Journal of Experimental Psychology:
Learning, Memory, and Cognition
1989, Vol. 15, No. 2, 241-245

On the Course of Forgetting in Very Long-Term Memory

Larry R. Squire
Veterans Administration Medical Center, San Diego, and Department of Psychiatry, University of California School of Medicine, San Diego, California

For practical reasons, little information is available about memory across very long learning-retention intervals. To determine the time course of forgetting in very long-term memory, a memory test was constructed that sampled events (former one-season television programs) that had occurred during a single year from 1 to 15 years ago. Updated versions of the test were administered each year for 9 consecutive years, and a forgetting curve was then calculated by superimposing the results from the nine tests. The best fitting function to the curve was monotonic and nonlinear. The findings show that forgetting in very long-term memory can be gradual and continuous for many years after learning.

Since the classic work of Ebbinghaus (1885/1964), experimental studies of memory have sought to characterize the lawful properties of learning and forgetting. For practical reasons, learning and retention sessions are usually separated by a rather short interval, ranging from seconds to several weeks and up to a year or two at most. Forgetting functions derived from such studies are nonlinear and negatively accelerated (Wickelgren, 1972, 1974). Very little information is available about the course of forgetting across longer retention intervals. Tests of remote memory have been developed (Albert, Butters, & Levin, 1979; Squire, 1974; Warrington & Silberstein, 1970), but these methods are limited by the difficulty of comparing results across time periods. For any time period it is difficult to know whether the score reflects the age of the memory or the difficulty of the questions.

Using cross-sectional methods designed to overcome some of these problems, Bahrick and his colleagues tested different groups of subjects from 1 to 50 years after learning. In one instance, memory was assessed for the names and faces of former high school classmates (Bahrick, Bahrick, & Wittlinger, 1975), and in another instance, memory was assessed for the Spanish learned in school (Bahrick, 1984a). After corrections for the effects of differential original learning and rehearsal, forgetting functions either scarcely declined for 25 or more years after learning, or they declined within a few years to a stable asymptote, after which retention remained unchanged for up to 30 years. In a third study involving memory for the names and faces of former students, relatively little memory survived beyond 8 years (Bahrick, 1984b). These studies therefore raised the question of whether gradual forgetting in long-term memory ever extends beyond a few years after learning. That is, maintenance of knowledge in very long-term memory could depend on quickly reaching a stable asymptote from which no further decline occurs.

Somewhat different conclusions have been reached in studies of forgetting in autobiographical memory (see Crovitz & Schiffman, 1974; Rubin, 1982). These studies calculated retention functions by determining the frequency with which memories of different ages were recalled in response to a set of cue words. The derived functions were monotonic and nonlinear across a 20-year period, that is, fewer memories were recalled as a function of the time elapsed since the remembered event occurred. Such results raise the possibility that forgetting can be continuous across many years, but these methods for calculating retention functions are relatively indirect and difficult to compare with the findings from objective tests.

The present article uses a different method to describe forgetting in very long-term memory. In 1973 a test was developed that sampled events (former one-season television programs) that had occurred during a single year from 1 to 15 years ago (Squire & Fox, 1980; Squire & Slater, 1975). Updated versions of this test were prepared annually since that time. In the years immediately following 1973, the test was administered only occasionally. Subsequently, it became apparent that averaged data derived from multiple, separate administrations of this test should provide an improved method for assessing remote memory. Accordingly, annual and systematic testing was initiated in 1978 and was continued each year for 9 consecutive years (1978–1986). Each annual test sampled programs that had been broadcast from 1 to 15 years ago. Finally, after making relatively minor adjustments for the effects of differential exposure to the programs, a forgetting curve was calculated by superimposing the results from the nine tests.

Method

Subjects

For 9 consecutive years (1978–1986), a test based on former one-season television programs was administered annually to different groups of 20 to 32 subjects (total = 231), who were volunteers or employees at the San Diego Veterans Administration Medical Center. Testing occurred during the months of November and December. The nine groups were about equally divided between male and female subjects (46% male). They were 22 to 70 years of age (overall M = 41.4 years, SD = 11.5), and the average age of the nine separate

This research was supported by the Medical Research Service of the Veterans Administration, the National Institute of Mental Health, and the Office of Naval Research.

I thank K. Rivero-Frink, P. Clopton, J. Zouzounis, A. Shimamura, A. Bernheim, K. Edwards, and the students of St. James Academy for their assistance.

Correspondence concerning this article should be addressed to Larry R. Squire, Veterans Administration Medical Center (V-116A), 3350 La Jolla Village Drive, San Diego, California 92161.

subject groups ranged from 37.5 to 47.2 years. The nine groups had an average of 14.0 to 14.8 years of education (overall $M = 14.4$). All 231 subjects reported having a television in their household during the 15 years sampled by the test. They had lived in San Diego for an average of 14.3 years.

Materials and Procedure

The television test was multiple-choice with four alternatives and asked subjects to recognize the names of one-season television programs that had been broadcast nationally between 7 p.m. and 11 p.m. during only 1 year within the past 15 years. For each item on the test, subjects were asked to decide "which of the following was a T.V. show?" and to circle their choice. The correct answer was a former program, and the three incorrect answers were taken randomly from a pool of fabricated titles (Table 1). The order of the items on the test was random with respect to the year that a program was broadcast. All subjects were tested individually.

The test was originally constructed in 1973 (Squire & Slater, 1975). Each November since 1973, the test was updated by identifying all programs that began in the previous autumn and that did not reappear in the following autumn season (10 to 18 programs/year). Program names that were also the names of real persons, films, or plays were excluded. The list of candidate programs was circulated each year to raters (mean group size = 29, range = 19–33; mean age = 39.6 years), and the five programs most often identified as familiar were added to the test. The raters were also employees and volunteers at the San Diego Veterans Administration Medical Center, but they were not the same subjects who took the annual test in that year.

Using this method, 115 programs were selected from the 23 years 1963 through 1985. Five programs were available from each year, and programs remained on the test until they were more than 15 years old. In other words, the 1963 programs became ineligible after the 1978 test was given, the 1964 programs became ineligible after the 1979 test was given, and so on. Each annual test consisted initially of 75 different programs (5 programs × 15 years). However, 10 programs were eventually discarded because they returned to regular programming in a subsequent year. As a result, the exact number of programs on any annual test varied slightly depending on how many of the 10 discarded programs had been broadcast during the 15-year period covered by that test. The average number of programs on the nine annual tests was 70.

Selecting television programs for a memory test using the methods just described avoids many sources of sampling bias. However, in order to compare one time period with another, one must also ask whether popular exposure to the programs selected for each time period was similar or different. First, the number of weeks that each program was broadcast was determined by consulting back issues of *T.V. Guide*. The 105 one-season programs that appeared on the nine annual tests had been broadcast weekly for an average of 27.3 weeks. However, decisions to cancel programs were made sooner in recent years than in earlier years. Thus, the 44 programs contributing to the score for 1-year-old programs had been broadcast for an average of 22.5 weeks, whereas the 37 programs contributing to the score for the 15-year-old programs had been broadcast for an average of 35.5 weeks. Accordingly, the number of weeks of exposure was used as a covariate when calculating the effect on recognition memory of how many years ago a program had been broadcast.

Table 1
Sample Questions from the Remote Memory Test

Year	Question
1974	Which of the following was a T.V. show? (a) Mandrake, (b) Shipmates (c) Private Nelson, (d) *Lucas Tanner*
1978	Which of the following was a T.V. show? (a) Gaslight Alley, (b) Cutting Corners (c) Black Knight, (d) *Kaz*
1981	Which of the following was a T.V. show? (a) Dateline Miami, (b) The Conductor (c) Discovery, (d) *McClaine's Law*
1985	Which of the following was a T.V. show? (a) Dog Talk, (b) *Our Family Honor*, (c) In Danger's Wake, (d) The Ronsard File

During the weeks that programs remained on the air, the audience size was similar for programs selected from the different time periods. Data pertinent to this issue were kindly provided by the A.C. Nielsen Company. First, the percentage of U.S. households having a television set was 92% in 1963, 97% in 1973, and 98% since 1977. Second, the percentage of U.S. adults viewing evening television has not varied by more than 3% in any year from 1965 to 1986. (Comparable data were not available prior to 1965.) In households with a television set, adults over the age of 18 watched 9.6 hours per week of evening television (8 p.m. to 11 p.m.) during the period 1965 to 1986. That is, from 8 p.m. to 11 p.m., 45.7% of adults in television households were watching television during any given minute.

Individual Nielsen ratings were available for all but 3 of the 105 programs that appeared on the nine annual tests. (A program's Nielsen rating represents the percentage of U.S. households with a television set that were viewing the program. The Nielsen rating for a program is calculated by sampling television viewing of that program between September and January of each year. By knowing the average number of persons per household that were viewing the program, one can estimate the national audience size. In 1986, a Nielsen rating of 15.0 reflected an estimated audience of 22,300,000.) Ratings were similar across the 15 years sampled by the test. For example, the 44 programs contributing to the score for 1-year-old programs had an average Nielsen rating of 14.4, and the 37 programs contributing to the score for 15-year-old programs had an average Nielsen rating of 14.7. The overall average for all 15 sets of programs was 15.1 (range = 14.4–15.5).

Results

To begin the data analysis, an average percentage correct score was calculated for each of the previous 15 years by combining the appropriate scores from all nine tests. For example, an overall score was obtained for all the programs that were 1-year-old at the time of testing by averaging the scores for the 1985 programs that appeared on the 1986 test, the 1984 programs that appeared on the 1985 test, and so on, ending with the 1977 programs that appeared on the 1978 test. To construct a forgetting curve, these 15 scores were then adjusted for the effects of differential exposure across years, as described below.

To determine the effect on recognition memory of how many years ago a program was broadcast, the 630 scores (resulting from having tested an average of 70 programs in each of 9 years) were submitted to an analysis of covariance, using as covariates Nielsen popularity ratings and the number of weeks each program was broadcast. The independent variable, number of years since a program was broadcast, had 15 levels. The predictive power of the covariates was also assessed. The effect on recognition memory of how many years ago a program was broadcast was highly significant, $F(14, 583) = 3.60$, $p < .0001$, as was the effect of the weeks-of-exposure covariate, $F(1, 583) = 23.1$, $p < .0001$. Nielsen

ratings, however, provided no unique information about the recognition-memory scores, $F(1, 583) = 0.27$, $p > .10$. Consequently, the Nielsen ratings were discarded from the analysis, and the percentage correct scores for each program were adjusted for weeks of exposure.

These adjusted scores were then used to calculate a forgetting curve (Figure 1A). The corrected forgetting curve was similar to the uncorrected curve, differing by less than 4% at each of the 15 time points (mean of absolute values = 1.8%).

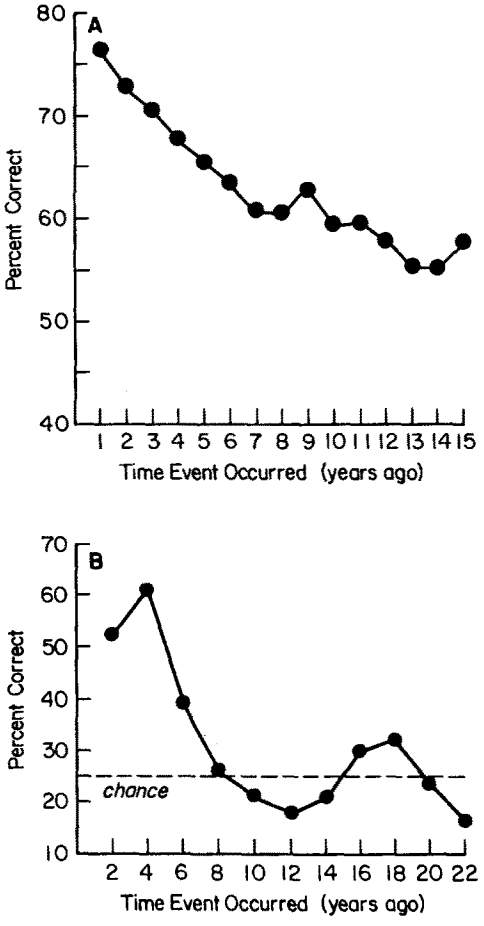

Figure 1. (A) Each year for 9 consecutive years (1978–1986) a different group of subjects was given a memory test involving former, one-season television programs. The raw data were then adjusted to correct for the total public exposure of each program on the test, and the results of the 9 test administrations were superimposed, so as to obtain an average measure of memory for events that occurred 1 to 15 years ago. For example, the data point for events that occurred one year ago shows percent recognition of programs that were broadcast during each of nine years 1977 through 1985, as measured one year later (1978 through 1986). (B) Junior high school students were tested for their knowledge of all the programs that had previously appeared on the annual tests. The students could recognize the names of programs that broadcast during the previous 6 years, but they performed close to 25% on questions about programs that broadcast before they were 7 years old. These data show that the incorrect choices on this multiple-choice test were effective foils for the correct choice. Chance performance can therefore be taken as 25% when the test is given to adults (A above).

The best fitting function was monotonic and nonlinear. Either an exponential or a power function can account for more than 94% of the variance in the 15 data points, whereas a linear function accounts for 88.3% of the variance.

How many years ago a program had been broadcast accounted for 8.5% of the total variance in the recognizability of individual program titles, and the number of weeks of exposure accounted for 5.3%. Additional variance is presumably attributable to the wording of the program titles and to remembered associations with these titles. This finding underscores the importance of estimating the forgetting function from the results of multiple tests given in different years. In this way individual test items contribute to the score for many different years, and the idiosyncratic influence of individual test items tends to be averaged out.

To explore further the shape of the forgetting function in Figure 1A, a separate analysis was carried out for the time period 6 to 15 years ago. The recognition-memory scores exhibited a significant linear trend in this portion of the forgetting function, $F(1, 398) = 4.71$, $p < .05$. When the same analysis was repeated for the time period 7 to 15 years ago, the test for linear trend fell short of significance, $F(1, 356) = 2.68$, $p = .10$, as it did for the time period 8 to 15 years ago, $F(1, 313) = 2.41$, $p = .12$. These results are consistent with the suggestion that forgetting is gradual and continuous across 15 years, but a negatively accelerated forgetting function is difficult to distinguish unambiguously from a function that reaches a plateau at some point within 15 years.

Because the subjects who participated in the nine tests varied widely in age, three separate forgetting functions were also examined, which were constructed from the scores obtained by the younger third ($M = 30.3$ years old), the middle third ($M = 39.8$ years old), and the older third ($M = 54.1$ years old) of each year's sample. There were a total of 77 subjects in each of these three groups. The forgetting curves for the two older groups were parallel. For the older group, the adjusted recognition score was 66.7% correct for the most recent 5 years covered by the test, 59.9% for the time period 6 to 10 years ago, and 55.9% for the period 11 to 15 years ago. For the middle group, the corresponding scores were 69.6%, 63.1%, and 55.8%. The younger group differed from the other two groups in that the forgetting function was steeper (74.4%, 62.8%, and 55.8%). To determine whether forgetting continued in each age group throughout the 15 years covered by the test, separate analyses were carried out in each case for the time period 6 to 15 years ago. For the younger subjects, the recognition-memory scores exhibited a significant linear trend in this portion of the forgetting function, $F(1, 398) = 8.2$, $p < .01$. The same test for linear trend approached significance for the middle group of subjects, $F(1, 398) = 3.04$, $p = .08$, and was not significant for the older subjects, $F(1, 398) = 1.0$, $p > .1$.

Some additional evidence that also bears on the shape of the forgetting function came from other test data, based on programs that were broadcast more than 15 years ago. In 1986, 29 subjects (mean age = 51.1 years, mean education = 13.7 years) were given all the test items for the years 1963 to 1969. The average raw score for these programs, which on average were broadcast 20 years earlier, was 52.7%, and the adjusted score was 47.7%. Note that in Figure 1A, the average

adjusted score for 11- to 15-year-old programs was higher (57.1%). Also, subjects in the main study, who were tested in 1978, obtained an adjusted score of 53.7% for the 1963 to 1969 test items. Thus, in 1978, when the 1963 to 1969 programs were 9 to 15 years old, subjects obtained an adjusted recognition score of 53.7%. In 1986, when the same programs were 17 to 23 years old, the adjusted recognition score was 47.7%. Although the differences are small, these additional findings support the idea that forgetting is continuous.

Some programs (32 out of 105) reappeared on television for short runs during the middle of one or more subsequent years after the year in which they first appeared. These brief reappearances did not measurably affect recognition memory. The average recognition-memory score for a program was about the same in the test year that immediately preceded its first reappearance ($M = 4$ broadcasts) as in the following year (64.0% vs. 65.4%). Moreover, deleting these 32 programs from the data analysis altogether did not affect the shape of the calculated forgetting curve. The average change in the 15 data points of Figure 1A was 1.3%.

An important question concerns to what extent the forgetting function in Figure 1A depends on having averaged scores across many subjects and many items. To ask whether the data collected during the nine annual test administrations reflected the contributions of similar groups of subjects on each occasion, 15 scores for each of the 231 subjects (i.e., each subject's average score for each year sampled by the test) were submitted to an analysis of variance. Whereas the effect of how many years ago the programs were broadcast was highly significant, $F(14, 3108) = 13.3$, $p < .0001$, the effect of the year in which the test was given was not significant, $F(8, 222) = 1.78$, $p > .05$. Thus, the overall recognition score obtained from each of the nine annual tests was similar. A separate analysis showed that the shapes of the nine separate forgetting functions were also similar (differences in linear and log trends, Fs < 1.0, ps $> .10$).

The distribution of scores around each of the 15 means was also examined. The standard deviations of each distribution were similar, ranging from 24.2% to 28.3%. All the distributions appeared continuous and roughly symmetrical around the mean, except for the distributions of 1-year-old and 2-year-old memories. These were positively skewed, with 36.3% of subjects obtaining a perfect (100%) score on the items that assessed memory for programs that were broadcast 1 year ago and 29.9% of subjects obtaining a perfect score on items that assessed memory for programs that were broadcast 2 years ago. It therefore seems likely that were it not for this ceiling imposed on the scores, the forgetting function would have been a little steeper during the first year or two than the averages suggest.

Because the incorrect titles were fabricated, one cannot assume that 25% represents chance performance on this four-alternative, multiple-choice test. If test takers could eliminate some of the fabricated titles because they did not appear credible, chance performance could be much higher than 25%. To determine chance performance experimentally, junior high school students were tested in early 1986 for their knowledge of all 100 programs that had ever appeared on the test up to that time (100 programs from the years 1963 through 1984). Each test item was identical to an item also given to adult subjects, that is, a correct choice was always matched to the same three incorrect choices. The test was administered in two parts, covering alternating years, to two groups of 19 students (mean age = 13.5 years; 56% male). Figure 1B shows the raw data. The students scored close to 25% for questions about programs that were broadcast when they were younger than 7 years of age (means of these data points were all within 0.6 standard deviations of 25%). Accordingly, if one assumes that junior high school students are similar to adults in their ability to judge the plausibility of television program titles, a score of 25% can be taken as chance on this test. A score of 50% to 60% correct, which the adult subjects obtained for the older programs on the test, would therefore represent considerable recognition memory. Also, because students scored close to chance on questions about the older programs, there must be limited opportunity to learn about them in the years after they appear.

Another way to ask whether the curve in Figure 1A fairly describes forgetting is to examine the recognition-memory scores obtained for each year's programs, noting how the scores changed as the same programs were tested repeatedly in successive years. Figure 2 shows all 135 means (average scores for the programs from each of 9 test years × 15 years of programs per test) as a function of years since the programs were broadcast. The scores decreased as the interval increased between when programs were broadcast and when they were tested. In other words, subjects who were asked about the programs from a recent year recognized those programs better than subjects who were asked about the same programs a few years later. The only reasonable conclusion would seem to be that memory for the individual items on the test, that is, memory for individual program names, fades as time passes. The function in Figure 1A appears to be a reasonable description of this process across 15 years.

Discussion

The names of one-season television programs are learned during a period of many weeks and are then forgotten gradually and continuously for many years in a negatively accelerated fashion. Although considerable loss from memory occurs, extrapolations based on best fitting nonlinear functions suggest that a considerable amount of this information may persist for a lifetime. The best fitting functions are exponential and power functions, the latter having been used previously to describe long-term forgetting functions for laboratory memory data and autobiographical memory data (Rubin, 1982; Wickelgren, 1972). Although rehearsal is certainly one mechanism for maintaining information in memory, it seems unlikely that information about failed one-season television programs was rehearsed to a significant extent in the years after the programs appeared. The durability of many memories may depend, not on rehearsal, but on a sufficient level of original learning coupled with gradual forgetting over the years.

Previous work has suggested that forgetting functions in very long-term memory either decline toward zero rather quickly (Bahrick, 1984b) or reach a stable asymptote within

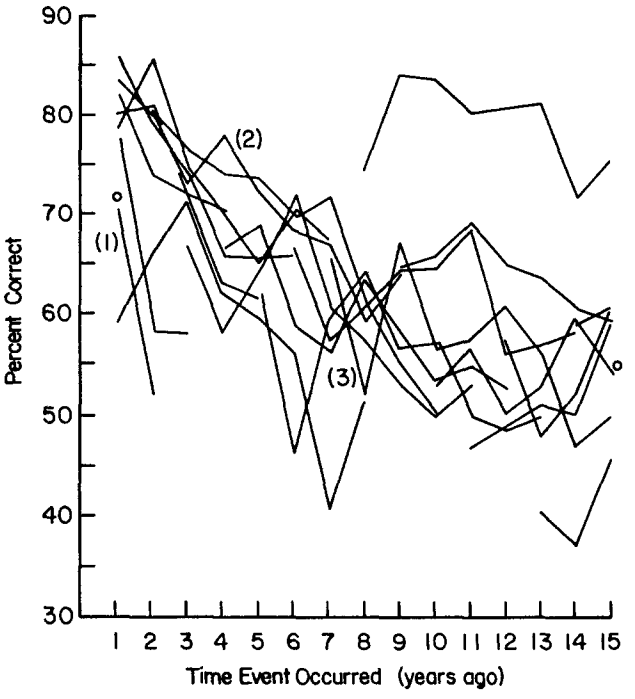

Figure 2. Average corrected recognition scores for the names of T.V. programs, which broadcast during only one year (1963 through 1985). The separate lines show how recognition memory of each group of annual programs changed across the years that these programs were tested. For example, [1] shows the scores for the 1984 programs, which were tested twice, in 1985 and 1986, when they were 1 and 2 years old, respectively; [2] shows the scores for the 1977 programs, which were tested 9 times, in 1978 to 1986, when the programs were 1 to 9 years old; [3] shows the scores for the 1971 programs, which were also tested 9 times, 1978 to 1986, when they were 7 to 15 years old. The single points show scores for the 1985 programs, which were tested only once (in 1986), and the 1963 programs which were also tested only once (in 1978). The 1972 programs (the atypical line in upper right) were well recognized each time they were tested. Averaging all the points in each time period yields the function in Fig. 1A. These data show that memory for individual items on the test is fading as time passes.

a few years after learning (Bahrick, 1984a; Bahrick et al., 1975). The present study appears to be the first to establish the intermediate case; namely, that forgetting can sometimes continue gradually for many years. Additional studies with other kinds of materials will be needed to determine the generality of these findings. In instances where retention was found to be stable across 25 years or more, the information was either strongly learned and acquired during several years of repeated exposure (Bahrick et al., 1975), or the material had a high degree of internal organization (e.g., a foreign language; Bahrick, 1984a). Thus, some kinds of material might be forgotten as Bahrick described, but this description may not be a general one for all long-term memory. Indeed, the gradual and continuous forgetting function observed here may be more typical, insofar as remembered information involves different facts, names, and events to which one has limited exposure.

References

Albert, M. S., Butters, N., & Levin, J. (1979). Temporal gradients in the retrograde amnesia of patients with alcoholic Korsakoff's disease. *Archives of Neurology, 36*, 211–216.

Bahrick, H. P. (1984a). Memory and people. In J. Harris (Ed.), *Everyday memory, actions, and absent-mindedness* (pp. 19–34). New York: Academic Press.

Bahrick, H. P. (1984b). Semantic memory content in permastore: "Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General, 113*, 1–47.

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

Crovitz, H. F., & Schiffman, H. (1974). Frequency of episodic memories as a function of their age. *Bulletin of the Psychonomic Society, 4*, 517–518.

Ebbinghaus, H. (1964). *On memory* (H. A. Ruger & C. E. Bussenius, Trans.). New York: Dover. (Original work published 1885)

Rubin, D. C. (1982). On the retention function for autobiographical memory. *Journal of Verbal Learning and Verbal Behavior, 21*, 21–38.

Squire, L. R. (1974). Remote memory as affected by aging. *Neuropsychologia, 12*, 429–435.

Squire, L. R., & Fox, M. M. (1980). Assessment of remote memory: Validation of the television test by repeated testing during a seven-year period. *Behavior Research Methods and Instrumentation, 12*, 583–586.

Squire, L. R., & Slater, P. C. (1975). Forgetting in very long-term memory as assessed by an improved questionnaire technique. *Journal of Experimental Psychology: Human Learning & Memory, 104*, 50–54.

Warrington, E. K., & Silberstein, M. S. (1970). The fate of old memories. *Quarterly Journal of Experimental Psychology, 22*, 508–512.

Wickelgren, W. A. (1972). Trace resistance and the decay of long-term memory. *Journal of Mathematical Psychology, 9*, 418–455.

Wickelgren, W. A. (1974). Single-trace fragility theory of memory dynamics. *Memory & Cognition, 2*, 775–780.

Received January 4, 1988
Revision received May 9, 1988
Accepted June 15, 1988 ■