

## INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

**The quality of this reproduction is dependent upon the quality of the copy submitted.** Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

# U·M·I

University Microfilms International  
A Bell & Howell Information Company  
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA  
313/761-4700 800/521-0600



**Order Number 1345306**

**Reconciling the work of Ballard and Bartlett: An investigation  
of repeated testing**

**Wheeler, Mark Allen, M.A.**

**Rice University, 1991**

**U·M·I**  
300 N. Zeeb Rd.  
Ann Arbor, MI 48106



RICE UNIVERSITY

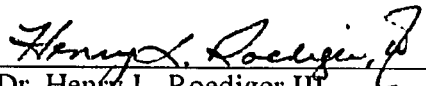
RECONCILING THE WORK OF BALLARD AND BARTLETT:  
AN INVESTIGATION OF REPEATED TESTING

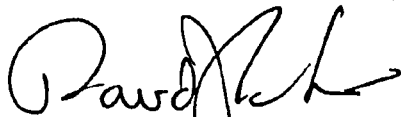
by


MARK A. WHEELER

A THESIS SUBMITTED  
IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE  
MASTER OF ARTS

APPROVED, THESIS COMMITTEE

  
\_\_\_\_\_  
Dr. Henry L. Roediger III,  
Professor of Psychology, Chair

  
\_\_\_\_\_  
Dr. David J. Schneider,  
Professor of Psychology

  
\_\_\_\_\_  
Dr. Michael J. Watkins,  
Professor of Psychology

Houston, Texas

April, 1991

**RECONCILING THE WORK OF BALLARD AND BARTLETT:  
AN INVESTIGATION OF REPEATED TESTING**

**by**

**MARK A. WHEELER**

**Abstract**

Three experiments investigated the ways memories change over repeated tests. A tradition of previous work started by Ballard (1913) revealed net improvement (or hypermnesia) over tests, while Bartlett (1932) showed intertest forgetting. Differences between the paradigms include the type of material learned (lists of items vs. prose) and the delay interval between recall tests (several days for Bartlett and up to one day for Ballard). Both factors were examined in three experiments. In all experiments, hypermnesia was obtained across repeated memory tests when tests occurred within a single, experimental session. However, intertest forgetting resulted after a one-week delay, replicating Bartlett. Performance on repeated tests was discovered to be a function of the number of prior tests and the delay between tests, while response criterion (lenient or strict) and the nature of the learned material (pictures or prose passages) were not important.

**Table of Contents**

<b>An Investigation of Repeated Testing</b>	<b>1</b>
<b>Reminiscence and Hypermnnesia</b>	<b>1</b>
<b>Reconstruction over Repeated Tests</b>	<b>6</b>
<b>EXPERIMENT 1</b>	<b>12</b>
<b>Method</b>	<b>12</b>
<b>Subjects</b>	<b>12</b>
<b>Materials</b>	<b>12</b>
<b>Design</b>	<b>13</b>
<b>Procedure</b>	<b>14</b>
<b>Results</b>	<b>17</b>
<b>Discussion</b>	<b>22</b>
<b>EXPERIMENT 2</b>	<b>26</b>
<b>Method</b>	<b>28</b>
<b>Subjects</b>	<b>29</b>
<b>Materials</b>	<b>29</b>
<b>Design</b>	<b>29</b>
<b>Procedure</b>	<b>29</b>
<b>Results</b>	<b>31</b>
<b>Discussion</b>	<b>35</b>
<b>EXPERIMENT 3</b>	<b>36</b>
<b>Method</b>	<b>37</b>
<b>Subjects</b>	<b>37</b>
<b>Materials</b>	<b>38</b>
<b>Design</b>	<b>38</b>
<b>Procedure</b>	<b>38</b>
<b>Results</b>	<b>39</b>

Discussion	44
GENERAL DISCUSSION	47
REFERENCES	52
Appendix A	56
Appendix B	61
Appendix C	63



**List of Tables**

<b>Table 1</b>	<b>Mean Number of Pictures Recalled as a Function of Presentation Context and Testing Schedule in Experiment 1</b>
<b>Table 2</b>	<b>Mean NC and CN Scores Across the Three Immediate Tests as a Function of Stimulus Materials and Test Number in Experiment 1</b>
<b>Table 3</b>	<b>Mean Pictures Correctly Recalled, Confidently Recalled, and False Alarms as a Function of Response Criterion and Test Number in Experiment 2</b>
<b>Table 4</b>	<b>Mean Idea Units Recalled by Condition and Test in Experiment 3</b>
<b>Table 5</b>	<b>Mean Errors Reported as a Function of Recall Criterion and Test in Experiment 3</b>
<b>Table 6</b>	<b>Mean NC and CN Scores for Correct Idea- Units as a Function of Response Criterion and Test in Experiment 3</b>

**List of Figures****Figure 1****Hypothetical Cumulative Recall Curves**

## AN INVESTIGATION OF REPEATED TESTING

Psychologists have used many different tests to investigate human memory, with the intent to assess retention as accurately as possible as a function of interacting variables.

Another interesting issue involves the extent to which these various tests not only measure memories, but also influence remembering of the prior episodes later. Several different traditions have employed the procedure of repeated testing, which refers to experiments in which subjects take two or more tests over the same stimulus material without any intervening study of the material. This procedure, combined with appropriate controls, affords a look at the consequences that a memory test has on subsequent tests.

Surprisingly, there is evidence for two opposing results of repeated testing. Research in the area of reminiscence and hypermnesia (beginning at least with Ballard, 1913) has investigated memory improvement over time with repeated tests, while the reconstructive approach of Bartlett (1932) implicated the forgetting, or distortion, of information over tests. These literatures are briefly reviewed.

### *Reminiscence and Hypermnesia*

Experimental research in this area began early in this century with a study by Ballard (1913). He presented London schoolchildren with a long poem to memorize, and then tested their retention of the poem at various intervals. Children often recalled passages of the poem on later tests that they could not recall on prior tests. Overall retention improved with each test through the third recall trial (which was three days after presentation of the poem). Ballard explained that if someone "remembers one or more lines that were not remembered on a (prior) test. . .reminiscence has taken place" (pp.17-18). The term reminiscence was not used in reference to a net increase in retention over time, but rather to the recovery on later tests of information forgotten on earlier tests. Ballard also pointed out that reminiscence could occur even when overall performance decreased over tests.

These results seemed contrary to the classic Ebbinghaus (1885/1964) forgetting curve, and therefore the phenomenon of reminiscence was further investigated for a number of years. Ballard's finding was replicated with such stimulus items as U. S. states (Brown, 1923) and common objects (Nicolai, 1922, cited in Woodworth & Schlossberg, 1955). Improvement in overall performance was not demonstrated, however, using abstract words (Williams, 1926) or nonsense syllables (Luh, 1922, in Woodworth & Schlossberg, 1955). It should be noted that there was confusion about Ballard's (1913) original definition of reminiscence, which is actually the remembering of material not remembered on a previous test. Both the Luh (1922, in Woodworth & Schlossberg) and Williams (1926) articles considered reminiscence to be a net increase in retention on a later test (what is now called hypermnesia.) Largely as a result of this confusion over the intended definition of reminiscence, empirical support for the phenomenon was mixed in the subsequent years (Payne, 1987). The study of reminiscence declined in popularity and vanished from the learning and memory literature for many years.

Within the last two decades, interest in the area has been rekindled, largely by a series of articles by Matthew Erdelyi and his colleagues. Erdelyi and Becker (1974) sought to demonstrate a significant increase in memory over time with repeated tests. In two experiments, they found a reliable net improvement over the successive tests for pictures, but not for words. The term hypermnesia was introduced to describe these net improvements in recall, while reminiscence was reserved for the later recall of previously unretrieved items (Roediger & Thorpe, 1978).

Erdelyi, Finkelstein, Herrell, Miller, and Thomas (1976) replicated the finding of hypermnesia for pictures, but not for words. They also demonstrated, however, that words could lead to a net increase in retention across tests if they were encoded using visual imagery. Erdelyi and Becker (1974) proposed that imagery plays the major role in the

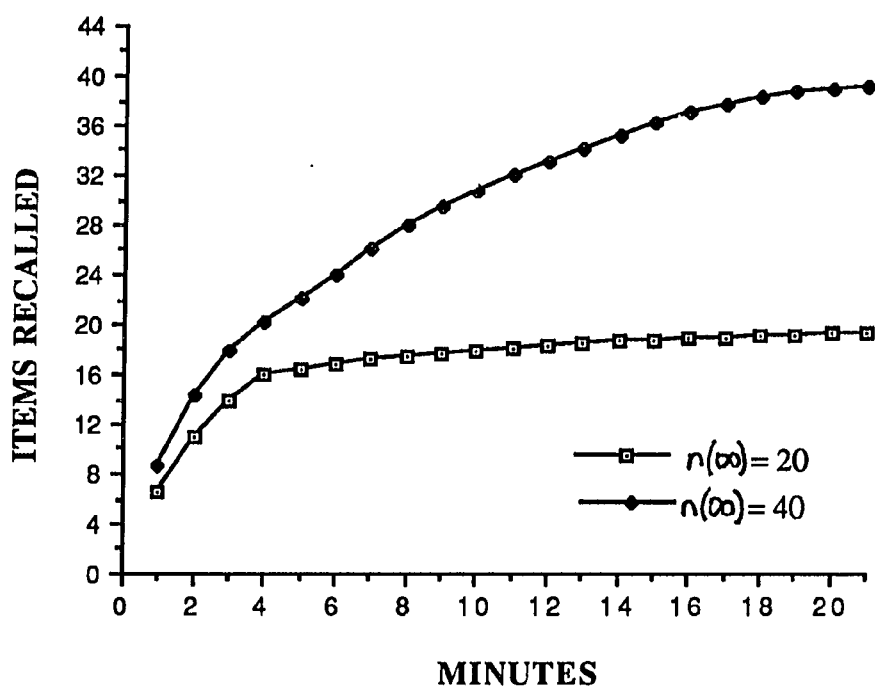
production of hypermnesia, and the absence of imaginal encoding manipulations was probably the important factor in previous failures to replicate Ballard, whose poetry likely aroused imagery in his subjects.

A study by Roediger and Thorpe (1978, Experiment 2) showed that semantic encoding instructions led to equivalent levels of hypermnesia as imaginal instructions. They also demonstrated significant hypermnesia for both pictures and words, although the effect was greater for pictures. Roediger, Payne, Gillespie, & Lean (1982) used nonsense syllables as stimulus materials which were presented either one or three times to subjects. When the syllables were presented three times, hypermnesia was observed across consecutive recall tests; this was not the case if the syllables were presented only once. Since it is difficult to explain these effects in terms of imaginal encoding, the authors concluded that there must be other critical determinants of hypermnesia. They suggest that a more reasonable hypothesis might state that any beneficial encoding strategy (imaginal encoding, semantic elaboration, etc.) is conducive to hypermnesia.

Roediger et al. (1982) proposed a recall level hypothesis to explain and predict hypermnesia. The hypothesis made use of the cumulative recall function (see Roediger, Stellan, & Tulving, 1977), which assumes that hypermnesia is predicted by the level of item recovery in situations in which only a single, long test is given. Roediger and Thorpe (1978), as well as Roediger et al. (1982) reported no differences in item recall when comparing multiple (three 7-minute) tests to a single, longer (21-minute) test. The cumulative recall function is described in the equation  $n(t) = n(\infty)(1 - e^{-\lambda t})$ , where  $n(t)$  is the number of items recalled by time  $t$ ,  $n(\infty)$  is the asymptotic estimate of the number of accessible items,  $e$  is the base of the natural logarithm, and  $\lambda$  is the rate of approaching the asymptotic level (Bousfield & Sedgewick, 1944).

Roediger et al. (1982) report that the asymptotic level of recall,  $n(\infty)$ , and the rate of approaching this level,  $\lambda$ , are negatively correlated. In other words, the higher the overall asymptote, the slower it will be reached. An illustrative example is given in Figure 1. In a multiple-test situation, those conditions leading to a higher level of overall recall should be farther from asymptotic performance after one test than should conditions leading to lower levels of recall. Therefore, we should expect more hypermnesia (improvement on later tests) for groups demonstrating a high overall level of recall, because after the first test, these conditions are the farthest from their asymptote. Differences in hypermnesia then, can be predicted directly from knowledge of asymptotic recall level. If the asymptote is achieved during the first recall test, then no net improvement can be expected.

Though the cumulative recall level hypothesis does account for many of the findings left unexplained by the imagery hypothesis, there are problems with this theory as well. Greater improvement is often shown between Tests 2 and 3, than Tests 1 and 2 (examples are in Erdelyi et al., 1977 and in Roediger et al., 1982, Experiments 1 and 2). This finding runs opposite to the recall level hypothesis, since subjects should be closer to asymptote after Test 2 than after Test 1 and subsequently show the least amount of improvement on Test 3. While this finding does not support the recall level hypothesis, neither is it a fatal blow because intertest forgetting is often different between Tests 1 and 2 than between Tests 2 and 3. If each recall trial can be considered as a study period for subsequent recall trials, then there should be less intertest forgetting on Test 3, as those words have been "studied" longer. Therefore, greater hypermnesia between Tests 2 and 3 does not necessarily reflect increased reminiscence. The recall level hypothesis does not take intertest forgetting into account, so knowledge of the asymptotic level of recall alone is insufficient to predict hypermnesia. It is also difficult for the recall level hypothesis to explain why hypermnesia is more readily obtained with pictorial than with verbal stimuli, if final recall level between

**FIGURE 1. HYPOTHETICAL CUMULATIVE RECALL CURVES.**

the two classes of material is experimentally equated (Payne, 1986). Less interest forgetting seems to occur for pictures than for words, a phenomenon not addressed by the hypothesis. Similarly, the imagery hypothesis proposed by Erdelyi and colleagues faces contradictory results. Hypermnnesia for nonsense syllables (Roediger et al., 1982, Experiment 2b) and abstract words (Hasher, Riebmman, and Wren, 1976, Experiment 2b) are not easily explained by the imaginal hypothesis. A possible explanation (proposed by Payne, 1986) is that both pictures and words show reminiscence, with pictures leading to lesser interest forgetting. This phenomenon has not been studied extensively, but it is consistent with many of the findings in the area.

#### *Reconstruction Over Repeated Tests*

Hypermnnesia is interesting because it is (at least seemingly) contrary to the classic Ebbinghaus forgetting curve. There have also been experiments using a repeated tests procedure which demonstrated the forgetting and distortion of memories over time. In 1932, Sir Frederic Bartlett published the classic book Remembering, in which he investigated (among other things) how people's recollections change across time. He employed the method of repeated reproduction, which involved presenting subjects information to be learned, and then giving each subject several recall tests over the materials at varying delay intervals. Bartlett was concerned that stimuli such as nonsense syllables often induced the strategy of rote recapitulation, and was interested in the processes by which a complex situation becomes accommodated into memory.

In his most famous experiment, he gave subjects a North American folk-tale, "The War of the Ghosts" (p. 65), and had them attempt to recall it immediately, and then again after varying delays. The folk tale (which was far too long to be reproduced verbatim) was from a culture different from that of the subjects and was not completely rational; it contained two



rather cryptic references to the supernatural. Bartlett in particular wanted to discover how these elements would be retained by the subjects.

Probably the most obvious result from the experiment is that the reproductions grew shorter with time. Each reproduction reported by Bartlett showed omissions of some details that had been present in the prior attempts at recall, and nothing resembling hypermnesia was reported. Recall in the experiment was attempted infrequently; the first reproduction came after an interval of fifteen minutes, and subsequent reproductions were at varying delays, which were always at least one day and usually two or more weeks. In contrast to recent hypermnesia research, there was never a single day in which more than one recall test was given.

Another change noticed throughout the experiment was that the stories became more rational with time. As subjects attempted to reconstruct the story, they altered the cryptic passages which were present in the original folk-tale. In addition, the entire tale became more cohesive and logical. Subjects unwittingly transformed the original story in their reproductions, filling in details to make their stories more connected. Bartlett concluded that in delayed recall, material is rationalized, and reconstructed by each subject's past experiences, which operate on the story and alter its reproduction (p. 94).

Though not nearly as well-known as his "War of the Ghosts" study, Bartlett (also in Remembering, 1932, Chapter 6) studied the repeated reproduction of picture-sign material. Picture-signs are relatively simple sketchings that were associated with a label or "sign". Some pictures had a direct and obvious relation to their corresponding labels ("eye" was paired with a drawing that looked like an eye), others could be tangentially associated ("string" was paired with a wavy line), yet others appeared entirely unrelated to their labels. After studying the associations, Bartlett gave each subject a piece of dictation. Their instructions were to wait until a previously-studied sign appeared in the dictation, and then

to sketch it as quickly as possible. A similar test was given two weeks later, and later still whenever possible.

Results were similar to those in the "War of the Ghosts" experiment. Subjects forgot more and more picture-signs over time, and those that were reported often became distorted. Those pictures that were tangentially related to their given labels gradually became more representative of the label, and less similar to the actual picture. Also, those items with no association to their name were the most frequently forgotten. Bartlett reports that all subjects showed some evidence of blending and confusion for several of the pictures, in which the salient features of two or more of the signs were combined into a single representation. And aside from those pictures that were very representative of their given labels, those pictorial symbols that were the simplest were the most likely to be retained over repeated trials.

None of Bartlett's findings in this experiment are counterintuitive, but they are consistent with his findings from the "War of the Ghosts" recall. Those details that were not consistent with subjects' previous ideas were the first to be omitted, and any detail not relevant to the central structure of the stimulus material quickly disappeared. Also, transformations and distortions which were evidenced always served to make the material more rational to the subject.

A comparison of the two differing traditions (Ballard's reminiscence and Bartlett's reconstruction paradigms) might provide some insight as to the ways our memories for episodes change over time. A closer examination of the two traditions reveals two differences, which may be responsible for the opposing results. One involves the type of material learned. All 172 recent hypermnnesia experiments reported by Payne (1987, pp. 18-22) in his review of the field utilized materials that were a series of words, pictures, or nonsense syllables. Although the orienting tasks varied widely, these experiments all involved the eventual recall of discrete units, such as words or pictures.

The Bartlett (1932) experiments, in contrast, required subjects to recall complex, interrelated information. The "War of the Ghosts" folk-tale could be considered a series as a series of connected (albeit sometimes loosely) passages, and the passages did not completely make sense to the subjects. It is possible that the very structure of the stimulus materials was a critical determinant of the types of errors that were made. Omissions and distortions occurred largely as subjects forgot about passages not centrally related to the "theme" or "schema" of the folk-tale, and modified those passages which did not make sense to them.

The other relevant experiment by Bartlett, the repeated reproduction of picture-sign material, produced declining performance over time, despite the fact that the stimulus items were not connected as in the "War of the Ghosts". Like recall of the story, picture-sign recall became both more distorted from the original and more coherent over time. A major difference between the Bartlett findings and modern hypermnesia research lies in the characteristics of the stimulus information. Bartlett's materials (both the folk-tale and the picture-signs) were complex stimuli which were capable of being distorted, or only partially recalled. A discrete unit, such as a single word in isolation, is unlikely to be recalled incompletely. It is difficult to conceive of a subject partially remembering a "shoe", for example, and instead reporting "hoe" or "shot". The individual pictures or words in a typical, recent hypermnesia experiment are not dependent upon one another the way the different passages in a story are dependent.

This example of the differences between Bartlett's materials and those used in recent hypermnesia experiments is surely not the only cause for the opposing conclusions drawn from these two traditions. However, the fact that Bartlett's materials are conducive to reconstruction while discrete stimuli (individual pictures or words) are not may be a factor that interacts with the time delay between memory tests. Over time, one could expect more reconstruction to occur, as subjects operate upon Bartlett's materials and alter them, making

them more coherent. This reconstruction is likely to make recall worse over time, while the discrete units employed as materials in recent hypermnesia research are far less susceptible to this deleterious effect.

A second, perhaps more obvious, difference between the two traditions involves the delays between the memory tests. A typical hypermnesia experiment from the Erdelyi or Roediger laboratories (e. g., Erdelyi & Becker, 1974; Roediger et al., 1982) involves a series of recall tests separated by only a few minutes or less. Bartlett's subjects always had at least one day between recall attempts and usually much longer. While the length of these delays was certainly a factor, it should be noted that early reminiscence research did demonstrate net improvement over repeated tests with long delay intervals. Spitzer (1939) found a net improvement in schoolchildren with a delay of over two weeks, Ballard (1913) and McGeoch (1931) both demonstrated improvement over a one-day interval, and Ammons and Irion (1954) report reminiscence (actually hypermnesia) after a delay of two, but not seven days. Bartlett's (1932) discovery of intertest forgetting and distortion may partially be the result of the lengthy delays between reproduction attempts. Even after delays of one day, however, he never reported a case of what would now be called hypermnesia for the materials.

Together, these two differences between hypermnesia research and the work of Bartlett (the use of different types of stimulus materials and the delay between tests) may represent the reason that the two traditions came to contrasting conclusions regarding effects of repeated testing. In order to disentangle the effects of these manipulations, the first experiment investigated these methodological differences within a single experiment. This experiment asked these three questions about repeated testing:

1. How does connecting the to-be-remembered stimuli within a story affect recall across repeated tests? To answer this question, all subjects viewed the same 60 pictures, but

in two different ways. Half simply saw the pictures in a series, but the others viewed the pictures in the context of a story. The story was presented from a cassette player and mentioned the name of the object in each picture as it was being viewed by the subjects. Several subsequent recall tests served to measure how memory for the pictures changed across time for the two presentation contexts (pictures plus names and pictures plus story).

2. Can both hypermnesia and intertest forgetting be shown in one experiment with the same subjects and the same stimulus materials? To answer the question, the time delays between memory tests were manipulated. Subjects in some conditions received three memory tests separated by a delay of only a minute, and then, after an interval of one week, took three more tests in succession. From past research, one might predict hypermnesia within each experimental session over the three tests and reliable intertest forgetting for the one-week interval. However, this outcome might interact with the type of stimulus material, as subjects viewing the pictures within the context of the story show increased intrusion errors over time.

3. How much do prior tests help successive tests? After presentation of the stimuli, subjects immediately took either zero, one, or three memory tests. Seven days later, all subjects returned to take another three tests. As demonstrated by Roediger and Payne (1982) and Spitzer (1939), it should be expected that the greater the number of prior recall tests, the better subjects' retention of the stimuli will be, with the interval between stimulus presentation and memory test held constant. The study will also look for hypermnesia across the three consecutive memory tests in the second experimental session (after the one-week delay).

4. What is the course of reminiscence and intertest forgetting across consecutive tests? The recall level hypothesis (Roediger et al., 1982) predicts greater item recovery between a first and a second test than a second and third (provided that asymptotic recall is not reached

on the first test). Many studies, however, have demonstrated greater improvement in net recall between Tests 2 and 3 than between Tests 1 and 2 (Erdelyi et al., 1977, Roediger et al., 1982, Experiments 1 and 2). A possible explanation is decreased intertest forgetting between Tests 2 and 3. Item analyses were performed on the picture recall of the three immediate tests to discover whether or not intertest forgetting did, in fact, decrease between Tests 2 and 3. These analyses also looked for an effect of the presentation condition. Klein, Kihlstrom, Loftus, and Aseron (1989) reported that encoding manipulations providing item-specific information will produce hypermnnesia by facilitating reminiscence, while manipulations producing relational, or organizational encoding lead to hypermnnesia largely because they depress intertest forgetting. Therefore, it was hypothesized that the addition of the story might influence subjects to organize the pictures together, while viewing the pictures accompanied by only their names could lead to item-specific encoding. If so, it would be expected that both manipulations would lead to reliable hypermnnesia, with the names promoting reminiscence, and the story depressing intertest forgetting.

### Experiment 1

#### Method

**Subjects.** Subjects were 120 Rice undergraduates 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. They represented pictures that could be easily named. In order, the names of the pictures were *football, dog, lightswitch, couch, house, clock, television, shirt, car, piano, bridge, doorknob, button, broom, refrigerator, grapes, eyeglasses, lion, belt, garbage can, lamp, dinosaur, rocking chair, hat, faucet, flower, pencil, saw, duck, ear, traffic light, pretzel, ambulance, whistle, sailboat, umbrella, kite, guitar, bicycle, picnic basket, banana, balloon,*

*turtle, newspaper, airplane, wristwatch, comb, fly, bed, leaf, sun, well, elevator, cannon, pipe, window, telescope, key, ironing board, and telephone.* When possible, the slides were taken from the Snodgrass and Vanderwart (1980) norms. Eight other slides were used (*ambulance, bridge, dinosaur, elevator, faucet, lion, pretzel, telescope*). These were constructed by tracing simple sketches of these objects onto transparencies to make the slides.

Half of the subjects listened to a children's story as they watched the slides. (The story is reproduced in Appendix A.) The story was written so that, as a slide was being presented, the object within the slide was being mentioned in the story. For example, subjects first saw the slide of a football for seven seconds. During this time, a football was part of the children's story. The story was also constructed so that it would take about 7 mins to be read from the cassette player. The other half of the subjects, not given the story, listened to the name of each picture from the cassette player as its corresponding picture was being viewed.

Design. Subjects viewed the picture series under one of two conditions. Half experienced the presentation while hearing only the name of each picture being read from a cassette player. (This was the P + N condition.) For the remaining subjects, the slides were accompanied by a children's story, which was read from the cassette player during presentation (the P + S condition). Subjects differed in the number of recall tests they received on the first day of testing, 0, 1, or 3. (All subjects later took three recall tests after a one week delay.)

Therefore, there were six between-subjects conditions, with two types of presentation (P + N and P + S) and three different possible schedules of recall tests. 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, to which they should pay close attention. They were informed only that they would "later be asked some questions" about the pictures. One-half of the subjects (in the P + N conditions) were told that they would also hear the name of each picture being read from a cassette player as it was viewed, so that no ambiguity would exist as to the identity of each object. The other subjects (in the P + S conditions) were told that, as they watched the slides, they would hear a children's story being read from a tape recorder. The experimenter explained that the object represented in each slide would also be an object appearing within the story when the slide was viewed. Subjects were given the example that, if they happened to be watching the slide of a mouse, then a mouse would be mentioned in the story as that particular slide was viewed. Subjects were not told how many slides to expect or how long the entire presentation would last.

After the instructions had been given to the subjects, the series of slides began. They were projected by a Kodak Ektagraphic III E projector at a rate of 7 secs each, with each slide immediately following the preceding one (with .75 sec between slides). There were 60 slides and the presentation lasted 7 mins, 3 secs. During the slide presentation, subjects in the P + N conditions heard the name of each slide spoken in a male voice from a cassette player. The other half of the subjects listened to the children's story (reproduced in Appendix A) as they watched the slides. The first picture appeared after the story had been playing for one second, so that the objects would appear in the story at the same time as they were viewed.

Immediately following presentation of the slides, all subjects spent 5 mins 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 three minutes to complete the questionnaire.

Although the questionnaire was given to all subjects in all conditions, its primary purpose was to justify the slide presentation to those subjects who would not be taking a memory test for the slides in the first experimental session. Without any justification, these subjects might have suspected a memory test during the second session, and rehearsed the slides during the one-week delay. Following the questionnaire, one-third of the subjects in both the pictures plus names and pictures plus story conditions were told that the first part of the experiment had been finished, and they were reminded to return for the next week's session.

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 sheets of paper numbered 1 to 60. Subjects were told that they would have 7 mins to recall the names of as many of the pictures as they could. If subjects were unable to remember the 60 objects that were in the slides, then they were to fill in the remaining spaces with their best guesses as to what the slides might have been. The experimenter told subjects that they should keep trying to think of new slides throughout the test, even though they would reach a point where it became difficult to do so. He also stressed that all 60 spaces on the test sheet must be filled in before the experiment could continue, and so subjects should pace themselves to complete the memory test by the end of the 7-min period. Pilot tests had indicated that subjects were often uneasy about following the forced-recall instructions, so the experimenter assured all subjects that they should not feel uncomfortable about making guesses of which they were not confident.

Subjects were also told that they would be given a cue after each minute of the recall session, at which time they should circle the last word they had written. The experimenter

said he was interested at the rate at which the words were being remembered. These test instructions took just over 2 mins to read, and all subjects who took at least one recall test in the first experimental session began their first memory test 11.5 mins after the end of the slide presentation.

Following these instructions, subjects began their forced-recall test. After each minute, the experimenter asked them to circle the last word that had been recalled. Also, at the end of the fifth and sixth minute, the experimenter gave subjects a time cue, reminding them how much time they had left to complete the task. An additional time cue was given with 20 secs remaining in the test. At the end of the 7 mins, subjects were asked to pass their test sheets back to the experimenter. If all 60 spaces had not yet been completed, they were instructed to fill in the remaining spaces as quickly as possible with guesses.

When all test sheets had been returned, those subjects assigned to the conditions which were to receive only one immediate test were told that the first experimental session was over and reminded when to return the following week. Remaining subjects were given new blank test sheets and informed that they would be taking an identical recall test under the same instructions. They were encouraged to keep trying to think of new slides, even though it would become difficult. The second test began one minute after the first was completed, and was procedurally identical to the first.

Upon completion of the second test, new test sheets were distributed and subjects were told that they would take an identical test again. They were also told that the third test would be the last task in the experimental session. Again the experimenter reminded them to keep thinking about the slides throughout the seven minutes although it would seem difficult to remember slides which had not previously been recalled. After a one-minute interval between tests, the third recall test was begun. After the test, subjects were reminded when to return for the second session of the experiment.

Seven days later, the subjects returned to the same room at the same time of day as the first session. Subjects in all conditions were treated identically in this session. They were told that they would be taking a memory test for the pictures that they had seen the previous week. All subjects were read the same set of test instructions that had been given the previous week, which explained the forced-recall procedure and the direction to circle the last word recalled after each minute. These instructions took just over 2 mins. Then subjects all took the forced-recall test, which was procedurally identical to the tests given in the first experimental session. They later took a second and third test, with a 60-sec interval separating the tests. Subjects were reminded before each test that they should keep thinking about the slides throughout the entire test and attempt to recall new slides, even though it would seem difficult. In addition, the experimenter told subjects before the third recall test that it was the last memory test they would have to take over the slides. After the final test, subjects were debriefed and thanked for their participation.

### Results

Several analyses were conducted with the recall data. Basic data are reported in Table 1. For both types of presentation (pictures plus names and pictures plus story), analyses-of-variance were conducted to compare the scores on the first recall tests of subjects in the conditions taking either one or three tests in the first session. Since these subjects had been treated identically through the first test, no differences were expected. The analyses confirmed this. For subjects that had been presented pictures plus names,  $F(1,38) = .14$ ,  $MSe = 51.61$ , and for those that received the pictures plus the story,  $F(1,38) = .17$ ,  $Mse = 47.65$ .

Analyses were conducted to determine whether or not hypermnesia had occurred for those subjects who had received three consecutive memory tests in the first experimental session, and whether or not the story aided recall of the pictures. Those subjects who heard

the story retained more information than those who did not,  $F(1, 38) = 7.62$ ,  $MSe = 208.85$ ,  $p < .01$ . There was also reliable hypermnnesia demonstrated when the data were collapsed across the presentation condition,  $F(2,76) = 17.20$ ,  $MSe = 4.32$ ,  $p < .001$ . The interaction between level of hypermnnesia and presentation context fell short of significance,  $F(2,76) = 2.66$ ,  $MSe = 4.32$ ,  $p < .08$ .

Item analyses were performed on the immediate tests taken by those subjects receiving three tests in the first session. Analyses compared both reminiscence and intertest forgetting as a function of both test number and presentation condition (P + N vs. P + S). These data are reported in Table 2. Looking at reminiscence, no significant differences were found due to presentation condition,  $F(1,38) = 1.82$ ,  $MSe = 5.39$ ,  $p > .18$ . There was also no effect of test number; as much reminiscence occurred between Tests 2 and 3 as between Tests 1 and 2,  $F(1,38) = 1.14$ ,  $MSe = 5.30$ . Neither did these two variables interact,  $F(1,38) = .08$ ,  $MSe = 5.30$ .

Intertest forgetting showed a similar lack of significant effects. Pictures were forgotten at a similar rate whether they had been accompanied by their names or the children's story,  $F(1,38) = .48$ ,  $MSe = 3.15$ . Likewise, there was no difference in intertest forgetting between Tests 1 and 2, as compared with Tests 2 and 3,  $F(1,38) = .88$ ,  $MSe = 4.11$ , and the two variables did not interact,  $F(1,38) = 1.90$ ,  $MSe = 4.11$ ,  $p > .17$ .

Analyses were also conducted to verify that intertest forgetting occurred over the one-week delay. Analyses-of-variance showed reliable forgetting within each of the four relevant conditions when the final test of the first experimental session was compared with the first test of the second session. Therefore, both hypermnnesia and intertest forgetting were demonstrated in the same experiment with the same materials through manipulation of the delay intervals between the tests. There was also a significant interaction between the

Table 1

Mean Number of Pictures Recalled as a Function of Presentation Context and Testing Schedule in Experiment 1

Pictures plus Names								
Group	Immediate Tests				Delayed Tests			
	T1	T2	T3	T3-T1	T1	T2	T3	T3-T1
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 plus Story								
Group	Immediate Tests				Delayed Tests			
	T1	T2	T3	T3-T1	T1	T2	T3	T3-T1
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. Conditions labeled 3-3 received three tests in the immediate session, 1-3 received one test in the immediate session, while 0-3 took no immediate memory tests.

\*These conditions demonstrated reliable hypermnesia across the three consecutive tests.

Table 2

Mean Reminiscence and Inter-test Forgetting Scores Across the Three Immediate Tests as a Function of Stimulus Materials and Test Number in Experiment 1

---

Group	Change T1-T2				Change T2-T3			
	T1	NC	CN	T2	T2	NC	CN	T3
P + N	26.6	4.2	3.6	27.2	27.2	3.8	2.6	28.4
P + S	32.7	5.0	2.7	35.0	35.0	4.3	2.9	36.4
Mean	29.7	4.6	3.2	31.1	31.1	4.1	2.8	32.4

---

Note. NC refers to the number of pictures correctly recalled on the later test, but not the earlier test. CN is the number of pictures forgotten between the two tests.

number of prior tests and the amount of forgetting,  $F(1,78) = 13.01$ ,  $MSe = 7.62$ ,  $p < .001$ . Subjects who had taken three immediate tests forgot less than those who had taken only one.

Data were also analyzed comparing the first test in the immediate recall session with the first test in the delayed session. If less forgetting is shown after the two additional recall tests are given in the first session, relative to only a single test, then these subsequent tests have enhanced subjects' retention of the pictures, despite the absence of additional study time. This hypothesis was supported. Reliable overall forgetting was obtained,  $F(1,76) = 107.22$ ,  $MSe = 6.23$ ,  $p < .001$ , but this finding is qualified by the forgetting X testing schedule interaction,  $F(1,76) = 55.37$ ,  $MSe = 6.23$ ,  $p < .001$ . Again, even with the first test used as the basis for comparison, subjects who took three tests in the first session showed significantly less forgetting than those taking only a single test.

The mean number of pictures correctly recalled after the one-week interval was analyzed using a 2 (presentation context) X 3 (testing schedule) X 3 (test order: 1, 2, or 3). Means are presented in Table 1. Analyses of this delayed recall indicate a significant main effect for the number of prior tests,  $F(2,114) = 33.79$ ,  $MSe = 124.02$ ,  $p < .001$ . The more immediate memory tests subjects had taken, the better was their recall on the delayed tests. A reliable main effect for presentation context (P + N vs. P + S) was also demonstrated,  $F(1,114) = 8.61$ ,  $MSe = 124.02$ ,  $p < .01$ . The addition of the children's story increased retention of the pictures. The interaction between these two variables was of borderline significance,  $F(2,114) = 2.60$ ,  $MSe = 124.02$ ,  $p < .08$ .

In a within-subjects analysis of the delayed test scores, there was an overall effect for the number of the test (first, second, or third), indicating that hypermnnesia occurred,  $p < .001$ ,  $MSe = 5.03$ . This hypermnnesia did not interact with the type of presentation, however,  $F(4, 228) < 1$ ,  $p > .7$ , nor with the schedule of the tests taken in the first session,

$F(4, 228) = 1.9, p > .10$ . The three-way interaction of these variables also was not significant,  $F(4,228) < 1$ .

Next, simple main effects were conducted upon each of the six between-subjects conditions, examining them for net improvement across the three delayed tests. For those conditions that employed the pictures plus story manipulation, no delayed hypermnesia was demonstrated for those subjects who had taken three prior memory tests,  $F(2,38) = 1.45$ ,  $MSe = 4.46$ . (Presence or absence of delayed hypermnesia is reported in Table 1 for all six conditions.) Subjects taking only one prior test, however, showed a significant overall improvement,  $F(2,38) = 7.71$ ,  $MSe = 4.92$ ,  $p < .01$ . Subjects with no prior tests demonstrated no hypermnesia,  $F(2,38) < 1$ ,  $MSe = 5.82$ .

For subjects who saw the pictures and heard the picture names, hypermnesia was discovered across the memory tests after the one-week interval for subjects who had completed three prior memory tests,  $F(2,38) = 4.45$ ,  $MSe = 2.97$ ,  $p < .025$ . Subjects given one test in the first session also demonstrated reliable improvement,  $F(2,38) = 5.21$ ,  $MSe = 5.46$ ,  $p < .025$ . Subjects who were taking their first recall tests for the slides after the one-week delay did not show any reliable hypermnesia,  $F(2,38) = 1.33$ .

### Discussion

Results replicated the basic findings of hypermnesia across tests separated by brief delays (e.g., Erdelyi & Becker, 1974; Roediger & Thorpe, 1978). Whether subjects were presented the children's story or the names along with the to-be-remembered pictures, net improvement occurred across the three recall tests in the first session. Also, reliable intertest forgetting was demonstrated in all four conditions in which tests were taken during both experimental sessions. The separation of two recall tests by a seven-day interval precluded any net improvement over the tests; whatever factors are responsible for hypermnesia are only evident when the delay between tests is less than one week. (This was true for both P



+ S and P + N conditions.) The interval between tests is a likely determinant of whether or not hypermnesia will be obtained. Over a one-day interval, net improvement has been found (Ballard, 1913; Ward, 1937, in Woodworth & Schlosberg, 1955, Scrivner & Safer, 1989). Hypermnesia has been reported (Erdelyi & Kleinbard, 1978) over a period of many days if subjects are tested multiple times within each day. On some occasions (though not all), there was net improvement between tests separated by several hours (Erdelyi & Kleinbard, 1978, Experiment 1), and on average, the results demonstrated hypermnesia in excess of 50% a week after the initial recall trial. Future manipulation of this delay could provide further insight into the phenomenon of hypermnesia.

Experiment 1 also confirms the finding that prior tests facilitate later ones. Overall level of recall on the delayed tests was a direct function of the number of prior recall tests for the pictures. Subjects who had taken three immediate tests recalled reliably more than subjects who had taken only one (a mean of 28.5 relative to 21.8), and these subjects retained more of the pictures than the subjects who took no immediate tests (17.1), replicating a number of experiments (Myers, 1914; Spitzer, 1939; Ammons & Irion, 1954; Glover, 1989). The amount of forgetting between experimental sessions also depended on the testing schedule manipulation, as forgetting was negatively correlated with the number of prior recall trials. The differences in forgetting over the week-long interval can be explained by retrieval practice and additional study afforded by the first test. The retrieval practice and additional study interpretation cannot, however, explain reminiscence. If a particular item was not recalled on a first test, then obviously subjects did not practice retrieval of that item or have an opportunity to restudy it.

Results demonstrated that connecting the pictures within a children's story produced greater recall than did presentation of the pictures with only their names. There are several hypotheses as to why this might be the case. The story might provide a "schema" within

which the pictures can be remembered. At test, subjects use their schema-related knowledge to retrieve the relevant pictures. Note that the schema would be especially helpful under conditions of forced-recall, as in this experiment. When subjects are required to guess additional items to complete the memory test, the schema provides a framework from which items can be guessed. This schema-consistent guessing was evident on many of the response sheets, as subjects who heard the story often guessed other items which had been mentioned in the story, even if those items were not actually presented as pictures.

While this explanation seems reasonable, other causes for the differences between the pictures plus names and the pictures plus story conditions cannot be ruled out. It is possible that, rather than provide a concrete schema, the story served only to enhance each item by providing more retrieval cues. In other words, each picture name was embedded within a sentence, which provided elaborations and associations for each name. These cues might enhance recall, even without an overall story schema. Each picture was contained not only within the story, but also within an individual sentence, which could provide the picture with some semantic associations. It could be this lower level of elaboration that enhanced the recall scores, rather than a more global schema.

It is also possible that the addition of the children's story made the initial picture presentation more interesting to the subjects. All subjects viewed the same 60 pictures in the same order for a total of seven minutes, yet subjects who did not hear the story may not have attended as carefully to the presentation. Without the story, subjects may have had less motivation to attend to the pictures (although they had, of course, been instructed by the experimenter to pay close attention), and with the story there was the incentive to attend carefully to the stimuli, as they were mentioned in the accompanying story.

One curious finding in this experiment is that, while the children's story raised the overall level of recall, this was not the case for subjects who completed memory tests only

after the one-week delay (see Table 1). An analysis of these two conditions revealed no effect for the story manipulation, and these conditions also demonstrated no reliable improvement across their three tests.

The lack of hypermnesia for these subjects is consistent with the cumulative-level of recall hypothesis (Roediger et al., 1982). If subjects reached their asymptotic level of recall during the first test, hypermnesia would not be possible during the remaining two tests. Analyses also demonstrated net improvement across the three delayed tests in three of the four conditions in which subjects had received a prior test. Subjects who viewed the pictures accompanied by the children's story and took three prior recall tests did not show hypermnesia during the second session. This was surprising, considering that three other conditions did show hypermnesia across the second set of recall trials. Future research will be required to determine if this finding can be replicated. Overall, these recall data are not entirely predicted by the recall level hypothesis. Subjects who received the story along with the pictures did not show more hypermnesia overall than subjects who were presented pictures only, despite the fact that their level of recall was significantly higher. Similarly, subjects receiving three tests should have shown more delayed test hypermnesia than those receiving one test, which did not occur. However, the imaginal coding hypothesis (Erdelyi & Becker, 1974) is insufficient to explain the results, as all conditions viewed the same pictures, yet hypermnesia was demonstrated only in some conditions. The conditions leading to hypermnesia on delayed tests must await future research.

A disappointing finding was that the item analyses led to no significant results of either presentation context or test number, probably because the P + S condition did not provide strong enough organizational cues for the pictures to be relationally encoded. Although they were joined by the story, many of the pictures were tangential to the overall "schema" or "theme" of the story. Unlike Klein et al. (1989) and Payne (1986), intertest forgetting did

not significantly decrease between Tests 2 and 3, although the overall results were in the expected direction. Future research will be required to determine the generality of this result.

### Experiment 2

The results of Experiment 1 clearly demonstrate both hypermnesia and interest forgetting within a single experiment, and indicate that the delay between tests is a critical determinant of the extent to which information is obtained across tests. At least two important questions, however, were not addressed by Experiment 1. First, the pictures plus story condition showed higher overall recall than pictures plus names, and at least two possible interpretations of this finding exist. One is that the pictures led to a richer, more elaborative encoding (through the addition of retrieval cues), thereby resulting in a greater number of the pictures being correctly retained. Another possibility is that the pictures in both presentation contexts were equally well “remembered” and differences between the two conditions were solely a result of the forced-recall procedure. The addition of the story might have simply allowed subjects to make better guesses. This would be the case if subjects in the pictures plus story conditions made guesses which were story, or schema, related.

Another issue not resolved by Experiment 1 is the extent to which subjects’ errors changed over the repeated tests. Of course, there were many errors made due to the forced-recall instructions. As Bartlett (1932) noted, in long-term remembering an invented or reconstructed detail might suddenly become part of one’s memory for an event. Some work by Loftus, Miller, and Burns (1978) has demonstrated that subjects exposed to misleading information often integrated this information into their memory for the event, suggesting people often cannot distinguish between accurate and inaccurate sources of memory. In Experiment 1, subjects may have created their own misleading information through

guessing. Since most incorrect responses were mere guesses, there was no measure of the extent to which subjects were actually confused.

A manipulation employed in this experiment should be helpful in resolving these and other issues. By varying the response criteria by which subjects attempt to recall the pictures, a better understanding can be attained as to why the story aided the recall of pictures, and also whether or not there are actually a greater number of reconstructive errors made over time. One-half of the memory tests were conducted under free recall instructions, under which subjects were warned against guessing. In other words, subjects reported only those pictures of which they were confident. The other tests used forced recall, identical to those in Experiment 1. If forced recall leads to significantly more correct responses than free recall, then subjects are simply making appropriate guesses, which are likely based upon their knowledge of the story. If the forced and free recall conditions show similar recall levels, then guessing was not a major factor in Experiment 1.

This forced vs. free recall manipulation has been utilized in past experiments to investigate response criteria and their effect on recall scores. The hypermnesia studies authored by Erdelyi (e.g. 1970; Erdelyi & Becker, 1974) adopted a forced-recall procedure to eliminate the potential problem of response bias. With a constant number of responses per test, any net improvement or decrement cannot be the result of a change in overall responsiveness; any observed hypermnesia must indicate that new, correct responses are taking the place of incorrect responses (Erdelyi & Kleinbard, 1978). Erdelyi's procedure assumed that, in many cases, criterion shifts occur naturally (i.e., more responses are produced on later tests), so forced recall should keep criteria identical for all subjects across all tests.

Their assumption has been challenged by a series of experiments (Roediger & Payne, 1985) in which the recall criteria (forced and free) had no effect on recall level, nor did it

produce a difference in hypermnnesia (but see Erdelyi, Finks, and Feigin-Pfau, 1989). A conclusion that may now be drawn from these studies that forced-recall instructions reliably improve recall only when the stimulus items can be easily guessed. Erdelyi et al. (1989, Experiments 2 and 2a, 4 and 4a) found a greater hit rate following forced-recall instructions, but the advantage was evident only to the extent that the stimulus items were guessable.

One method of determining what is actually “remembered” by subjects is to have them perform confidence ratings on each item reported. If there is truly an advantage after forced recall instructions, then subjects in these conditions should not only produce more correct responses, but be confident about a greater number of these responses. Some data of Roediger and Challis (1989) suggests that this is not the case. Though forced recall led to more stimulus items being correctly reported, this advantage did not remain after recall was conditionalized in terms of confidence. (Strangely, when the first memory test was delayed by one week, free recall was inferior to forced recall, even after conditionalization.)

The utility of forced recall instructions is questionable if the procedure produces performance no different than free recall, with the exception of guessing. It is possible, however, that requiring subjects to generate a large number of guesses might actually have a deleterious effect on later recall. This could occur if, on later tests, subjects confused their earlier responses with memory for the actual stimulus items. In other words, subjects might confidently report those incorrect guesses which they had previously produced.

This experiment examined these questions, by using all combinations of forced and free recall for both immediate and delayed tests. Also, two conditions received only a single delayed test, in an attempt to replicate the Roediger and Challis (1989) finding of a forced recall advantage for delayed tests.

### Method

**Subjects.** The subjects were 120 Rice University undergraduates enrolled in lower division psychology courses. They participated in exchange for partial course credit.

**Materials.** Subjects were presented with the same sixty slides used in Experiment 1, in an identical order. As the subjects viewed the pictures, they also heard the story from Experiment 1 (reproduced in Appendix A).

**Design.** Two between-subjects factors and one within-subjects factor comprised the experimental design. After receiving the stimulus presentation in the first experimental session, subjects received either two forced-recall tests, two free-recall tests, or no tests. After a seven-day delay, half of the subjects in each of the three immediate test conditions took a single forced-recall test; others received a single free-recall test. There were twenty subjects in each of the between-subjects conditions (designated as FR-FR, FR-FO, FO-FR, FO-FO, N-FR, and N-FO to identify their test type during the immediate and delayed sessions). The within-subjects variable was test number (first, second, or third).

**Procedure.** At the beginning of the experiment, all subjects received the slide presentation under conditions identical to the pictures plus story subjects in the first experiment. After the presentation (which lasted 7 min, 3 sec), subjects spent five minutes recalling names of U. S. presidents. Then the experimenter passed out a questionnaire asking subjects to provide information about the slides (the number of slides, length of the presentation, etc.), and to recall the instructions they heard before viewing the slides. Subjects receiving either free recall or no recall in the first session were given 3 min 45 sec to complete the questionnaire, while those taking forced recall tests were allowed 3 min. (Pilot testing had shown that the forced-recall instructions took 45 sec longer to read than free-recall instructions, so total time between study and test was equated across conditions.)

After completing the questionnaire, subjects in the N-FR and N-FO conditions were thanked for their attendance and reminded when to return the following week. Subjects in

the two initial forced-recall conditions (FO-FR and FO-FO) received instructions identical to those given all subjects in Experiment 1. Subjects in FR-FR and FR-FO conditions were told that they would be taking a memory test for the pictures that they had seen earlier, which would last for 7 min. The experimenter informed them that before the 7 min were over, they would reach a point where they would think that they had reported all of the pictures that they would be able to remember; however, they should keep trying to think of new pictures for the entire 7 min, even though it would become difficult. Subjects were also told to report pictures if they were reasonably confident of their response, and were warned against guessing wildly. All subjects took their recall tests, which lasted seven minutes. After the test, all the sheets were collected and then identical, blank sheets were passed out to each subject. The experimenter told subjects that they would now take another, identical memory test over the same pictures. Appropriate instructions were briefly repeated in summary to all subjects. After the second test, subjects were reminded about when to return the following week.

Exactly one week later, subjects returned to the identical room in which they had completed the first session. One-half of the subjects in each of the three first-session treatments (FR, FO, and N) were given a single forced-recall test for the pictures, whereas others received a single free-recall test. Complete recall instructions were read to all subjects before the test. Following the seven-minute test, subjects were asked to reread all of their responses and to rate their confidence that each reported item was actually a picture that had been viewed in the original slide presentation. Ratings were made on a six-point scale ranging from 6 (very confident that the picture had appeared in the slide presentation) to 1 (very unconfident that the picture had appeared in the presentation). Subjects were given as much time as they needed to complete the ratings, then they were debriefed and thanked for their participation.



## Results

Recall data are summarized in Table 3. Data from the two sessions will be conducted separately.

### *First Experimental Session.*

There were two pairs of two conditions each treated identically in the first session (FR-FR and FR-FO, as well as FO-FR and FO-FO). First, analyses were conducted to determine that groups treated identically did not differ from one another. In forced recall, a 2 X 2 ANOVA (condition X test number) revealed no overall differences in recall level,  $F(1, 38) = 1.93$ ,  $MSe = 59.76$ ,  $p > .17$ . These conditions did show overall hypermnesia,  $F(1, 38) = 5.24$ ,  $MSe = 3.82$ ,  $p < .05$ . Similarly for free recall, there was no reliable difference in overall level of recall,  $F(1, 38) = .07$ ,  $MSe = 101.24$ . Testing for hypermnesia, net improvement across tests did occur,  $F(1,38) = 34.20$ ,  $MSe = 3.80$ ,  $p < .001$ . For neither response criterion did condition interact with amount of hypermnesia. In forced recall,  $F(1,38) = 0$ ,  $MSe = 3.82$ , and for free recall,  $F(1,38) = .64$ ,  $MSe = 3.80$ .

Since no reliable differences were found between conditions treated identically, their scores were combined for subsequent analyses. Further analysis of data from the immediate recall tests were conducted to determine potential differences between the forced- and free-recall instructions on correct recall. A 2 X 2 ANOVA (recall criterion X test number) found no overall difference between the two recall conditions,  $F(1,78) = .36$ ,  $MSe = 80.00$ , confirming the results of Roediger & Payne, 1985). Overall hypermnesia was obtained,  $F(1,78) = 33.67$ ,  $MSe = 3.74$ ,  $p < .001$ , and the level of hypermnesia significantly interacted with recall criterion,  $F(1,78) = 6.42$ ,  $MSe = 3.74$ ,  $p < .05$ . Free recall subjects demonstrated greater improvement between tests 1 and 2 (2.6 compared to 1.0 items gained). This result is unlike that of Roediger & Payne (1985).

Table 3

Mean Pictures Correctly Recalled, Confidently Recalled, and False Alarms as a Function of Response Criterion and Test Number in Experiment 2

	T1	T2		T3	Conf.	FA
Immediate			Delayed			
Forced	30.8	31.8	Forced	26.1	21.8	4.1
			Free	24.9	23.0	2.0
Free	29.1	31.7	Forced	27.9	24.6	4.6
			Free	25.2	24.2	0.7
None			Forced	16.7	10.8	7.7
			Free	10.3	7.9	2.2

Conf. is the number of correct responses for which subjects were confident.

FA (false alarms) are the number of incorrect responses for which they were confident.

Item analyses were performed to determine how the level of reminiscence and interest forgetting between Tests 1 and 2 was affected by the recall criteria manipulation. There was an equivalent amount of reminiscence for the two conditions. Forced-recall subjects recovered 4.6 items on average, while free-recall subjects recovered 4.8 items,  $F(1,78) = .14$ ,  $MSe = 5.67$ . For interest forgetting, however, free recall subjects forgot significantly fewer pictures (2.2) than did forced-recall subjects (3.6),  $F(1,78) = 10.72$ ,  $MSe = 3.40$ ,  $p < .005$ . Therefore, the greater hypermnesia demonstrated in free-recall conditions was a function of decreased interest forgetting.

#### *Delayed Test Results*

Recall scores on Test 3 for the six different conditions were analyzed using a 3 X 2 ANOVA with test type during the first session (FR, FO, or N) and delayed test type (FR or FO) as independent variables. A main effect of the testing condition in the first session was demonstrated,  $F(2,114) = 55.32$ ,  $MSe = 37.92$ ,  $p < .001$ . A follow-up TUKEY test revealed that this effect was because of the depressed recall level of subjects taking no immediate recall tests (13.5 items). The forced- and free-recall manipulations on the immediate test led to similar performance after the delay (26.5 and 25.5 items, respectively).

There was also a significant main effect for test instructions on the final test,  $F(1,114) = 9.51$ ,  $MSe = 37.92$ ,  $p < .005$ . Forced recall led to a greater number of correct responses (23.55 to 20.08 for free recall). Interaction between the two variables, test conditions during the immediate and the delayed sessions, fell short of significance,  $F(2,114) = 1.90$ ,  $MSe = 37.92$ ,  $p > .15$ .

Despite this lack of significance, follow-up analyses were conducted to discover the presence of any simple main effects of the delayed test manipulation across each of the three levels of immediate recall (FR, FO, and N). For those subjects receiving no immediate tests, N-FO recalled significantly more pictures (16.7 correct) than N-FR (10.3 correct),

$F(1,38) = 19.71$ ,  $MSe = 21.10$ ,  $p < .001$ . If forced-recall instructions had originally been given, there was no effect of recall criterion on the delayed test, with FO-FO subjects recalling 26.1, and FO-FR, 24.9,  $F(1,38) = .38$ ,  $MSe = 40.75$ . The simple main effect was also short of significance when free recall instructions had been given in the first session, with FR-FO reporting 27.9 correct items and FR-FR reporting 25.2,  $F(1,38) = 1.40$ ,  $MSe = 51.92$ ,  $p > .24$ .

The same set of analyses were conducted on those correct responses which were confidently recalled (those rated as a 4, 5, or 6 by subjects) in 3 X 2 ANOVA (FR, FO, or N recall in the first session X FR or FO in the delayed session). Once again, there was a main effect of recall condition during the first session,  $F(2,114) = 68.62$ ,  $MSe = 38.97$ ,  $p < .001$ . A TUKEY test revealed that this difference occurred because subjects taking no immediate tests confidently reported fewer correct pictures (9.3) than those who had received free or forced tests (24.4 and 22.4, respectively).

There was no overall effect of the forced vs. free manipulation after the one-week delay,  $F(1,114) = .40$ ,  $MSe = 38.97$ , unlike the case for overall correct recall. Conditionalizing in terms of confidence (counting only those correct items which were rated as a 4, 5, or 6) wiped out the forced-recall advantage, replicating the pattern reported by Roediger and Challis (1989). Once again, the interaction of test condition during the first and second sessions was not significant,  $F(2,114) = 1.12$ ,  $MSe = 38.97$ .

More follow-up analyses were conducted to look for simple main effects of delayed recall criterion on the number of pictures confidently reported for each level of immediate test condition (FR, FO, and N). As with overall correct responses, there was no significant effect of the free vs. forced manipulations for those conditions which had taken two immediate tests of either kind. For subjects receiving only the delayed test, however,

forced-recall instructions led to more correct, confident responses (10.80) than did free recall (7.85),  $F(1,38) = 4.59$ ,  $MSe = 18.94$ ,  $p < .05$ .

### Discussion

Hypermnesia across the two immediate tests was demonstrated, replicating much of the contemporary literature in the area (e. g., Erdelyi & Becker, 1974; Roediger & Thorpe, 1978). Somewhat surprisingly, free recall led to reliably greater hypermnesia than did forced recall. The reasons for this are unclear; Roediger and Payne (1985) reported no difference in hypermnesia for free and forced recall over three tests. There was no significant effect of recall criterion for overall recall level, however, which is consistent with Roediger and Payne's (1985) findings, and not consistent with the claims of Erdelyi et al (1989, Experiments 3 and 4). The lack of a forced-recall advantage here likely means one of two things. One hypothesis is that the target stimuli are not easily guessed. If so, then the additional responses generated by subjects would be of no advantage (see Erdelyi et al., 1989). However, given the story schema and the pictorial materials, another explanation is that forced-recall subjects made less of an effort to retrieve the items because of a negative reaction to the instructions. Erdelyi et al (1989) suggest that there is a "processing bias" which forced-recall procedures may not successfully control. This bias is sensitive to motivational factors which accompany different response criteria, so forced-recall subjects may be less motivated to give their maximal performance (to what is a more difficult task). This explanation would conclude that forced and free recall led to a similar number of correct responses because forced recall benefitted from a response bias (a lower criterion for responding), yet suffered from a processing bias (they were less motivated to achieve maximal performance). This explanation cannot be refuted by Experiment 2, although it certainly suffers from a lack of parsimony (two conflicting, unseen processes are being hypothesized to explain a null effect).

Like the Roediger and Challis (1989) results, forced-recall instructions led to more correct responses after the week delay. However, when responses are conditionalized in terms of confidence, the difference between groups vanishes for those subjects which had taken the two immediate tests. On the other hand, N-FO outperformed N-FR. While these data are not easily explainable, they do replicate the Roediger and Challis (1989) finding.

With the exception of these two delayed recall results, forced recall had no advantage over free recall at any level of analysis in this experiment. This finding (or lack of a finding) would indicate that the forced-recall procedure may be an unnecessary one in many cases.

### Experiment 3

An original goal of this series of experiments was to examine repeated testing across different types of stimulus materials. All contemporary hypermnesia research has used discrete pictures or words as to-be-remembered items, whereas Bartlett (1932) gave his subjects longer, more complicated materials. Neither of the prior two experiments have addressed the question of whether or not this difference (prose vs. discrete items) might be a factor for the disparate results (forgetting over time vs. net improvement). While Experiment 1 examined recall of pictures in the context of a story versus pictures with only their names, the actual to-be-remembered items have been the same in all conditions of both experiments so far. This experiment was designed to investigate memory for prose passages, much as Bartlett did in Remembering. The story selected was not “The War of the Ghosts”, but a passage from a John Updike short story called “The Kid’s Whistling”. The passage was edited to be of similar length to “The War of the Ghosts” but it does not contain many of the more confusing elements of the North-American folk-tale. Bartlett selected “The War of the Ghosts” because it was of a social environment different from those of the subjects and contained events with no obvious connection. Also, he was

curious as to the ways that subjects would treat the supernatural elements of the story (Bartlett, 1932, pp. 64-65).

While these features led to much confusion and reconstruction, it is not an accurate sample of the type of prose to which people are commonly exposed. Cofer, Chmielewski, and Brockway (1976) reported that when passages are less unusual than “The War of the Ghosts”, then transformations of content should be less frequent and bizarre. “The Kid’s Whistling” (reproduced in Appendix B), in contrast, is a more typical passage, so results from this experiment should provide better evidence for the way memories for prose change over repeated tests.

A manipulation similar to that used in Experiment 2 was employed. Story recall was attempted under either a strict or relaxed criterion. Subjects in the Relaxed criterion condition were simply asked to report the story as best they could, while subjects in the Strict condition were told only to report a fact if they could remember it being in the story. These subjects were warned not to invent new material simply for the purpose of making a coherent story. A similar manipulation by Gauld and Stephenson (1967) found that subjects significantly reduced their errors (in recalling “The War of the Ghosts”) if given strict recall instructions. The article does not report how these instructions affected the level of material reported correctly. It is still useful, however, for its evidence that many of the reconstructive errors commonly found in memory are not true memory failures or distortions, but simply material that has been introduced for the purpose of telling a more rational or coherent story.

### Method

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

Materials. A passage excerpted from the John Updike short story “The Kid’s Whistling” was used as study material. The passage was broken down into forty-one idea units. (The story, separated into its idea units, is presented in Appendix B.)

Design. Subjects were randomly assigned to a strict or a relaxed recall condition. Each subject also took three separate memory tests, providing for a mixed design (condition X 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. They were to keep this side down until all instructions had been read. Subjects were instructed to read the passage through twice at their normal reading rate. The experimenter warned them that they should not spend an excessive amount of time reading the passage, even if doing so might slightly improve their memory for the story. All subjects then read the passage twice at their normal reading rate. This took subjects anywhere between one and one-half minutes to two and three-quarters minutes.

When subjects finished the passage, the experimenter collected the pages. Subjects in the Relaxed recall condition then spent five and one-half minutes recalling U. S. presidents (the Strict condition received only five minutes for this task, to equate overall retention interval). Next, the experimenter passed out blank sheets of paper for story recall. In the Relaxed condition, subjects were instructed to write down the story as best 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 the Strict condition, subjects heard all of the above instructions, plus:

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 nine minutes for story recall. The experimenter gave them a time cue when there was one minute remaining. Following the recall trial, subjects spent three minutes recalling the names of U. S. states. Then the experimenter passed out more blank pages for a second story recall. Relevant instructions were briefly summarized for all subjects, and then they had another nine minutes to report the story as best they could, under the identical response criterion as in Test 1. When the second recall trial was over, subjects were reminded when to return the next week and then dismissed.

After a seven-day delay, subjects returned to the same room and were given one additional memory test for the story. All subjects were tested using the strict response criterion. After the nine-minute 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. Subjects were reminded that sometimes people invent new facts or events just because those events make sense to them. Subjects were asked to reread their stories and, 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

Protocols were scored for three categories of responses. Subjects were given credit for correct idea-units if their report included a phrase which was judged to be identical with, or very close to, one of the 41, pre-established idea-units (which are reproduced in Appendix

B). Before the experiment, exact criteria were established for each individual idea-unit, identifying what information would be required for acceptance. As a general rule, verbatim reproduction was not necessary. If a subject correctly reported the information in an idea-unit, then credit was allowed.

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. 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. Another possible minor error might be reporting an incorrect name for one of the tunes whistled by Jack in the story. The information is not correct, yet it is not inconsistent with the events in the story and does not change the passage's meaning. It was also considered to be one minor error if a subject reported an incorrect name for a character, even if it happened consistently throughout a protocol. By dividing errors into these two types, an understanding might be reached concerning exactly what kinds of reconstructive errors people make across repeated tests. In Appendix C, a representative protocol is reproduced and scored according to its correct idea-units and major and minor errors.

Correct idea-units, plus major and minor errors, were originally analyzed as a function of both recall criterion and test number. After completing the full set of analyses (relevant data are summarized in Tables 4 and 5), it was discovered that recall criterion had no significant effect or interaction on any measure. Data were collapsed across recall criterion for subsequent analyses.

First, the scores were analyzed for hypermnnesia across the first two tests. Significant net improvement was found,  $F(1,37) = 15.27$ ,  $MSe = 1.31$ ,  $p < .001$ .

Table 4

Mean Idea Units Recalled by Condition and Test in Experiment 3

	Immediate		Delayed	
	T1	T2	T3	Conf.
Condition				
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. Conf. refers to those idea units of which subjects were confident.

Table 5

Mean Errors Reported as a Function of Recall Criterion and Test in Experiment 3

		Immediate		Delayed	
		T1	T2	T3	Conf.
Group	Error Type				
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. Conf. refers to those errors about which subjects expressed confidence.

Table 6

Mean Reminiscence and Intertest Forgetting for Correct Idea-Units as a Function of Response Criterion and Test in Experiment 3

Group	Change T1-T2				Change T2-T3			
	T1	NC	CN	T2	T2	NC	CN	T3
Strict	12.1	2.3	1.2	13.2	13.2	1.2	3.7	10.7
Relaxed	13.4	2.1	1.1	14.4	14.4	0.5	4.1	10.8
Mean	12.8	2.2	1.2	13.8	13.8	0.9	3.9	10.8

Note. Rem. refers to reminiscence, or the number of idea-units correctly recalled on the later test, but not the earlier test. IF (or intertest forgetting) is the number of idea-units forgotten between the two tests.

Although correct idea-units increased on Test 2, this was not the case for major errors. There were .66 major errors committed on Test 1 and .71 during Test 2,  $F(1,37) = .18$ ,  $MSe = .30$ . Minor errors, however, did increase between the two tests,  $F(1,37) = 14.51$ ,  $MSe = .26$   $p < .001$ .

Also analyzed was the trend across the one-week delay. A comparison of Test 2 with Test 3 showed that forgetting did occur over this interval,  $F(1,37) = 65.23$ ,  $MSe = 2.62$ ,  $p < .001$ . Neither the major nor minor errors, however, showed any change over the delay. For major errors,  $F(1,37) = .28$ ,  $MSe = .19$ , and for minor errors,  $F(1,37) = .95$ ,  $MSe = .50$ .

Item analyses were conducted with the data across all three tests to examine intertest recovery (reminiscence) and forgetting. Results are presented in Table 6. Again, response criterion did not significantly affect results, nor did it interact with test (the difference between Tests 1 and 2 and that difference between Tests 2 and 3). Further item analyses were collapsed across this manipulation. As might be expected, test number had a significant effect on both reminiscence and intertest forgetting. More words were recovered between Tests 1 and 2 than between 2 and 3,  $F(1,74) = 19.03$ ,  $MSe = 1.73$ ,  $p < .001$ . Also, more forgetting was demonstrated between Tests 2 and 3,  $F(1,74) = 58.93$ ,  $MSe = 2.37$ ,  $p < .001$ .

After Test 3, subjects were given an opportunity to indicate the information that they had reported about which they were not certain. Results indicate that subjects identified 33% of their major errors and 47% of minor errors. Only 5% of the correct idea-units were marked as questionable by subjects.

### Dicussion

Results from this experiment suggest that net improvement over tests occurs for prose recall as well as for discrete words or pictures if the interval between tests is short.

Forgetting between Tests 2 and 3 was also demonstrated, confirming Bartlett's findings. Therefore, Bartlett's discovery of decreasing performance over repeated tests can be considered a function of the delay between the tests rather than of stimulus materials. There were not many major reconstructive errors made in the experiment (only about .70 per subject per reproduction), which is also inconsistent with "The War of the Ghosts" recall. The nature of the two passages is the likely cause of this disparity. While "The War of the Ghosts" contained many passages that were difficult for the reader to comprehend (because of lack of connection between events, cultural differences, references to the supernatural, etc.), "The Kid's Whistling" is a passage which is both coherent and comprehensible to college undergraduates, and there is no reason for them to wildly reconstruct and commit the type of major factual errors found in Remembering. Both types of passages reveal important qualities about human memory, materials of the sort that people might read everyday (like the "The Kid's Whistling") may have more applicability to memory as it (memory) is commonly used.

Also unlike the Bartlett results, reconstructive errors did not increase over the week delay. This is probably because the story had no striking or bizarre events around which subjects could expound and reconstruct. It is also possible that subjects adopted a stricter criterion for this test. Test 3 was given under strict instructions, while the half of the subjects took Tests 1 and 2 after hearing relaxed instructions.

Type of recall criterion, somewhat surprisingly, had no significant effect throughout the experiment, despite the fact that the results came out in the expected direction. (Subjects in the Relaxed condition reported both more correct idea units and more major and minor errors than did subjects in the Strict condition on the immediate recall tests.) There are at least two possible reasons for this lack of an effect. One is that the manipulation was not strong enough and therefore, simply did not work. Although it certainly did not work in the sense

that no significant differences were obtained, this is probably not a complete explanation. Gauld and Stephenson (1967) used these identical instructions and obtained a reliable difference in errors on “The War of the Ghosts” recall. Their study measured words recalled, however, and many of the “mistakes” recorded by Gauld and Stephenson would not have been considered errors in this experiment. As demonstrated in Appendix C, it was possible for subjects to report sentences that were scored as neither correct idea-units nor errors. This occurred if facts were reported that were true, yet not explicit enough to be scored as a correct idea-unit. Also, it is possible that the strict vs. relaxed manipulation would have a greater influence on “The War of the Ghosts” than “The Kid’s Whistling”, because of the nature of the two passages.

Another potential reason for the lack of differences between the two recall criteria conditions is that subjects may be self-imposing a reasonably strict criterion upon themselves, even while under relaxed instructions. There is no *a priori* reason for subjects to willingly reconstruct. Although Bartlett claimed that his subjects wanted to report “The War of the Ghosts” in ways which were more familiar to them, subjects (even those in the relaxed condition) may not have felt this motivation given “The Kid’s Whistling”. Therefore, relaxed criterion instructions may not have led to a recall criterion which was truly “relaxed”. These relaxed instructions are not comparable to forced-recall instructions (like those in Experiments 1 and 2), which forced subjects to guess wildly and confabulate. Had there been a similar, guessing condition in Experiment 3, there would have surely been a great increase in the error rate. It is not obvious from this previous experiments, however, that correct idea-units would have significantly increased.

Subjects' abilities to monitor their own reports and decide which information was not accurate were impressive. They indicated that they were not certain about 33% of their major errors and 47% of their minor ones, while questioning only 5% of their correct idea



units. Not only were there few reconstructive errors, but subjects seemed to know which information was accurate and which was reconstructed on the delayed test.

### General Discussion

There are three important new results to discuss from this research:

1. **Hypermnesia vs. forgetting.** A tradition of hypermnesia and reminiscence research dating back at least to Ballard (1913) has documented improvement in memory over repeated tests. The method of repeated reproduction, utilized by Bartlett (1932), however, showed forgetting and distortion across consecutive tests. Two major differences were proposed to explain the opposing results. One was the delay interval between consecutive tests. Hypermnesia research generally involves short delays (of only a few minutes) between recall periods (but see Erdelyi & Kleinbard, 1978), while Bartlett's subjects attempted recall after much longer intervals, with usually at least two weeks separating the first and second test. The other major difference was the nature of the to-be-remembered material. Of the 172 hypermnesia experiments documented by Payne (1987, pp. 18-22) used lists of words or pictures as stimuli. Bartlett, of course, had subjects recall "The War of the Ghosts". Both of these differences (delay interval and stimulus material) were tested within this set of experiments and results show that length of the delay between tests is the critical variable in the determination of hypermnesia or forgetting across the tests. In all three experiments, significant net improvement was demonstrated across consecutive, identical recall tests when these tests occurred soon after the stimulus materials had been presented, while forgetting occurred over a week delay. In Experiment 3, a prose passage was the to-be-remembered stimulus yet, unlike in Remembering, hypermnesia was a function of the delay interval between recall tests.

2. **Recall criteria.** There was little or no effect of recall criteria on correct recall scores on immediate tests. In Experiment 2, subjects receiving two immediate tests under free

recall instructions reported as many correct pictures as subjects taking forced-recall tests, replicating Roediger and Payne (1985). A surprising finding was that free recall instructions led to significantly greater hypermnesia between the first two tests than did forced recall instructions, which is inconsistent with Roediger and Payne's (1985) findings. Future research will be required to determine whether this is a replicable finding.

Another finding of Experiment 2 showed that, in those conditions in which no immediate tests were given (N-FO and N-FR), there was an effect of recall criterion. Forced recall led to significantly more correct pictures being reported. This was not the case for those subjects that had received recall tests of either type during the first experimental session. While this phenomenon is not easily explained, it does replicate Roediger and Challis (1989).

It is important to note that these findings (or lack of findings) do not refute the original reasoning behind the use of forced recall. The procedure was introduced to control for the possibility of a laxer production criteria in later trials. That is, if this happened, then increased recall may not indicate enhanced "memory" rather than the actual relaxation of criterion (Klatzky & Erdelyi, 1985). Forced-recall instructions require subjects to adopt an identical strategy for each successive test, therefore eliminating this potential bias.

Erdelyi and his colleagues (e.g., Erdelyi, 1984) did not claim that the forced-recall procedure is a better indicator of memory, but rather that it allows for a more objective method of assessing hypermnesia, than does free recall. This point gains credibility with evidence that people actually do occasionally relax their criteria over repeated tests (Erdelyi, 1970). However, some results of Experiment 2, as well as those of Roediger and Challis (1989), question this technique. If the first memory test is taken after a delay of one week, forced recall leads to both a higher number of correct pictures recalled with confidence and a higher number of false alarms (incorrect pictures recalled with confidence). This would

indicate that it is something about the forced-recall procedure itself that made subjects more confident about what they have reported. While it is difficult to explain such a phenomenon, it has now happened in two experiments, using two different sets of stimulus materials. This would suggest that forced recall is not only measuring, but also affecting the extent to which subjects are confident about their memories. The fact that this phenomenon does not occur on immediate tests further attests to the complexity of this issue.

3. The extension of repeating testing to prose recall. Of course, Bartlett (1932) originated this very method, and King and Cofer (1960) have also employed repeated reproduction using "The War of the Ghosts". Neither of these investigations, however, involved consecutive, repeated tests closely following presentation of the story. Perhaps as a consequence of this, neither found a net improvement across tests. Also, King and Cofer (1960) reported the number of words correctly recalled, while Bartlett simply looked for general patterns in the reproductions without performing any quantitative analyses. Through the use of idea units, a different and probably more reliable measure of recall can be obtained. Subjects' protocols showed that, while recall was often accurate, verbatim reproduction of even one entire sentence was very much the exception. Therefore, the most appropriate level of analysis is one which emphasizes ideas, rather than individual words. A method which scored by correct words recalled, even if it accepted reasonable synonyms, would often miss the essence of what subjects remembered.

Experiment 3 was useful in that it allowed prose recall to be directly compared to research involving the repeated testing of pictures or words. Results show that the two types of materials are similar: both lead to either hypermnesia or intertest forgetting depending upon the delay interval. Within the story recall literature, Experiment 3 provided evidence that reconstructive errors may not be as frequent as they had once been considered. Although subjects did commit errors, more than half were of the minor variety, and did not

affect overall events or facts in the story. There was also an impressive level of memory-monitoring, as subjects recognized many of their reconstructive additions, while not questioning that correct information that they had reported.

While the three experiments constituting this study address some reasonably diverse issues, they are all anchored around the notion of repeated testing, and the ways in which our memories change as a function of the number of tests, time delays, response criteria, and stimulus materials. Of these variables, the first two were shown to be important. In Experiment 1, the level of improvement (or decrement) demonstrated across tests was a direct result of the length of the delay interval elapsed since the most recent prior test. Also, memory was enhanced as a function of the number of prior tests, with delay between stimulus presentation held constant. Delayed picture recall in Experiments 1 and 2 was positively correlated with the number of prior tests taken the week before. The manipulation of stimulus materials in this study led to few, if any, interesting effects or interactions. In Experiment 1, connecting the pictures with a children's story led to significantly higher recall than did the presentation of the pictures accompanied only by their names, but this manipulation did not lead to differing levels of reminiscence or intertest forgetting, as might have been expected. Also, repeated testing was extended to prose recall and, based on Experiment 3, the nature of the stimulus materials does not appear to be critical to the discovery of hypermnesia.

The effects of recall criteria on repeated testing were also minimal. Although greater hypermnesia was discovered following free- than forced-recall instructions, this effect was not expected and it is questionable whether or not future research pursuits will verify this phenomenon. While there were some differences found as a result of criteria manipulation

in Experiment 2, they only occurred in conditions in which no immediate tests had been given, and were therefore not a result of repeated testing.

Since much of the remembering we do in life involves repeated retrievals of the same information, investigations of repeated testing may lead to future insights which can be extended, or applied, outside of psychology laboratories. Through manipulation of the variables described above (and others), a understanding of the ways our memories operate can be attained.

### References

- Ammons, H., & Irion, A. I. (1954). A note on the Ballard reminiscence phenomenon. Journal of Experimental Psychology, 48, 184-186.
- 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.
- Bousfield, W. A., & Sedgewick, C. H. W. (1944). An analysis of sequences in restricted associative responses. Journal of General Psychology, 30, 149-165.
- Brown, W. (1923). To what extent is memory measured by a single recall trial? Journal of Experimental Psychology, 6, 377-382.
- 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. (1970). Recovery of unavailable perceptual input. Cognitive Psychology, 1(2), 99-113.
- 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., Buschke, H., & Finkelstein, S. (1977). Hypermnesia for Socratic stimuli: The growth of recall for an internally generated memory list abstracted from a series of riddles. Memory & Cognition, 5, 283-286.
- Erdelyi, M. H., Finkelstein, S., Herrell, N., Miller, B., & Thomas, J. (1976). Coding modality vs. input modality in hypermnesia: Is a rose a rose a rose? Cognition, 4, 311-319.

- Erdelyi, M. H., Finks, J., & Feigin-Pfau, M. B. (1989). The effect of response bias on recall performance, with some observations on processing bias. Journal of Experimental Psychology: General, 118(3), 245-254.
- Erdelyi, M. (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. New York: Academic Press, pp. 95-127.
- Gauld, A., & Stephenson, G. M. (1967). Some experiments relating to Bartlett's theory of remembering. British Journal of Psychology, 58, 39-49.
- Glover, J. A. (1989). The testing phenomenon: Not gone but nearly forgotten. Journal of Educational Psychology, 81(3), 392-399.
- Hasher, L., Riebmman, B., & Wren, F. (1976). Imagery and retention in free-recall learning. Journal of Experimental Psychology: Human Learning & Memory, 2, 172-181.
- King, D. J., & Cofer, C. N. (1960). Exploratory studies of stories varying the adjective-verb quotient. Journal of General Psychology, 62, 199-221.
- Klatzky, R. L., & Erdelyi, M. H. (1985). The response criterion problem in tests of hypnosis and memory. International Journal of Clinical and Experimental Hypnosis, 33, 246-257.
- Klein, S. B., Kihlstrom, J. F., Loftus, J., & Aseron, R. (1989). Effects of item-specific and relational information on hypermnesia recall. Journal of Experimental Psychology: Learning, Memory, and Cognition, 15(6), 1192-1197.
- Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. Journal of Experimental Psychology: Human Learning and Memory, 4(1), 19-31.

- McGeoch, G. O. (1931). The intelligence quotient as a factor in the whole-part problem. Journal of Experimental Psychology, 14, 333-358.
- Myers, G. C. (1914). Recall in relation to retention. Journal of Educational Psychology, 5, 119-130.
- Payne, D. G. (1986). Hypermnnesia for pictures and words: Testing the recall level hypothesis. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 16-29.
- Payne, D. G. (1987). Hypermnnesia and reminiscence in recall: A historical and empirical review. Psychological Bulletin, 101, 5-27.
- Roediger III, H. L., & Challis, B. H. (1989). Hypermnnesia: Improvements in recall with repeated testing, in C. Izawa (Ed.), Current Issues in Cognitive Processes: The Tulane Floweree Symposium on Cognition. Hillsdale, N. J.: Lawrence Erlbaum Associates, pp. 175-199.
- Roediger III, H. L., & Payne, D. G. (1982). Hypermnnesia: The role of repeated testing. Journal of Experimental Psychology: Learning, Memory, and Cognition, 8, 66-72.
- Roediger III, H. L., & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnnesia: A puzzle for generate/recognize theories. Memory & Cognition, 13, 1-7.
- Roediger III, H. L., Payne, D. G., Gillespie, G. L., & Lean, D. S. (1982). Hypermnnesia as determined by level of recall. Journal of Verbal Learning and Verbal Behavior, 21, 635-655.
- Roediger III, H. L., Srinivas, K., & Waddill, P. (1989). How much does guessing influence recall? Comment on Erdelyi, Finks, and Feigin-Pfau. Journal of Experimental Psychology: General, 118(3), 255-257.



- Roediger III, H. L., Stellan, C. C., & Tulving, E. (1977). Inhibition from part-list cues and rate of recall. Journal of Experimental Psychology: Human Learning and Memory, 3, 174-188.
- Roediger III, H. L., & Thorpe, L. A. (1978). The role of recall time in producing hypermnesia. Memory & Cognition, 6, 296-305.
- Scrivner, E., & Safer, M. A. (1988). Eyewitnesses show hypermnesia for details about a violent event. Journal of Applied Psychology, 73(3), 371-377.
- Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Human Learning and Memory, 6, 174-215.
- Spitzer, H. F. (1939). Studies in retention. Journal of Educational Psychology, 30, 641-656.
- Williams, O. (1926). A study of the phenomenon of reminiscence. Journal of Experimental Psychology, 9, 368-387.
- Woodworth, R. S., & Schlosberg, H. (1954). Experimental Psychology. New York: Henry Holt and Company.

## Appendix A

### Story Accompanying the Pictures in Experiments 1 and 2

Sam was in the middle of his summer vacation from school. He had planned to spend this day throwing the **football** with his friend James, but James had just called to tell Sam that he was sick. Disappointed, Sam went into the living room to play with his **dog**. A few minutes later, he watched as his mother came out of her bedroom and into the living room, flicking on the **lightswitch** as she entered. Sam suddenly had a feeling he was going to be given housework to do, so he laid back on the **couch**, pretending to be asleep.

"I have an idea. Why don't you visit your Uncle George," said Sam's mother. "Living by himself in that **big house** must make him depressed, and I'm sure he'd appreciate the company." Sam looked up at the **clock** and let out a sigh.

"I don't feel like it," he said. "I never seem to have a good time there. Besides, I had planned on watching a baseball game on **television**."

"But your uncle has so many wonderful drawings," said Sam's mother. "And you don't really have anything to do today. So change your **shirt** and we'll leave in a half hour."

So Sam began getting ready to spend the day with his Uncle George. George had a nice house, but his **car** was old, and he never kept the inside of the house straight. He spent almost all his time either drawing things around the house, or playing his **piano**.

Uncle George lived in the middle of the city, yet rarely left home. On the way to his uncle's house, Sam and his mother drove across the **bridge** leading into the city. Sam looked down at the water and into the city below and wondered if his uncle had at least fixed the broken **doorknob** on his front door.

Uncle George greeted Sam at the door. He was wearing a faded, crumpled shirt that had a **button** missing. Then Sam's uncle led him into the kitchen and said that it was time for lunch. Looking around the house, Sam noticed a **broom** in the corner, and was relieved that his uncle was keeping the place clean. Uncle George took some milk from the **refrigerator** and poured two glasses. On his uncle's dining room table were two small dishes, and on each was some spaghetti, green beans, and some **grapes**.

Sam and his uncle talked some about the summer, but mostly they were quiet. Uncle George seemed to be having trouble seeing through his **eyeglasses**, and Sam wondered if they were polished. After lunch, Sam sat down to watch a TV show about wild animals. Some man was on TV with a **lion**, but Sam didn't think it was very interesting. Then his Uncle George, who was putting on a **belt**, approached him.

"Sam, I don't want you to be so bored," he said. "Let's go and take a walk around the city. I just put out the **garbage can** and noticed that the weather is nice."

Sam looked up at his uncle, with light from a single **lamp** casting weird shadows around the house. "Okay," he replied, "We'll go to the museum. They have this new exhibit with fossils and a **dinosaur**."

"Good," said his uncle. "I don't make it out of the house much, and this will give me the chance to draw something new. I can only enjoy sitting in that old **rockingchair** for so long."

Uncle George thought for a minute, then went to the closet and put on an old **hat**, then he picked up some drawing supplies and headed to the front door. As they exited, Sam realized that the water **faucet** in the kitchen was dripping, but his uncle did not notice.

The weather was hot, but tolerable. Uncle George quickly spotted a woman a few houses away, picking a **flower** in her front yard. He said, "What a great drawing this would make," and walked towards her with a sketching pad and a **pencil**.

"Maybe you shouldn't do that," answered Sam. The woman's husband, who was cutting limbs from a tree with a **saw**, had spotted the two and was staring curiously. Sam said, "I know. Let's go to the park. You can draw a **duck** down at the pond."

Uncle George seemed to like this idea. "Alright, but this means we'll have to walk through the city." He put the pencil behind his **ear** and continued walking.

The park was several blocks away, and Sam and his uncle were forced to walk through downtown. While waiting at a **traffic light**, Sam began to get hungry again. He saw several vendors on the sidewalk, but the only one selling food had **pretzels** for sale, and this did not appeal to Sam. He decided that they should stop at a restaurant after the park.

Just then, an **ambulance** drove by quickly with its lights flashing. "I knew there was a reason I didn't come here often," said Uncle George. "Too much noise and hassle." They heard the **whistle** of a policeman. "There, you can see what I mean," he said.

"It will be much nicer in the park," replied Sam, "and maybe there will be a **sailboat** out on the lake for you to sketch. Besides, we're almost there."

Sam noticed that clouds were forming above. He saw a woman carrying an **umbrella** and hoped that his uncle didn't spot her. He didn't want his uncle to complain about the weather.

In the park, they spotted two boys flying a **kite**. Uncle Georger sat down, to draw this scene, and Sam wandered off to explore the park. A man was playing his **guitar** and Sam stopped to listen. He wasn't a very good guitarist and didn't seem to want any

company either, so Sam moved on. "If only I had my **bicycle**," he thought, "then I could take a nice ride and see much more of the park."

Sitting next to the lake Sam spotted a family, and they were opening a **picnic basket**. Sam had nothing else to do, so he sat down to watch. The family included two children, one of which was a little girl, who was eating a **banana**. The girl was watching her sister as she ate. The sister was walking around the lake, holding what looked like a **helium balloon**. Sam watched as the sister ran back to her parents, telling them that she had spotted a **turtle**. Her parents replied that turtles were very exciting, but they didn't really seem too interested. In fact, the father kept reading his **newspaper**.

Sam laid back on the grass and stared into the suddenly cloudy sky. He spotted an **airplane** behind the clouds and wished that he were on it. Uncle George really wasn't such bad company, but Sam wanted to be with friends his own age. His **wristwatch** showed that it was getting late, so he walked back, to find his uncle in the same spot as before. His uncle's hat was pulled over his eyes, and his **comb** had fallen to the ground.

Sam tapped Uncle George on the shoulder to wake him, and noticed that a **fly** was sitting on his uncle's shoulder. Uncle George woke with a stir. "I'm surprised I fell asleep," he said. "This bench sure isn't as comfortable as that **bed** back at the house."

"Let's go to the museum," suggested Sam. "It's right next to the park." He brushed away a **leaf** that had fallen onto his uncle's hat.

"Okay, but we shouldn't stay for too long," replied Uncle George. "It's getting late and the **sun** will be down soon."

As they left the park, many fewer people were in the park than had been there before. By an old **well**, Sam spotted the two boys still flying the kite. In the museum, they found that the dinosaur exhibit was closed, so they took the **elevator** to the third floor to see an exhibit on the American Revolution. There was a crowd around one of the displays, and

when Sam got closer he saw that it was a **cannon** from some old battle, but he couldn't read about it because of all the people.

He started to take a closer look, but the man in front of him was smoking a **pipe**, so Sam decided not to bother. He walked back to find Uncle George, but he was nowhere to be seen. Finally, Sam spotted him staring out a **window** which overlooked the park. Uncle George would look out the window for a while, and then draw on his sketchpad. Sam saw that there was a **telescope** overlooking the park. His uncle was drawing something outside, despite the oncoming darkness.

Sam felt for his house **key**. "Let's take a taxi to my house. It's dark outside and I'm getting hungry," he said.

At Sam's house, his mother was standing over the **ironing board**, pressing clothes. "I was wondering where you had taken my son, George" she said, smiling. "I couldn't reach you on the **telephone**, so I figured you had taken Sam out." But Sam did not want to listen to the grown-ups talk. He went into the kitchen and started to eat.

---

Note. Pictures seen by the subjects are in bold-face.

## Appendix B

"The Kid's Whistling". Separated into Idea-Units

Things were nearly perfect:<sup>1</sup> Christmas was three weeks away.<sup>2</sup> Roy worked late and was doubling his salary in overtime,<sup>3</sup> and tonight rain was falling.<sup>4</sup> Rain was Roy's favorite sensation,<sup>5</sup> and he never felt more at rest than when working on his signs<sup>6</sup> in his hot little room in Herlihy's department store,<sup>7</sup> with the rain tapping on the black skylight.<sup>8</sup>

The one trouble was the kid's whistling.<sup>9</sup> For ten months a year Roy had the Display Department to himself.<sup>10</sup> But at the beginning of November, the store manager hired a high school kid<sup>11</sup> to come in on weekday evenings and on Saturday's.<sup>12</sup> This year's helper was called Jack,<sup>13</sup> and he whistled all the time.<sup>14</sup> At the moment, Jack was printing counter cards<sup>15</sup> and whistling "Summertime".<sup>16</sup> He seemed to feel the tune needed a cool, restrained treatment,<sup>17</sup> for which Roy was grateful;<sup>18</sup> he was all set to begin the Toy Department sign<sup>19</sup> and he wanted things to go well.<sup>20</sup>

Jack switched to "After You're Gone,"<sup>21</sup> doing it loud, tapping his foot.<sup>22</sup> It got so trumpety that,<sup>23</sup> in the middle of putting the hairline on the Y,<sup>24</sup> Roy turned and stared burningly at Jack's spine.<sup>25</sup> It made no impression.<sup>26</sup> Jack was tall, about six inches taller than Roy,<sup>27</sup> and his neck led to a muff of uncut hair.<sup>28</sup> Just then, the kid leaned back and let fly four enormous, jubilant notes.<sup>29</sup>

"Hey, Jack," Roy called.<sup>30</sup>

The boy turned and said, "Beg pardon?"<sup>31</sup> He wasn't one of those mean kids actually.<sup>32</sup>

"What about a Coke?"<sup>33</sup>

"Sure. If you're having one."<sup>34</sup>

Roy didn't want a soft drink; he wanted quiet.<sup>35</sup> But he had worked himself into a position where there was nothing to do but go to the dark hall<sup>36</sup> and buy two wet bottles,<sup>37</sup>

then bring them back to the Display Department.<sup>38</sup> When Roy returned and offered Jack one,<sup>39</sup> the boy offered him a nickel and five pennies.<sup>40</sup> "Keep it," Roy told him, "and buy yourself a saxophone."<sup>41</sup>



## Appendix C

### Sample Protocol from Experiment 3 Scored According to Correct Idea-Units.

#### Major and Minor Errors

Roy was making signs in Hirlegy's department store. He was doubling his salary by working overtime.<sup>3</sup> For ten months a year he had the window<sup>Mi</sup> to himself until the store hired a high school kid who started in November.<sup>11</sup> He worked weekday nights and Saturdays.<sup>12</sup> His name was James.<sup>Mi</sup> It was raining out that particular night.<sup>4</sup> This was Roy's favorite kind of weather. James was humming a tune<sup>Ma</sup> which started to annoy Roy. He stared at James's spine<sup>25</sup> but this made no impression.<sup>26</sup> James switched to another song. Finally, Roy asked him if he wanted a Coke.<sup>33</sup> James said only if you are going to have one.<sup>34</sup> Roy wasn't really thirsty, he just wanted peace and quiet.<sup>35</sup> But now he was in a situation where he had to have one. So he went down the hall and came back with two bottles. James proceeded to give Roy a nickel and five pennies,<sup>40</sup> but Roy told him to keep it and buy himself a saxophone.<sup>41</sup>

---

Note. The numbers refer to the numbers of the correct idea-units that were scored. (These idea-units are listed in Appendix B.) "Mi" and "Ma" refer to minor and major errors, respectively.