing the difference between the results of Ballard (1913) and Bartlett (1932) is the interval of time between successive tests. When the intervals are relatively short (only 1 min in our experiment, but up to 5 or more min in other experiments), one generally obtains improvements in recall over repeated tests. When the intervals between tests are long (1 week here, but up to months in Bartlett's experiments), forgetting occurs. This conclusion held whether pictures were studied in a list or were embedded in the context of a children's story, indicating that the type of material may not be crucial to the outcome. The story generally boosted recall of the pictures, so apparently schematic or other elaborative processes were invoked by the story. However, Experiment 2 is also directed at the issue of whether the conclusions of Experiment 1 generalize to prose material. The other notable result from Experiment 1 is the power of testing in improving later recall. For example, the three groups of subjects in the Pictures + Story condition studied pictures under identical conditions, presumably encoding similar amounts of information, and then took three successive tests a week later. Subjects who received no immediate tests recalled an average of 17.6 words a week later, those who had one test recalled 24.6, and those who had initially taken three tests averaged 32.7. We consider these results in the General Discussion.

EXPERIMENT 2

The main conclusion drawn from Experiment 1 is that whether one obtains hypermnesia or forgetting between successive tests depends primarily on the amount of time between tests, and that the type of material studied is relatively unimportant. However, in some sense the material was the same in the two conditions of Experiment 1, as pictures were used in both conditions (albeit in a story context in one case).

The purpose of Experiment 2 was to replicate the primary results of Experiment 1—hypermnesia between tests with short intervals—with a prose passage. As a pilot study, we had 52 students in a cognitive psychology course read "The War of the Ghosts" twice at a comfortable rate. Then the students spent 5 min recalling U.S. presidents before receiving instructions on how to recall the story (instructions were similar to those outlined below for Experiment 2). We tested them twice on the story, at 8.5 min per recall test. There was an 8-min interval between study and Test 1, and a 5-min delay occurred between Tests 1 and 2. (Between tests, subjects recalled names of U.S. states.) Performance was scored in terms of the number of idea units recalled, as suggested in Mandler and Johnson's (1977) analysis of the story, in which they subdivided "The War of the Ghosts" into 42 propositions. Subjects recalled 21.4 idea units on the first test and 22.9 on the second. Although the effect appears small, 37 out of the 47 subjects whose level of recall changed showed improvement on Test 2, making the effect highly significant by a sign test. Each proposition is composed of an average of about eight words, so the difference would appear greater if scored in terms of the number of words recalled. Experiment 2 confirmed this preliminary finding of hypermnesia with a different prose passage.

Method

Subjects

Subjects were 38 Rice University undergraduate students enrolled in lower division psychology courses. They participated in exchange for partial course credit.

Materials

A passage excerpted from John Updike's short story "The Kid's Whistling" was used as study material. The passage was broken down into 41 idea units. Each idea unit consisted of a major event or description within the story, and the idea units were an average of about eight words long. (The story, separated into its idea units, can be obtained from the authors.)

Design

Subjects were randomly assigned to a Strict or a Relaxed Criterion recall condition (see below). Each subject also took three separate memory tests, providing for a mixed design (condition × test number).

Procedure

At the beginning of the experiment, the experimenter told subjects that this was an investigation of memory for short stories. All subjects were handed a sheet of paper with "The Kid's Whistling" typed on one side. Subjects were instructed to read the passage through twice at their normal reading rate, which took between 90 and 165 s.

Subjects in the Relaxed recall condition then spent 5.5 min recalling U.S. presidents. Those in the Strict condition received only 5 min for this task, to equate overall retention interval, because their instructions were longer. In the Relaxed condition, subjects were instructed to write down the story as well as they could, using exact wording when possible. When this was not possible, subjects were to describe the events of the story as accurately as they could in their own words. In the Strict Criterion condition, subjects heard all of the above instructions, plus the following:

> Please write down a fact only if you can explicitly remember it from the story. Do not write something down just because it makes sense to you. Make sure you can remember it being in the passage that you read. Of course you do not have to remember it verbatim, but please report only the facts of which you are confident.

When it was clear that all subjects understood the instructions, they were given 9 min for story recall. Following the recall trial, subjects spent 3 min recalling the names of U.S. states. Relevant instructions were briefly summarized for all subjects, and then they had another 9 min to report the story again, under the identical response criterion as in Test 1.

After a 7-day delay, subjects returned to the same room and were given one additional memory test for the story. This time, all subjects were tested using the strict response criterion. After the 9-min test, subjects were asked to reread what they had written. The experimenter told them that it was important that everything they had reported was something they remembered from the story; if there was any information about which they were not certain, they were to put that information in parentheses. The experimenter warned them not to cross out anything that was already written, and just to put parentheses around questionable information. After this part was completed, subjects were debriefed and thanked for their participation.

Results and Discussion

Protocols were scored for three categories of responses. Subjects were given credit for recall of a correct idea unit if their report included a phrase that was judged to be identical with, or very close to, 1 of the 41 preestablished idea units. Two types of errors were also scored. A major error was marked if a subject reported a fact which was clearly incorrect and inconsistent with the story. For example, one subject reported that "the boy was humming" when he was actually whistling. A minor error occurred if a subject wrote information which was not quite correct, yet was consistent with each action and event within the story. For example, if it was reported that "the boy offered him a nickel and two pennies," rather than the correct "nickel and five pennies," then a minor error was recorded. A minor error also occurred if a subject reported an incorrect name for a character, even if it happened consistently throughout a protocol. Correct idea units, plus major and minor errors, were originally analyzed as a function of both recall criterion and test number. The full set of analyses (relevant data are summarized in Tables 2 and 3) revealed that recall criterion had no significant effect or interaction on any measure. Data were collapsed across recall criterion for subsequent analyses, which are discussed below. Although the Strict-Relaxed instructional manipulation at test was ineffective, its inclusion serves the purpose of providing a replication of the main finding.

The issue of interest is whether hypermnesia can be obtained in prose recall with short intervals between tests. The answer (in Table 2) is yes: On average, 12.8 idea units were recalled on the first test and 13.8 on the second, $F(1, 37) = 15.27$, $MS_e = 1.31$, $p < .001$. In addition, the drop in recall owing to the week's delay between the second and third tests was significant, $F(1, 37) = 65.23$, $MS_e = 2.62$, $p < .001$.

Table 2. *Mean idea units recalled as a function of recall criterion and test in Experiment 2*

Recall criterion	Initial tests T1	Initial tests T2	Delayed test T3
Strict	12.1	13.2	10.7 (10.4)
Relaxed	13.4	14.4	10.8 (10.1)
Mean	12.8	13.8	10.8 (10.3)

Note. Figures in parentheses represent the number of idea units of which subjects were confident in Test 3.

Table 3. *Mean errors reported as a function of recall criterion and test in Experiment 2*

Recall criterion	Error type	Initial tests T1	Initial tests T2	Delayed test T3
Strict	Major	0.7	0.7	0.6 (0.4)
	Minor	0.5	1.1	0.9 (0.5)
	Total	1.2	1.8	1.5 (0.9)
Relaxed	Major	0.6	0.7	0.7 (0.5)
	Minor	1.2	1.5	1.3 (0.9)
	Total	1.8	2.2	2.0 (1.4)

Note. Figures in parentheses represent the number of errors about which subjects expressed confidence in Test 3.

The error analyses did not produce much of interest. There was no reliable change in major errors across the three tests, although minor errors did increase between the first and second initial tests, $F(1, 37) = 14.51$, $MS_e = 0.26$, $p < .001$. After Test 3, subjects were given an opportunity to indicate the reported information about which they were not certain. Results indicate that subjects identified 33% of their major errors and 47% of their minor errors. Only 5% of the correct idea units were marked as questionable by subjects. Clearly, we did not find the dramatic distortions in repeated recall of prose that were reported by Bartlett (1932).

The results of Experiment 2 are fully in accord with those of Experiment 1: With short intervals between repeated tests, recall of a short story improved over tests; with a week between tests, forgetting occurred. However, subjects did not show dramatic errors over the week's delay to implicate the reconstructive processes that Bartlett (1932) claimed in his subjects. The facts that the passage was quite different and that the interval between the second and third tests was only a week may account for this disparity. In sum, both "The War of the Ghosts" pilot experiment and Experiment 2 indicate that hypermnesia can be

obtained with prose passages if tests are separated by short delays.

GENERAL DISCUSSION

The aim of the present experiments was to solve the puzzle of why formally similar repeated-testing paradigms have produced discrepant results: The repeated-reproduction paradigm introduced by Bartlett (1932) revealed dramatic forgetting and distortion over tests, whereas the reminiscence and hypermnesia paradigms initiated by Ballard (1913) and Erdelyi and Becker (1974) showed better recall on later tests than earlier tests. Results of the present experiments implicate the interval between successive tests as the critical factor. When this interval was short, on the order of minutes, performance on successive noncued tests improved in the present experiments. The type of information studied (pictures in a list or in a story, or prose passages) was immaterial.

In some sense, it is no surprise that the interval between repeated tests played a large role in determining intertest forgetting. To take a limiting case, if subjects learned a prose passage and were tested shortly after study and then a year later, recall would surely decline. The more interesting point is that reminiscence and hypermnesia can be obtained even in recall of prose passages. As our pilot study for Experiment 2 shows, if Bartlett had chosen to test his subjects repeatedly over shorter intervals of time, the conclusions he drew from his famous "The War of the Ghosts" experiments might have been dramatically different. Of course, in the modern era of cognitive psychology, the role of reconstructive processes has been investigated with methods other than repeated reproduction (e.g., Bransford & Johnson, 1972; Loftus, 1979; Spiro, 1980).

Another interesting point to emerge from Experiment 1 is the powerful effect that testing has on later recall. After comparable study conditions, subjects who were given three successive tests recalled about twice as much a week later as subjects who received no tests in between; subjects who received one test immediately after study performed at an intermediate level. This positive effect of taking tests on later retention has been long known (Gates, 1917; Raffel, 1934; Spitzer, 1939), but the educational implications have not been developed, as noted by Glover (1989). Because taking tests has such powerful positive effects on memory, perhaps the introduction of more frequent tests in classrooms would boost educational achievement.

To return to the main point, the paradox of why repeated testing of memories sometimes reveals dramatic forgetting and in other cases surprising recoveries has, to our knowledge, never been raised, despite its existence since Bartlett's studies. The contribution of this article is to raise the puzzle and to provide its tentative solution: Repeated testing without external recall cues reveals hypermnesia (or at least reminiscence) when the tests occur close together in time, but forgetting when they are more widely spaced.

Acknowledgments—This research was reported at the 1990 meetings of the Psychonomic Society in New Orleans and at the 1991 Texas Cognition Conference in College Station, Texas. The experiments constituted part of a master's thesis submitted to Rice University by the first author under the direction of the second author. We thank committee members David J. Schneider and Michael J. Watkins for their advice, and Stacey Bourland for her assistance testing subjects.

REFERENCES

Ballard, P.B. (1913). Oblivescence and reminiscence. *British Journal of Psychology Monograph Supplements, 1*, 1–82.

Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, England: Cambridge University Press.

Bransford, J.D., & Johnson, M.K. (1972). Contextual prerequisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior, 11*, 717–726.

Brown, W. (1923). To what extent is memory measured by a single recall trial? *Journal of Experimental Psychology, 6*, 377–382.

Buxton, C.E. (1943). The status of research in reminiscence. *Psychological Bulletin, 40*, 313–340.

Ebbinghaus, H. (1964). *Memory: A contribution to experimental psychology* (H.A. Ruger & C.E. Bussenius, Trans.). New York: Dover. (Original work published 1885)

Erdelyi, M.H. (1984). The recovery of unconscious (inaccessible) memories: Laboratory studies of hypermnesia. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 18, pp. 95–127). New York: Academic Press.

Erdelyi, M.H., & Becker, J. (1974). Hypermnesia for pictures: Incremental memory for pictures but not for words in multiple recall trials. *Cognitive Psychology, 6*, 159–171.

Erdelyi, M.H., & Kleinbard, J. (1978). Has Ebbinghaus decayed with time? The growth of recall over days. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 275–289.

Gates, A.I. (1917). Recitation as a factor in memorizing. *Archives of Psychology, 6*(40).

Glover, J.A. (1989). The testing phenomenon: Not gone but nearly forgotten. *Journal of Educational Psychology, 81*(3), 392–399.

Klein, S.B., Loftus, J., Kihlstrom, J.F., & Aseron, R. (1989). Effects of item-specific and relational information on hypermnesic recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 1192–1197.

Loftus, E.F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.

Mandler, J.M., & Johnson, N.S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9*, 111–151.

Osgood, C.E. (1953). *Method and theory in experimental psychology*. New York: Oxford University Press.

Payne, D.G. (1986). Hypermnesia for pictures and words: Testing the recall level hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 16–29.

Payne, D.G. (1987). Hypermnesia and reminiscence in recall: A historical and empirical review. *Psychological Bulletin, 101*, 5–27.

Payne, D.G., & Roediger, H.L. (1987). Hypermnesia occurs in recall but not recognition. *American Journal of Psychology, 100*, 145–156.

Raaijmakers, J.G.W., & Shiffrin, R.M. (1980). SAM: A theory of probabilistic search of associative memory. In G.H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 14, pp. 207–262). New York: Academic Press.

Raaijmakers, J.G.W., & Shiffrin, R.M. (1981). Search of associative memory. *Psychological Review, 88*, 93–134.

Raffel, G. (1934). The effect of recall on forgetting. *Journal of Experimental Psychology, 17*, 828–838.

Roediger, H.L., & Challis, B.H. (1989). Hypermnesia: Improvements in recall with repeated testing. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Floweree Symposium on Cognition* (pp. 175–199). Hillsdale, NJ: Erlbaum.

Roediger, H.L., Payne, D.G., Gillespie, G.L., & Lean, D.S. (1982). Hypermnesia as determined by level of recall. *Journal of Verbal Learning and Verbal Behavior, 21*, 635–655.

Roediger, H.L., & Thorpe, L.A. (1978). The role of recall time in producing hypermnesia. *Memory & Cognition, 6*, 296–305.

Spiro, R.J. (1980). Accommodative reconstruction in prose recall. *Journal of Verbal Learning and Verbal Behavior, 19*, 84–95.

Spitzer, H.F. (1939). Studies in retention. *Journal of Educational Psychology, 30*, 641–656.

Tulving, E. (1967). The effects of presentation and recall in free recall learning. *Journal of Verbal Learning and Verbal Behavior, 6*, 175–184.

Williams, O. (1926). A study of the phenomenon of reminiscence. *Journal of Experimental Psychology, 9*, 368–387.

(RECEIVED 7/8/91; ACCEPTED 9/20/91)