# REMEMBERING, KNOWING, AND RECONSTRUCTING THE PAST

*Henry L. Roediger III*
*Mark A. Wheeler*
*Suparna Rajaram*

## I. Introduction

A central issue in the psychology of learning and memory is how memories change over time. This issue was first addressed experimentally by Ebbinghaus (1885/1964, Chap. 7) and has been pursued assiduously ever since. A related issue, raised only briefly by Ebbinghaus, is our state of awareness concerning our memories. For some memories, even distant ones, we feel a sense of reexperiencing the events during their recollection. In remembering other past events, we feel less engaged: We know the event occurred, but the act of recollection does not carry with it the strong feeling of reliving it. In other cases we may know that events happened to us, but we cannot remember their occurrence at all. In still other cases, experiences may have considerable impact on our behavior, but we cannot recollect them at all (see Jacoby, Kelley, & Dywan, 1989, for examples). These two topics—the course of retention over time and the phenomenological experience of the rememberer—represent two focuses of the present chapter. The third is an issue that did not much exercise Ebbinghaus, but which has interested researchers ever since: How accurate are our memories for distant events?

THE PSYCHOLOGY OF LEARNING
AND MOTIVATION, VOL. 30

When Ebbinghaus conducted his pioneering experiments on the temporal course of retention, he learned lists to the criterion of two perfect recitations and then, at some later point in time, relearned the lists. Retention was measured by savings in relearning. Ebbinghaus advocated this method for studying retention because it obviated the need to consider states of conscious awareness; people might still show savings for the events even if they were unaware that the events had ever occurred. All readers are familiar with the beautiful logarithmic forgetting function that Ebbinghaus discovered when he plotted savings against time. Virtually all studies confirm the general shape of the function in a variety of situations, with the main point of dispute being whether the function flattens out for general knowledge obtained years previously (e.g., Bahrick, H. P., Bahrick, P. O., & Wittlinger, 1975) or instead continues to decline very slowly even over long intervals (e.g., Squire, 1989).

We bother to recount the familiar tale of Ebbinghaus's research here because we want to remind readers of a methodological detail that was introduced by Ebbinghaus and that has been carried forward in almost all later research concerned with forgetting. To wit, when Ebbinghaus learned lists and tested himself on them later, he always tested different lists at the varying retention intervals. Thus, a list was studied and tested only once. Today we would say that Ebbinghaus used a within-subject, between-lists design. Most researchers following in his footsteps have made a similar choice; indeed, the overwhelming preference of researchers has been to use between-subjects designs in which different groups of subjects learn the same material but are then tested at different points in time from the learning episode. The reason that researchers have chosen between-subjects (or at least between-lists) designs probably has to do with their beliefs concerning effects of repeatedly testing the same material, as would occur in a within-subjects, within-lists design in which the same subjects learn and are repeatedly tested on the same material. In this latter design, the first test given may carry over and affect the later tests in unknown ways. Therefore, researchers have sensibly avoided the repeated-testing design as involving interpretive difficulties in tracing the course of forgetting.

Although the dominant mode of investigating forgetting has been the between-subjects or between-lists design, the repeated-testing design has not been entirely overlooked. Typically, subjects study lists of words or pictures and are tested repeatedly, without feedback or intervening study opportunities. Indeed, many celebrated investigations in cognitive psychology have used this arrangement, including those of Ballard (1913), Brown (1923), Bartlett (1932), Estes (1960), Tulving (1967) and Erdelyi and Becker (1974), to take some cases in their chronological order of occurrence. The first part of this chapter will seek to unravel some puzzles and paradoxes raised by such experiments in which subjects' memories are tested repeatedly for the same events.

We can think of several good reasons to examine performance in repeated testing paradigms, even though they have been avoided by most psychologists interested in learning and memory. First, the alleged contaminating influence of a test is of interest in its own right. Having subjects successfully recall material on a first test powerfully enhances performance on later tests (e.g., Hogan & Kintsch, 1971; Thompson, Wenger, & Bartling, 1978). A test of memory is not a neutral event to gauge subjects' knowledge, but affects that knowledge when it is displayed on later tests (e.g., Brown, 1923; Spitzer, 1939; Tulving, 1967). This testing effect, as it is sometimes called, may have wide educational impliciations (Glover, 1989) and therefore seems worthy of study. On the other hand, a second important consideration is that sometimes a test may have the opposite effect on the later test, causing interference and forgetting of the original material (e.g., Hastie, Landsman, & Loftus, 1978). We shall report experiments in this chapter revealing this effect and will then attempt to answer the critical question of when a test will facilitate or interfere with later recall. A third reason to study how memories change when they are repeatedly tested is the external validity of such a situation. People repeatedly recollect the important events of their lives either to relate them to others or to reflect on them privately. Therefore, if systematic changes in the content of memories occur as a function of such repeated recollections, the proper way to discover them is through experiments with repeated tests. Between-subjects or between-lists designs will not answer the important questions in this case.

As the astute reader will by now have surmised, this chapter is largely about the operation of retrieval processes as revealed through repeated testing of subjects on the same material. In the next section of the article we describe a puzzle that arose from prior work using this paradigm—why some experiments have shown improvements from repeated testing whereas others showed the opposite—and report experiments that solve the puzzle. We also show a powerful positive effect of taking one test on performance on a similar test a week later. In the next section of the chapter we consider the question of when an earlier test can have a deleterious effect on a later test and report experiments documenting this effect. These experiments show that a test conducted under the appropriate circumstances can create interference and harm performance on a later test. The effect seems similar to the misinformation effect in eyewitness testimony investigated by Loftus (1979, 1991). After this section, we consider the phenomenological experience of recollection by investigating two types of awareness referred to as *remembering* and *knowing* by Tulving (1985) and Gardiner (1988). In both cases people are aware that they are making reports about their past experiences, but in the former the recollection is accompanied by a feeling of traveling back in time, in some sense, to reexperience the events as they are recollected. In the latter case, that of knowing, people are confident of their recollections of events as being in their past, but the retrieval is not accompanied by the rich experience of remembering the original event. The experiments reported here add to those of Gardiner and his colleagues (summarized by Gardiner & Java (1993) in showing that these two bases of recollection are separable and can be dissociated by experimental variables. We conclude the chapter by suggesting directions for future research.

## II. Disparate Effects of Repeated Testing

Psychologists have used repeated testing paradigms for many different purposes; therefore, several different traditions of this research have become established, often with little cross-referencing. In this section we describe research from early in this century in which investigators tested memory repeatedly and reached diametrically opposed conclusions about its operation. Curiously, to our knowledge no one noticed the paradox posed by this early research until two of the current authors raised it in 1992 (Wheeler & Roediger, 1992).

In some of the most well known research in cognitive psychology, Bartlett (1932) reported his famous experiments in which he had English college students read an American Indian story, "The War of the Ghosts," and then recall it several times. Typically, the first recall attempt occurred 15 min or so after initial study, and later tests might occur days, weeks, or months later. On later tests Bartlett found dramatic distortions in recall of the story, with many omissions, alterations of meaning, and occasional additions. Bartlett emphasized the constructive nature of memory and argued that his subjects likely used the schema of a fairy tale, a common form to these students, in encoding and reconstructing the story. Supernatural elements were deemphasized, and the story was often made more consistent and rational. Therefore, from his repeated-reproduction technique, Bartlett (1932) concluded that memories often become more error prone over repeated tests.

It is worth noting that Bartlett (1932) produced no aggregate data in support of his conclusions, but rather presented sample protocols and anecdotes to bolster his conclusions. Interestingly, we can cite only one attempted replication of Bartlett's (1932) pioneering research—one published by Gauld and Stephenson (1967) and discussed below—that tried to confirm his claims using the repeated-reproduction technique. "The War of the Ghosts" has been used in a great deal of later research, but this work rarely involved repeated testing and was usually conducted for purposes other than examining the reconstructive nature of memory.[1] By the same token, much research has been conducted on the reconstructive nature of memory, but rarely has this interesting work involved repeated testing; it is more customary to assess memory via a single recognition test for information that might have been inferred but not actually stated in a prose passage (e.g., Johnson, Bransford, & Solomon, 1973) or via misinformation given in a narrative after subjects have witnessed some event (e.g., Loftus, 1979, 1991).

Curiously, Bartlett (1932) did not mention that his repeated testing research conflicted with other research dating back at least 20 years, also conducted in England. Ballard (1913) gave schoolchildren passages of poetry to memorize and then tested them repeatedly for intervals of up to one week later. Ballard found that children would often recall lines of poetry on later tests that they could not recall on earlier tests, a phenomenon he termed *reminiscence*. The basic observations of recall on a later test of material that had been missed on earlier tests were confirmed in later research (e.g., Brown, 1923) conducted well before Bartlett (1932) published his book, so it is curious that he did not at least cite it. However, the basic findings of Ballard and Brown were directly contrary to Bartlett's observations and conclusions; rather than dramatic forgetting and distortion of memory, Ballard and Brown had reported actual improvements over time in subjects' abilities to recollect their experiences. In fairness, Bartlett's celebrated book was intended to report his new experiments, which apparently began in 1913 (Bartlett, 1932, p. v). However, he might have added to the social factors that influence historical memory the tendency to disregard published evidence at odds with one's conception.

The observations of Ballard (1913) and Brown (1923) indicating positive effects of repeated testing on overall recall were examined for some years before being abandoned, for a time, as an object of serious study. Buxton (1943) reviewed the literature and concluded that reminiscence was an ephemeral phenomenon that failed to occur about as often as it appeared. However, Payne (1987) argued that Buxton (1943) reached his conclusion because the phenomenon of reminiscence had been redefined over the

[1] Many of Bartlett's (1932) experiments used a serial reproduction technique in which one subject read "The War of the Ghosts" and recalled it later, and then a second subject read the first subject's protocol and later recalled it, and so on. This technique has been used in later research (e.g., Paul, 1959), and Bartlett's observations were generally confirmed.

years. Ballard's (1913) original definition of the term was of material that could not be recalled on a first test that was recovered on a second (or later) test. Whenever total recall improves between two tests, reminiscence must have occurred; therefore, Ballard sometimes used overall improvement between tests as an index of the occurrence of reminiscence. This now seems a mistake, because it led later researchers such as Buxton (1943) to redefine reminiscence as overall improvement in recall between tests. However, it is perfectly possible to have reminiscence (defined as "intertest recovery") without having overall improvement between tests, because forgetting between tests can offset the reminiscence or recovery. Therefore, when Buxton (1943) concluded that the phenomenon was unreliable, he referred to an overall improvement between tests, not to reminiscence defined as intertest recovery, which was widely obtained. Nonetheless, his review is generally credited with diminishing research in this area for several decades.

Research on the topic of improvements in recall across repeated tests was resuscitated by Erdelyi and Becker (1974). They presented subjects with pictures or concrete words and had subjects recall them on three successive tests, each lasting 7 min. They also used a new procedure, forced recall, in which subjects basically engaged in a free-recall test but were forced to produce a preset number of responses that was greater than the number of items subjects could recall. This procedure was used to overcome the argument that gains observed on later recall tests should be attributed to relaxed recall criteria on these tests. (We examine the effects of forced recall on memory in a later section). Erdelyi and Becker reported overall improvements across tests in recall of pictures but not of words. They labelled this improvement *hypermnesia* (the opposite of amnesia, or forgetting).[2]

It is worth noting that virtually all experiments employing free or forced recall have reported strong reminiscence in recall of both pictures and words (e.g., Erdelyi, Finkelstein, Herrell, Miller & Thomas, 1976), although in the case of words the improvement between tests is offset by intertest forgetting in many experiments. (However, some researchers have reported reliable hypermnesia for words; e.g., Payne & Roediger, 1987). There is by now a large literature on hypermnesia (see Erdelyi, 1984; Payne, 1987; and Roediger & Challis, 1989 for reviews). The point

[2] Today the term *reminiscence* is used to refer to recall of events on a later test that were not recalled on an earlier test; reminiscence is intertest recovery, which agrees with Ballard's (1913) original definition of the term. However, reminiscence can be offset by intertest forgetting. Hypermnesia refers to overall gains in recall between two tests, that is, when reminiscence reliably outweighs intertest forgetting. Thus one can obtain reminiscence without hypermnesia, but the reverse case is impossible.

we want to establish here, however, is simply that the phenomena of reminscence and hypermnesia are real and often replicated. In one interesting experiment, Scrivner and Safer (1988) showed subjects a videotape of a burglary and then gave them repeated tests in recalling critical details of the event. Recall improved steadily across four repeated tests. The generality of the phenomenon of hypermnesia leads us back to the initial question as to why two different traditions of research, both employing repeated testing, can arrive at such disparate conclusions regarding retrieval processes in memory.

Wheeler and Roediger (1992) examined past research and isolated two likely factors as potential causes of the different findings and conclusions: the type of material used and the length of the interval between tests. Bartlett (1932) used prose passages in most of his memory experiments, such as "The War of the Ghosts." On the other hand, much of the research showing improvements between tests used lists of words, pictures, or similar materials. (Payne, 1987, reviewed 172 such experiments documenting the phenomenon, all using lists.) Another potential factor is the intertest interval; Bartlett (1932) used rather lengthy intervals between tests, often days and sometimes months, whereas researchers studying hypermnesia usually interpose only five minutes or less between tests.

It is not obvious from prior research which of these factors should be more important, or whether both are critical. For example, Ballard (1913) obtained hypermnesia for passages of poetry, which might be considered connected discourse like prose. Similarly, Roediger, Payne, Gillespie and Lean (1982) obtained hypermnesia in recall of categories (presidents, birds, sports), which are also in well-structured sets. On the other hand, some researchers have obtained hypermnesia across long intervals. In the experiment by Scrivner and Safer (1988) described above, hypermnesia was obtained between tests over a 48-hr interval. Similarly, Erdelyi and Kleinbard (1978) obtained hypermnesia for a list of pictures over a week by testing subjects three times a day with a forced-recall procedure.

Wheeler and Roediger (1992) examined both retention interval and type of material as the possible factors underlying prior discrepant results in several experiments. They had subjects study 60 pictures under one of two conditions before taking forced-recall tests on the pictures. In one case, subjects heard a story and the names of the 60 pictured objects occurred in the story. They were told to learn both the story and the names of the pictures. Other subjects saw the same 60 pictures presented in the same order, but they heard the names of the pictures as they were being presented. These two conditions were meant to simulate, to some degree, the difference between Bartlett's materials (schematic processing of prose) in the pictures + story condition, on the one hand, and the list-learning conditions of the typical hypermnesia experiment in pictures + names conditions, on the other. Wheeler and Roediger's procedure arranges this comparison with the target material held constant between conditions, so one need not make a comparison between prose recall on the one hand and recall of some completely different material presented in lists, on the other.

The other main variable in the Wheeler and Roediger (1992) experiment was the schedule of testing subjects received after seeing the 60 pictures in one of the two conditions. All subjects received a brief questionnaire asking them about various features of the experiment, such as estimating the number of pictures presented. One third of the subjects were dismissed at this point and told to return a week later. (The questionnaire was created to give subjects in this condition a plausible rationale for having participated.) Another third of the subjects received one test for the pictures; they were given sheets numbered 1–60 and were told to recall the names of as many of the 60 pictures they had previously studied as possible, but that they should guess to fill up the 60 spaces. Seven minutes were permitted for recall. The final third of the subjects were treated the same way, except that they were given three 7-min forced-recall tests, with 1-min breaks between tests. Finally, all subjects returned a week later and then received three consecutive forced-recall tests for the 60 pictures that had been studied the previous week.

In summary, subjects studied either pictures in a list or in the context of a story, then took zero, one, or three forced-recall tests on the pictures, and then returned the next week and took three more tests. We anticipated that subjects taking the three immediate tests would show hypermnesia (improved recall over tests), at-least in the condition where they studied pictures and their names (replicating Erdelyi & Becker, 1974, among many others). However, we expected that forgetting (not improvement) would occur between tests with a week lag between them, and that this forgetting might be more pronounced in the pictures + story condition with its schema-driven processing.

The basic results are shown in Table I. The six groups of subjects are labelled on the left, with the first number indicating the number of tests taken on the day subjects studied the pictures (0, 1, or 3) and the second number indicating the three tests taken a week later (always 3). Recall of pictures was greater in the pictures + story conditions than in the pictures + names conditions; collapsing across all other conditions, the difference was about four items. However, the type of material is irrelevant to the main point here, so let us focus on the picture + story recall in the bottom of Table I to answer four questions of interest. First, did recall improve in the three immediate tests? The answer is clearly *yes:* Recall

TABLE I

NUMBER OF PICTURES RECALLED AS A FUNCTION OF PRESENTATION CONTEXT AND TESTING SCHEDULE[a]

| Context and group | Immediate tests 1 | 2 | 3 | $T_3 - T_1$ | Delayed tests 1 | 2 | 3 | $T_3 - T_1$ |
|---|---|---|---|---|---|---|---|---|
| Pictures plus names | | | | | | | | |
| 3–3 | 26.6 | 27.2 | 28.4 | 1.8[b] | 25.2 | 26.3 | 26.0 | 0.8 |
| 1–3 | 25.7 | — | — | — | 20.2 | 21.7 | 23.0 | 2.8[b] |
| 0–3 | — | — | — | — | 16.7 | 17.5 | 17.5 | 0.8 |
| Pictures plus story | | | | | | | | |
| 3–3 | 32.7 | 35.0 | 36.4 | 3.8[b] | 31.8 | 33.0 | 33.4 | 1.6[b] |
| 1–3 | 31.8 | — | — | — | 23.3 | 25.0 | 25.6 | 2.3[b] |
| 0–3 | — | — | — | — | 17.4 | 17.2 | 18.4 | 1.0 |

[a] Data are from Wheeler and Roediger (1992) and are reprinted by permission of the Cambridge University Press.

[b] These conditions demonstrated reliable hypermnesia across the three tests.

improved by 3.8 items across the tests, and the hypermnesia was even greater in the pictures + story condition than in the more typical pictures + names condition. Clearly, hypermnesia can be obtained in recall following a story (albeit of pictures embedded in the story).

The second question of interest is whether forgetting occurred between tests when a week rather than few minutes elapsed between them. Again, the answer is *yes*. In Group 3–3 in the pictures + story condition, recall dropped from 36.4 items recalled to 31.8 over the week; in Group 1–3, the drop was from 31.8 to 23.3. Successive tests with a week between tests produce forgetting, not hypermnesia. These results show that it is likely the delay between tests and not the type of material that produced the disparate outcomes in Ballard's (1913) and Bartlett's (1932) experiments.

The results in Table I can also be used to address two other issues that are of some interest. Can hypermnesia be obtained after a week retention interval if short intervals occur between successive tests after the week's delay? The answer given from the six conditions in the right side of Table I seems to be *yes*. In all six cases subjects recalled more on the third test than on the first test and an analysis of variance on only the delayed test data produced a significant effect for the number of tests, $F(2, 114) = 14.35$, $MS_e = 5.03$, $p < .001$. However, the effect is not particularly robust because it was significant on only three of the six conditions in individual ANOVAs. Nonetheless, in all likelihood, hypermnesia can be reliably obtained after a week's delay.

The final point to draw from Table I is the power of a test in aiding later recall. Subjects who took three immediate tests recalled more pictures a week later than did subjects who took only one test, but these subjects in turn recalled the pictures much better than did subjects who had no tests after studying the pictures initially. These results are portrayed graphically in Fig. 1, where performance on the three delayed tests has been averaged and plotted as a function of the number of immediate tests. The three groups of subjects in either the pictures + names and pictures + story condition were treated identically up to the point of the first test and then when they returned a week later. The only difference between conditions that could affect performance on the later tests was the number of tests taken during the initial session. Nonetheless, as Fig. 1 shows, delayed recall increased monotonically with the number of prior tests and the enhancing effect of testing affected delayed recall much more powerfully in the pictures + story than in the pictures + names condition. We have no ready interpretation of this last outcome, but our point here is mainly to document the powerful effect that taking a test has on later retention.

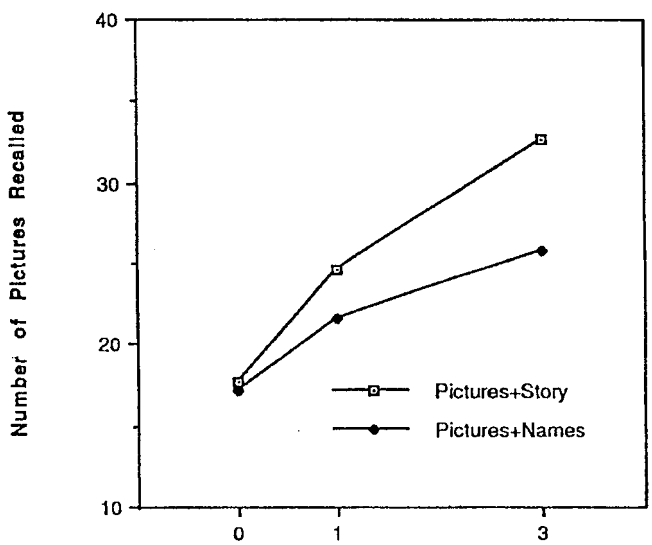


Fig. 1. The testing effect. The number of tests taken soon after studying pictures greatly affected recall one week later.

Many others have made this point, too (e.g., Glover, 1989; Izawa, 1971; Spitzer, 1939; Thompson et al., 1978).

To return to the main point of the experiment, Wheeler and Roediger (1992) argued that their experiment resolved the puzzle posed by prior inconsistent studies of Ballard (1913) and Bartlett (1932). The answer is rather straightforward: If there are short intervals between tests, then one usually finds hypermnesia over repeated testing. If the intervals are long (a week, in our experiment), then one obtains forgetting between tests. Of course, this latter point must be true in the limiting case—say, with 5 years between tests—but also occurs with intervals as short as a week.

We conclude that type of material played little role in the earlier discrepancies between outcomes, because we found no interaction with other variables between pictures presented in a list (the pictures + names condition) and those presented in a story context (the pictures + story condition). However, this conclusion can be challenged, because in a sense we used the same materials—a series of 60 pictures—in both conditions. Wheeler and Roediger (1992) conducted two further experiments to see whether the basic findings could be obtained with prose materials. Can hypermnesia be obtained with short intervals between tests and forgetting be obtained when the interval is lengthened to one week?

The two experiments were similar, except for the types of materials and the fact that one was conducted as a classroom demonstration and the other was conducted under more controlled laboratory conditions. Nonetheless, the outcomes were very much alike. In the classroom experiment, students in a cognitive psychology course at Rice University read "The War of the Ghosts" twice at a comfortable rate and then spent 5 min recalling U.S. presidents. They were then given 8.5 min to recall the story as well as possible; following this free-recall attempt, they recalled U.S. states for 5 min and then recalled the story again for 8.5 more minutes. One week later students were given a surprise test and asked to recall the story again. The laboratory experiment was conducted under generally similar conditions with an excerpt from a John Updike short story, "The Kid's Whistling," serving as the target material.

The results of the two experiments are shown here in Table II. For both types of materials, subjects showed modest but statistically significant improvements between the first two tests. Following others (e.g., Mandler & Johnson, 1977), we scored the results in terms of the number of idea units (significant phrases or ideas in the passage) that were recalled. The improvement between the two initial tests looks modest, but each idea unit is composed of seven to eight words, on average, so the improvement would look greater if scored in this manner (which, however, is difficult with prose materials). In both cases the improvement was quite consistent

TABLE II

NUMBER OF IDEA UNITS RECALLED IN TWO EXPERIMENTS, USING DIFFERENT PROSE PASSAGES[a]

| Material | Test 1 | Test 2 | Delayed test |
|---|---|---|---|
| "The War of the Ghosts"[b] | 21.4 | 22.9 | 19.0 |
| "The Kid's Whistling"[c] | 12.1 | 13.2 | 10.7 |

[a] Data are from Wheeler and Roediger (1992) and are reprinted by permission of the Cambridge University Press.
[b] 42 idea units.
[c] 41 idea units.

across subjects (see Wheeler & Roediger, 1992 for details). Furthermore, in both experiments recall decreased between the second test and the third one a week later. However, in neither case did subjects display gross confusion and inaccuracy on the delayed test, which might have been expected from Bartlett's (1932) results. We return to this point later.

We conclude from the experiments described thus far that, for recently learned material, recall improves between tests when short intervals separate them but that forgetting occurs when the interval is increased to one week. Bahrick and Hall (1993) have argued that this conclusion may hold only for episodic memory situations, because when subjects are repeatedly tested on relatively permanent knowledge (e.g., of public events or famous faces), they show improvements over long delays between tests. Bahrick and Hall (1991) reported improvements over one month, as did Hermann, Buschke, and Gall (1987). In a rather different paradigm, Squire, Haist, and Shimamura (1989) reported significant improvements with one year between tests. These reports indicate that hypermnesia may be obtained with long intervals between tests, but Roediger and Wheeler (1993) noted that one possible interpretation of these gains in knowledge between tests is that subjects may be exposed to the relevant material from magazines, newspapers, television, or books during this time. (In some cases, such as research by Squire et al. [1989], the test procedure exposed subjects to the correct answers.) It is inherently difficult to test general knowledge with widely spaced intervals and not to have intervening study opportunities for the material. Indeed, the first test may sensitize subjects to relevant information and lead them to pay more attention if they are exposed to it later (but see Bahrick & Hall, 1991). Nonetheless, the results described by Bahrick and Hall (1993) are interesting and deserve further study.

To conclude this section, we think we have resolved the paradox posed by the disparate results from repeated testing in episodic memory by showing that the interval between tests is the critical variable. For recently learned material such as pictures or prose passages, repeated testing with short intervals between tests produces improvement in overall recall (hypermnesia). (If Bartlett, 1932, had used short intervals between tests he might have reached very different conclusions about the reconstructive nature of memory.) When the interval between tests is extended to one week, then recall declines. However, even with this longer interval, most of the forgetting occurred as omissions of material; there were few errors and confabulations introduced, and most of these were minor. This was true in the classroom experiment with "The War of the Ghosts" and with the other types of material, too. Does this evidence indicate that remembering is not as reconstructive as Bartlett (1932), among many others, would have us believe? We turn to this issue in the next section.

## III. Interference and Reconstruction

Bartlett (1932) concluded from his research that remembering was highly constructive, because his subjects' recollections of "The War of the Ghosts" were so distorted during delayed tests. The distortions were not caused by misleading suggestions planted in other material, but were created by the subjects themselves in their previous recall attempts, or so the story goes. As previously noted, Bartlett's (1932) experiments have never been successfully replicated, to our knowledge. This seems a curious state of affairs for some of the most widely cited observations in cognitive psychology, especially because the original observations themselves were almost anecdotal. In light of more recent evidence we might wonder indeed whether the findings of reconstructed recall *can* be replicated.

In an interesting series of experiments, Gauld and Stephenson (1967) explored the possibility that Bartlett's (1932) instructions to his subjects may have led them to construct rather than to remember. Bartlett never says exactly what instructional set he encouraged in his subjects and indeed seemed to be deliberately vague in his instructions: "I thought it best, for the purposes of these experiments, to try to influence the subjects' procedure as little as possible" (1932, p. 78). Gauld and Stephenson (1967) note that if subjects took their task as retelling a story rather than as remembering it, Bartlett's results showing invention and construction would be expected. As they say:

> Most people who retell a story are unlikely to care very much whether the story they retell is the same, detail by detail, as the story they originally heard. In other words, they are most unlikely to take pains that what they come out with is always what they remember rather than what they guess at or even consciously invent. Now if the changes and inventions in reproductions of stories . . . are to serve as the foundation for a theory of remembering, [then it should be established that the subjects] were indeed seriously trying to remember, and were not more or less consciously guessing or romancing in order to fill in gaps in their memories." (Gauld & Stepenson, 1967, p. 40)

Gauld and Stephenson (1967) conducted three experiments that led to the conclusion that Bartlett (1932) must have used rather loose instructions, ones encouraging subjects to tell a good story rather than to remember conscientiously. When subjects recalled "The War of the Ghosts" under instructions to remember the story as well as possible, there were relatively few errors, just as Wheeler and Roediger (1992) reported 25 years later. In addition, subjects were generally good at recognizing their own errors when asked to judge them. If Bartlett's (1932) instructions did emphasize that his subjects should tell a good story rather than try to accurately recall the story, "reconstructive memories" would be expected; if subjects are told to construct a story, they will. However, they may well not confuse this process with remembering if they were asked to make such a judgment. We report some observations below that bear on this issue.

In the modern era, the study of reconstructive memory has employed different techniques from Bartlett's (1932) method of repeated reproduction. Perhaps the most widely used is the so-called misinformation paradigm pioneered by Elizabeth Loftus and now widely studied (e.g., Loftus & Palmer, 1974; Loftus, 1979; Loftus, 1991). This paradigm capitalizes on a form of retroactive interference of the sort obtained in the A–B, A–D paradigm, in which subjects are taught two different responses to the same stimulus situation. Not surprisingly, people become confused about how to respond in Situation A if they have learned conflicting responses. If subjects are told to produce the response that they originally learned (B), but instead they produce D by mistake, they have wrongly reconstructed the original event because of the retroactive interference created by the interpolated learning.

In the standard form of the misinformation paradigm introduced by Loftus, subjects witness a series of events in which some critical items are embedded. For example, they may see a wrench. Later, subjects in an experimental condition read a narrative in which the suggestion is made that the tool in the original slide sequence was a hammer. Control subjects read the same passage except that the misinformation about a hammer appearing is not included (a neutral word such as *tool* is used). Later subjects are given a test in which they are asked to recognize which item appeared in the original slide sequence, *wrench* or *hammer*. Subjects in the experimental (misinformation) condition erroneously pick *hammer* much more often than control subjects, demonstrating the misinformation effect (or retroactive interference). Loftus (1979) interprets the results as indicating that subjects' memories are altered by the misinformation so that, at test, they believe that their recoded version of the memory is the actual event they saw. This interpretation is akin to the notion that the original information is unlearned and then replaced by the new information, one of the two factors in classic interference theory (Melton & Irwin, 1940).

McCloskey and Zaragoza (1985) have argued that Loftus' standard paradigm is not the appropriate one to determine whether subjects' memories of the original event (the appearance of the wrench, in the above example) has been altered. They suggested that a modified recognition test, in which subjects are given a choice between the original item (*wrench*) and a new exemplar from the same general category (say, *screwdriver*), is a more appropriate test of whether the misinformation reduced accessibility to the original target item. In a series of experiments McCloskey and Zaragoza (1985) found that subjects in the experimental (misinformation) and control conditions performed about equally well on the modified recognition test (see, too, Zaragoza & Koshmider, 1989; Zaragoza, McCloskey & Jamis, 1987). They interpreted this outcome as indicating that the misinformation produced no alteration in the representation of the original event. The standard effect, they argued, was due to response competition when subjects were given two plausible competing responses that were both familiar. Of course, response competition is the second major factor in classic two-factor interference theory, the one that later caused so much debate on how to eliminate it from measures of retention (e.g., Barnes & Underwood, 1959).

In more recent experiments Loftus (1991) and others (e.g., Ceci, Ross & Toglia, 1987) have found the misinformation effect even on the modified recognition test, but its magnitude has been quite small relative to the outcomes on the standard recognition test employed by Loftus and others. Typically, subjects are only 5–6% worse on the modified test in the experimental condition than in the control condition, when the effect is obtained at all. In some sense, failing to find the misinformation effect on recognition tests is not too surprising, because in an earlier era researchers also had trouble finding retroactive interference in the standard A–B, A–D paradigm on recognition tests (e.g., Postman & Stark, 1969). Curiously, there are relatively few reports in which recall tests are used in the misinformation paradigm, although both prior research in the interference tradition and considerations of external validity would seem to argue that such tests should be preferred. The preference on the latter grounds is that witnesses to crimes are frequently asked to recall the information, in response either to open-ended questions with general contextual cues ("Tell me what happened to you") akin to free recall, or to specific questions, as in cued recall procedures. Comfortingly, some experiments that have used recall procedures do tend to show misinformation effect (e.g., Lindsay, 1990).

The next experiments we report here can be thought of as a new way of studying reconstructive memory that contains features of both Bartlett's (1932) and Loftus' (1979) procedures, using recall tests. Bartlett's subjects in some sense provide their own misinformation because (he argued) on later recall attempts they were recalling their own prior recollections into which errors had been introduced. Loftus gained experimental control over the misinformation by planting the conflicting information in a narrative read after the original events were seen. However, misleading information from someone else may not pose as potent a source of interference as that generated by oneself. We have developed a technique in which we ensure that subjects are exposed to misleading information about recently experienced events and that they produce this information themselves. We can then examine whether this misinformation harms memory for the original events on later tests. The experiments in which we employed this technique were originally devised for a different purpose, so first we provide the relevant background.

Roediger and Payne (1985) were interested in the question of why some researchers routinely found hypermnesia for words (e.g., Roediger & Thorpe, 1978; Roediger et al., 1982) and others did not (e.g., Erdelyi & Becker, 1974; Erdelyi & Kleinbard, 1978), even with comparably short intervals between tests. One difference between experiments conducted in the different laboratories was that Erdelyi and his colleagues used forced-recall procedures whereas Roediger and his colleagues used free recall, with a warning against guessing. (Because intrusion rates did not increase over tests, they assumed that increased guessing was not a problem.) Roediger and Payne (1985) performed the straightforward experiment to find out whether the differences in the recall test led to the discrepant findings. Subjects studied a list of 70 words and then received three tests under one of three instructional conditions. Free-recall subjects were asked to recall the material as well as possible and were warned against guessing; forced-recall subjects were told to recall as many items as possible and were told that, before the test period was over, they needed to fill at least 50 spaces on their response sheets, guessing when necessary. A third group of subjects were given what Bousfield and Rosner (1970) called "uninhibited" recall instructions: They were told to recall the information as well as possible, but to hold nothing back and to produce every response that occurred to them. (Recall was not to be inhibited by withholding responses for any reason. The instruction is essentially one to free associate during recall.)

Roediger and Payne's (1985) basic results are given in Table III. Briefly, the type of instruction had no effect on the amount of hypermnesia obtained, and reliable improvements were obtained in all three conditions. Thus they did not solve the puzzle that had been the reason for the experiment—to explain why some researchers found hypermnesia for words and others did not. However, they noted that the results posed a more interesting puzzle for students of learning and memory: why huge variations in recall criteria had no effect on the amount recalled. An assumption ingrained in our theories of remembering is that guessing should produce benefits in performance, even if such benefits would not reflect true remembering but rather chance hits due to guessing (e.g., Klatzky & Erdelyi, 1985). However, Roediger and Payne's (1985) results do not show this: Response criteria were manipulated by almost a factor of 10 (2.5 intrusions in free recall and 23.8 in forced recall), and yet the number of items correctly produced did not vary (recall was actually 0.4 items worse for forced-recall subjects, though this slight difference was not significant). Apparently the assumption—implicit in most theories but explicit in certain versions of generate/recognize theories (e.g., Kintsch, 1970)—that subjects recall so little under free-recall conditions because they adopt a strict criterion for responding and therefore do not produce some correct responses that they generate, is wrong. If subjects are encouraged to produce every item they think of (the uninhibited-recall

TABLE III

MEAN RECALL, INTRUSIONS, AND HYPERMNESIA FOR THREE TYPES OF TEST[a]

| Type of recall test | Test 1 | Test 2 | Test 3 | Hypermnesia $T_3 - T_1$ | Mean recall[b] | Mean intrusions |
|---|---|---|---|---|---|---|
| Free | 25.1 | 26.6 | 28.2 | 3.1 | 26.6 | 2.5 |
| Uninhibited | 24.3 | 26.4 | 28.4 | 4.1 | 26.4 | 9.6 |
| Forced | 25.1 | 26.0 | 27.4 | 2.3 | 26.2 | 23.8 |
| Mean | 24.8 | 26.3 | 28.0 | 3.2 | 26.4 | |

[a] Data are from Roediger and Payne (1985) and are reprinted from *Memory & Cognition*, *13*, pp. 1–7, by permission of the Psychonomic Society, Inc.

[b] Means are averaged across the three tests.

condition) or are even forced to write reponses that are guesses (the forced-recall condition), they produce no more correct targets than under free-recall conditions.

Of course, there must be limits to the conclusion that guessing does not affect recall. Imagine that subjects are presented with the names of six of the months of the year tachistoscopically, so that they cannot be perceived very well. One group of subjects now engages in forced recall (being instructed to respond with 12 month names), whereas the other group is tested by free recall, with the customary warning against guessing. It would hardly be a surprise that the forced-recall group "recalled" correctly more presented months than the free-recall group. (Of course, the greater number of hits would be offset by more intrusions, too, so memory for the six presented months might be equal in the two conditions if an appropriate correction for guessing were applied to the data). In short, if subjects are given a constrained or guessable set of material, guessing is likely to help correct responding, but at the expense of many more errors, so that some correction must be used to try to estimate accurate recall.

Roediger and Payne (1985) used lists of more or less random words that would be difficult to guess and showed that instructions to guess did not benefit raw recall hit rates, even without worrying about corrections for guessing. Erdelyi, Finks, and Feigen-Pfau (1989) compared free and forced recall for a set of pictures and found that forced-recall subjects showed a small benefit relative to free-recall subjects, but at the expense of more intrusions. However, picture lists tend to be categorized lists (some animals, some plants, some toys, etc.) and so may be more guessable, a result that, in fact, Erdelyi et al. (1989) showed. Although overall responding was greater for the forced-recall condition than for the free-recall condition, if performance were somehow corrected for guessing, then the difference might evaporate, as pointed out by Roediger, Srinivas and Waddill (1989).

This concern with recall criteria led us to the current experiments, although now they seem more interesting from the perspective of the reconstructive approach to memory, as a new way of studying the interfering effects of misinformation supplied by the subjects themselves. In the forced-recall procedure subjects are told to try to recall as many target items as possible, but to guess to fill up the preset number of required responses before the recall period is over. We have found that subjects do not like to comply with the request for forced recall, because they know they are engaged in sheer guessing toward the end of the recall period. However, our interest in forced recall for present purposes is how this procedure will affect subjects' memories when they are tested later. If subjects study a set of material and then engage in either forced recall or free recall, we know that (1) the two groups will produce about the same number of target items, and (2) the forced-recall group will produce many more intrusions than the free-recall group. The question of present interest is what will happen when these two groups are asked to judge the accuracy of their responses. Will forced-recall subjects falsely recognize some of their confabulated responses as having occurred in the list? Will this tendency increase over repeated tests? That is, will subjects who took a forced-recall test initially later show poor retention for the original list because they supplied their own misinformation in the form of the intrusions generated on the first test? Subjects' own erroneous responses, their confabulations in attempting to recall the material, may provide a potent source of interference relative to the free-recall conditions where intrusions are rare. If so, then the phenomenon might be a workable addition to our techniques for studying reconstructive aspects of remembering, blending a strength of Bartlett's (1932) approach of having subjects be misled by their own prior recalls while still gaining some control over the nature of the misleading information, as in Loftus (e.g., 1979, 1991).

We describe here one experiment, in a series to be reported in full elsewhere (Roediger, Challis & Wheeler, in preparation), that used this paradigm. Subjects in this experiment studied 60 pictures in the context of a story, as in the Wheeler and Roediger (1992) experiment described above (although these pictures and the story were different). This story was 805 words long and revolved about a central character, "Sam the Alligator." The story had a plot but was somewhat eclectic, with an unpredictable theme, to miminize guessing from a stereotyped script (such as a visit to the zoo). Subjects were told that they would hear a story and would see pictures representing objects in the story. Their task was to remember as many of these objects as possible for a later test. A slide representing each picture was presented while the name of the object was mentioned in the story. During some of the time, when no critical objects were being mentioned in the story, the screen was blank. The presentation of the story and pictures lasted about 6 min.

After the presentation, half the subjects were dismissed from the experiment with instructions to report back the following week. The other half of the subjects took a test, either a free-recall test with a warning against guessing, or a forced-recall test. The instructions for this latter test said that subjects should try to recall as many items (names of the pictures) as possible, but that they should produce 60 responses before the 7-min recall period ended. They were told that they could produce more than 60 responses or guesses if they wanted to. A week later all subjects returned to the laboratory; half now took a free-recall test and half a forced-recall test. Subjects who had taken an immediate test received the same type of test a week later. Therefore, the experiment employed four between-subjects test conditions: One group received a free-recall test both immediately after study and after a week delay; the second group received forced-recall tests at both times; and two other groups received only delayed tests, either free recall or forced recall. The contrast between the groups taking two tests and those taking a single delayed test permits examination of the effects of the immediate test on a delayed test.

A final important detail of the procedure is that subjects were required, after taking each test, to go back and to rate each item they had produced on a 1–6 scale in terms of their confidence that it had appeared in the original list of pictures. Ratings of 4, 5, or 6 represented increasing levels of confidence that the items had appeared in the original set of pictures; ratings of 3, 2, or 1 represented increasing confidence that the item was new and had not appeared in the list. Therefore, a response of 1 indicated that subjects were certain that the response was new, and a rating of 6 indicated that subjects were certain the item was from the original set of pictures they had seen.

Consider first the immediate-recall results under free- and forced-recall conditions, shown in Table IV. The number of target items correctly produced is about the same in the two conditions, 35.9 and 35.6 produced, despite the large difference in the number of intrusions (0.9 and 25.9). These results replicate those of Roediger and Payne (1985) in showing that a huge variation in recall criterion has no effect on the number of responses correctly produced. However, these results go beyond those, because subjects were asked to rate each produced word on a 6-point confidence scale as to whether they believed it was from the list. Ratings of 4, 5, or 6 indicated that subjects judged the produced item as from the list. As can be seen from Table IV, subjects in the free-recall condition judged that virtu-

TABLE IV

NUMBER OF TARGET ITEMS PRODUCED AND RECOGNIZED, AND THE NUMBER OF INTRUSIONS AND THOSE FALSELY RECOGNIZED, IN THE IMMEDIATE TEST CONDITIONS[a]

| Type of recall test | Targets produced | Targets "recognized"[b] | Intrusions produced | Intrusions "recognized"[b] |
|---|---|---|---|---|
| Free | 35.9 | 35.6 | 0.9 | 0.8 |
| Forced | 35.6 | 32.8 | 25.9 | 5.0 |

[a] Data are from Roediger, Challis, and Wheeler (in preparation).
[b] Subjects gave a rating of 4, 5, or 6 on the recognition test.

ally all their produced items were from the list (the items received ratings of 4 or higher). However, this was not the case for subjects who engaged in forced recall. They produced 2.8 items that they failed to recognize as studied; apparently, they thought that these were guesses they had generated in complying with the experimenter's request to produce a large number of responses. In addition, they falsely recognized 5.0 of the 25.9 erroneous responses as having actually occurred in the list. Thus, unlike free recall, the forced-recall procedure leads subjects into two types of errors: recognition failure of recallable words and erroneous recognition as old of items that were not studied but that were generated on the tests.[3] The requirement for subjects to confabulate a large set of responses creates confusion or a reality-monitoring problem (Johnson & Raye, 1981) on the test. In complying with the forced-recall instruction, subjects generate their own misinformation (or set of interfering responses). If these responses create interference as to which items occurred in the set of studied pictures on the immediate test, just after the responses have been produced, the interference is likely to be even more potent after a delay.

The results in Table V, showing performance on the delayed test, indicate this to be the case for at least one type of error. First, note that on the delayed free-recall tests there is no strong evidence of the two sorts of confusion just defined, although some errors are apparent. Items free recalled by subjects (whether target items or the few intrusions) are generally judged to be from the list. This pattern occurs both for subjects who received an initial test a week earlier and for those who did not. However, data from both of the forced-recall test conditions show evidence of great confusion; both types of errors seen in forced recall on the immediate test appear on the delayed test. For example, subjects in both the forced-recall conditions produced about five more items actually from the target list than they recognized, producing 30.9 and 22.4 and recognizing 25.9 and 17.3, respectively. Thus they showed heightened recognition failure of recallable words relative to the free-recall test conditions. Similarly, forced-recall subjects frequently falsely recognized their generated responses that were not from the set of pictures as actual list members. Subjects who had a prior test falsely recognized 7.6, or 20%, of their incorrect responses as "old"; those who had not had a prior test averaged 12.1, or 28% of the generated (incorrect) responses as this sort of error.

[3] We are obviously assuming that it is appropriate to analyze absolute numbers of errors in the two conditions. A different set of conclusions would emerge if proportions were calculated.

TABLE V

NUMBER OF TARGET ITEMS PRODUCED AND RECOGNIZED, AND THE NUMBER OF INTRUSIONS AND THOSE FALSELY RECOGNIZED IN THE DELAYED TEST CONDITIONS[a]

| Type of recall test | Targets produced | Targets "recognized"[b] | Intrusions produced | Intrusions "recognized"[b] |
|---|---|---|---|---|
| Free | | | | |
| Prior test | 29.6 | 29.5 | 2.1 | 1.9 |
| No prior test | 15.4 | 14.3 | 3.6 | 2.5 |
| Forced | | | | |
| Prior test | 30.9 | 25.9 | 38.6 | 7.6 |
| No prior test | 22.4 | 17.3 | 43.9 | 12.1 |

[a] Data are from Roediger, Challis, and Wheeler (in preparation).
[b] Subjects gave a rating of 4, 5, or 6 on the recognition test.

We find these figures interesting on two counts. First, the absolute number of these false recognitions seems quite high. Subjects were required to produce these interfering responses, yet tended to judge over 20% as "memories" of external happenings. To reiterate, having subjects generate their own responses constitutes a potent source of interference. Second, we find it interesting that the interference on the delayed test was no greater for subjects who were tested a week earlier than for those who took a single test after a week. A plausible scenario is that subjects who engaged in forced recall on the first test would be especially prone to interference a week later, and to confuse the studied pictures with their own responses generated soon after. The fact that both groups' performances on the delayed test were so similar suggests that processes operating during forced recall create the false recognitions, rather than that they carry over from a prior test. One other feature of the data in Table V is worth noting, even though it does not play a role in our present discussion: In both free and forced recall, the positive effect of taking a prior test was observed on delayed-recall performance, in terms of both the total number of correct responses produced and their correct recognition. Thus we see the power of taking a test on later retention yet again, as in much earlier work (Spitzer, 1939; Thompson et al., 1978; Wheeler & Roediger, 1992).

The main point to be gleaned from the experiment just described is that subjects' own responses provide a potent source of interference. The forced recall procedure requires subjects to confabulate; later they confuse their confabulations with items presented in the target set. One type of error is generating items that were actually in the target set, but failing to recognize them. Production of these items may constitute a type of priming akin to that observed in implicit memory studies where subjects are asked to free associate to a cue (e.g., Srinivas & Roediger, 1990); the failure to recognize such generated responses as members of the target set qualifies as an instance of the recognition failure of recallable (or at least producible) words (e.g., Watkins & Tulving, 1975). The second sort of error is the one more commonly expected in studies of reconstructive memory: Subjects generate and then recognize as old (or remember as their own experiences) events that never happened in the target set but that were instead produced as confabulations. If numerous forced recalls were given, this tendency might be expected to increase, as generated items became more familiar over time.[4] This procedure of repeated recall accompanied by confabulation, which is induced by the forced-recall technique, should prove useful in future studies of reconstructive memory. As previously remarked, it seems to blend the best aspects of Bartlett's repeated-reproduction procedure (of having subjects generate their own interfering responses) with Loftus's control over the interfering events, at least to some degree.

One might complain that the experiment described here is artificial in several ways, and indeed it is. People rarely engage in forced recall in the sense of being required to generate material they believe to be erroneous. However, we think that the experiment just described (the basic findings of which have been replicated; see Roediger et al., in preparation) actually captures the nature of reconstructive memory quite well. When people see a complex event such as a crime and then try to recall it later, they probably respond to tacit demands from others to go beyond the information that they can accurately retrieve and to tell a plausible story, just as our subjects are asked to respond with plausible candidates for the pictures in the set. Once people respond with plausible guesses, they may later recall their guesses as actual occurrences. These self-generated guesses may provide a more potent source of misinformation or interference than would information planted in a narrative, as in the standard misinformatoin paradigm. If Gauld and Stephenson's (1967) analysis of Bartlett's (1932) experiments is correct and Bartlett's subjects were instructed to create plausible renderings (rather than accurate memories) of "The War of the Ghosts," then his claims of massive reconstruction and erroneous recall are more plausible. Indeed, we would predict that after creating

[4] Of course, as noted, there was no greater tendency for false recognitions for forced-recall subjects on a second test than on a first test in this experiment. We suspect that the tendency to falsely recognize generated events will increase when they are generated many times, although this is an empirical question. Note that the difference in Table V is in the opposite direction from that predicted.

plausible stories related to the general theme, subjects would be unable, even if they tried, to accurately recall the story. Just as in our experiments with forced recall, subjects would have trouble sorting out fact from fiction after repeated retellings of plausible stories related to the target story. That such repeated callings to mind of invented "facts" can establish the inventions as memories might be illustrated in Jean Piaget's famous anecdote of his vivid memory of being narrowly saved from a kidnapping when he was a young child, which later turned out to be a fabrication of his nurse at the time. He credited the false memory to the family's repeated recollections of the event during his childhood, which he then projected into the past as an actual memory. Having subjects repeatedly tell themselves such erroneous information seems a natural way to study the reconstructive aspects of remembering. We plan to use this method in future investigations of this topic.

We end this section with a few methodological points. First, to reiterate the most obvious, we believe that the method of having subjects confabulate plausible candidates for recall on a test and then examining whether these creations are falsely recognized, and whether they will appear in later reconstructions of events, to be a good method to study the reconstructive aspects of remembering. The emphasis on recall helps broaden the arsenal of techniques used to study reconstructive aspects of memory. Various recognition procedures have been standard in the literature, but it has always been notoriously difficult to observe retroactive interference with certain recognition procedures (McCloskey & Zaragoza, 1985; Postman & Stark, 1969; Loftus, 1991).

A second implication of these results is somewhat less obvious yet just as clear: the forced recall procedure introduced by Erdelyi and Becker (1974) is flawed in regard to accurately estimating what is remembered. This is especially true on a delayed test, as the data in Table V show, but the same confusions observed there occur in lesser degree even on an immediate test (see Table IV). Forced recall was introduced as a praiseworthy way to control possible response-criterion effects by requiring subjects to produce the same number of responses across tests. The procedure does this, but the production of a correct response does not indicate that it is a product of conscious recollection. Responding is not remembering. On implicit memory tests even amnesic patients may respond with a recently studied item, but they do not remember the experience of having studied the item before (e.g., Graf, Squire, & Mandler, 1984). Jacoby (in press) points out that responding even on ostensibly explicit-memory tests, those involving conscious recollection, may involve correct response production through some automatic process (*priming*, in the jargon of implicit-memory researchers; see Roediger, 1990;

Schacter, 1987), and not through conscious recollection. Apparently, forced recall produces considerable responding of this sort relative to free recall, as the data in Tables IV and V reveal. Items are often produced that are not recognized as being from the list.

The objections to using forced recall are similar to those raised against using hypnosis to aid memory retrieval (e.g., Smith, 1983). Witnesses who are hypnotized and who then attempt to retrieve events often confabulate during their recall attempts, and then, when testifying later, may confuse the events created under hypnosis with real events. Like hypnosis, forced recall forces people to confabulate. Several reports have shown that highly hypnotizable subjects are likely to become increasingly confident that their errors represent true memories across repeated recall tests (Nogrady, McConkey, & Perry, 1985; Whitehouse, Dinges, E. C. Orne, & M. T. Orne, 1988). The experiment just reported shows that it is not necessary to have highly hypnotizable subjects or to use hypnosis to produce false recognitions of confabulated responses: The forced-recall procedure creates a similar effect. Poole and White (1991) reported similar results with repeated questioning. When asked about an event of which they had been given little information, subjects sometimes speculated about the correct answer. On a later test, their answers to the question became more certain.

The practical implication is that we should reexamine prior studies that use forced recall so that we can determine whether effects observed are based on remembering or on response production that combines both remembering and more automatic production of responses. Many basic findings obtained with forced recall, such as hypermnesia, seem quite secure, because they have been obtained in free-recall procedures during which subjects are more purely engaged in conscious recollection (at least judging by the recognition responses in Tables IV and V). However, forced recall experiments such as those of Wheeler and Roediger (1992) described above may deserve further scrutiny, but we suspect that the main conclusions are secure. If anything, correcting response production in forced recall for items actually remembered (recognized) would likely reveal greater forgetting over the week retention interval than that shown now in Table I, whereas the hypermnesia with short intervals between tests would remain intact (see the data in Table III from Roediger & Payne, 1985).

In the forced-recall procedure subjects produce responses that may be in the target set but that may not actually be remembered, and they may remember items that they produce as being from the target set when in fact they are not. In the Roediger et al. (in preparation) experiments, we used the procedure of having subjects give recognition-confidence ratings for their produced responses to ascertain their status as either products of conscious recollection or products of some more automatic process. However, subjects are capable of providing more detailed judgments than confidence ratings. In the next section we describe experiments in which subjects try to apportion their correct responses in memory tests on more qualitative bases.

## IV. Remembering and Knowing Past Events

In 1970 Tulving and Madigan remarked that theories and models of memory had neglected to incorporate "one of the truly unique characteristics of human memory: its knowledge of its own knowledge" (p. 477). The situation in the intervening 23 years has evidenced a remarkable change in this state of affairs, with the advance in the study of metamemory (see Nelson's, 1992, excellent collection of readings in this area). Investigators have pursued the study of the feeling of knowing (e.g., Nelson, 1984; Schacter, 1983), reality monitoring (Johnson & Raye, 1981), and the overconfidence of judgments (e.g., Fischhoff, 1975). In each case researchers make use of subjects' abilities (or, sometimes, inabilities) to evaluate the status of their own knowledge.

Tulving (1985) proposed that people may be able to differentiate their past experience in two ways, or that there are two ways of accessing knowledge about the personal past. When we say we remember past events, we usually mean that we can actually recollect their occurrence—we can remember vivid details, we can (in some sense) reexperience the events as we remember them. However, Tulving noted that our memories for past events can also be more impersonal: We can know, with certainty, that something happened to us but not be able to recall the event in the sense just described. Tulving (1985) reported a demonstration experiment that showed that subjects could reliably make "Remember" and "Know" judgments about their experiences; further, the more powerful the retrieval cues given during a test, the greater the preponderance of Know judgments relative to Remember judgments do subjects provide. More recently, Gardiner and his colleagues (1988; Gardiner & Java, 1990, 1993) have used the Remember/Know paradigm to separate different bases of responding in recognition memory.

Several researchers have proposed that recognition may have two separate bases for positive responses. Mandler (1979, 1980) proposed a distinction between integration of an experience as a perceptual unit (intra-item integration) and elaboration on the unit by relating it to other knowledge. These different encoding processes give rise to two bases for recognition:

familiarity-based responding (based on intraitem integration) and retrieval (based on elaborative processing). Jacoby (1983a, b; Jacoby & Dallas, 1981) proposed a similar distinction between responding on the basis of perceptual fluency (similar to familiarity in Mandler's proposal) or by conscious recollection (see Jacoby, 1991). Further, Jacoby (1983b) proposed that the fluency-based responding should be affected by data-driven variables, or those that affect basic perceptual processing (such as modality of presentation), whereas manipulation of meaning (or conceptual manipulations) affect conscious recollection. There is considerable evidence in agreement with recognition memory having two bases, although there is contrary evidence in the literature, too (Watkins & Gibson, 1988).

The distinction between experiences that are remembered and those that are known to have occurred but are not remembered maps naturally onto this distiction between two bases of responding in recognition. Briefly, Remember judgments should reflect elaborative or conceptual processing, whereas Know judgments would indicate responding on the basis of familiarity or perceptual fluency. Evidence provided by Gardiner (1988) and Gardiner and Java (1990) generally agrees with this proposal, although one problem will be noted below. The general paradigm they have used involves presenting subjects with words under various study conditions known to affect recognition memory and then giving subjects a recognition-memory test in which, for each item deemed "old" or previously studied, subjects are required to judge whether they remember the occurrence of the word in the list or rather know that it was there on some other basis. Subjects are told to judge an item as remembered if they can recollect its actual presentation in the list (e.g., they can recollect some feature of the context, or their own reaction). If they are sure it was in the list, but cannot remember its actual occurrence, then they are to provide a Know response. Although this distinction seems difficult, with proper instructions subjects provide sensible and reliable judgments. In most of Gardiner's experiments, variables that strongly affect recognition are shown to have their basis in Remember responses. That is, when the overall recognition responses are decomposed into the two components, the independent variable is shown to affect the Remember responses but to leave Know responses unaffected. This result is in accord with the idea that recognition largely reflects conscious recollection, but it also can be partly driven by perceptual fluency or familiarity that is reflected in Know responses.

One worry that emerges from the pattern of results described above is that Know responses are rather insensitive; perhaps subjects use this category only when they have low confidence that an event occurred. Gardiner and Java (1993) have argued that Remember and Know judgments are not to be equated with high- and low-confidence recognition judgments, respectively. The most compelling sort of evidence to document the claim that these judgments reflect distinct means of knowing the personal past is to show that a variable has opposite effects on the two judgments (rather than affecting one but not the other). Gardiner and Java (1990) accomplished this in one experiment by having subjects study words and nonwords and then receive a recognition memory test on old and new words and nonwords. Not surprisingly, words were recognized better than nonwords overall. However, when responding was decomposed into its components, words received more Remember responses than did nonwords, but nonwords received more Know responses than did words.

The experiments we report here are a selection from those that Rajaram (1990, 1993) reported as part of her PhD thesis. The general goal of the dissertation was to examine further factors that might affect Remember and Know judgments differently; the more specific goal was to investigate Gardiner's (1988) claim that Remember judgments are more affected by conceptual factors and that Know judgments are more driven by perceptual factors. Rajaram (1993) reasoned that factors affecting perceptual (or data-driven) implicit-memory tests should also affect Know responses. One such variable that greatly affects verbal implicit-memory tests such as word-fragment completion and word-stem completion is the symbolic form in which items are studied (i.e., items are given as words or pictures). When subjects study mixed lists of words and pictures, pictures are better recalled (e.g., Paivio, Rogers & Smythe, 1968) and better recognized (Madigan, 1983) than are words, even when the mode of responding is verbal (i.e., with subjects recalling words—the names of the pictures—or recognizing the picture names). This picture-superiority effect is generally credited to conceptual factors, either dual coding of pictures (e.g., Paivio, 1986) or greater analysis of pictures' meaning (Nelson, 1979). On the other hand, words produce much more priming than do pictures on verbal implicit-memory tests such as word-fragment completion, word-stem completion, and perceptual identification (e.g., Weldon & Roediger, 1987; Rajaram & Roediger, in press).

Given these considerations, Rajaram (1993) reasoned that the symbolic format variable might produce opposite effects on Remember and Know judgments. Specifically, overall recognition should be greater for pictures than for words (replicating Madigan, 1983, and others), but when recognition responses were decomposed into Remember and Know components, opposite effects would appear. Specifically, Remember judgments should be greater for pictures than for words, but Know judgments (driven by perceptual fluency) should be greater for words than for pictures.

Subjects studied 60 items in the first phase of Rajaram's (1993) Experiment 2, half as words and half as pictures, at the rate of 5 sec per item. After studying the list, subjects solved unrelated word fragments for 15 min prior to the test. During the testing phase subjects were given 120 words, half studied (30 as pictures, 30 as words) and half not studied. (Items were counterbalanced across studied and nonstudied conditions and, within studied conditions, across presentation as pictures or words.) The test words were given in a booklet, and subjects were instructed to decide if each item had been studied in the prior list as either a word or a picture, placing a "Y" for *yes* or an "N" for *no* next to each item. Further, if subjects placed a Y next to the item, signaling its recognition, they were further instructed to write an "R" for Remember or a "K" for Know next to the other response. The instructions for this response were modeled closely after those of Gardiner (1988). Instructions are critical in making subjects understand the distinction and those used in the experiments described here are presented in an appendix of Rajaram (1993). In brief, subjects were told to judge the recognized item as Remembered when they could recall the actual experience of its presentation. On the other hand, Know responses should be made:

> when you recognize that the word was in the study list but you cannot consciously recollect anything about its actual occurrence, or what happened or what was experienced at the time of its occurrence. In other words, write "K" (for "know") when you are certain of recognizing the words but these words fail to evoke any conscious recollection from the study list. (p. 102)

Rajaram's (1993, Experiment 2) results are presented in Table VI, where it can be seen that a picture-superiority effect occurred in recognition: The hit rate for pictures was .90 and for words was .69, with a .09 false alarm rate.[5] However, when the overall hit rate was decomposed into Remember and Know responses, then opposing patterns occurred. Remember responses in recognizing pictures greatly exceeded those in recognizing words, .81 to .51, consistent with the idea that the picture-superiority effect is conceptual in nature and therefore shows up in Remember responses. On the other hand, Know responses were greater following recognition of words (.18) than of pictures (.09), consistent with the notion that Know responses are driven by the same sort of perceptual operations that produce priming on perceptual implicit-memory tests. Interestingly,

[5] Subjects were actually recognizing words, so the match in physical features between study and test tokens was greater when subjects studied words than when they studied pictures. Despite this, pictures were recognized better than words, as reported by many others (e.g., Madigan, 1983).

TABLE VI

PROPORTIONS OF HITS AND FALSE ALARMS FOR REMEMBER AND KNOW JUDGMENTS FROM RAJARAM'S EXPERIMENT 2[a]

| | Target items | | |
|---|---|---|---|
| | Pictures | Words | Lures |
| Recognition "Yes" | .90 | .69 | .09 |
| Remember | .81 | .51 | .01 |
| Know | .09 | .18 | .08 |

[a] Data are from Rajaram (1993), Experiment 2, and are reprinted from *Memory & Cognition*, *21*, pp. 89–102, by permission of the Psychonomic Society.

false alarms in recognition were almost entirely Know responses; the subjects felt that the item was familiar, but could not remember its occurrence in the list because it had not occurred. Rajaram's data shown in Table VI strongly implicate Remember and Know judgments as independent components of recognition that are affected by different factors, in agreement with Gardiner's (1988) hypothesis.

The research reviewed thus far has revealed two patterns of dissociation between Remember and Know responses. Variables either affect Remember judgments and not Know judgments, or variables have opposite effects on the two types of responses, as in the data of Table VI. If the two bases of responding are independent, then it should be possible to find the remaining type of interaction: A variable that would affect Know responses but not Remember responses. Rajaram (1993) sought to find this sort of dissociation in a further experiment (Experiment 3). The trick was to increase the perceptual fluency of the target word without affecting the ability to recollect it consciously.

Rajaram (1993) used the masked repetition paradigm (Forster, 1985) in order to enhance perceptual fluency of the target. In this paradigm, a target item is preceded by its own masked presentation, so during a test trial subjects receive a forward mask (a series of ampersands), a very brief exposure to the target, and then the target is presented until the response is made. Subjects are unable to see the first presentation of the target, but this presentation is presumed to enhance the perceptual fluency of the target when it is presented clearly. Several previous experiments have shown that when a target is immediately preceded by its own masked presentation in the course of a recognition memory experiment, subjects produce more hits, relative to the case in which an unrelated word is the masked prime (Forster, 1985; Jacoby & Whitehouse, 1989; Rajaram & Neely, 1992). Jacoby and Whitehouse (1989) also reported that masked repetition increased the false-alarm rate, which would be expected if the prime increased perceptual fluency and this fluency made recognition more likely. Rajaram (1993) reasoned that masked repetitions should therefore affect Know responses (believed to be driven by perceptual operations) and not Remember responses.

In Rajaram's (1993) Experiment 3, subjects studied words in a list and then took a recognition test in which studied and nonstudied words were intermingled. Subjects were instructed to write Y or N to signal their recogniton of items as old or new and, in the case of items receiving a Y, to respond with Remember and Know judgments. Half the items were tested in the masked-repetition condition where the target was briefly preceded by its own copy; half were tested with an unrelated masked item preceding the target. The results are shown in Table VII. Masked repetition of targets increased both the hit rate for targets and the false-alarm rate for lures, by 7% and 5%, respectively. When Rajaram decomposed subjects' overall responding into Remember and Know components, the masked repetition variable enhanced the Know responses for both targets and lures and did not affect Remember responses. Thus, she had found the type of interaction missing from the literature: a variable that affected Know responses and not Remember responses. Further, the effect of masked repetition was predicted to affect only Know responding on the basis of Gardiner's (1988) hypothesis that perceptual operations affect such responses. In another experiment, Rajaram (1993, Experiment 4) showed that masked-repetition

TABLE VII

PROPORTIONS OF HITS AND FALSE ALARMS FOR REMEMBER AND KNOW JUDGMENTS FROM RAJARAM'S EXPERIMENT 3[a]

| | Targets | | Lures | |
|---|---|---|---|---|
| | Masked repetition | Unrelated primes | Masked repetition | Unrelated primes |
| Recognition "Yes" | .67 | .60 | .23 | .18 |
| Remember | .43 | .42 | .05 | .05 |
| Know | .24 | .18 | .18 | .13 |

[a] Data are from Rajaram (1993), Experiment 3, and are reprinted from *Memory & Cognition*, *21*, pp. 89–102, by permission of the Psychonomic Society.

priming did not have the same effect on confidence judgments as on Remember and Know judgments, further supporting the idea that Know responses are not simply low-confidence recognition responses.

The evidence we have just reviewed strongly supports the idea that Know responses are driven by the same factors that are responsible for perceptual priming. However, at least one source of evidence contradicts this idea. Rajaram (1993, Experiment 1) manipulated the modality of presentation of words (auditory or visual) and then later gave subjects a recognition test in which they were required to make Remember and Know judgments for items judged "old" in recognition. Modality of presentation consistently produces large differences in priming on perceptual implicit-memory tests (e.g., Blaxton, 1989; Rajaram & Roediger, in press), so the expectation is that modality should similarly affect recognition responses and that the difference should be localized in the Know component rather than the Remember component. However, Rajaram's (1993) results failed to verify this prediction because she found that modality had no effect on overall recognition and did not differentially affect either Remember or Know judgments. Gardiner and Java (1993) reported the same null finding. At this writing it is unclear how to reconcile the null results of modality with the contrasting results from manipulations of symbolic form (picture or word) and other variables. The preponderance of the evidence supports the idea that Know responses are driven by the same perceptual factors that produce priming on perceptual implicit-memory tests, but the clear failure of manipulations of modality to produce the expected results indicates that the story must be more complicated than the rather straightforward account provided here. In addition, the general failure of modality to affect recognition responses seems problematic for theories such as Mandler's (1980) or Jacoby's (1983a, b), which argue that perceptual factors increase fluency of responding and hence familiarity.

Remembering and Knowing are the mental states produced by two independent means of accessing the past. Knowing seems to tap some more automatic feature of memory, perhaps akin to priming on implicit-memory tests. If this is so, then perhaps Know responses can be driven by conceptual factors as well as perceptual factors, because conceptual implicit-memory tests show robust priming in memory-impaired patients as well as in normal subjects (e.g., Graf, Squire, & Mandler, 1984; Rappold & Hashtroudi, 1991; Srinivas & Roediger, 1990). Indeed, Blaxton (in press) has used the Remember/Know technique to study the mnemonic abilities of memory-impaired patients. However, as useful as the Remember/Know technique is, bear in mind that its use omits the possibility that responding may occur that is affected by past experience but in which responses may not be recognized. In the forced recall conditions of the Roediger et al. (in preparation) experiment described earlier, subjects often produced responses that had appeared in the target list (and therefore were unlikely to have been produced at random) and yet failed to recognize the responses as list members. If responses are not recognized as old, then of course the Remember/Know technique cannot be applied.

## V. Conclusions

This article has been concerned with the processes of remembering, knowing and reconstructing past events. Memory refers to all these processes, and others, but we can separate different bases and mental states in our knowledge of the past. Remembering, as used here, refers to the recollection of past events when details of these events can be retrieved; the events can be mentally reexperienced (Tulving, 1985; Gardiner, 1988). Subjects claim to have such "remembering" experiences more frequently after deep as opposed to shallow levels of processing of material, after generating rather than reading words, and following the study of pictures rather than words (Gardiner & Java, 1993; Rajaram, 1993).

Subjects can know that events happened in their past without being able to remember them. We can know that we returned to our home for a Christmas vacation a decade ago without remembering the trip. In the lab, Know responses occur more often for words than for pictures and nonwords relative to words (Rajaram, 1993; Gardiner & Java, 1990). Knowing the past, in this sense, represents an interesting challenge for theory. We know something about an episode in our lives, but the knowledge seems impersonal, like the knowledge of any fact from semantic memory, because we cannot recollect the detailed occurrences that would provide the basis for true remembering. Our knowledge of the trip home is like our knowledge that William McKinley was President of the United States; both memories refer to true occurrences, but neither carries the features—the remembrance of details, the warm feeling of reexperiencing the event—that are the signature of remembering.

Reconstruction of the events of our lives involves our weaving together episodes and facts that we remember and those that we know into a plausible story of our past. In the process we may add details that did not actually occur. We described the new method of studying reconstructive memory by having subjects engage in forced recall during a memory test. In forced recall, subjects try to remember the events as well as possible, but then must also confabulate responses. The interest centers on subjects' tendencies to recognize their confabulated responses as actual memories. We used confidence ratings to estimate their subjective experience of remembering, but it would be interesting to ask subjects on a delayed forced-recall test to provide Remember and Know responses for the information they claim to recognize as having occurred (when in fact it did not). Would subjects claim to know that the events occurred in the target set but not to be able to remember them, as they do to false alarms in recognition tests? Or would subjects exhibit a propensity to say that they actually remember the occurrence of events which, in fact, they had only produced themselves? This latter category of response would be the clearest indication of reconstructive memory: Subjects claim to remember the experiencing of an event when in fact they produced it themselves on an earlier test. The methods described here—the forced-recall technique, making Remember and Know judgments—permit a new attack on the problem of reconstructive remembering, one that is just beginning.

## Acknowledgments

Some of the research reported here was supported by Grants 91-0253 and F49620-92-J-0437 from the Air Force Office of Scientific Research. The authors are grateful to Brad Challis, Beth Loftus, Doug Medin, and Mary Sue Weldon for commentary on an earlier version of the article and to Ulric Neisser for helpful correspondence.

## References

Bahrick, H. P., Bahrick, P. O., & Wittlinger, R. P. (1975). Fifty years of memory for names and faces: A cross-sectional approach. *Journal of Experimental Psychology: General, 104*, 54–75.

Bahrick, H. P., & Hall, L. K. (1991). Preventive and corrective maintenance of access to knowledge. *Applied Cognitive Psychology, 5*, 1–18.

Bahrick, H. P., & Hall, L. K. (1993). Long intervals between tests can yield hypermnesia: Comments on Wheeler and Roediger. *Psychological Science, 4*, 206–207.

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

Barnes, J. M., & Underwood, B. J. (1959). "Fate" of first-list associations in transfer theory. *Journal of Experimental Psychology, 58*, 97–105.

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

Blaxton, T. A. (1989). Investigating dissociations among memory measures: Support for the transfer-appropriate processing framework. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15*, 657–668.

Blaxton, T. A. (in press). The role of the temporal lobes in recognizing nonverbal materials: Remembering versus knowing. *Neuropsychologia*.

Bousfield, W. A., & Rosner, S. R. (1970). Free vs. uninhibited recall. *Psychonomic Science, 20*, 75–76.

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

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

Ceci, S. J., Ross, D. F., & Toglia, M. P. (1987). Suggestibility of children's memory: Psychological implications. *Journal of Experimental Psychology: General, 116*, 38–49.

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

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

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

Erdelyi, M. H., 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 effects of response bias on recall performance, with some observations on processing bias. *Journal of Experimental Psychology: General, 118*, 245–254.

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

Estes, W. K. (1960). Learning theory and the new "mental chemistry." *Psychological Review, 67*, 207–233.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1*, 288–299.

Forster, K. I. (1985). Lexical acquisition and the modular lexicon. *Language and Cognitive Processes, 2*, 87–108.

Gardiner, J. M. (1988). Functional aspects of recollective experience. *Memory & Cognition, 16*, 309–313.

Gardiner, J. M., & Java, R. I. (1990). Recollective experience in word and nonword recognition. *Memory & Cognition, 18*, 23–30.

Gardiner, J. M., & Java, R. (1993). Recognizing and remembering. In A. Collins, S. Gathercole, & P. Morris (Eds.), *Theories of memory* (pp. 168–188). Hillsdale, NJ: Erlbaum.

Gauld, A., & Stephenson, G. M. (1967). Some experiments related 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*, 392–399.

Graf, P., Squire, L. R., & Mandler, G. (1984). The information that amnesic patients do not forget. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10*, 164–178.

Hastie, R., Landsman, R., & Loftus, E. F. (1978). Eyewitness testimony: The dangers of guessing. *Jurimetrics Journal, 19*, 1–8.

Hermann, D. J., Buschke, H., & Gall, M. B. (1987). Improving retrieval. *Applied Cognitive Psychology, 1*, 27–33.

Hogan, R. M., & Kintsch, W. (1971). Differential effects of study and test trials on long-term retention and recall. *Journal of Verbal Learning and Verbal Behavior, 10*, 562–567.

Izawa, C. (1971). The test trial potentiating model. *Journal of Mathematical Psychology, 8*, 200–224.

Jacoby, L. L. (1983a). Perceptual enhancement: Persistent effects of an experience. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 21–38.
Jacoby, L. L. (1983b). Remembering the data: Analyzing interactive processing in reading. *Journal of Verbal Learning and Verbal Behavior, 22,* 485–508.
Jacoby, L. L. (1991). A process dissociation framework: Separating automatic and intentional uses of memory. *Journal of Memory and Language, 30,* 513–541.
Jacoby, L. L. (in press). Measuring recollection: Strategic vs. automatic influences of associative context. In C. Umilta & M. Moscovitch (Eds.), *Attention and Performance.* Vol. 15. Cambridge, MA: Bradford.
Jacoby, L. L., & Dallas, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General, 110,* 306–340.
Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger & F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Erlbaum.
Jacoby, L. L., & Whitehouse, K. (1989). An illusion of memory: False recognition influenced by conscious perception. *Journal of Experimental Psychology: General, 118,* 126–135.
Johnson, M. K., Bransford, J. D., & Solomon, S. K. (1973). Memory for tacit implications of sentences. *Journal of Experimental Psychology, 98,* 203–205.
Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88,* 67–85.
Kintsch, W. (1970). *Learning, memory, and conceptual processes.* New York: Wiley.
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.
Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1077–1083.
Loftus, E. F. (1979). *Eyewitness Testimony.* Cambridge, MA: Harvard University Press.
Loftus, E. F. (1991). Made in memory: Distortions in recollection after misleading information. In G. H. Bower (Ed.), *The psychology of learning and motivation.* (Vol. 27, pp. 187–215). New York: Academic Press.
Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior, 13,* 585–589.
Madigan, S. (1983). Picture memory. In J. C. Yuille (Ed.), *Imagery, memory, and cognition: Essays in honour of Allan Paivio* (pp. 65–89). Hillsdale, NJ: Erlbaum.
Mandler, G. (1979). Organization and repetition: Organization principles with special reference to rote learning. In L. Nilsson (Ed.), *Perspectives on memory research* (pp. 293–327). Hillsdale, NJ: Erlbaum.
Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87,* 252–271.
Mandler, J. M., & Johnson, N. S. (1977). Remembrance of things parsed: Story structure and recall. *Cognitive Psychology, 9,* 111–151.
McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology: General, 114,* 1–16.
Melton, A. W., & Irwin, J. M. (1940). The influence of degree of interpolated learning in retroactive inhibition and the overt transfer of specific responses. *American Journal of Psychology, 53,* 173–203.
Nelson, D. L. (1979). Remembering pictures and words: Appearance, significance, and name. In L. S. Cermak & F. I. M. Craik (Eds.), *Levels of processing in human memory* (pp. 45–76). Hillsdale, NJ: Erlbaum.

Nelson, T. O. (1984). A comparison of current measures of the accuracy of feeling-of-knowing predictions. *Psychological Bulletin, 95,* 109–133.

Nelson, T. O. (Ed.). (1992). *Metacognition: Core readings.* Needham Heights, MA: Allyn and Bacon.

Nogrady, H., McConkey, K. M., & Perry, C. (1985). Enhancing visual memory: Trying hypnosis, trying imagination, and trying again. *Journal of Abnormal Psychology, 94,* 195–204.

Paivio, A. (1986). *Mental representations: A dual coding approach.* New York: Oxford University Press.

Paivio, A., Rogers, T. B., & Smythe, P. C. (1968). Why are pictures easier to recall than words? *Psychonomic Science, 11,* 137–138.

Paul, I. H. (1959). Studies in remembering: The reproduction of connected and extended verbal material. *Psychological Issues, 1* (Monograph 2).

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

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

Poole, D. A., & White, L. T. (1991). Effect of question repetition on the eyewitness testimony of children and adults. *Developmental Psychology, 27,* 975–986.

Postman, L., & Stark, K. (1969). The role of response availability in transfer and interference. *Journal of Experimental Psychology, 79,* 168–177.

Rajaram, S. (1990). *Components of recollective experience: Remembering and knowing.* Unpublished doctoral dissertation, Rice University, Houston.

Rajaram, S. (1993). Remembering and knowing: Two means of access to the personal past. *Memory & Cognition, 21,* 89–102.

Rajaram, S., & Neely, J. H. (1992). Dissociative masked repetition priming and word frequency effects in lexical decision and episodic recognition tasks. *Journal of Memory and Language, 31,* 152–182.

Rajaram, S., & Roediger, H. L. (in press). Direct comparison of four implicit memory tests. *Journal of Experimental Psychology: Learning, Memory, and Cognition.*

Rappold, V., & Hashtroudi, S. (1991). Does organization improve priming? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17,* 103–114.

Roediger, H. L. (1990). Implicit memory: Retention without remembering. *American Psychologist, 45,* 1043–1056.

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

Roediger, H. L., Challis, B. H., & Wheeler, M. A. (in preparation). Effects of confabulation on later recall.

Roediger, H. L., & Payne, D. G. (1985). Recall criterion does not affect recall level or hypermnesia: A puzzle for generate/recognize theories. *Memory & Cognition, 13,* 1–7.

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

Roediger, H. L., Srinivas, K., & Waddill, P. (1989). How much does guessing influence recall? Comment on Erdelyi, Finks, and Feigin-Pfau. *Journal of Experimental Psychology: General, 118,* 255–257.

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

Roediger, H. L., & Wheeler, M. A. (1993). Hypermnesia in episodic and semantic memory: Response to Bahrick and Hall. *Psychological Science, 4,* 207–208.

Schacter, D. L. (1983). Feeling-of-knowing in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 39–54.

Schacter, D. L. (1987). Implicit memory: History and current status. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 501–518.

Scrivner, E., & Safer, M. A. (1988). Eyewitnesses show hypermnesia for details about a violent event. *Journal of Applied Psychology, 73,* 371–377.

Smith, M. C. (1983). Hypnotic memory enhancement of witnesses: Does it work? *Psychological Bulletin, 94,* 387–407.

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

Squire, L. R. (1989). On the course of forgetting in very long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 241–245.

Squire, L. R., Haist, F., & Shimamura, A. P. (1989). The neurology of memory: Quantitative assessment of retrograde amnesia in two groups of amnesic patients. *The Journal of Neuroscience, 9,* 828–839.

Srinivas, K., & Roediger, H. L. (1990). Testing the nature of two implicit tests: Dissociations between conceptually-driven and data-driven processes. *Journal of Memory and Language, 28,* 389–412.

Thompson, C. P., Wenger, S. K., & Bartling, C. A. (1978). How recall facilitates subsequent recall: A reappraisal. *Journal of Experimental Psychology: Human Learning and Memory, 4,* 210–221.

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

Tulving, E. (1985). Memory and consciousness. *Canadian Psychologist, 26,* 1–12.

Tulving, E., & Madigan, S. A. (1970). Memory and verbal learning. *Annual Review of Psychology, 21,* 437–484.

Watkins, M. J., & Gibson, J. M. (1988). On the relation between perceptual priming and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 477–483.

Watkins, M. J., & Tulving, E. (1975). Episodic memory: When recognition fails. *Journal of Experimental Psychology: General, 104,* 5–29.

Weldon, M. S., & Roediger, H. L. (1987). Altering retrieval demands reverses the picture superiority effect. *Memory & Cognition, 15,* 269–280.

Wheeler, M. A., & Roediger, H. L. (1992). Disparate effects of repeated testing: Reconciling Ballard's (1913) and Bartlett's (1932) results. *Psychological Science, 3,* 240–245.

Whitehouse, W. G., Dinges, D. F., Orne, E. C., & Orne, M. T. (1988). Hypnotic hypermnesia: Enhanced memory accessibility or report bias? *Journal of Abnormal Psychology, 97,* 289–295.

Zaragoza, M., & Koshmider, J. W. (1989). Misled subjects may know more than their performance implies, *Journal of Experimental Psychology: Learning, Memory, and Cognition, 15,* 246–255.

Zaragoza, M., McCloskey, M., & Jamis, M. (1987). Misleading postevent information and recall of the original event: Further evidence against the memory impairment hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 13,* 36–44.