Developmental Psychology
1996, Vol. 32, No. 3, 467–478

0012-1649/96/$3.00

Mere Memory Testing Creates False Memories in Children

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

It has been reported that initial recall tests inoculate true memories against forgetting without creating false memories. This is not true of recognition tests. In 2 experiments with 5- and 8-year-olds, initial recognition tests elevated children's false-memory responses (false alarms to unpresented distractors) on delayed tests. In Experiment 1, false-memory creation exceeded true-memory inoculation in 5-year-olds, producing net losses in accuracy over time. In Experiment 2, false-memory creation exceeded true-memory inoculation at both age levels.

How does memory testing influence the accuracy of children's subsequent memory reports? This question has been extensively investigated in the literature on children's eyewitness testimony, with most studies concentrating on the effects of tests that incorporate suggestive or misleading language (for a review, see Ceci & Bruck, 1993b). It has been found that under various conditions, such tests impair subsequent recollection of both incidental details and central thematic information (Bruck, Ceci, Francoeur, & Barr, 1995; Cassel, Roebers, & Bjorklund, in press). Consequently, there is a growing consensus that suggestive or misleading language is to be avoided in forensic interviewing of children (Bruck & Ceci, in press; Ceci & Bruck, 1993a).

A closely-related issue on which far less evidence exists concerns the effects of mere memory testing; the administration of tests that neither suggest nor mislead, but are intended only to diagnose the contents of memory. In an early discussion of this issue, Brainerd and Ornstein (1991) pointed out that although we have little control over the circumstances in which children's memories are originally stored, we have considerable control over the conditions under which they are tested. This allows us to exploit the results of laboratory research so as to select testing conditions that should optimize accuracy in the long run (Reyna, 1992). One result to which Brainerd and Ornstein drew attention is the so-called *inoculation effect.* On a long-term retention test, recall of studied items is enhanced if children (or adults) receive a prior recall test, with the amount of enhancement increasing as the interval between memory storage and initial testing decreases (for a review, see Brainerd, Reyna, Howe, & Kingma, 1990). Brainerd and Ornstein concluded that as long as questioning is not suggestive or misleading, "children's testimony would be expected to be facilitated as a consequence of . . . repeated questioning . . . " (p. 18), but they added that even neutral questioning "may contribute to an alteration in the content of the reported memories" (p. 18).

The practical question, then, is, Do mere memory tests only inoculate true memories against forgetting or do they also create false ones? Of course, mere memory testing can have other consequences, emotional ones for instance (Goodman, Bottoms, Schwartz-Kenney, & Rudy, 1991; Saywitz, 1995), but we shall confine attention to its effects on memory. If memory tests also create false memories, their net impact on subsequent accuracy will depend on their respective rates of true-memory inoculation and false-memory creation. Several recent recall studies provide relevant findings (Cassel & Bjorklund, 1995; Poole & White, 1991, 1993; Warren & Lane, 1995). These studies, along with the earlier literature on this question, were reviewed by Poole and White (1995). They concluded that "laboratory simulations of eyewitness testimony offer strong corroborating evidence that multiple testing sessions preserve memories over time. . . We are aware of no studies in which multiple interviews with nonsuggestive questions were associated with increases in the amount of inaccurate information recalled by children" (p. 34).

This conclusion can be illustrated by two studies, an early one by Dent and Stephenson (1979) and a later one by Baker-Ward, Hess, and Flannagan (1990). The magnitudes of the inoculation effects in these experiments were in the 5%–10% range. In Dent and Stephenson's study, children were assigned to three testing conditions after viewing a film of a theft: (a) immediate, 1-day, 2-day, 2-week, and 2-month tests; (b) 2-week and 2-month tests only; and (c) 2-month test only. All tests involved free recall. Prior testing inoculated true memories. Children in the first condition recalled more correct information than children in the other two conditions on the last two tests. Prior testing did not create false memories. The number of intrusions (recalled information that was not present in the film) on the last two tests did not vary as a function of amount of prior testing. Turning to the Baker-Ward et al. study, children were assigned to two testing conditions after they participated in a series of activities: multiple testing (immediate, 1 week, 2 weeks, and 3 weeks) or single testing (3 weeks). Children were asked to recall the target activities during each session. At 3 weeks, recall of target activities was better in the multiple-testing condition, but intrusions did not differ between conditions.

C. J. Brainerd and V. F. Reyna, Department of Educational Psychology, University of Arizona.

We are indebted to D. A. Poole for her critique of a draft of this article and for much sound advice on the connections between our research and children's testimony.

Correspondence concerning this article should be addressed to C. J. Brainerd, Department of Educational Psychology, University of Arizona, Tucson, Arizona 85721. Electronic mail may be sent via Internet to brainerd@ccit.arizona.edu.

The studies reviewed by Poole and White (1995) support the contention that mere memory testing inoculates true memories without creating false ones. However, those studies share a design limitation—they all involved some form of recall.[1] This is worrisome for three reasons. First, there is the general concern that recall is only one of the two basic methods of testing memory, the other being recognition. The mainstream memory literature is replete with dissociations between recognition and recall; effects that are observed with one type of test but not with the other (e.g., Baddeley, 1976). It is quite conceivable, therefore, that recognition tests could create false memories even if recall tests do not. Indeed, we have argued that fuzzy-trace theory provides multiple mechanisms to produce such effects (Reyna, 1995; Reyna & Brainerd, 1995). Those mechanisms are examined later.

Second, as is well known (e.g., Baddeley, 1976), recall tests are less sensitive to the contents of memory than recognition tests are. Following exposure to a series of memory targets, target recognition (when corrected for guessing) is typically 2–3 times better than target recall (e.g., Mandler, Pearlstone, & Koopmans, 1969). The same is presumably true of false memories; if prior recall tests create false memories, later recall tests may be unable to detect them. A datum that supports this claim is that, in adults, a prior recall test increases false acceptance of distractors on later recognition tests (Brown & Packham, 1967; Roediger & McDermott, 1995).

Third, as is also well known, intrusion rates (the false-memory measure in recall) are usually near-floor. This is true for initial tests administered immediately after study (Bjorklund & Muir, 1988) and for readministrations of those tests after retention intervals of 1 or 2 weeks (Brainerd et al., 1990).[2] This creates a statistical power problem inasmuch as the minimum precondition for detecting between-conditions differences in any measure is that it should exhibit substantial variability. In other words, it is hard to detect differences in intrusion rates as functions of prior testing simply because there is little variability to work with.

In short, general conclusions about true-memory inoculation versus false-memory creation require detailed findings on recognition. Some suggestive results have already been reported by Reyna and Kiernan (1994). The children in their experiments were exposed to some target material, received an immediate recognition test, and then received a repetition of that test 1 week later. Unlike intrusions in recall, there were substantial increases in false-alarm rates (the false-memory measure in recognition) between the immediate and 1-week tests, as much as 126% in one condition. Comparable findings were obtained by Reyna (in press) and by Reyna and Kiernan (1995). Such results are only suggestive, however, because previously untested items were not included on the delayed tests. Thus, the extent to which increases in false alarms were due to the initial recognition test per se could not be determined. In the experiments that we report below, delayed tests were modified to include previously untested items.

A purely pragmatic motivation for investigating the effects of recognition testing is the prevalence of such tests in forensic interviews of child witnesses. Although questions requiring some form of recall figure in most interviews conducted by attorneys, police, social workers, and therapists, so do questions requiring a simple yes or no. With young children, who are not very responsive to open-ended recall, yes-no questions may predominate, and they often provide the bulk of the usable responses. In a recent case of alleged sexual abuse in Arizona, for instance, 60% of the questions posed in a police interview of a 28-month-old child were yes-no, and the only recall questions that produced interpretable responses were concerned with the child's age.

Experiment 1

The aim of this experiment was to measure an initial recognition test's rates of true-memory inoculation and false-memory creation. There were two sessions. The first involved a standard recognition design. Children from two age levels (5- and 8-year-olds) were exposed to a series of memory targets, followed by a recognition test. Half the items on the test were ones that the children had been exposed to (targets), and the other half were new (distractors).

During the second session, 1 week later, rates of true-memory inoculation and false-memory creation were measured by administering another recognition test. As before, half the items were targets and half were distractors. However, both targets and distractors were further subdivided into items that had been tested 1 week earlier and items that had not been tested. The true-memory inoculation rate is the difference between the hit rates for previously tested versus previously untested targets. The false-memory creation rate is the difference between the false-alarm rates for previously tested versus previously untested distractors.

We also manipulated the degree of target-distractor resemblance. Some of the distractors that appeared on the immediate test shared salient meaning with targets that had been presented during the exposure phase, whereas other distractors did not. Meaning-sharing distractors were of two sorts: names of categories to which presented targets belonged (e.g., animal would be a distractor if dog had been presented) and typical exemplars of categories to which presented targets belonged (e.g., iron would be a distractor if steel had been presented).

This manipulation was included for two reasons. First, we thought that if mere recognition testing creates false memories, the effect should be more pronounced for related distractors than for unrelated ones. We thought this because (a) children normally exhibit higher false-alarm rates for meaning-sharing distractors than for unrelated distractors on initial tests (e.g., Brainerd, Reyna, & Kneer, 1995; Reyna & Kiernan, 1994, in press) and (b) initial false alarms for meaning-sharing distractors are more likely to be preserved across a forgetting interval than initial false alarms for unrelated distractors (Brainerd, Reyna, & Brandse, 1995).

---

[1] One of the studies reviewed by these authors included yes-no questions in the design (Poole & White, 1991). However, those questions measured children's inferences about observed events rather than their memory for the events themselves.

[2] We are still focusing on the standard situation in which the task is simply to recall previously presented targets. Intrusion rates can be much higher when suggestive or misleading information is interpolated prior to the recall test, especially when the test occurs several weeks after target presentation (e.g., Poole & Lindsay, 1995).

second motivation for manipulating semantic overlap be-
n targets and distractors was forensic relevance. In forensic
views of child witnesses, most questions that require yes-
nswers revolve around some topic that is the focus of inves-
on, which means that they share gist with the presumed
ts. Hence, forensic interest attaches strongly to the issue of
her recognition tests create false memories for information
is meaningfully related to presumed events.

hod

rticipants. A total of 80 children participated, 40 younger chil-
and 40 older children. The younger children were kindergartners
in age = 5 years 7 months), and the older children were third grad-
mean age = 8 years 11 months). This age range was chosen for the
of comparability with the larger literature on children's eyewitness
nory. Most of that literature deals with the preschool to mid elemen-
school years. Half the children at each age level were boys and half
girls.

'aterials and procedure. Children were tested individually in
ll, quiet rooms within their schools. The procedure involved three
s, the first two taking place during the acquisition session and the
r one taking place during the retention session.

Target exposure. Following initial instructions, which described
experiment as a study of memory for vocabulary, the child listened
tape recording of 100 familiar concrete nouns being read at a rate
s per word (i.e., total presentation time = 5 min). Five tapes, each
n a different order, were used, with 16 children (eight 5-year-olds and
nt 8-year-olds) being randomly assigned to each tape. These memory
gets were selected from a pool of 250 nouns that had been drawn
n the Paivio, Yuille, and Madigan (1968) and the Battig and Mon-
ue (1969) norms. The meanings of these words are well known to
ldren in this age range (cf. Brainerd et al., 1990). Forty items (e.g.,
, rose, steel, table) were critical targets that were typical exemplars
oduction frequencies 1–8 on the Battig and Montague, 1969, norms)
'amiliar categories (e.g., animal, flower, metal, furniture). Each cat-
ry was represented by only one exemplar. The remaining 60 items
e noncritical targets that had been selected from the word pool and
t were unrelated to each other and to the critical targets. (The dis-
ction between critical and noncritical targets is that the former served
sources of target-related distractors on the immediate and delayed
ognition tests.) The exposure phase also included a learning manip-
tion. Half of the items (20 of the critical targets + 30 of the noncrit-
l targets) were read once (low learning), and half were read 3 times
gh learning). This manipulation was included because, on the basis
prior research (Brainerd, Reyna, & Kneer, 1995), it was thought that
th true-memory inoculation and false-memory creation might vary
a function of level of learning. Presentation order was random, except
the constraint that at least three targets intervened between consec-
ve presentations of repeated targets.

2. Initial recognition test. After the last item was presented, there was
-min, attention-consuming buffer activity. A naturalistic hidden fig-
es test called *Where's Waldo?* (Handford, 1987) was administered. In
here's Waldo?, a series of colored drawings of large numbers (>100)
people in familiar crowd situations (e.g., at the beach, at a sporting
ent) is presented. The child attempts to locate a particular person
'aldo) in the crowd. Following this activity, a recognition test was
ministered. Standard recognition instructions were given in which
children were told to say "yes" to any word that they had heard on
tape recording and to say "no" to all other words. The instructions
essed that half the items would be words that they had heard, that
lf would be words that they had not heard, and that they should say
o" to words that had not been heard even if they resembled presented
rds. To ensure understanding, examples of hypothetical targets and
related distractors that should be rejected were given. Following instructions, the experimenter read a test list composed of a total of 60 items. Half of these items were old (15 high-learning targets and 15 low-learning targets) and half were new (distractors). The 30 targets were selected from the 60 noncritical targets. There were 3 groups of distractors, with ~~10 distractors per group. A category group was generated by replacing~~ 10 of the critical targets with their corresponding category names (e.g., replacing cat with animal). Five of the replaced critical targets were high learning and five were low learning. An exemplar group was generated by replacing 10 other critical targets with other typical exemplars from the same categories (e.g., replacing steel with iron). Again, five of the replaced critical targets were high-learning and five were low-learning. (With respect to both the category and exemplar groups, it should be emphasized that the critical targets were always exemplars, never category names.) An unrelated group was generated by sampling 10 items from the word pool that were not related to any of the presented targets. Thus, on the immediate recognition test, children responded to 60 items, 30 targets and 30 distractors (10 category names, 10 new category exemplars, and 10 unrelated nouns).

3. Delayed test. One week later, children participated in a delayed recognition test. The instructions stated that children were to say yes to targets that they had heard on the audio tape a week earlier (as opposed to items from the initial recognition test) and to say no to all other items. The experimenter then read a test list of 120 items (60 targets and 60 distractors), and children again responded in a self-paced manner. This list was composed of the 60 items (30 targets and 30 distractors) that had been tested 1 week earlier plus 60 new items (30 targets and 30 distractors) that had not been tested. The 30 new targets were the remaining 30 noncritical targets that had not appeared on the immediate test. The 30 new distractors were again divided into 3 groups of 10 each, a categories group, an exemplars group, and an unrelated group. The categories group was generated by replacing 10 of the 20 critical targets that had not been used on the immediate test with their corresponding category names (e.g., replacing rose with flower). The exemplars group was generated by replacing the other 10 critical targets that had not been used on the immediate test with typical exemplars from the same categories (e.g., replacing table with chair). The unrelated group was generated by selecting 10 more unrelated items from the word pool.

A summary of the list design is given in Figure 1. Examples of studied targets and recognition items (targets, category distractors, exemplar distractors, unrelated distractors) are given for both the immediate and delayed tests. As can be seen, targets and distractors that appeared on the immediate test also appeared on the delayed test, but so did additional targets and distractors that had not been previously tested.

Results and Discussion

Although we are chiefly interested in performance on the delayed test, results for the immediate test should be briefly considered. At both age levels, performance was quite good in the sense that target recognition (hits) was well above the nominal chance level of .5. For kindergartners, the hit rate was .93 for high-learning targets and .68 for low-learning targets. For third graders, the corresponding hit rates were .95 and .75.

We report the findings for the delayed test in three sections. First, we consider data on true-memory inoculation. Next, we consider data on false-memory creation. Finally, we consider whether the initial test produced net gains or losses in accuracy on the delayed test.

True-memory inoculation. To assess true-memory inoculation rates, we analyzed hit rates (target acceptances) on the 1-week test as functions of prior memory testing. The major anal-

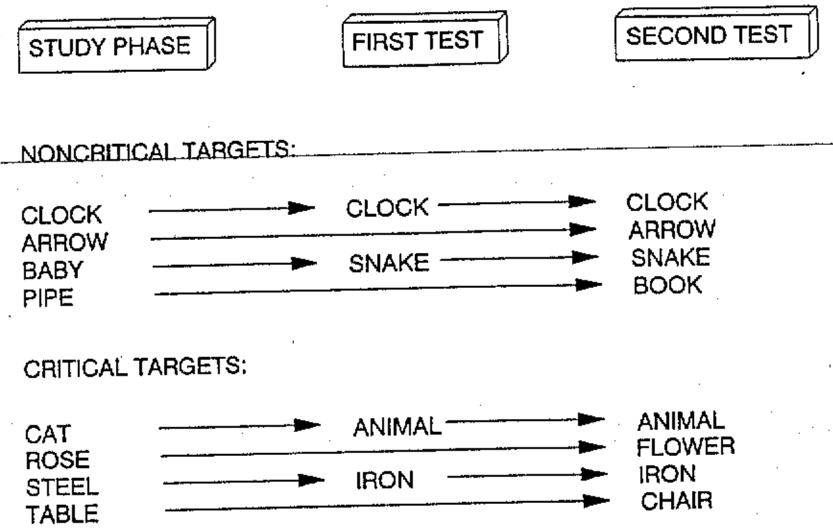

Figure 1. Structure of study lists, immediate test lists, and delayed test lists in Experiments 1 and 2. Some noncritical targets (clock, arrow) appeared on the recognition tests, and others (baby, pipe) were replaced by unrelated distractors (snake, book). Some tested noncritical targets appeared on both tests (clock), and others appeared only on the delayed test (arrow). Some unrelated distractors appeared on both tests (snake), and others appeared only on the delayed test (book). All the critical targets were replaced by related distractors. Some of these distractors were the names of categories to which the targets belonged (animal, flower), and others were exemplars from categories to which the targets belonged (chair, iron). Some distractors appeared on both tests (animal, iron), and others appeared only on the delayed test (chair, flower).

ysis was a 2 (age: 5 vs. 8) × 2 (degree of learning: high vs. low) × 2 (prior testing status: tested vs. untested) analysis of variance (ANOVA) of participants' hit probabilities on the 1-week test. This ANOVA revealed main effects for all three factors, an Age × Prior Testing Status interaction, and a Degree of Learning × Prior Testing Status interaction (all Fs > 5.2 and all ps $< .005$). The overall hit rates were .64 for kindergartners and .69 for third graders, which were still well above the nominal chance level of .5.

The main effect for prior testing status is the critical datum. It shows that there was a true-memory inoculation effect. Hit rates were higher on the delayed test for previously tested targets than for previously untested targets in both 5-year-olds (.67 vs. .62) and 8-year-olds (.62 vs. .76). A Newman-Keuls analysis of the Age × Prior Testing Status interaction showed that the tested-induced increase in hit rates was greater in older children than in younger children (i.e., the true-memory inoculation effect increased with age).

The other main effects were predictable and therefore less interesting. The age main effect was due to the fact that hit rates were higher for older children and for high-learning targets. However, the Degree of Learning × Prior Testing Status interaction produced instructive findings. Newman-Keuls analyses of this interaction showed that prior testing produced more inoculation for high-learning targets than for low-learning targets at both age levels. Consequently, retention of high-learning targets benefited more from an immediate memory test.

Finally, it is well known from prior studies of children's recognition (Brainerd, Reyna, & Kneer, 1995; Reyna & Kiernan, 1994) that there are strong dependencies between hits on immediate tests and hits on 1-week delayed tests; targets that produce hits on delayed tests are more likely to have also produced hits on immediate tests than are targets that produce misses on delayed tests. The same was true in this experiment: Depending on age level and type of target, the conditional probability that an item that produced a delayed hit had also produced an initial hit varied from .61 to .87.

False-memory creation. To assess false-memory creation, we examined false-alarm rates (distractor acceptance probabilities) on the 1-week test as functions of prior memory testing. The major analysis was a 2 (mean age: 5-year-olds vs. 8-year-olds) × 2 (degree of learning: high vs. low) × 2 (prior testing status: tested vs. untested) × 2 (type of distractor: categories vs. exemplars) ANOVA of false-alarm rates on the 1-week test for the two groups of distractors that shared meaning with presented items. (Analyses for unrelated distractors are reported separately below.) This ANOVA produced main effects for age, prior testing status, and type of distractor, plus an Age × Prior Testing Status interaction and a Degree of Learning × Prior Testing Status interaction (all Fs > 5.9 and all ps $< .005$). These results are plotted in Figure 2.

Again, the main effect for prior testing status is of primary interest, this time because it determines whether distractor acceptance rates were elevated by a prior test. They were. As can be seen in Figure 2, both 5-year-olds (Panel A) and 8-year-olds (Panel B) falsely accepted category names and same-category exemplars at higher rates if those distractors had been previously tested than if they had not been. The overall increase in rates of false-memory acceptance were .17 for category names and .18 for same-category exemplars. The rate of false-memory

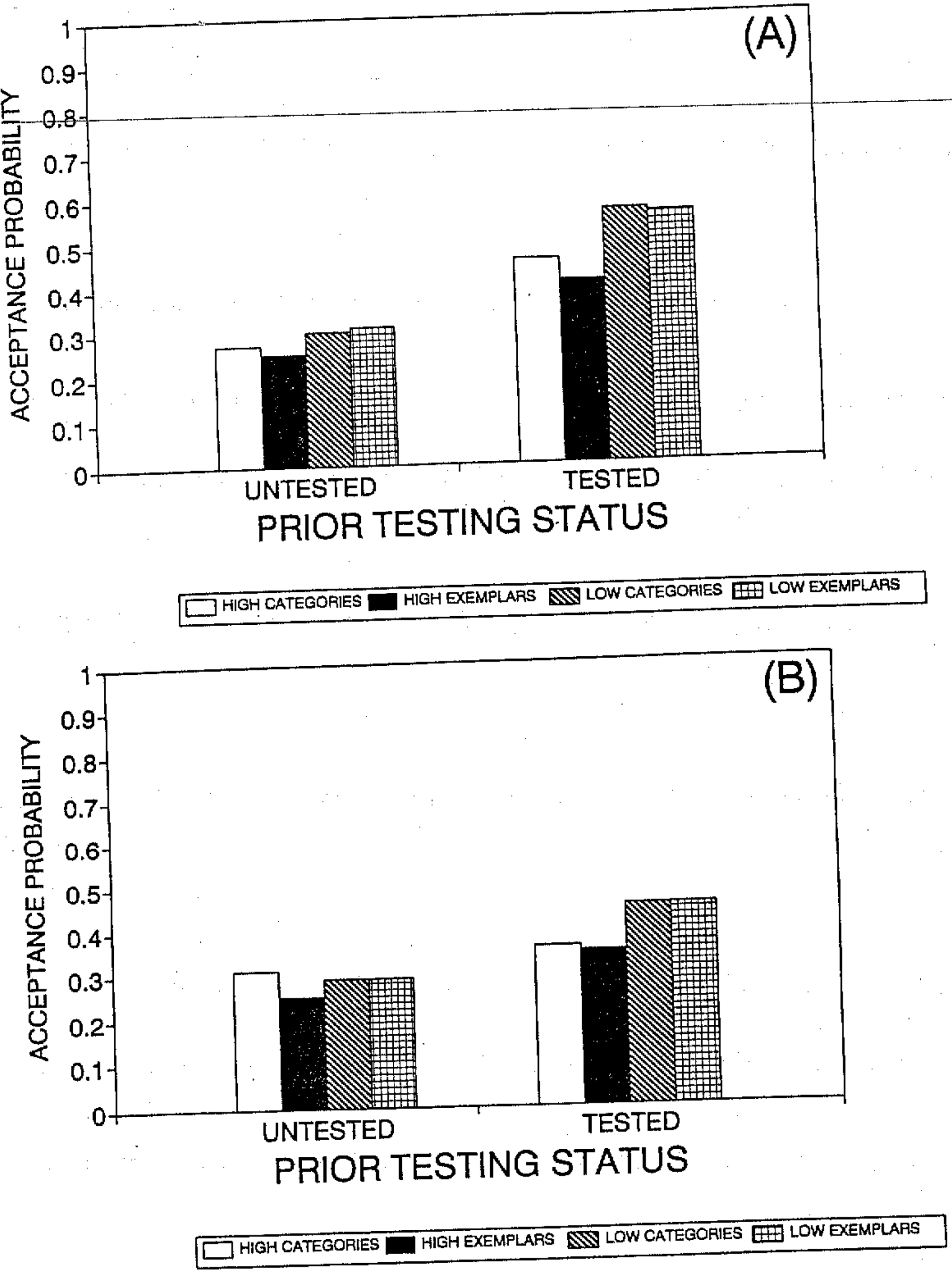

Figure 2. Relationships between distractor acceptance probability (false-alarm rate) and prior testing status in Experiment 1. A: Results for 5-year-olds. B: Results for 8-year-olds.

creation was also higher in kindergartners (.22) than in third graders (.13). Newman-Keuls analysis of the Age × Prior Testing Status interaction showed that this age difference was reliable.

The other main effects were less interesting: False-alarm rates were lower for third graders than for kindergartners and lower for categories than for exemplars. The Degree of Learning × Prior Testing Status interaction was informative, however. The rate of false-memory creation was higher for distractors that shared meaning with low-learning items (.22 overall) than for distractors that shared meaning with high-learning targets (.13 overall). Newman-Keuls analysis of the interaction showed that this difference was reliable.

Turning to unrelated distractors, these items were analyzed separately because the degree-of-learning factor was not defined over them. The analysis was a 2 (age) × 2 (prior testing status) ANOVA. Neither main effect nor the interaction was significant. Thus, although a prior memory test increased false alarms when distractors shared meaning with targets, it did not affect false-alarm rates for unrelated distractors. This confirms the earlier prediction that if initial recognition tests create false memories, the effect will be more pronounced for distractors that share salient meaning with targets than for distractors that do not.

Between the immediate and delayed tests, what was the pattern of increase in false-alarm rates for previously tested distractors? Were distractors that had been falsely accepted on the initial test then falsely accepted on the delayed test, with some previously rejected distractors now being falsely accepted too, or was there little relationship between the groups of distractors that were falsely accepted on the two tests? The answer is that the former pattern prevailed. We have reported elsewhere (Brainerd, Reyna, & Brandse, 1995) that false alarms on an immediate test tend to be preserved on a 1-week delayed test. This was also true in the present experiment: For the two types of distractors that showed false-memory creation effects, the conditional probabilities of delayed false alarms given immediate false alarms were quite high, ranging from a low of .80 to a high of .94.

Gains–losses in accuracy. We have just seen that when distractors shared meaning with presented items, the immediate memory test increased false acceptance rates on the delayed test. For these distractors, the question therefore becomes, Did the immediate test produce net gains or net losses in accuracy on the delayed test? This question can be answered by computing d', the standard measure of the difference between hit rates (true-memory responses) and false-alarm rates (false-memory responses).[3] We computed d' values for each participant for each of the eight possible combinations of learning, prior testing, and distractor type (2 Types of Learning × 2 Types of Prior Testing × 2 Types of Distractors). Mean values of d' for these eight combinations appear by age level in Table 1. The fact that the d' values for previously tested distractors were all positive means that, 1 week after target exposure, children could still distinguish targets that had been presented on the tape recording from distractors that had been presented on the recognition test.

In Table 1, prior testing produces a net gain in accuracy (true-memory inoculation exceeded false-memory creation) for a particular combination of targets and distractors whenever the value in Column 2 was larger than the value in Column 3. Conversely, prior testing produces a net loss in accuracy (false-memory creation exceeded true-memory inoculation) if the value in Column 2 was smaller than the value in Column 3. We tested each of these pairs of d' values for significance using correlated *t* tests.

Table 1
Estimated Values of d' for Experiment 1

	Prior testing status	
Age–distractor–learning	Tested	Untested
Kindergarten		
Categories		
High	0.96	1.35
Low	0.22*	1.15*
Exemplars		
High	1.15	1.12
Low	0.22*	0.98*
Third grade		
Categories		
High	2.12	1.93
Low	1.09	1.44
Exemplars		
High	2.47	2.25
Low	0.74	1.09

* $p < .05$.

Unlike recall, prior recognition testing did not produce significant net gains in accuracy on the delayed test. However, the results were different for younger and older children. Concerning 5-year-olds, prior testing produced a net loss in accuracy in all four comparisons—that is, the increase in false-memory responses exceeded the increase in true-memory responses. As can be seen, the loss was significant in two of the four comparisons. Concerning older children, prior testing produced a net loss in accuracy in two comparisons and a net gain in accuracy in the other two comparisons. None of these differences was significant, however.

Summary. The present experiment showed that an initial recognition test, like an initial recall test, inoculates true memories against forgetting, but unlike an initial recall test, it also creates false memories. As predicted, false-memory creation was more pronounced for distractors that shared meaning with targets than for distractors that did not. There was also an interaction between age and amount of false-memory creation. Among younger children, the immediate test produced significant declines in accuracy on the delayed test in two of the four comparisons. Among older children, the immediate test produced neither net gains nor net losses. Finally, the immediate test elevated false-alarm rates more for distractors that shared meaning with low-learning targets than for distractors that shared meaning with high-learning targets.

Experiment 2

The aim of this experiment was to determine how the preceding results vary when the initial recognition test does not occur until after a forgetting interval. This issue is of more than passing interest because delays are the norm in forensic interviewing (Brainerd & Ornstein, 1991). In Experiment 1, the initial test occurred a few minutes after target exposures, and the delayed test occurred 1 week later. In Experiment 2, there were three

---

[3] The statistic d' is $(p_T - p_D)/SD(p_D)$. Related statistics that give very similar results are reviewed in McNicol (1972).

s, spaced 1 week apart. Target exposures occurred dur-
sion 1. The initial recognition test was administered dur-
ssion 2, 1 week later. The delayed test, which was again
o measure true-memory inoculation and false-memory
n, was administered during Session 3, one week after Ses-
 Apart from these modifications, the designs of the two
ments were identical.

v should the 1-week delay affect the initial test's tendency
culate true memories and create false ones? The answer
 former case is suggested both by empirical results and
tical predictions. Empirically, inoculation effects in recall
s have been found to decrease as the delay between target
ures and initial testing increases (Brainerd et al., 1990;
& White, 1995). Theoretically, these results are predicted
zzy-trace theory's assumption that early testing selectively
ves verbatim memories of target exposure events (Reyna,
 Reyna & Brainerd, 1995). Because verbatim memories
ne rapidly inaccessible over forgetting intervals, the ability
recognition test to preserve such memories should also
ne.

hat about distractors? If initial recognition tests create false
ories, the theoretical expectation is that this effect should
ase after a forgetting interval (Brainerd, in press; Reyna,
; Reyna & Brainerd, 1995). The obvious reason is that
ren's ability to access verbatim memories of target presen-
ns is known to decline as the interval between presentation
test increases (e.g., Brainerd & Gordon, 1994; Brainerd,
na, & Kneer, 1995; Reyna & Kiernan, 1994). Conse-
tly, on the initial test, it will be harder for children to ascer-
 that particular distractors were not presented when that
occurs a week after target presentation than when it occurs
ediately. Moreover, distractors that share gist with targets
 be falsely recognized at higher rates after a delay because
 memories are forgotten more slowly than verbatim
nories.

thod

articipants. A total of 80 children participated in this experiment.
re were 40 kindergartners (mean age = 5 years 9 months) and 40
d graders (mean age = 8 years 11 months). These children were
vn from the same sources as those in Experiment 1. As before, half
children at each age level were boys and half were girls.

ists and procedures. The methodological details (lists, instruc-
s, etc.) of this experiment were identical to those of Experiment 1,
ept for four modifications. First, there were three sessions, rather
n two, spaced 1 week apart. Second, Session 1 ended after the target
osure phase had been completed. Third, the initial recognition test
 administered during Session 2, 1 week later. Fourth, the delayed
ognition test, which was used to measure true-memory inoculation
 false-memory creation, was administered during Session 3, 1 week
r Session 2.

sults and Discussion

As in Experiment 1, we report the results for true-memory
oculation first, the results for false-memory creation second,
d the results for overall gains–losses in accuracy last. Before
oceeding, however, we note that although the initial recogni-
n test occurred one week after target presentation, mean hit rates (.54 for kindergartners and .59 for third graders) were above the nominal chance level of .5.

True-memory inoculation. Hit rates during Session 3 (2-week delay) were used to measure the true-memory inoculation rate. As in Experiment 1, the major analysis was a 2 (age) × 2 (degree of learning) × 2 (prior testing status) ANOVA of participants' hit probabilities on the 2-week test. This ANOVA revealed main effects for all three factors and an Age × Prior Testing Status interaction (all $Fs > 3.9$ and all $ps < .05$).

As before, the main effect for prior testing status is the critical result because it shows that there was a true-memory inoculation effect in both 5-year-olds (overall hit rates of .54 vs. .50 for tested vs. untested targets) and 8-year-olds (overall hit rates of .60 vs. .53). This effect was much smaller than in Experiment 1, which is consistent with fuzzy-trace theory's verbatim-preservation explanation. The overall increase in hit rate on the 2-week test as a function of prior testing was significant, but it was only about half as large as in Experiment 1 (.06 vs. .10). Further, Newman-Keuls analysis of the Age × Prior Testing Status interaction showed that the inoculation effect was confined to 8-year-olds; younger children did not show significant inoculation.

The only other results for targets, the main effects for age and for degree of learning, were predictable. Hit rates on the 2-week test were higher for 8-year-olds than for 5-year-olds, and they were higher for targets that had been presented three times than for targets that had been presented once.

False-memory creation. These analyses focused on false-alarm rates on the 2-week test as a function of whether distractors had appeared on the 1-week test. The main analysis was a 2 (age) × 2 (degree of learning) × 2 (prior testing status) × 2 (type of distractor) ANOVA of participants' false-alarm probabilities for category names and exemplars on the 2-week test. This ANOVA revealed main effects for the first three factors, an Age × Prior Testing Status interaction, and a Degree of Learning × Prior Testing Status interaction (all $Fs > 5.7$ and all $ps < .005$). False-alarm rates are plotted by age and condition in Figure 3.

The main effect of prior testing status continues to be of central interest because it determines whether there was statistically significant false-memory creation as a function of prior testing. There was. Both 5-year-olds (Figure 3, Panel A) and 8-year-olds (Figure 3, Panel B) displayed higher false-alarm rates for category names and same-category exemplars on the 2-week test for items that appeared on the 1-week test than for items that did not. Like Experiment 1, the overall rate of false-memory creation was greater in kindergartners (.38) than in third graders (.29). A Newman-Keuls analysis of the Age × Prior Testing Status interaction showed that this age difference was reliable. Also, note that the average rate of false-memory creation in this experiment (.34) was nearly twice the average rate (.18) in Experiment 1.

As in Experiment 1, false-memory creation was inversely related to learning level. The rate for distractors that shared meaning with low-learning targets (.41) was higher than for distractors that shared meaning with high-learning targets (.26). Newman-Keuls analysis of the Degree of Learning × Prior Testing Status interaction showed that this difference was reliable.

The other two main effects, age and degree of learning, were

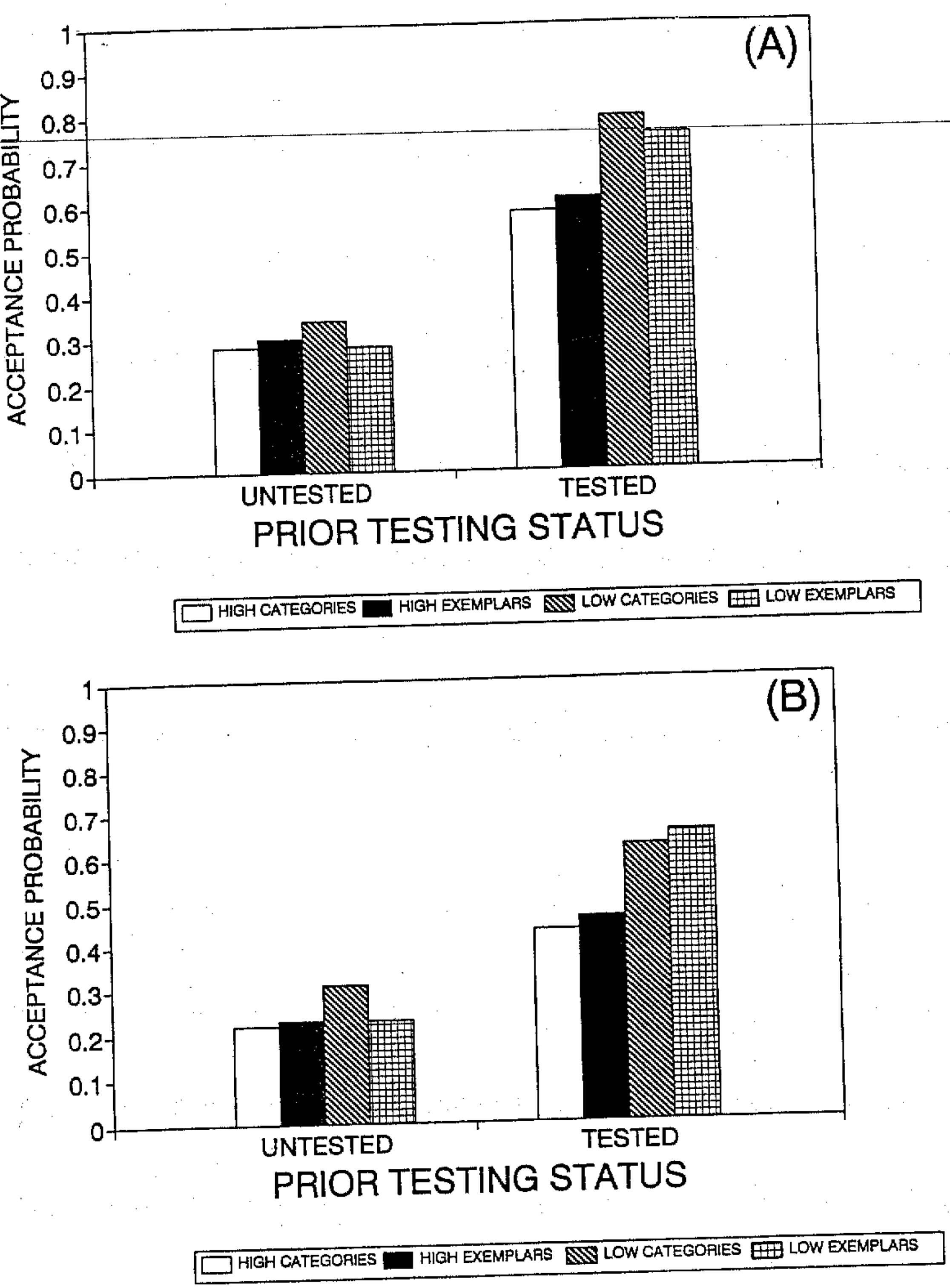

Figure 3. Relationships between distractor acceptance probability (false-alarm rate) and prior testing status in Experiment 2. A: Results for 5-year-olds. B: Results for 8-year-olds.

less interesting. As can be seen in Figure 3, false-alarm rates were higher overall for younger children and for distractors that shared meaning with targets that had been presented three times.

Finally, the results for unrelated distractors were somewhat different than those of Experiment 1. In Experiment 1, the immediate recognition test had no effect on the false-alarm rates for unrelated distractors on the 1-week test; false-memory cre-

alues of d' for Experiment 2

r–learning	Prior testing status Tested	 Untested
arten		
ories		
gh	−0.08*	0.89*
w	−0.96*	-.48*
plars		
gh	−0.19*	0.96*
w	−0.72*	0.68*
lated	0.79*	1.41*
rade		
gories		
igh	1.00*	1.68*
ow	−0.29*	0.78*
mplars		
igh	0.86*	1.48*
ow	−0.39*	1.12*
elated	1.92	1.60

confined to meaning-sharing distractors. In this ex-
however, a 2 (age) × 2 (prior testing status) ANOVA
umbers of acceptances for unrelated distractors on the
st produced main effects for both factors and an Age
Testing Status interaction. Newman-Keuls analysis of
ction showed that for kindergartners, the false-alarm
previously tested distractors (.35) was significantly
an the false-alarm rate for previously untested distrac-
). This was not true for third graders. It should be
at, in kindergartners, the magnitude of the false-mem-
ion effect was still much larger for gist-sharing distrac-
for unrelated distractors.

–losses in accuracy. Relative to Experiment 1, the *d'*
for this experiment produced stronger evidence of net
losses as a function of mere memory testing. As in
ent 1, we computed *d'* to measure differences in the
etween hits (true-memory responses) and false alarms
emory responses) as a function of prior memory test-
cifically, *d'* values were computed for each participant
1 possible combinations of learning, prior testing, and
distractor. (There were only 11 combinations, rather
, because the level-of-learning factor was not defined
related distractors.) Unlike Experiment 1, we included
ed distractors, as well as category names and same-cate-
emplars, in this analysis because, as mentioned, there
false-memory creation effect for unrelated distractors.
d' values for the treatment combinations appear by age
Table 2.

, consider the results for kindergartners. In Experiment
ound that prior testing produced statistically significant
n accuracy for meaning-sharing distractors in two out of
omparisons (cf. Table 1). In this experiment, it can be
Table 2 that there were significant losses in all four com-
ns for meaning-sharing distractors. Also, the average ab-
difference between the *d'* values for tested versus untested
ns of these distractors was larger than it had been in Experiment 1; a delayed initial test produced larger net accuracy losses for these distractors than an immediate initial test. Note, further, that all four *d'* values in Column 2 are negative, which means that false-alarm rates for previously tested distractors were higher than hit rates for previously targets. For unrelated distractors, unlike Experiment 1, a prior memory test produced a net loss in accuracy.

Second, consider the results for third graders. In Experiment 1, we found that prior testing produced neither net gains nor net losses in accuracy for meaning-sharing distractors (cf. Table 1). In the present experiment, however, it produced net accuracy losses in all four comparisons. Moreover, the average absolute difference between the *d'* values for tested versus untested versions of these distractors was more than four times the average difference in Experiment 1. For unrelated distractors, a prior memory test had no net effect on accuracy.

Summary. Administering the initial memory test 1 week after target exposures inoculated true-memory responses against forgetting on a second test that was administered 2 weeks after target exposures. However, the effect was smaller than in Experiment 1, and it was not significant for younger children. The 1-week test also increased false-alarm rates for meaning-sharing distractors on the 2-week test. Like Experiment 1, this false-memory creation effect was present at both age levels, but it was more marked for 5-year-olds. Importantly, increases in false-alarm rates as a function of prior testing were larger than in Experiment 1. As a result, when distractors shared meaning with targets, a prior memory test produced net losses in accuracy at both age levels, with the losses being much larger for 5-year-olds than for 8-year-olds. Another difference between the results of this experiment and those of Experiment 1 is that there was a false-memory creation effect for unrelated distractors among 5-year-olds.

General Discussion

It has been previously reported that initial recall tests increase the incidence of true-memory responses on delayed recall tests, but not the incidence of false-memory responses (Brainerd & Ornstein, 1991; Poole & White, 1995). Although this finding has been consistently obtained, recognition tests could produce different results. This possibility was explored in the present experiments.

On the one hand, we found that initial recognition tests, like initial recall tests, inoculated true-memory responses (hits) against forgetting on delayed tests. Also like recall, inoculation was positively related to goodness of target memories: The effect was larger for older children (better memory) than for younger children (poorer memory), larger for targets that had been presented three times (better memory) than for those that had been presented once (poorer memory), and larger when the inoculating test was administered immediately after target exposures (better memory) than when it was administered 1 week later (poorer memory). Because these variables are known to affect verbatim memory more than gist memory (Brainerd, Reyna, & Kneer, 1995), such results are congruent with the verbatim-preservation interpretation of true-memory inoculation (Reyna, 1995; Reyna & Brainerd, 1995).

On the other hand, we found that initial recognition tests, unlike initial recall tests, increased false-memory responses (false alarms) on delayed tests. When a recognition test was administered immediately after target exposures (Experiment 1), false-alarm rates for tested distractors that shared meaning with targets (category names and same-category exemplars) were elevated on a 1-week delayed test, relative to previously untested distractors. In addition, this false-memory creation effect was inversely related to goodness of target memory: It was larger for younger children than for older children, and it was larger for distractors that shared meaning with low-learning targets than with high-learning targets. When a recognition test was administered 1 week after target exposures (Experiment 2), again, false-alarm rates for tested distractors that shared meaning with targets were elevated on a 2-week delayed test, relative to previously untested distractors. However, these effects were larger than they had been in Experiment 1, and a false memory-creation effect was now observed for unrelated distractors in younger children.

Because prior recognition tests elevate both false-memory responses and true-memory responses on subsequent tests, their overall impact on accuracy depends on which effect is larger in a particular context. In our experiments, we answered this question in the obvious way by computing *d′* values. The results for younger children were consistent across the two experiments. Regardless of whether the prior recognition test was administered immediately or after a 1-week delay, false-alarm rates for meaning-sharing distractors were elevated more than hit rates for the corresponding targets, the difference being significant only for the low-learning condition in Experiment 1 but for both the low- and high-learning conditions in Experiment 2. The results for older children depended on how soon the initial test was administered. When it was administered immediately, the test produced neither net gains nor net losses in accuracy for meaning-sharing distractors. But when it was administered after a 1-week delay, the test produced net losses in accuracy, the difference being significant for both the low- and high-learning conditions.

To conclude this article, we consider two further questions: What are the potential implications of our findings for children's testimony? How are the effects of mere recognition testing to be explained in theoretical terms?

Children's Testimony

That prior recognition testing increases children's false alarms on later tests, sometimes producing net losses in accuracy, is of considerable forensic interest. Indeed, these results, together with those of Reyna and Kiernan (1994), provide a scientific basis for recent recommendations in the psycholegal literature concerning the types of questions that are appropriate in forensic interviews of children. As mentioned earlier, recognition questions are common in such interviews. Lately, however, a consensus has formed, in both Great Britain (e.g., Bull, 1992) and North America (e.g., Yuille, Hunter, Joffe, & Zaparniuk, 1993), that the use of recognition questions should be minimized.

Unfortunately, this recommendation is not data-based, but is predicated instead on rationales of the following sort. As previously noted, the questions in a forensic interview deal with possible events that are related to some focal issue that is under investigation (e.g., a robbery that a child witnessed). The interviewer lacks knowledge of the actual events, which means that recognition questions will necessarily state things that did not happen (e.g., Was the robber holding a knife?) as well as things that did (e.g., Was the robber wearing a mask?). Because false events, like true ones, fit with the core gist of the interview, recognition questions may contaminate subsequent testimony by encouraging false memories of gist-consistent events (e.g., *Memorandum of Good Practice,* 1992).

In line with such rationales, Reyna and Kiernan (1994) reported large increases in false-alarm rates on a 1-week delayed test for distractors that had received a prior recognition test. The target memories in their experiments were similar to those investigated in forensic interviews in that they dealt with everyday objects and actions. Specifically, children (first and third graders) listened to a series of short vignettes describing everyday situations in which familiar objects (e.g., foods, animals) and relationships (e.g., in front of, bigger than) figured. Objects and relationships that had been presented in the vignettes (targets) and objects and relationships that had not been presented (distractors) then appeared as items on an immediate recognition test and on a delayed test. Pooling across age levels, the increase in false-alarm rates between the immediate and delayed tests ranged from 23% to 126%, depending on condition, with a mean increase of 55%. Reyna (in press) and Reyna and Kiernan (1995) observed similar increases in false-alarm rates for sentences that described familiar objects and events.

Although such findings are consistent with the hypothesis that prior recognition testing contaminates subsequent testimony by creating false memories, the support is indirect because Reyna and Kiernan's (1994) design did not allow delayed false-alarm rates for previously untested distractors to be compared to delayed false-alarm rates for previously tested distractors. Our design permitted such comparisons, and the results confirmed the hypothesis. In Experiment 1, pooling across the two age levels, the percent increases in false-alarm rates for previously tested distractors were in the neighborhood of 50% for both categories and exemplars, whereas the corresponding false-alarm rates for untested categories and exemplars did not increase significantly. In Experiment 2, there were again large increases in false-alarm rates for previously tested category and exemplar distractors, but false-alarm rates for the corresponding untested distractors again failed to increase significantly.

There is a residual methodological difference between recognition questions in forensic interviews and the present recognition tests that deserves mention. That difference concerns children's awareness that many of the items are not memory targets. In our experiments, and in Reyna and Kiernan's (1994), children were cautioned that many test items would be distractors to which they should respond "no." In forensic interviews, children typically do not receive such instructions. From the perspective of false-memory creation, therefore, our results are conservative; these tests should be less likely to create false memories than forensic interviews because children are explicitly warned to anticipate false items.

Explaining the Effects of Mere Memory Testing

What theoretical mechanisms are responsible for true-memory inoculation and false-memory creation? In the latter conon, we mentioned that fuzzy-trace theory provides iple mechanisms whereby mere recognition testing could ce false memories (Reyna, 1995; Reyna & Brainerd,). Those mechanisms come from the theory's general anal- of the development of recognition memory (Brainerd & na, 1995; Brainerd, Reyna, & Kneer, 1995; Reyna, 1992; na & Brainerd, 1995; Reyna & Kiernan, 1994, 1995). A ral feature of that analysis is the notion that the processes underlie true-memory inoculation and false-memory cre- n are different. The former, as previously noted, are as- ed to involve the preservation of verbatim traces of target sure events, whereas the latter are assumed to involve ei- the formation of verbatim memories of distractors for h source information subsequently becomes inaccessible e retrieval of gist memories that distractors share with pre- ed targets (Reyna, 1995).

his interpretation is based on three considerations. First, nerd and Gordon (1994) and Reyna and Kiernan (1994, 5) demonstrated that when targets are presented for study, dren store dissociated representations of the surface forms e presentation events (verbatim traces) and of the various es and meanings that those events instantiate (gist traces). nd, on immediate recognition tests, it has been found that h targets are correctly accepted, it is predominantly because cue the retrieval of verbatim traces of their presentation ng the study phase (Reyna & Kiernan, 1994). Third, in rast, it has been found that when distractors are falsely ac- ed, it is predominantly because they cue the retrieval of gist es that they share with presented targets (Brainerd & na, 1995; Brainerd, Reyna, & Kneer, 1995).

ecause targets cue the retrieval of verbatim traces, recogni- probes for targets strengthen such traces (e.g., through ntegration) and therefore help to preserve them across for- ng intervals (Reyna, 1995). A number of findings that fa- this verbatim-preservation interpretation have been re- ed by eyewitness-memory researchers (Poole & White, , 1995; Warren & Lane, 1995). This interpretation is also istent with developmental studies of a closely related phe- enon, reinstatement (Howe, Courage, & Bryant-Brown, ; Rovee-Collier & Shyi, 1992). In the reinstatement para- , cuing selected surface features of a previously studied t improves subsequent recognition of that target. Like true- ory inoculation, the reinstatement effect varies with age as a function of the delay between target exposure and g.

e situation is more complex for distractors. On the one , because distractors cue the retrieval of gist traces, false ns can be supported by gist memories that were stored dur- e original study phase (Brainerd, Reyna, & Kneer, 1995; na & Kiernan, 1994). On the other hand, once a distractor e has been presented on an initial recognition test, a verba- race of that event will be in storage. Later, on a delayed presentation of the same probe could cue retrieval of that tim trace. If source information about it has been lost dur- e forgetting interval (i.e., if children can no longer tell that ace was stored during the test phase), such retrieval would te false-alarm rates (relative to previously untested ctors).

important objective for future research, therefore, is to determine whether the false-memory creation effect is gist-based, verbatim-based, or both. Although definitive answers must await further experimentation, the gist method of false-memory creation would at least appear to be implicated by the findings for target-resembling distractors (categories and exemplars). Because such distractors had meanings that overlapped with those of targets, they would have been especially likely to cue the retrieval of shared gist on the initial test. Under the gist method of false-memory creation, these distractors should therefore have produced larger increases in false-alarm rates as a consequence of prior testing than distractors that did not share gist with targets. This result was obtained at both age levels in both experiments. Indeed, only gist-sharing distractors displayed such increases in Experiment 1.

Evidence bearing on the verbatim method of false-memory creation is more mixed. On the negative side, there is the fact that no false-memory creation effect occurred in Experiment 1 for distractors whose meanings did not overlap with those of targets. Clearly, under the verbatim method, previously tested distractors that do not share salient meaning with targets should exhibit higher delayed false-alarm rates. Another result that argues against the verbatim method can be found in studies concerned with the effects of physical similarity between the study phase and the initial test. Increasing physical similarity makes source information about the test phase less discriminable from source information about the study phase (cf. Ackerman, 1994), which should translate into higher delayed false-alarm rates if verbatim traces of prior distractor presentations are being retrieved. In fact, increasing the physical similarity between the study phase and the initial test has been found to lower delayed false-alarm rates (Brainerd, Reyna, & Kneer, 1995).

Despite such negative evidence, other findings suggest that the verbatim method may contribute to the false-memory creation effect. For instance, although this effect was not observed for unrelated distractors at either age level in Experiment 1, younger children exhibited the effect for unrelated distractors in Experiment 2. In addition, Reyna (1995) reviewed evidence that the increases in delayed false-alarm rates that are usually obtained for distractors that were previously presented with accompanying misinformation is due to retrieval of verbatim traces of those presentation events. So, although the gist method of false-memory creation was most directly implicated by our results, the verbatim method remains a possibility.

References

Ackerman, B. P. (1994). Children's source errors in referential communication. *Journal of Experimental Child Psychology, 58,* 432–464.

Baddeley, A. D. (1976). *The psychology of memory.* New York: Basic Books.

Baker-Ward, L., Hess, T. M., & Flannagan, D. A. (1990). The effects of involvement on children's memory for events. *Cognitive Development, 5,* 55–69.

Battig, W. F., & Montague, W. E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph, 80* (3, Pt. 2).

Bjorklund, D. F., & Muir, J. E. (1988). Children's development of free recall memory. *Annals of Child Development, 5,* 79–123.

Brainerd, C. J. (in press). Children's forgetting, with implications for memory suggestibility. In N. L. Stein, P. A. Ornstein, B. Tversky, & C. J. Brainerd (Eds.), *Memory for everyday and emotional events.* Hillsdale, NJ: Erlbaum.

Brainerd, C. J., & Gordon, L. L. (1994). Development of verbatim and gist memory for numbers. *Developmental Psychology, 30,* 163–177.

Brainerd, C. J., & Ornstein, P. A. (1991). Children's memory for witnessed events: The developmental backdrop. In J. Doris (Ed.), *The suggestibility of children's recollections* (pp. 10–20). Washington, DC: American Psychological Association.

Brainerd, C. J., & Reyna, V. F. (1995). Autosuggestibility in memory development. *Cognitive Psychology, 28,* 65–101.

Brainerd, C. J., Reyna, V. F., & Brandse, E. (1995). Are children's false memories more persistent than their true memories? *Psychological Science, 6,* 359–364.

Brainerd, C. J., Reyna, V. F., Howe, M. L., & Kingma, J. (1990). The development of forgetting and reminiscence. *Monographs of the Society for Research in Child Development, 55,* (3–4, Serial No. 222).

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995). False-recognition reversal: When similarity is distinctive. *Journal of Memory and Language, 34,* 157–185.

Brown, J., & Packham, D. W. (1967). The effect of prior recall on multiple response recognition. *Quarterly Journal of Experimental Psychology, 19,* 356–361.

Bruck, M., & Ceci, S. J. (in press). The nature of applied and basic research on children's suggestibility. In N. Stein, P. A. Ornstein, B. Tversky, & C. J. Brainerd (Eds.), *Memory for everyday and emotional events.* Hillsdale, NJ: Erlbaum.

Bruck, M., Ceci, S. J., Francoeur, E., & Barr, R. (1995). "I hardly cried when I got my shot!": Influencing children's reports about a visit to their pediatrician. *Child Development, 66,* 193–208.

Bull, R. (1992). Obtaining evidence expertly: The reliability of interviews with child witnesses. *The International Digest of Human Behaviour Science and Law, 1,* 5–12.

Cassel, W. S., & Bjorklund, D. F. (1995). Developmental patterns of eyewitness memory and suggestibility: An ecologically based short-term longitudinal study. *Law and Human Behavior, 19,* 507–532.

Cassel, W. W., Roebers, C. E. M., & Bjorklund, D. F. (in press). Developmental patterns of eyewitness responses to repeated and increasingly suggestive questions. *Journal of Experimental Child Psychology.*

Ceci, S. J., & Bruck, M. (1993a). Child witnesses: Translating research into policy. *SRCD Social Policy Report, 7*(3), 1–30.

Ceci, S. J., & Bruck, M. (1993b). The suggestibility of the child witness: A historical review. *Psychological Bulletin, 113,* 403–439.

Dent, H. R., & Stephenson, G. M. (1979). An experimental study of the effectiveness of different techniques of questioning child witnesses. *British Journal of Social and Clinical Psychology, 18,* 41–51.

Goodman, G. S., Bottoms, B. L., Schwartz-Kenney, B., & Rudy, L. (1991). Children's memory for a stressful event: Improving children's reports. *Journal of Narrative and Life History, 1,* 69–99.

Handford, M. (1987). *Where's Waldo?* Boston, MA: Little, Brown.

Howe, M. L., Courage, M. L., & Bryant-Brown, L. (1993). Reinstating preschoolers' memories. *Developmental Psychology, 29,* 854–869.

Mandler, G., Pearlstone, Z., & Koopmans, H. S. (1969). Effects of organization and semantic similarity on a recall and recognition task. *Journal of Verbal Learning and Verbal Behavior, 8,* 410–423.

McNicol, D. (1972). *A primer of signal detection theory.* London: Allen & Unwin.

Memorandum of good practice (1992). London: Her Majesty's Stationary Office.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph, 76,* (Number 1, Pt. 2).

Poole, D. A., & Lindsay, D. S. (1995). Interviewing preschoolers: Effects of nonsuggestive techniques, parental coaching, and learning questions on reports of nonexperienced events. *Journal of Experimental Child Psychology, 60,* 129–154.

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

Poole, D. A., & White, L. T. (1993). Two years later: Effects of question repetition and retention interval on the eyewitness testimony of children and adults. *Developmental Psychology, 29,* 844–853.

Poole, D. A., & White, L. T. (1995). Tell me again and again: Stability and change in the repeated testimonies of children and adults. In M. Zaragoza, J. R. Graham, G. N. N. Hall, R. Hirschman, & Y. S. Ben-Porath, (Eds.), *Memory, suggestibility, and eyewitness testimony in children and adults* (pp. 24–43). Thousand Oaks, CA: Sage.

Reyna, V. F. (1992). Reasoning, remembering, and their relationship: Social, cognitive, and developmental issues. In M. L. Howe, C. J. Brainerd, & V. F. Reyna (Eds.), *Development of long-term retention* (pp. 103–127). New York: Springer-Verlag.

Reyna, V. F. (1995). Interference effects in memory and reasoning: A fuzzy-trace theory analysis. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 29–61). San Diego, CA: Academic Press.

Reyna, V. F. (in press). Meaning, memory, and the interpretation of metaphors. In J. Mio & A. Katz (Eds.), *Metaphor: Pragmatics and applications.* Hillsdale, NJ: Erlbaum.

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7,* 1–75.

Reyna, V. F., & Kiernan, B. (1994). The development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic content, encoding instructions, and retention interval. *Developmental Psychology, 30,* 178–191.

Reyna, V. F., & Kiernan, B. (1995). Children's memory and interpretation of psychological metaphors. *Metaphor and Symbolic Activity, 10,* 309–331.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21,* 803–815.

Rovee-Collier, C., & Shyi, G. (1992). A functional and cognitive analysis of infant long-term retention. In M. L. Howe, C. J. Brainerd, & V. F. Reyna (Eds.), *Development of long-term retention* (pp. 5–55). New York: Springer-Verlag.

Saywitz, K. J. (1995). Improving children's testimony: The question, the answer, and the environment. In M. S. Zaragoza, J. R. Graham, G. C. N. Hall, R. Hirschman, & Y. S. Ben-Porath (Eds.), *Memory and testimony in the child witness.* Thousand Oaks, CA: Sage.

Warren, A. R., & Lane, P. L. (1995). Effects of timing and type of questioning on eyewitness accuracy and suggestibility. In M. Zaragoza, J. R. Graham, G. N. N. Hall, R. Hirschman, & Y. S. Ben-Porath, (Eds.), *Memory, suggestibility, and eyewitness testimony in children and adults.* (pp. 44–60). Thousand Oaks, CA: Sage.

Yuille, J. C., Hunter, R., Joffe, R., & Zaparniuk, J. (1993). Interviewing children in sexual abuse cases. In G. S. Goodman & B. L. Bottoms (Eds.), *Child victims, child witnesses.* New York: Guilford Press.

Received November 9, 1994
Revision received March 24, 1995
Accepted June 14, 1995 ■