Can Age × Learnability Interactions Explain the Development of Forgetting?

C. J. Brainerd and V. F. Reyna
University of Arizona

Although recent developmental studies have identified robust age differences in forgetting rates, these differences may be artifacts caused by confounding age with an underlying relationship between forgetting rates and children's learning abilities. This possibility was examined in some experiments that implemented path-analytic causal modeling. None of the path analyses supported the artifactual hypothesis. On the contrary, there were strong causal paths between development and forgetting rates when the potentially confounding influences of learning ability were eliminated, and the relationship between learning ability and forgetting rates was both weak and inconsistent. Coincidentally, these results also ruled out another artifactual hypothesis based on age variability in overlearning.

We explore this possibility in the present article. To set the stage for our experiments, we briefly discuss three issues: methodological objections to early developmental studies of forgetting, results and interpretations of more recent studies, and Age × Learnability artifacts. Next, we exhibit a path analysis procedure that is capable of detecting Age × Learnability artifacts. Last, we report some experiments that were designed to generate data to which this path analysis could be applied.

Background

Early Studies: Design Dilemmas

In the long-term retention paradigm, subjects are administered one or more study trials on a target set of items, usually words or pictures, during an acquisition session. Some hours, days, or weeks later, during a retention session, they receive a series of memory tests without further opportunities to study the items. The slope of the line connecting performance at the acquisition terminus to performance during the retention session is the measure of forgetting rate; the steeper the slope, the faster the forgetting. The memorial events that occur between

---


Despite this monolithic consistency, reviewers have identified certain methodological variables, common across most early studies, that would be apt to occlude developmental trends. Evidence of robust ontogenetic changes in forgetting rates has now accumulated in long-term retention experiments in which these variables were controlled (Brainerd et al., 1985; Howe & Hunter, 1986). However, these latter studies pose a new methodological problem of their own—Age × Learnability confounds. The age differences that were reported might have been spurious consequences of (a) the fact that the initial memorization of an item will normally be harder for younger children than for older children, in conjunction with (b) the fact that forgetting rates may be governed by learning rates (i.e., the harder an item is to memorize, the sooner it is forgotten).

We thank Mark Howe, Geoff Loftus, and three anonymous reviewers for their comments.

Correspondence concerning this article should be addressed to C. J. Brainerd or V. F. Reyna, Division of Educational Psychology, University of Arizona, Tucson, Arizona 85721.
the end of acquisition and the start of the retention session that produce declines in memory-test performance constitute the process definition of forgetting. The word declines is crucial because some processes operating during the retention interval have the opposite effect (Brainerd & Reyna, 1988b; Howe & Brainerd, in press).

Students of memory have traditionally classified forgetting events as being of two basic sorts—retirement failures (decreases in the accessibility of intact traces) and storage failures (losses of the traces themselves). The question of whether forgetting consists primarily of one or the other type of event has long been a contentious issue; a recent exchange between E. F. Loftus and Hoffman (1989), Tversky and Tuchin (1989), and Zaragoza and McClosky (1989) illustrates that this issue remains as contentious as ever. However, research on the development of forgetting has not been greatly concerned with such questions of theoretical interpretation (Brainerd et al., 1985).

Instead, the focus of research has simply been whether it is possible to produce age differences in forgetting rates, without regard to whether forgetting is chiefly retrieval-based or chiefly storage-based (Howe & Brainerd, in press). Most studies have relied on straightforward long-term retention designs, and as we have said, null age effects have been a routine feature of their results. Lately, an important variant of such designs, E. F. Loftus’s (1979) leading questions procedure, has also been studied with children, and some of these experiments have included age comparisons (e.g., Ceci, Ross, & Toglia, 1987a, 1987b). The control conditions of such experiments supply data that bear directly on developmental trends in forgetting rates (Brainerd & Reyna, 1988a; Ceci, Toglia, & Ross, 1988), and these data have also failed to show that forgetting develops.

Although null age effects have been ubiquitous, reviewers have drawn attention to three design factors that would tend to mask developmental trends: recognition insensitivity, low forgetting rates, and stages-of-learning confounds (see Brainerd et al., 1985; Howe & Brainerd, in press; Howe & Hunter, 1986). Concerning recognition insensitivity, if one seeks to diagnose developmental changes in any attribute of memory, it is essential to avoid measurement procedures that are not optimally sensitive to such changes. There is a large body of memory development research using recognition tests. Null age effects are the rule with recognition (e.g., Ornstein & Corsale, 1979), even though ceiling effects in performance are not usually a problem on such tests. Although the reasons for the developmental insensitivity of recognition are not fully understood, one may say, purely as a matter of methodology, that such tests stack the deck against observing age differences in any memory ability. Reviewers have noted that recognition tests were commonly used to measure forgetting in early developmental studies.

The second problem, low forgetting rates, refers to the fact that acquisition-to-retention declines were small, in some cases unreliable, in most early studies (see Kagan et al., 1973; Morrison et al., 1980; Rogoff et al., 1974). For instance, Morrison et al. (1980) reported two experiments involving four age levels in which most conditions did not display statistically reliable forgetting. One group of reviewers has concluded that the low forgetting rates in early studies were caused by a confluence of factors, such as short retention intervals, extremely memorable items, and memory tests with high false alarm rates (Brainerd, Reyna, Kingma, & Howe, 1989). Whatever the reasons, age changes in forgetting cannot be observed when the forgetting rate is zero, and are correspondingly difficult to detect when the rate is low.

The last problem, stages-of-learning confounds, arises from the fact that small, fixed numbers of study trials were administered during the acquisition sessions of most studies, with one-trial designs predominating. Because older children learn most things more rapidly than do younger children, the older children in these studies should have reached more advanced stages of learning than the younger children. Furthermore, these discrepancies should have become more severe as the number of study trials decreased and should have been maximal with one-trial designs because learning curves are negatively accelerated (Brainerd & Howe, 1982; Salthouse & Kausler, 1985). Now, imagine that forgetting rates decrease during birth-to-adult development and that the rate at which an item is forgotten (in either the sense of storage failure or the sense of retrieval failure) varies inversely with completeness of original learning. That is, the initial “primary” information that is learned about an item is preferentially conserved relative to the “secondary” information that is subsequently learned about it. Under this scenario, it would be difficult to measure age changes in forgetting rates because stages-of-learning confounds and the developmental trend in forgetting have opposite effects on retention data.

Later Studies: Data and Theory

Modifications that provide a systematic attack on these problems have been incorporated into certain experiments published since 1985. Recognition insensitivity has been dealt with by switching to some variant of recall as the retention measure. Forgetting rates have been increased by imposing longer retention intervals and by administering less memorable items. The remaining obstacle, stages-of-learning artifacts, has proven to be a more knotty matter, but procedural and analytical refinements are now in place that eliminate these confounds. Procedurally, children of different ages are required to meet identical, stringent acquisition criteria (typically, two or three consecutive errorless passes through the items). The rationale is that because learning curves are negatively accelerated, age disparities in completeness of learning should become negligible as learning approaches asymptote. This assumption will hold as long as the underlying theoretical process (learning) is negatively accelerated and is related to the performance measure by a monotonic transformation. The analytical refinement consists of fitting Markov models of learning stages (Brainerd, Howe, & Desrochers, 1982; Brainerd, Howe, & Kingma, 1982) to both acquisition and retention protocols. These models deliver esti-
that forgetting rates develop. Furthermore, a new theory has been advanced, the disintegration/redintegration hypothesis, which assumes that storage and retrieval are largely matters of depositing traces whose constituent features are so well integrated that they stand out sharply against the background of memory noise. In line with fuzzy-trace theory's emphasis on reconstructive aspects of memory (e.g., Brainerd & Reyna, in press), it is also assumed that children have the ability to redintegrate faded traces on retention tests.

**Age × Learnability Artifacts**

Unfortunately, the five experiments that have produced clear age effects pose a potentially worrisome methodological problem—Age × Learnability artifacts. If items that are harder to learn tend to be forgotten more rapidly, as is widely claimed (e.g., Slamecka & Katsaiti, 1988; Underwood, 1964), age differences in forgetting rates can be manufactured by merely confounding age with item learnability. A subtle confound of this sort was present in the Brainerd et al. (1985) and Howe and Hunter (1986) experiments. All of the subjects in these experiments memorized homogeneous sets of items that had been precisely equated for average learnability along a number of standard difficulty dimensions. Even with such homogeneous items, however, there will still be idiosyncratic differences in learnability (i.e., the word furniture may be hard for some subjects and easy for others). The existence of such intersubject variability carries in its wake the possibility of spurious developmental differences in forgetting rates if (a) the variability is correlated with age (i.e., subjects from some age levels find more of the items harder to learn than subjects from other age levels) and (b) this variability is also correlated with forgetting rates.

Concerning (a), younger children will normally find more of the items harder to learn than older children will. This is the standard developmental trend in learning ability. Concerning (b), suppose that item learnability is related to forgetting, so that items are forgotten more rapidly as their learnability decreases. If these two assumptions hold, developmental studies of long-term retention in which children of different ages memorize homogeneous sets of items to criterion will automatically show age changes in forgetting rates.

Because the truth of Assumption a has been established many times over in studies of acquisition, it is Assumption b that is critical. One might reasonably ask, therefore, if there is any familiar memory process that might confer validity on this assumption. Although various candidates might be mentioned, overlearning is the process that would probably occur first to most memory theorists (e.g., see Slamecka & Katsaiti, 1988). Some years ago, Underwood (1964) argued that although learning to criterion at acquisition may equate initial learning levels across conditions, it does so only at the expense of introducing discrepancies in degree of overlearning. More particularly, he argued that at least some overlearning of some items will take place under stringent acquisition criteria and that the amount of such overlearning will increase as item learnability increases. If forgetting rates are related to overlearning, his argument continued, such that it is easier to retain overlearned items, between-conditions differences in forgetting rates may be artifacts of correlated differences in overlearning. Wang (1988) has extrapolated this argument to developmental studies of long-term retention that involved criterion learning, proposing that the lower forgetting rates observed in older children were consequences of greater amounts of overlearning.

**Causal Models**

The situation that we have just described, observed age trends in forgetting rates coupled with the possibility that they may be rooted in Age × Learnability confounds, adumbrates a textbook problem in path-analytic causal modeling involving one independent variable (age) and two dependent variables (learning rate and forgetting rate). There are three alternatives in this situation, and they are formalized in the models exhibited in Figure 1. Model 1 assumes that development produces changes in item learnability, which in turn produces changes in forgetting rates, but that development does not produce changes in forgetting per se. This is the no-development alternative. The other two models assume that there is a direct path between development and forgetting rates, but they vary in their assumptions about the path between learnability and forgetting rates. Model 2 assumes that development affects item learnability, that it also affects forgetting rates, but that there is no causal path between learnability and forgettablity. Like Model 2, Model 3 assumes independent causal paths between age and learnability and between age and forgetting rates. In addition,
Figure 1. Three causal models of the relationship between development, learning difficulty, and forgetting.

DEVELOPMENT OF FORGETTING

MODEL 1

DEVELOPMENT \rightarrow ITEM LEARNABILITY \rightarrow ITEM FORGETTING RATE

MODEL 2

DEVELOPMENT \rightarrow ITEM FORGETTING RATE \rightarrow ITEM LEARNABILITY

MODEL 3

DEVELOPMENT \rightarrow ITEM FORGETTING RATE \rightarrow ITEM LEARNABILITY

However, it assumes that there is a third causal path between learnability and forgetting rates.

Path analysis (e.g., Cohen & Cohen, 1983; Duncan, 1966; Kenny, 1979; Zimmerman & Blom, 1983) can be used to pit these three models against each other in any set of data for which independent measures of age, item learnability, and item forgetting rates are available. Path analysis is a form of partial correlational analysis in which different patterns of correlations lend support to different causal models. In other words, a series of possible causal models is postulated in advance, and partial correlations are used as goodness-of-fit tests for the models. The patterns of correlations that path analysis specifies for the models in Figure 1 follow (see also, Salthouse, Kausler, & Saults, 1988).

It is easiest to begin with Model 3. The parameters in any path analysis are path coefficients, which are standardized regression weights. Path coefficients indicate the number of unit changes in one variable that is produced by a unit change in another variable. To compute the path coefficients for Model 3, let A, L, and F be measures of age (e.g., months since birth), item learnability (e.g., total errors to criterion at acquisition), and item forgetting rate (e.g., total errors on a long-term retention test). Let \( r_{AL} \), \( r_{AF} \), and \( r_{LF} \) be the raw correlations between these variables in some set of data. First, the fact that item learnability has only one "cause" in Model 3—namely, age—means that the path coefficient from age to learnability is just the raw correlation \( r_{AL} \). The other two path coefficients are the calculated from these formulas:

\[
\text{Age} \rightarrow \text{Forgetting Rate} = \frac{r_{AF} - (r_{LF} r_{AL})}{1 - (r_{AL})^2} \quad (1)
\]

and

\[
\text{Learnability} \rightarrow \text{Forgetting Rate} = \frac{r_{LF} - (r_{AL} r_{AF})}{1 - (r_{AL})^2} \quad (2)
\]

Because age has both a direct effect on forgetting rate and an indirect effect that is mediated through learnability, the total effect of age on forgetting rate is the sum of the direct and indirect effects. The direct effect is given by Equation 1, and the indirect effect is found by multiplying \( r_{AL} \) by the right side of Equation 2.

Turning to Model 2, this model assumes independent causal paths between age and learnability, on the one hand, and between age and forgetting rate, on the other hand. Hence, the path coefficient between age and learnability is \( r_{AL} \), and the path coefficient between age and forgetting rate is \( r_{AF} \). Furthermore, the correlation between learnability and forgetting rate, when the correlations of these two variables with age have been partialled out (Equation 2), should be zero because there is no learnability \( \rightarrow \) forgetting-rate causal path.

Finally, Model 1 assumes independent causal paths between age and learnability, on the one hand, and between learnability and forgetting rate, on the other hand. Therefore, the path coefficient between age and learnability is again \( r_{AL} \), and the path coefficient between learnability and forgetting rate is \( r_{LF} \). The entire effect of age on forgetting rate is an indirect one that is mediated through learnability. Consequently, the path coefficient between age and forgetting rate should be zero because \( r_{AF} \) should equal \( r_{AL} \times r_{LF} \) in the numerator of Equation 1.

The fact that path coefficients are correlations is self-evident from these derivations. As such, the procedure for testing path coefficients for statistical significance is the same as it is for correlations. These significance tests provide the decision rules that are used to decide which model fits the data best. For Model 3 to be acceptable, the path coefficients given by \( r_{AL} \), Equation 1, and Equation 2 must all be significant. For Model 2 to be acceptable, the path coefficients given by \( r_{AL} \) and \( r_{AF} \) must both be significant, but the path coefficient given by Equation 2 must not be significant. For Model 3 to be acceptable, the path coefficients given by \( r_{AL} \) and \( r_{AF} \) must both be significant, but the path coefficient given by Equation 1 must not be significant. For Model 3 to be acceptable, the path coefficients \( r_{AL} \) and \( r_{AF} \) must both be significant, but the path coefficient given by Equation 2 must not be significant. For Model 3 to be acceptable, the path coefficients \( r_{AL} \) and \( r_{AF} \) must both be significant, but the path coefficient given by Equation 1 must not be significant. In addition to statistical significance, a criterion of practical importance or meaningfulness is often used in applications of path analysis. Generally speaking, path coefficients smaller than .10 are not regarded as meaningful even when they achieve conventional levels of statistical significance (see Salthouse et al., 1988).

**Experiment 1**

**Method**

**Subjects.** Children from two age levels participated in this experiment. The younger participants were a mixed sample of first- and second-grade children. There were 50 children in the younger group (25 girls and 25 boys). The mean age of the younger group was 7 years, 4
months, and the range was 6 years, 6 months to 8 years, 5 months. The older participants were a mixed sample of fifth- and sixth-grade children. There were also 50 children in the older group (25 girls and 25 boys). The mean age of the older group was 11 years, 8 months, and the range was 10 years, 5 months to 13 years, 0 months. All 100 children were pupils in public elementary schools, and their participation was secured via a letter of parental permission.

Items. All of the children were administered homogeneous lists of 16 concrete nouns. The Paivio, Yuille, and Madigan (1968) norms were used to establish an initial pool of concrete nouns. This pool was constructed by sampling all of the words from the Paivio et al. norms that met all of the following criteria: rated concreteness values of 6 or above, rated imagery values of 6 or above, rated meaningfulness values of 6 or above, and rated familiarity values of A or AA on the Thorndike-Lorge (Thorndike & Lorge, 1944) count. Examples include words such as animal, bowl, mountain, tree, and window. This pool was used extensively in our prior research with children from the target age levels (e.g., Braierd & Howe, 1982; Braierd, Howe, & Desrochers, 1982; Brainernd et al., 1984). It is important that we know that these words are in these children’s lexicons and that they are of homogeneous difficulty (i.e., there are no reliable differences in how many trials it takes to learn them in standard list-learning tasks).

Procedure. The procedure of criterion learning, followed by repeated long-term retention tests that has been used in other recent experiments on the development of forgetting (Brainernd et al., 1985; Howe & Hunter, 1986), was applied here. The children were tested individually in a small quiet room of their school that had been set up as a temporary laboratory. Each child participated in two sessions, acquisition and long-term retention.

1. Acquisition. During this session each child memorized a 16-word list to a strict performance criterion (two consecutive errorless test trials) under standard free recall conditions. The lists memorized by individual children were constructed by randomly sampling the items from the noun pool. Each list of this sort was memorized by one younger child and by a yoked older child, who was selected at random. Hence, the lists memorized by the two age levels were identical. The experimenter and the child sat beside each other at a small table. They faced the window of a memory drum that was positioned in the center of the table. Some simple free recall instructions were first read to the child. Next, the memory drum was activated, and the first study trial began. Each of the 16 words appeared in the window of the memory drum for 5 s. To control for possible age differences in reading ability, the experimenter pronounced each word aloud as soon as it appeared.

Following the first study trial, the child engaged in 30 s of buffer activity (letter shadowing on a work sheet supplied by the experimenter) to empty short-term memory. After the buffer activity, the child was administered an oral free recall test on which he or she was instructed to “tell me as many of the words as you can remember.” The test continued until 15 s had elapsed without the recall of a word.

Upon completion of the first recall test, the experimenter reactivated the memory drum and another study trial ensued. The study trial was followed by another 30 s of buffer activity, which was followed by a second free recall test. This cycle of study–buffer–recall was continued until the child had performed two consecutive errorless free recall tests. At this point, the acquisition session was terminated, and the child was returned to his or her classroom.

2. Retention. Fourteen days after the acquisition session, the child returned to the laboratory for a series of long-term retention tests. At the start of the session, the child was read some simple retention instructions. The child was then administered a free recall test on which the instructions were to “tell me as many of the words that you learned two weeks ago as you can remember.” As before, this test continued until 15 s had elapsed without recall of a word. The child then performed 30 s of letter shadowing to empty short-term memory, and this was followed by another free recall test. The cycle of recall–buffer continued until the child had performed a total of five free recall tests. The retention session was then terminated. Note that no opportunities to relearn the items were provided during the retention phase.

Some brief comments are in order concerning the length of the retention interval that we selected. Because our aim was to detect possible age differences in forgetting rates, it was obviously essential both to avoid intervals that were so short that neither age level forgot very much (floor effects) and to avoid intervals that were so long that both age levels forgot everything (ceiling effects). Our prior studies have shown that retention intervals of 1–2 weeks are optimal in these respects, at least for elementary-schoolers (Brainernd et al., 1985; Howe & Hunter, 1986). However, some readers may be concerned about the possibility that because forgetting functions are negatively accelerated (the rate is faster initially than subsequently, Wickens, 1972), the length of the retention interval that one chooses may determine whether there are age differences in forgetting rates (see also Footnote 1). Although this is a theoretical possibility, we have studied retention intervals between 2 days and 30 days in previous experiments, and we have not found that the effects of age (and other manipulations) are inclined to appear and disappear as a function of retention interval within this range.

Results and Discussion

We used two statistics to measure acquisition difficulty—namely, total number of errors to criterion per item and trial number of the last error per item. We used total recall errors per item during the retention session as the measure of forgetting.

There were clear age differences in both the acquisition and forgetting measures. (In the significance tests that we report throughout this article, the .05 level of confidence was always used.) Concerning acquisition, the mean number of errors per item committed by the younger children was 3.50 (SD = 3.42), and the mean number of errors committed by the older children was 1.63 (SD = 1.28). This difference was reliable, t(98) = 3.58. The results for the trial of last error were similar. The mean trial of last error was 5.67 (SD = 3.08) for the younger children and 2.45 (SD = 1.99) for the older children. This difference was also reliable, t(98) = 6.15. For the list as a whole, the median trial of last error was Trial 8 for the younger children and Trial 5 for the older children.

Because performance at the end of acquisition was always perfect in this experiment, the percentage of errors per item during the retention session is a direct measure of forgetting rates for individual items. As in other recent experiments, forgetting rates were much higher in younger children than in older children. The mean number of retention errors per item among younger children was 2.15 (SD = 2.11), for an average forgetting rate of 43% [100 x (2.15 / 5 retention tests)]. The mean number of retention errors per item among older children was 0.67 (SD = 1.26), for an average forgetting rate of 15%. The difference between these two forgetting rates was highly reliable, t(98) = 3.93.

To determine whether forgetting actually developed, we conducted path analyses using the three causal models in Figure 1. In a comparative model-fitting exercise such as this, the preferred strategy is to arrange the models in a simple (fewer parameters) to complex (more parameters) ordering (e.g., Brainernd, Howe, & Desrochers, 1982; Theios, Leonard, & Brelsford, 1977). One then begins at the simple end of such a continuum, performing goodness-of-fit tests on each successive model until
a model is located that agrees with the data. This simple-to-complex procedure ensures that the most parsimonious model (the one with the smallest number of free parameters that fits the data) will be selected. Regarding the models in Figure 1, Models 1 and 2 are simpler than Model 3 because each of them assumes that one of the three path coefficients is zero. So, we always evaluated Models 1 and 2 first and only evaluated Model 3 when neither of the other models fit the data. For the sake of generality, we conducted two parallel evaluations. In the first, the independent variable was A = age in months, and the dependent variables were L = total acquisition errors per item and F = total retention errors per item. In the second, the independent variable was A = age in months, and the dependent variables were L = trial of last acquisition error per item and F = total retention errors per item.

The results of the first analysis were as follows. Model 1 assumes that the A → L path \( r_{AL} \) and the L → F path \( r_{LF} \) are both reliable, but the A → F path \( r_{AF:L} \), Equation 1) is not reliable. When the relevant path coefficients were calculated, the A → L and L → F paths were both reliable, \( r_{AL} = -.61 \) and \( r_{LF} = .60 \), but so was the A → F path, \( r_{AF:L} = -.66 \). Hence, Model 1 was rejected. Model 2 assumes that the A → L path \( r_{AL} \) and the A → F path \( r_{AF} \) are both reliable, but the L → F path \( r_{LF:A} \), Equation 2) is not reliable. This model was unacceptable because \( r_{LF:A} = .20 \). Finally, Model 3 was acceptable because all three of its path coefficients were reliable, \( r_{AL} = -.61 \), \( r_{AF} = -.77 \), and \( r_{LF:A} = .20 \). These results are shown at the top of Figure 2.

In the second path analysis, Model 1 was also rejected because \( r_{AF:L} = -.70 \). However, Model 2 was accepted because \( r_{LF:A} \) was not reliable. These results are shown at the bottom of Figure 2.

Summarizing the path analyses, there was a strong causal relationship between age and forgetting rates in both analyses that could not be explained on the ground of confounding age with learnability. The model that embodies this particular hypothesis (Model 1) was rejected in both cases. Also, the A → F path was the strongest of the three paths in both analyses (see Figure 2), which is also inconsistent with the hypothesis. Although the two analyses seemed to give slightly different results in the sense that Model 3 was accepted in one case and Model 2 was accepted in the other case, the difference was more apparent than real. In the first place, the two L → F path coefficients, whose significance is the only difference between the two analyses, were very similar, differing by only .08. The value of .20 is slightly higher than the critical value of .16, whereas the .12 is slightly lower than the critical value. Second, when one calculates the .05 confidence interval around .20, .12 falls well within this interval, which means that the two analyses were equivalent statistically, although there was a difference in significance. The take-home messages are that both analyses revealed a definite A → F path and that, surprisingly, it was the strongest path of the three.

As a final check on these results, we studied the relationship between learnability and forgetting rates within each age level. We have seen that this was the weakest of the three paths in the overall data. There are two possible explanations. First, it could be that, as the path analyses suggest, the L → F relationship is weak. Second, it could be that the relationship is strong, but it varies with age in such a manner that its strength tends to be masked when data are pooled across age levels. For example, if there were a strong positive relationship in older children and strong negative relationship in younger children, or vice versa, the L → F association would be more pronounced when calculated within age levels than when calculated for the pooled data. But if the first hypothesis is correct and the relationship is simply weak, the degree of association should actually shrink when it is calculated within age levels because of the Spearman-Brown effect. According to the Spearman-Brown effect, the size of a correlation depends on the number of data points involved in its computation, with the exact relationship being

\[
r' = \frac{Nr}{1 + (N-1)r},
\]

where \( r \) is the value of the correlation coefficient for the target data and \( r' \) is the value of the correlation for a set of data involving \( N \) times as many pairs of scores as the target data.

The results of the analysis supported the first hypothesis rather than the second. For the pooled data, the raw correlation between learning difficulty and forgetting was .60 for the total errors measure and .54 for the trial-of-last error measure. For the younger children alone, the correlations were .27 (total errors) and .18 (trial-of-last error). For the older children alone, the correlations were .30 (total errors) and .25 (trial of last error). It appears, then, that the causal path between learning difficulty and forgetting was not very marked in these data.

These findings argue strongly against the claim that age
differences in performance on long-term retention tests are artifacts induced by the confounding of age with learnability. Although there was evidence that forgetting rates tend to increase as learning difficulty increases, there was an independent causal path between age and forgetting rates. Furthermore, of the three possible causal relationships, the one between learnability and forgetting rates was the weakest and was not reliable in one of the analyses. The relationship between age and forgetting rates was the strongest of the three relationships.

Experiment 2

The objective of Experiment 2 was to broaden the findings of Experiment 1 in two major respects. First, we sought to determine whether the same pattern of relationships stood up when the to-be-remembered materials were much more difficult. As free recall materials go, the very familiar and highly concrete nouns administered in Experiment 1 are among the easiest types of word lists for children to memorize. It is quite conceivable that age, learning difficulty, and forgetting rates would be related in a different manner if more difficult materials were memorized.

The second objective was to determine whether the specific relationship between learnability and forgetting rates is affected by the overall difficulty of the materials being memorized. A surprising outcome of Experiment 1 was that although words that were harder to learn tended to be forgotten more rapidly, this relationship was very weak. With more difficult words, however, harder items might be forgotten much more rapidly than easier items.

Method

Subjects. A mixed sample of 50 first and second graders and a mixed sample of 50 fifth and sixth graders served as the subject sample. The mean age of the younger sample was 7 years, 6 months, and the range was 6 years, 9 months to 8 years, 3 months. The mean age of the older sample was 11 years, 8 months, and the range was 10 years, 10 months to 13 years, 4 months. All of the children were pupils of public elementary schools, and their participation was again secured through letters of parental permission.

Lists. The lists administered in Experiment 2 were composed of abstract nouns that were known to be 2–3 times more difficult to memorize than the concrete nouns used in Experiment 1 (see Brainerd et al., 1984). An initial pool of abstract nouns was constructed by sampling all of the words on the Paivio et al. (1968) norms that met the following criteria: rated concreteness values of 3.5 or less, rated imagery values of 3.5 or less, and rated familiarity values of A or AA on the Thorndike-Lorge count. Examples include words such as effort, fact, history, idea, and quality. As was the case for Experiment 1, the items in this pool have been used in our prior research with children from this age range. These words are well-known to children, and they are homogeneous with respect to average learning difficulty. The lists memorized by individual children were then constructed through the sampling and yoking procedures described in Experiment 1.

Procedure. The procedure was the same as in Experiment 1, except for two changes. First, the children memorized abstract nouns rather than concrete nouns. Second, the buffer activity during both the acquisition and long-term retention sessions consisted of shadowing number pairs on work sheets provided by the experimenter.

Results and Discussion

Preliminary analyses. The same statistics as before were used to measure acquisition difficulty (total errors to criterion per item and trial of last error per item) and to measure forgetting (total errors per item on the five retention tests). As expected, these words were much more difficult to memorize than were the concrete nouns of Experiment 1. For younger children, the mean total errors per item was 7.76 (SD = 3.98), and the mean trial of last error was 10.59 (SD = 3.76). The corresponding values for Experiment 1 were 3.50 and 5.67, so the increase in difficulty was roughly 100%. For the older children, the mean total errors per item was 3.92 (SD = 2.63) and the mean trial of last error was 5.63 (SD = 3.44). Because the corresponding values in Experiment 1 were 1.63 and 2.45, the increase in difficulty was again in the neighborhood of 100%. For the list as a whole, the modal trial of last error was Trial 8 for the older children and Trial 13 for the younger children.

Also as expected, these items were more easily forgotten than those in Experiment 1. The mean number of retention errors was 2.81 (SD = 2.15) for the younger children and 1.90 for the older children (SD = 2.18), compared with 2.15 and .77 for the younger and older children, respectively, in Experiment 1.

The age trends for both acquisition and forgetting were reliable in Experiment 2. Younger children committed more acquisition errors than older children, t(98) = 8.27, and younger children forgot more than older children, t(98) = 2.08.

Path analyses. Path analyses involving the same variables as in Experiment 1 were used to decide whether the developmental decline in forgetting rate was an artifact of confounding age with item learning difficulty. As before, the results showed that there was a true decline in forgetting rate with age. The results of the path analyses are displayed in Figure 3.

As in Experiment 1, we conducted two parallel analyses—the first with A = age in months, L = total errors per item at acquisition, and F = total errors per item at retention, and the second with the same values of A and F, but L = trial number of the last error per item at acquisition. Also, as before, we tested the models for fit in a simple-to-complex order, testing Model 3 only if Models 1 and 2 were rejected. In the first analysis, Model 1 was rejected because r_{AF.L} = -.50, and Model 2 was rejected because r_{L.F.A} = -.19. Model 3 was accepted (see Figure 3). In the second analysis, Model 1 was rejected because r_{AF.L} = -.52, and Model 2 was rejected because r_{L.F.A} = -.28. Thus, the existence of an age-forgetting causal path was not affected by the fact that the present items were roughly twice as difficult as those in Experiment 1.

The major difference between the results of the two experiments lies in the path between learnability and forgetting rate. With the easier lists of Experiment 1, there was a weak positive path between these variables (i.e., slightly more forgetting for harder-to-learn items) in both analyses, although the path coefficient was not reliable in one case. Experiment 2 was partially motivated by the possibility that a much stronger positive path would materialize with more difficult lists. The opposite result was obtained, however. It can be seen in Figure 3 that there was a weak negative path in both analyses—harder-to-learn items were remembered slightly better. This outcome is especially interesting from the standpoint of our initial hypothesis that age
differences in forgetting rates are spurious consequences of confounding age with learnability. Because forgetting rate tended to decrease as learning difficulty increased in this study, the ironic conclusion is that the raw age differences in forgetting that were previously reported are underestimates of the true amount of age difference; the learnability–forgetting relationship tends to mask developmental differences in forgetting rates.

To sum up, the combined results of the path analyses for both experiments suggest the following tentative generalizations. First, the relationship between learnability and forgetting rate is not strong, regardless of whether learning is easy or hard on average. The relationship was so weak in our data that, for practical purposes, it can be ignored. Second, the nature of this relationship is inclined to vary with overall learning difficulty. With easier lists that produced smaller age differences in learnability, the relationship was weak and positive. But with harder lists that produced larger age differences in learnability, the relationship was weak and negative.

**General Discussion**

The ontogeny of forgetting is a topic of fundamental concern from both theoretical and applied perspectives. Theoretically, it is comparable in significance with such well-mined domains as the development of organization in semantic memory, the development of short-term memory capacity, and the development of mnemonic strategies. On the applied side, most questions about curriculum effectiveness in elementary school are questions about intersubject variability in long-term retention, as well as about intersubject variability in initial learning. Also on the applied side, questions about children’s testimony—questions that are sometimes matters of life and death—are grounded in implicit assumptions about age changes in forgetting (Brainerd & Reyna, 1988a; Ceci et al., 1988). Another motivation for studying the development of forgetting arises from repeated experimental demonstrations that acquisition and retention are asymmetrical processes; they are not opposite sides of the same memorial coin (Brainerd et al., 1985, 1989; Brainerd & Reyna, 1988a, 1988b; Howe & Brainerd, in press; Howe & Hunter, 1986). In this latter connection, it is often found that variables that have a certain effect on acquisition either have no effect on retention (e.g., Slamecka & McElree, 1983) or have opposite effects on retention (e.g., Brainerd et al., 1985).

Despite its importance, the development of forgetting is a subject on which little is known (e.g., see Bjorklund, 1989), a lacuna that can be traced to consistent failures to find age effects in early studies. Although large age effects were secured in more recent experiments that coupled criterion learning of items with Markovian analyses, these effects may have been by-products of Age \times Learnability confounds. Consequently, the overriding aim of our experiments was to evaluate the hypothesis that these effects were entirely due to (a) age increases in item learnability together with (b) decreases in forgetting rates as functions of increases in learnability.

The bottom line of our results is that Assumption b in this hypothesis is wrong and, hence, the hypothesis itself fails. The path analyses uniformly favored models (Figure 1, Models 2 and 3) that postulated independent causal paths from development to forgetting rates. In all cases, a robust age–forgetting-rate relationship was obtained when the effects of item learnability were stripped away. What is more, the learnability–forgetting-rate relationship proved to be both weaker and less consistent than the artifact hypothesis demands. This hypothesis is only credible if the relationship is powerful and if it is also positive (i.e., items that are harder to learn are harder to retain). Our data did not confirm either requirement. The learnability–forgetting-rate relationship was invariably the weakest one in the path analyses, and it was not always reliable. It was not consistently positive, either. With easier items (Experiment 1), there was a weak positive relationship, but with harder items (Experiment 2) there was a slightly stronger negative relationship. Although it would have been possible to hold out some hope for Assumption b if the relationship had at least been positive, the fact that the relationship was negative with harder items is a clear violation of the assumption.

Another striking disconfirmation of Assumption b emerges when one compares the respective age differences in forgetting rates in Experiment 1 versus Experiment 2. If there is a positive learnability–forgetting-rate relationship, then items that produce larger age differences in average learnability must necessarily produce larger age differences in forgetting rates. The reverse was true in our data. In Experiment 1, the age difference in mean acquisition errors was 1.87, and the age difference in mean trial of last error was 3.25. The age difference in mean number of retention errors was 1.40. In Experiment 2, the age difference in mean acquisition errors rose to 3.84, and the age
difference in mean trial of last error rose to 4.69. However, the age difference in mean number of retention errors shrank to 0.91. In short, although the age difference in item learnability was twice as great in the second experiment, the age difference in forgetting rates was only about one half as large.

To conclude this article, some retrospective comments about the general strategy that we adopted to study potential Age × Learnability artifacts will help establish the boundary conditions of our findings and will suggest near-term directions for future research. Essentially, our strategy involved two steps. First, children of different age levels memorized identical sets of items that were homogeneous with respect to average learning difficulty. Second, path analyses were used to determine whether forgetting rates declined with age when the ostensibly confounding influences of learnability were factored out. This design ensured maximum comparability with recent experiments in which developmental differences in forgetting rates were observed in children who had memorized identical sets of homogeneous items (Brainerd et al., 1985; Howe & Hunter, 1986). Thanks to such comparability, we are able to infer that the age trends in these other experiments were probably not rooted in Age × Learnability confounds, thereby increasing the total stock of published data pointing to the conclusion that forgetting rates develop.

Against our strategy, however, it might be argued that even though identical sets of homogeneous items were administered to the different age levels, older children's knowledge of these items—in the sense of richness of representations, number of associations, overall memory strength, and so forth—was undoubtedly greater than younger children's knowledge. Therefore, the age differences in forgetting rates might have been produced by age differences in preexperimental item knowledge and, the argument concludes, it would be desirable to study the development of forgetting using sets of items that equate for these preexperimental knowledge differences, such as those developed by Chechile and his associates (e.g., Chechile & Richman, 1982).

On the one hand, long-term retention studies in which "age-appropriate" items are administered would certainly be useful additions to the data base. However, there were three reasons for preferring the design that we adopted, at least in these initial experiments. First, an alternative design would not have achieved the desired comparability with other developmental studies of long-term retention. Second, age differences in preexperimenteral knowledge, although undoubtedly real, should have been minimal in our studies because of the types of items that we used. Recall, here, that the items in both experiments had Thorndike-Lorge counts of A or AA (e.g., doctor, mind) and were known to be familiar to children from the target age range. Third, and perhaps most important, the path analyses should also control for the possibility that preexperimental knowledge differences produced the age trends in forgetting rates. If there were developmental differences in such knowledge, then, naturally, they would translate into correlated differences in learnability—greater knowledge mean greater learnability and greater developmental knowledge differences mean greater developmental differences in learnability (e.g., Bjorklund, 1987). In fact, such knowledge differences might be the correct explanation of the overall age trend in learnability that we observed and of the fact that this trend was larger in Experiment 2 (in which preexperimental variability in item knowledge was presumably greater). Note, however, that the path analyses were designed to control the effects of learnability on forgetting rates. So, they should automatically control the effects of any memory process (e.g., item knowledge) that produces differences in learnability.

References
Fajnzstejn-Pollack, G. (1973). A developmental study of decay rate in


