Child Development, September/October 2002, Volume 73, Number 5, Pages 1363–1377

Are Young Children Susceptible to the False-Memory Illusion?

C. J. Brainerd, V. F. Reyna, and T. J. Forrest

False memories have typically been found to be more common during early childhood than during later childhood or adulthood. However, fuzzy-trace theory makes the counterintuitive prediction that some powerful forms of adult false memory will be greatly attenuated in early childhood, an important example being the Deese/Roediger/McDermott (DRM) illusion. Three developmental studies of this illusion ($N = 282$) found that (1) it was at near-floor levels in young children, (2) it was still below adult levels by early adolescence, and (3) the low levels of the illusion in young children may be due to failure to "get the gist" of DRM materials.

INTRODUCTION

Research on children's false memories has been intense in recent years, owing chiefly to concerns about false reports in certain types of legal cases (e.g., *State v. Michaels*, 1994). A core finding is that young children are often especially vulnerable to false memories, perhaps because verbatim memory for the actual events of their lives is poor. For instance, when false information is explicitly suggested to them (e.g., "Remember when you drank that Coke™ at the baseball game?"), young children are more likely than are older children (or adults) to incorporate it into their memory reports (for reviews, see Bruck & Ceci, 1997; Ceci & Bruck, 1995; Howe, 2000). Likewise, in developmental studies of spontaneous false memory, a standard finding is that mistaken reports of events that are congruent with the gist of experience (e.g., eating a hot dog at the baseball game) decline between early childhood and young adulthood (for reviews, see Brainerd & Reyna, 2001; Reyna, 1996). Such developmental decreases in spontaneous false memories of gist-consistent events have been reported for a wide variety of materials, such as narratives (Ackerman, 1992, 1994; Reyna & Kiernan, 1994, 1995), live event sequences (Pipe, Gee, Wilson, & Egerton, 1999; Poole & White, 1991), and word lists (Brainerd & Reyna, 1996; Brainerd, Reyna, & Kneer, 1995). Further, it has long been known that intrusions of unpresented items in free recall of word lists are more frequent in young children than in older children or adults (for a review, see Bjorklund & Muir, 1988).

Despite extensive evidence of developmental reductions in false-memory reports, there are theoretical grounds for predicting that some important forms of false memory will exhibit dramatic developmental increases and will therefore be of less concern with younger children than with older children or adults. Such predictions fall out of fuzzy-trace theory's dual-process analysis of false memory (Ceci & Bruck, 1998; Reyna & Brainerd, 1995). Fuzzy-trace theory assumes that children store dissociated representations of both the exact surface form of their experiences (verbatim traces: "drank 7-Up™," "ate a hamburger") and their understanding of the meaning of those experiences (gist traces: "went to a baseball game"). Children's later recountings of their experiences on memory tests are mixtures of verbatim and gist processing. According to fuzzy-trace theory, false-memory reports are of two basic types (Brainerd & Reyna, 1998a; Reyna & Titcomb, 1997): (1) recalling or recognizing false events because alternative verbatim traces are retrieved (e.g., "drank a Coke") rather than verbatim traces of actual experiences (e.g., "drank a 7-Up"), and (2) recalling or recognizing false events (e.g., eating a hot dog) because gist traces of experience are retrieved (e.g., "went to a baseball game") rather than verbatim traces of actual experiences (e.g., "ate a hamburger"). Concerning the first type of false report, memory suggestions ("Remember when you drank that Coke at the baseball game?") and episodic confusions (drinking a Coke at home after the game) are typical sources of alternative verbatim traces (Reyna & Titcomb, 1997). Concerning the second type of false report, when children process the gist of experience, false-but-gist-consistent events seem very familiar (e.g., eating hot dogs is commonplace at baseball games), which makes it seem as though those events were part of the index experience (Brainerd & Reyna, 1998b).

It follows from this analysis that false reports can exhibit developmental increases when two conditions are met (see Brainerd & Reyna, 1998a; Ceci & Bruck, 1998; Reyna & Brainerd, 1998). The first and more obvious one is that the target false reports depend on gist-memory abilities that develop slowly, so that false-but-gist-consistent events do not seem as

 0009-3920/2002/7305-0003

familiar to young children as they do to older children or adults because the memories that would make them seem so are less likely to be processed. The other, less obvious, condition is that it is difficult to suppress the target false reports by retrieving verbatim traces of actual experiences, so that normal developmental improvements in verbatim memory fail to neutralize the effects of developmental improvements in gist memory. Of course, various lines of evidence confirm that verbatim and gist abilities both improve with age (Reyna, 1996). Concerning verbatim memory, research shows that the ability to remember the exact surface form of experience (e.g., the voice, loudness, font, color, or size of the individual words on a word list) undergoes substantial improvement between early childhood and early adolescence (e.g., Brainerd & Mojardin, 1998; Brainerd & Reyna, 2002). Concerning gist memory, research shows that (1) young children sometimes fail to extract the meaning of individual events as well as do older children, and (2) they have particular limitations when it comes to spontaneously noticing conceptual relations that hold across a series of events whose individual meanings are understood.

With regard to the latter point, extensive documentation of young children's connecting-the-gist limitations can be found in the large literature on free recall of semantically related word lists (for reviews, see Bjorklund, 1987; Bjorklund & Muir, 1988). When adults study lists whose words belong to a few familiar categories (e.g., bird names, city names, color names, tree names), their free-recall protocols exhibit strong semantic clustering (i.e., same-category words are output together), indicating that they have connected the categories across individual exemplars. When young children recall the same types of lists, they do not spontaneously cluster their output by category (Bjorklund, 1987; Schneider & Pressley, 1989). Other findings demonstrate that this is not merely a retrieval deficit but rather is a more basic failure to notice and store conceptual relations among exemplars as they are encountered. For instance, Bjorklund and Hock (1982) found that although 9-year-olds have above-chance clustering scores, they do not spontaneously connect the gist among same-category exemplars as exemplars are presented for study. In their experiment, the latter ability was not present until age 13. A second finding that points to the same conclusion comes from developmental studies of proactive interference. It is well known that adults exhibit meaning-based proactive interference: When they study and recall a sequence of lists whose members belong to the same semantic category (e.g., each list is composed of color names), recall is poorer on each successive list (e.g., Underwood, 1957). In contrast, Bjorklund (1978) found that children do not exhibit meaning-based proactive interference until after age 9.

Returning to developmental trends in false memory, the straightforward prediction from fuzzy-trace theory is that false reports will increase with age when they are due to the processing of familiar gists, especially gists that must be connected across a series of events, as long as verbatim memory cannot readily be used to suppress those false reports. The obvious question, given the many studies that have detected the opposite trend, is whether there are any false-memory tasks that so unambiguously satisfy these conditions that age increases would be predicted. It turns out that the most extensively studied adult false-memory paradigm of recent years—the Deese/Roediger/McDermott (DRM) procedure (Deese, 1959; Roediger & McDermott, 1995)—is such a task.

The DRM procedure is a variant of tasks that have been used in the aforementioned studies of children's free recall of categorized word lists. Rather than study a word list whose members belong to a few common categories, participants study what is, in effect, a one-category list. Specifically, all the words are semantic associates of a critical unpresented word (e.g., BED, REST, AWAKE, TIRED, DREAM, WAKE, SNOOZE, BLANKET, DOZE, SLUMBER, SNORE, and NAP are all semantic associates of SLEEP). Participants then recall as many words as they can remember and are cautioned not to output words of which they are uncertain. The surprising outcome is very high levels of false recall of critical unpresented words. In Roediger and McDermott's (1995) experiments, overall false recall was 48%, which was only slightly lower than overall true recall (60%). When recognition tests were administered, false-alarm rates for critical unpresented words (76% overall) were not significantly lower than hit rates for studied words (78% overall). These high levels of false recall and false recognition have been replicated by many other investigators (e.g., Payne, Elie, Blackwell, & Neuschatz, 1996; Schacter, Verfaellie, & Pradere, 1996; Seamon, Luo, Schwartz, Jones, Lee, & Jones, 2002; Toglia, Neuschatz, & Goodwin, 1999).

Fuzzy-trace theory's analysis of this illusion (e.g., Payne et al., 1996; Reyna & Lloyd, 1997; Seamon et al., 2002) emphasizes both the processing of gist memories of list themes and the difficulty of using verbatim traces of studied words to suppress false reports of critical unpresented words, thereby making it a prime candidate for the production of dramatic developmental increases in false memory. Concerning gist processing, the words on DRM lists repeatedly cue certain meanings (e.g., "sleep"), resulting in very strong gist memories of those list themes. Because gist traces are so strong, they are apt to be retrieved on memory tests, which makes critical unpresented words likely to be recalled or recognized because they are especially good examples of list themes. Developmentally, gist-processing abilities are limited in young children, leading to a prediction of reduced susceptibility to the DRM illusion. With regard to verbatim processing, retrieval of verbatim traces of some studied items is possible, and the number of studied items whose verbatim traces can be accessed on recall or recognition tests will increase with age. However, verbatim retrieval is not particularly helpful in suppressing false reports of critical unpresented words. Because all studied words have similar meanings, the fact that the presentation of, for example, BED, TIRED, and SNORE can be vividly recollected but the presentation of SLEEP cannot be recollected is inconclusive evidence that SLEEP is a false memory.[1] Individuals who have connected the gist are aware that many additional sleep-related words were presented that cannot be clearly recollected. Thus, although older children's and adult's verbatim memory for list words will be superior to younger children's, this is not likely to neutralize the memory-falsifying effects of developmental increases in gist processing.

We report three experiments that were designed to determine, first, whether reduced susceptibility to the DRM illusion could be confirmed in young children and, second, the extent to which there are qualitative differences in the DRM illusion in children (as compared with known findings in adults). Experiment 1 was a free-recall design in which young children (5-year-olds) recalled lists from the original Roediger and McDermott (1995) materials, and levels of false recall were compared with known adult levels. Experiment 2 was also a free-recall design in which young children (5-year-olds) and somewhat older children (7-year-olds) recalled DRM lists that varied in their tendency to produce the false-memory illusion (Stadler, Roediger, & McDermott, 1999). Experiment 3 was a recall-recognition design in which young children (5-year-olds), young adolescents (11-year-olds), and young adults (undergraduates) performed free recall of DRM lists and then responded to a recognition test for those same lists.

[1] An additional complexity of the DRM illusion is that the prior "presentation" of critical unpresented words is often vividly recollected, much as are actual target presentations. However, these phantom recollective experiences occur less frequently than do true recollective experiences for targets (e.g., Brainerd, Wright, Reyna, & Mojardin, 2001; Mather, Henkel, & Johnson, 1997; Payne et al., 1996; Roediger & McDermott, 1995). Thus, despite this complexity, for a certain proportion of DRM lists at least, targets will be vividly recollected but critical unpresented words will not.

EXPERIMENT 1

Method

Participants. Sixty kindergarten children (age: M = 5,7; 30 males, 30 females) participated in this study. The children were pupils of preschools (that included kindergarten classes) and elementary schools serving middle-class residential areas of a southwestern city. Children's participation was secured via letters of parental permission.

Materials and procedure. The lists that were used in this experiment are provided in the first half of the Appendix. These 24 lists were originally developed by Deese (1959) and by Roediger and McDermott (1995). Note that the critical unpresented words for the lists are all familiar, high-frequency words (e.g., BLACK, CHAIR, FOOT) that are in receptive and productive vocabularies of young children.

For the sake of comparability with the adult literature on the DRM illusion, the procedures of this experiment replicated those of many prior adult experiments (e.g., Payne et al., 1996; Toglia et al., 1999). At the start of the experiment, children were given general memory instructions. They were told, in particular, that they would be listening to short lists of "vocabulary words" and that they should pay careful attention to each list as it was read so that they could recall as many words as possible on a later memory test. Individual children then studied and recalled a total of 10 lists that were randomly sampled from the 24 lists provided in the Appendix. Lists were presented in random order. Each child simply listened to an audio recording of List 1 being read in the highest-to-lowest associate order shown in the Appendix (2-s presentation rate). This was immediately followed by 1 min of oral free recall. Children were instructed to recall as many words as possible, and when they stopped, they were prompted to recall any further words that could be remembered. After the List 1 recall was completed, List 2 was presented and recalled, and so on, until List 10 had been presented and recalled. As children performed each free-recall test, the experimenter circled recalled list items on a prepared answer sheet and wrote intrusions in blank spaces that were provided on the answer sheet.

Results and Discussion

Summary statistics appear in the upper half of Table 1 for four types of responses: (1) recall of studied

Table 1 Proportions of Correctly Recalled Studied Words and Intrusions of Critical Unpresented Words, Unpresented Same-Meaning Words, and Words That Appeared on Earlier Lists

	Response			
Experiment	Correct Recall	Intrusions: Critical	Intrusions: Same Meaning	Intrusions: Earlier Lists
Experiment 1	.23	.06	.15	.22
Experiment 2				
5-year-olds				
High	.27	.04	.12	.17
Low	.34	.05	.14	.21
7-year-olds				
High	.36	.08	.11	.16
Low	.40	.09	.12	.19

words (e.g., BED, REST), (2) intrusions of critical unpresented words (e.g., SLEEP), (3) intrusions of unpresented words with similar meanings to words on the study list (e.g., DROWSY, YAWN), and (d) intrusions of words that appeared on earlier lists (e.g., GLASS or DOOR during free recall of the SLEEP list, when the WINDOW list was presented earlier). Concerning Categories 3 and 4, the intrusion data were scored as follows for each free-recall protocol. First, it was determined whether the critical unpresented word had been recalled. Second, when there were intrusions other than the critical unpresented word, intrusions that were words from previously unpresented lists were identified. Third, when there were any remaining intrusions, words with meanings that were similar to words on the just-presented list were identified. Identification of such words was quite easy: The coefficient of agreement between two independent scorers was .96. Only 1% of the intrusions did not fall into Categories 2 through 4; therefore, attention was restricted to these categories.

With regard to the main focus of Experiment 1—the intrusion data—the most dramatic finding was that 5-year-olds were not susceptible to the DRM illusion. Intrusions of critical unpresented words occurred following only 6% of the lists. For these same materials, critical unpresented words intruded following roughly half of the lists in adults (e.g., Payne et al., 1996). There were also striking findings for the other two types of false recall. In adults, intrusion rates are much higher for critical words than for all other unpresented words combined, and virtually all intrusions other than critical words also preserve the themes of studied lists (Payne et al., 1996; Toglia et al., 1999); intrusions of words from previously studied lists are extremely rare. Fuzzy-trace theory explains all of these findings as follows: Because adults readily connect the conceptual dots between list words, they "get the gist" of DRM lists (e.g., "medical words," "names of fruit"), which promotes false recall of gist-consistent words and suppresses false recall of words that violate that gist (Brainerd & Reyna, 1998b).

These standard adult patterns were absent in young children. Remarkably, the rate of false recall for critical words was lower than for either of the other two types of intrusions. The critical unpresented word was recalled for 6% of the lists. However, 15% of the lists produced one or more similar-meaning intrusions, and 22% of the lists produced one or more intrusions of words from earlier lists. (The percentages for the three intrusion categories do not sum to 100% because they are the percentages of lists that produced each type of intrusion.) Paired-sample t tests showed that false recall was lower for critical words than for similar-meaning words, $t(59) = -5.05$, $p < .001$, two-tailed, and for words from previous lists, $t(59) = -5.56$, $p < .001$, two-tailed. Further, intrusions that preserved list themes were not the predominant form of intrusion: The percentage of lists that produced false recall of gist-consistent words did not exceed the percentage of lists that produced false recall of gist-violating words. Words from previous lists intruded at a higher rate than did either critical words or similar-meaning words, which suggests that young children failed to "get the gist" of DRM lists. If they had, they should not have recalled words with wildly different meanings than words on the just-studied list.

Contrary to this interpretation, it might be argued that young children connected the gist across subsets of DRM words but not across entire lists (e.g., they noticed that BED, REST, AWAKE, TIRED, and DREAM were sleep words, but did not notice that WAKE, SNOOZE, BLANKET, DOZE, SLUMBER, SNORE, and NAP were sleep words), so that the metacognitive representation of the list would be as a group of words that specify a common theme plus other words that are not thematically connected. There are three difficulties with this argument. First and most obviously, it is inconsistent with young children's known limitations in noticing conceptual relations between items on word lists (e.g., Bjorklund, 1978; Bjorklund & Hock, 1982). Second, this argument fails to explain the present pattern of intrusions: If the core theme of a DRM list is identified, albeit for only a subset of words, this conceptual knowledge would presumably guide retrieval, again leading one to expect that intrusions that are consistent with the theme would predominate. Third, this argument seems inconsistent with the near-floor level of false recall that was observed: If the core theme of a DRM list is identified, albeit for only a subset of words, higher levels of false recall would surely be expected.

It should be added that in addition to the low overall level of false recall, there was no evidence of high levels of false recall in particular children. The largest number of critical unpresented words recalled by any of the 60 children was three, with most children recalling either zero or one of these words.

EXPERIMENT 2

The first experiment revealed both quantitative and qualitative differences in the false recall of young children versus adults. On the quantitative side, kindergartners displayed near-floor levels of false recall. On the qualitative side, children's false recall, unlike adults', was not dominated by critical unpresented words. In addition, false recall was not governed by an understanding of the themes of just-studied lists because the most common intrusions were ones that violated those themes.

The key purpose of Experiment 2 was to determine whether such differences could be replicated under conditions that are known to produce especially high levels of false recall in adults. Stadler et al. (1999) normed false recall for 36 DRM lists, and found that some lists produced high levels of false recall (in the neighborhood of 60%) whereas others produced much lower levels (in the neighborhood of 20%). In Experiment 2, both high and low lists were administered to determine whether high lists produced levels of false recall that more closely approximated adult levels.

A second purpose of the experiment was to begin to track developmental changes in the DRM illusion by including a group of second-grade children, as well as kindergartners. This age range was chosen so as to bracket the interval of the "5-to-7 shift," a period during which there are major changes in learning and memory that might produce increases in the DRM illusion.

Method

Participants. Fifty kindergarten children (age: M = 5,9; 25 males, 25 females) and 50 second-grade children (age: M = 7,10; 25 males, 25 females) participated, all of whom were pupils in elementary schools that served middle-class residential areas of a southwestern city. Children's participation was secured via letters of parental permission.

Materials and procedure. The methodology was the same as in Experiment 1, except for the lists. Each child listened to recordings of 16 DRM lists (presented in the second half of the Appendix), which were obtained from the Stadler et al. (1999) norms. Eight lists were the ones that yielded the highest levels of false recall in adults (the WINDOW, SLEEP, SMELL, DOCTOR, SWEET, CHAIR, SMOKE, and ROUGH lists), and the other eight were ones that yielded the lowest levels (the SHIRT, HIGH, ARMY, MAN, THIEF, LION, FRUIT, and KING lists). In Stadler et al.'s norms, the mean levels of false recall were 58% for the eight high lists and 22% for the eight low lists, but the mean levels of true recall (61% versus 62%) did not differ reliably.

As in Experiment 1, children were first given general memory instructions, which informed them that they would be listening to short lists of "vocabulary words" and that they should pay careful attention to each list so that they could recall as many words as possible on a later memory test. Each child then studied and recalled the 16 DRM lists, with lists being presented in random order. Each list was presented on an audio recording (2-s presentation rate) in the highest to lowest associate order shown in the Appendix, followed by 1 min of oral free recall. Children were instructed to recall as many words as possible, and when they stopped, they were prompted to recall any further words that could be remembered. After the List 1 recall was completed, List 2 was presented and recalled, and so on, until List 16 had been presented and recalled.

Results and Discussion

Summary statistics appear in the bottom half of Table 1 for four types of responses: (1) correct recall of studied words; (2) intrusions of critical unpresented words, (3) intrusions of unpresented words with similar meanings as the study list, and (4) intrusions of words from earlier lists. A 2 (age: kindergarten versus grade 2) × 2 (type of recall: true versus false) × 2 (list: high versus low) ANOVA was computed using recall proportions for targets and critical distractors as the dependent variable. There was a main effect for age, $F(1, 98) = 25.09$, $MS_E = .01$, $p < .001$; a main effect for true versus false recall, $F(1, 98) = 2196.06$, $MS_E = .003$, $p < .001$; a main effect for high versus low lists, $F(1, 98) = 17.70$, $MS_E = .01$, $p < .001$; an Age × True–False interaction, $F(1, 98) = 12.78$, $MS_E = .003$, $p < .001$; and a True–False × High–Low interaction, $F(1, 98) = 14.61$, $MS_E = .01$, $p < .001$.

Follow-up post hoc analyses (Tukey honestly significant difference tests, which were also used to analyze interactions in Experiment 3) were conducted of the interactions (using a confidence level of .05), which produced three major results. First, the age increase in recall was reliable for true recall but not for false recall. Second, true recall was higher than was false recall at both age levels (.34 versus .07, pooling across age levels). Third, true recall of low lists was superior to true recall of high lists, but there was no high–low difference in false recall.

There are three important differences between these findings and standard patterns of true and false recall that have been reported for adults (e.g., see Brainerd & Reyna, 1998b; Payne et al., 1996; Roediger & McDermott, 1995; Seamon et al., 2002). First, as in Experiment 1, the absolute level of false recall of critical unrepresented words was very low in 5- to 7-year-olds. Overall, 7% of the lists produced false recall of the critical unpresented word. Second, high lists did not increase children's false recall of critical unpresented words, although these same lists produce a threefold increase in adult's false recall. Third, high lists produced less true recall than low lists in 5- to 7-year-olds, but high and low lists produce equivalent levels of true recall in adults.

Two further differences between child and adult false recall were detected, both of which are theoretically informative. First, as in Experiment 1, DRM lists were more likely to produce intrusions of words from earlier lists or intrusions of other similar-meaning words than they were to produce intrusions of critical unpresented words. For kindergartners, it can be seen in Table 1 that it was more common for DRM lists to produce intrusions from prior lists than to produce either of the two meaning-based intrusions. Hence, as before, kindergartner's false recall was not dominated by critical unpresented words, and also as before, the fact that words from prior lists were the most common type of intrusion suggests that kindergartners failed to get the gist of DRM lists. For second graders, it can be seen in Table 1 that it was more common for DRM lists to produce either intrusions from prior lists or similar-meaning intrusions than to produce intrusions of critical unpresented words. Although the frequency of intrusions from prior lists was reliably higher than the frequency of intrusions of either critical unpresented words or other similar-meaning words, the frequencies of the latter two categories of intrusions did not differ reliably. However, the main implication is that because words from earlier lists were the predominant form of intrusion, it appears that second graders also failed to get the gist of DRM lists.

The other additional finding is that, taken together, the two types of intrusions that were consistent with the gist of DRM lists (critical words + similar-meaning words) did not reliably exceed intrusions from previously studied lists at either age level. Again, the fact that young children failed to suppress intrusions that, logically, could not have been present on the just-studied list reinforces the conclusion that they had difficulty connecting the gist across DRM targets.

EXPERIMENT 3

Experiment 2 replicated and extended the quantitative and qualitative discrepancies between the false recall of young children versus adults that were obtained in Experiment 1. On the quantitative side, both 5- and 7-year-olds displayed very low levels of false recall. On the qualitative side, children's false recall, unlike that of adults, was not dominated by critical unpresented words. Further, children's false recall was not constrained by an understanding of the themes of just-studied lists, because the most common intrusions were ones that violated those themes. Two other qualitative differences between children's performance and standard adult findings were that high DRM lists did not increase children's false recall (whereas they dramatically increase adults' false recall) and that high lists decreased children's true recall (whereas they do not decrease adults' true recall; Stadler et al., 1999).

Experiment 3 had three principal aims. One was to determine whether increased levels of false recall could be detected by age 11. Many prior studies of children's recall demonstrate that by age 11, children spontaneously use a range of semantic-processing strategies (e.g., elaboration, organization) that should aid in extracting the gist of individual words and in connecting gist across multiple words (e.g., see Bjorklund, 1999; Bjorklund & Muir, 1988). If such processing is critical to the DRM illusion, increased levels of false recall, which were not detected by age 7 in Experiment 2, should be detected by age 11.

The second aim of Experiment 3 was to measure children's false recognition of critical unpresented words. Generally speaking, recognition tests are more sensitive to the contents of memory than recall tests (Baddeley, 1976), with the sensitivity differential being especially marked in young children (Brainerd, Reyna, Howe, & Kingma, 1990). In the study of children's false memories, in particular, the effects of some manipulations have been detected with recognition but not with recall. Examples include the effects of certain forms of suggestion (cf. Ceci, Ross, & Toglia, 1987 and Howe, 1991) and the effects of mere memory testing (cf. Brainerd & Reyna, 1996 and Poole & White, 1995). Thus, given the greater sensitivity of recognition tests, it is conceivable that young children's low susceptibility to the DRM illusion on recall tests would be replaced by high susceptibility on recognition tests. To evaluate this possibility, a standard procedure in the adult literature (e.g., Payne et al., 1996) was followed, and a recognition test for several DRM lists following presentation and recall of those lists was administered.

The third aim of Experiment 3 was to effect direct comparisons between levels of false memory in children and levels of false memory in adults, under identical experimental conditions. This was done to eliminate the possibility that the high levels of false memory that have been reported in prior adult studies are due to differences between the methodologies of those studies and the methodologies of Experiments 1 and 2. More specifically, high levels of DRM false recall in adults have been obtained with a procedure in which participants receive an oral presentation of a DRM list, followed by a written free-recall test. In contrast, the children in Experiments 1 and 2 received an oral free-recall test following each DRM list. A standard finding in memory research is that mismatches between study and test modalities reduce the accuracy of memory performance (e.g., Jacoby, 1996), and the oral–written method may therefore be responsible for high levels of adult false recall. Similarly, high levels of false recognition in adults have been obtained with a procedure in which DRM lists are presented orally, followed by a written recognition test (e.g., participants receive a printed page of probe words and are instructed to respond "yes" to presented items and "no" to unpresented items). Thus, the oral–written method may also be responsible for high levels of adult false recognition of critical unpresented words.

To deal with this potential problem, a sample of young adults was added to the 5- and 11-year-old samples. In the procedure, oral memory tests (recall and recognition) followed oral presentation of DRM lists at all three age levels. Applying this procedure to adults as well as to children provided an opportunity to answer two questions: First, and most important, if there are developmental increases in DRM false memory between early childhood and early adolescence (as fuzzy-trace theory and the developmental literature on memory strategies would expect), are there further developmental increases between early adolescence and young adulthood (as fuzzy-trace theory and the developmental literature on memory strategies would also expect)? Second, to what extent are the high levels of false memory that have been reported in adults due to the oral–written procedure? To answer the latter question, the levels of adult false recall and false recognition that were obtained with the present oral–oral procedure were compared with reported norms for the oral–written procedure.

Method

Participants. Forty kindergarten children (age: M = 5,7; 20 males, 20 females) and 42 sixth-grade children (age: M = 11,11; 20 males, 22 females) participated, all of whom were pupils in elementary and middle schools that served middle-class residential areas of a southwestern city. Children's participation was secured via letters of parental permission. Forty undergraduate volunteers (12 males, 28 females) also participated to fulfill a course requirement.

Materials and procedure. The methodology was the same as in Experiment 2, except for the introduction of a poststudy manipulation (recall versus letter shadowing) after presentation of each list and the administration of a terminal recognition test. The 16 DRM lists from Experiment 2 (eight high lists and eight low lists) were presented in the same manner as before. Following half of the lists (4 randomly selected high lists and four randomly selected low lists), participants again engaged in 1 min of oral free recall, as in the first two experiments. Following the other half of the lists, however, participants engaged in 1 min of nonmemorial activity (letter shadowing). List presentation order was random. The ordering of the two activities was also random, except for the constraint that half of the lists of each type (high versus low) should be followed by recall and half should be followed by letter shadowing.

Following presentation of the sixteenth DRM list, a simple yes–no recognition test was administered. This test was a standard recognition procedure in which a list of test words was presented one at a time, and participants were instructed to say "yes" to old words that had previously appeared on the study lists (targets) and "no" to new words that had not appeared on the study lists (distractors). Whereas prerecorded audio tapes had been used to present the DRM lists during the study-recall phase of the experiment, the recognition test was read by the experimenter. The experimenter simply read each test word aloud, waited for the child's answer, and then moved on to the next test word. The test list was composed of 96 words: (1) 48 targets (3 selected at random from each of the 16 lists), (2) the 16 critical distractors for the presented lists, (3) 16 other distractors that were semantically related to the presented lists (1 per list), (4) 8 distractors that were selected from among the targets on the 20 unpresented lists from the Stadler et al. (1999) norms, and (5) 8 critical distractors from among the critical distractors for the 20 unpresented lists. Following a procedure from prior experiments (Brainerd et al., 2001), distractors in Group 3 were generated by presenting the first 14 words of each

DRM list as the study list and using the unpresented fifteenth word as a meaning-preserving distractor (e.g., PIE on the SWEET list). Distractors in Groups 4 and 5 were not semantically related to any of the presented lists. Such distractors are included on recognition tests so that true memory for targets (Group 1) and semantically based false memory (Groups 2 and 3) can be distinguished from response bias (yea-saying). Following another standard procedure (Payne et al., 1996), distractors in Group 4 provided the response-bias control for items in Groups 1 and 3 (both of which were DRM list targets), and distractors in Group 5 provided the response-bias control for items in Group 2 (DRM critical distractors).

Results and Discussion

Summary statistics for recognition and recall are presented by age level in Table 2. With regard to the recognition data, it can be seen that response bias levels, as measured by false alarms to unrelated distractors (Groups 4 and 5), were more than twice as high in young children ($M = .25$) as in young adolescents ($M = .10$), and were slightly higher in young adolescents than in young adults ($M = .07$). Because there were age differences in yea-saying, raw hit rates for targets were not an appropriate measure of true memory, and raw false-alarm rates for critical distractors and semantically related distractors were not an appropriate measure of meaning-based false memory. Instead, raw hit and false alarm rates were corrected for differences in yea-saying using signal detection statistics, such as A'. The relevant A' values for targets, critical distractors, and semantically related distractors are reported in parentheses in Table 2.[2]

[2] A' is the nonparametric counterpart of the familiar signal-detection statistic d' (Snodgrass & Corwin, 1988). For targets, $A' = .5 + [(H - FA_T)(1 + H - FA_T)] \div [4H(1 - FA_T)]$, where H is the proportion of hits for targets from presented lists and FA_T is the proportion of false alarms for targets from unpresented lists. This formula applies as long as $H \geq FA$. If $H < FA$, the replacement formula is $A' = .5 - [(FA_T - H)(1 - H + FA_T)] \div [4\,FA_T(1 - H)]$. For related distractors, $A' = .5 + [(FA_R - FA_T)(1 + FA_R - FA_T)] \div [4\,FA_R(1 - FA_T)]$, where FA_R is the proportion of false alarms for related distractors for presented lists and FA_T is the proportion of false alarms for targets from unpresented lists. This formula applies as long as $FA_R \geq FA_T$. If $FA_R < FA_T$, the replacement formula is $A' = .5 - [(FA_T - FA_R)(1 - FA_R + FA_T)] \div [4\,FA_T(1 - FA_R)]$. For critical distractors, $A' = .5 + [(FA_{CDP} - FA_{CDN})(1 + FA_{CDP} - FA_{CDN})] \div [4\,FA_{CDP}(1 - FA_{CDN})]$, where FA_{CDP} is the proportion of false alarms for critical distractors for presented lists and FA_{CDN} is the proportion of false alarms for critical distractors from unpresented lists. This formula applies as long as $FAC_{CDP} \geq FA_{CDN}$. If $FAC_{CDP} < FA_{CDN}$, the replacement formula is $A' = .5 - [(FA_{CDN} - FA_{CDP})(1 - FA_{CDP} + FA_{CDN})] \div [4\,FA_{CDN}(1 - FA_{CDP})]$.

Recall. Results are again reported for four types of responses: (1) recall of studied words, (2) intrusions of critical unpresented words, (3) intrusions of unpre-

Table 2 Proportions of Different Types of Responses on the Recall and Recognition Tests of Experiment 3

	Age		
Condition/Item	5 Years	11 Years	Adult
Recall test			
Correct recall			
High	.19	.39	.60
Low	.22	.42	.66
Intrusion: Critical			
High	.10	.27	.53
Low	.12	.18	.21
Intrusion: Same-meaning			
High	.08	.14	.04
Low	.14	.12	.01
Intrusion: Earlier lists			
High	.19	.01	0
Low	.24	.02	0
Recognition test: Previously recalled lists			
Target			
High	.45 (.72)	.59 (.85)	.75 (.92)
Low	.58 (.78)	.66 (.87)	.69 (.90)
Critical distractor			
High	.65 (.77)	.56 (.83)	.80 (.92)
Low	.64 (.76)	.51 (.81)	.51 (.82)
Same-meaning distractor			
High	.27 (.59)	.18 (.65)	.39 (.81)
Low	.33 (.64)	.17 (.64)	.13 (.68)
Unrelated target	.20	.09	.05
Unrelated critical distractor	.29	.11	.09
Recognition test: Previously unrecalled lists			
Target			
High	.69 (.83)	.73 (.90)	.67 (.90)
Low	.54 (.76)	.69 (.88)	.67 (.90)
Critical distractor			
High	.65 (.77)	.79 (.91)	.79 (.91)
Low	.59 (.73)	.46 (.79)	.43 (.79)
Same-meaning distractor			
High	.18 (.47)	.14 (.60)	.28 (.77)
Low	.28 (.60)	.18 (.65)	.15 (.69)
Unrelated target	.20	.09	.05
Unrelated critical distractor	.29	.11	.09

Note: Values in parentheses are A' values. A total of eight high lists and eight low lists, containing 14 words apiece, were presented, with only half of the lists of each type being followed by recall tests. Thus, the maximum number of words that could be correctly recalled for each list type was 56. With regard to intrusions, the maximum number of critical unpresented words that could be recalled was four per list type. There was no upper bound on the number of same-meaning words or the number of earlier list words that could intrude for either list type. With regard to recognition, the response proportion data for targets were hit proportions, whereas the corresponding data for distractors were false-alarm proportions.

sented words with similar meanings as words on the study list, and (4) intrusions of words that appeared on earlier lists. A 2 (age: kindergarten, grade 6, adult) $\times$ 2 (type of recall: true versus false) $\times$ 2 (list: high versus low) ANOVA was computed using recall proportions for targets and critical distractors as the dependent variable (i.e., only critical distractors were used to measure false recall in this ANOVA). The ANOVA produced a main effect for age, $F(2, 119) = 54.68$, $MS_E = .03$, $p < .001$; a main effect for type of recall, $F(1, 119) = 104.75$, $MS_E = .03$, $p < .001$; and a main effect for list, $F(1, 119) = 29.32$, $MS_E = .03$, $p < .001$. However, these main effects were qualified by a large Age $\times$ Type of Recall $\times$ List interaction, $F(2, 119) = 33.20$, $MS_E = .02$, $p < .001$.

Post hoc analyses of the Age $\times$ Type of Recall $\times$ List interaction (using a confidence level of .05) produced four findings that are readily apparent in Table 2. First, recall of high targets, high critical words, low targets, and low critical words all increased with age. This is the first evidence of developmental increases in false recall of critical unpresented words. Age increases were reliable between early childhood and early adolescence and were also reliable between early adolescence and young adulthood. Second, developmental increases were greater for true recall than for false recall between early childhood and early adolescence, and between early adolescence and young adulthood. Third, for high and low lists, true recall exceeded false recall among young children and young adolescents. In adults, true recall also exceeded false recall for both high and low lists, but the discrepancy was much smaller for high than for low lists (.11 versus .45). Fourth, levels of false recall did not differ for high versus low lists in young children, but the adult pattern of increased false recall for high lists had emerged by early adolescence. The gap between false recall for high and low lists then widened considerably between early adolescence and young adulthood.

The false recall of young children also exhibited four of the five discrepancies from adult false recall that were noted in Experiments 1 and 2: (1) absolute levels of false recall were very low, (2) high lists failed to produce more false recall than did low lists, (3) intrusions from earlier lists and of similar-meaning words were more frequent than intrusions of critical words, and (4) intrusions from earlier lists were more frequent than were either type of gist-preserving intrusion. These results suggest, once again, that young children are not spontaneously getting the gist of DRM lists. Young adolescents' false recall resembled that of adults' more than did young children's, especially with regard to the fact that false recall of critical unpresented words predominated over other types of intrusions. However, three notable differences remained: First, the percentage of high lists that produced false recall of the critical unpresented word (27%) was less than half that of the adult percentage (64%). Second, the difference between the intrusion rates for critical words and other similar-meaning words was much smaller than that of adults. Third, although high lists produced more false recall of critical words than did low lists, the difference between the two was much smaller than in adults.

Finally, consider the absolute levels of adult false recall that are reported in the last column of Table 2. Earlier, it was mentioned that written free-recall tests have been used in previous adult studies and that this may be responsible for high levels of false recall. The present data argue otherwise. Previous research showed that with written recall, high lists produced an average of 58% false recall, and low lists produced an average of 22% false recall (see Stadler et al., 1999). In the present experiment, which used oral recall, high lists produced an average of 53% false recall, and low lists produced an average of 21%.

Recognition. Owing to age differences in response bias, the present analyses of the recognition data focused on A' values rather than raw-hit and false-alarm rates (see Table 2). The formulas for computing A' (see Footnote 2) are such that a value of .5 indicates an absence of true recognition (higher acceptance rates for targets than for unrelated distractors) or an absence of semantic false recognition (higher acceptance rates for distractors that preserve some of the meaning content of studied material than for unrelated distractors). A value of 1 indicates perfect true or false recognition. Using this metric, true recognition increased with age between early childhood and early adolescence and between early adolescence and young adulthood, reaching a very high level in young adults ($M = .91$). False recognition of critical distractors also increased with age between early childhood and early adolescence, although not between early adolescence and young adulthood, reaching a high level in young adults ($M = .85$). Indeed, if one were to use the difference between mean A' values for targets and critical distractors as a measure of the overall accuracy of recognition memory, there was no improvement in overall accuracy between early childhood and young adulthood because increases in true recognition were compensated for by increases in false recognition. Finally, false recognition of semantically related distractors also increased with age, with a small increase between early childhood and early adolescence (.06) followed by a larger increase between early adolescence and young adulthood (.10).

A 2 (age: kindergarten, grade 6, adult) × 3 (type of item: target, critical distractor, related distractor) × 2 (list: high versus low) × 2 (poststudy activity: recall versus letter shadowing) ANOVA was computed, using A' values as the dependent variable. This ANOVA produced a main effect for age, $F(2, 118) = 29.20$, $MS_E = .09$, $p < .001$; a main effect for type of item, $F(2, 236) = 208.22$, $MS_E = .07$, $p < .001$; and a main effect for list, $F(1, 118) = 19.33$, $MS_E = .01$, $p < .001$. The reasons for these main effects are obvious from an inspection of Table 2. The main effect for age was due to the fact that mean A' values for all three types of items increased between early childhood and young adulthood. The main effect for item was due to the fact that mean A' values for semantically related distractors were considerably smaller than were the corresponding values for targets and critical distractors and the fact that mean A' values for critical distractors were somewhat smaller than were the corresponding values for targets. The main effect for list was due to the fact that mean A' values were higher for high lists than for low lists.

The main effect for list was qualified in important respects by four interactions—specifically, an Age × List interaction, $F(2, 118) = 13.01$, $MS_E = .03$, $p < .001$; an Item × List interaction, $F(2, 236) = 7.50$, $MS_E = .04$, $p < .001$; an Age × Item × List interaction, $F(4, 236) = 4.76$, $MS_E = .04$, $p < .001$; and an Item × List × Poststudy Activity interaction, $F(2, 236) = 9.75$, $MS_E = .03$, $p < .001$. Because the Age × List and Item × List interactions are subsumed by the Age × Item × List interaction, the latter was subjected to post hoc analysis. Post hoc tests produced the following findings: (1) there were no reliable list differences in A' values for young children, (2) there were reliable list differences in A' values for critical distractors but not for targets or semantically related distractors in young adolescents, and (3) there were reliable list differences in A' values for critical distractors and semantically related distractors but not for targets in young adults. Post hoc analysis of the Item × List × Poststudy Activity revealed that list differences in A' values for critical distractors were significantly larger for lists that had not been previously recalled than for lists that had been previously recalled.

Although the recognition data parallel the recall data in showing lower levels of the false-memory illusion in young children than in young adolescents and young adults, there is an important difference between the recognition and recall data that involves relative levels of true and false memory. Concerning recall, from early childhood to early adolescence, the consistent finding has been that true recall exceeds false recall, for both high and low lists. (The standard adult pattern is that true recall exceeds false recall for low lists but not high lists.) In contrast, on recognition tests, when critical distractor false alarms were the measure of false memory, true and false memory were roughly equivalent in young children, although true memory exceeded false memory for low lists in young adolescents. (The standard adult pattern is the same as the young adolescent pattern: Target A' values are higher than critical distractor A' values for low lists but not for high lists.) Thus, although the recognition version of the false-memory illusion was weaker in young children than in adults, young children's overall accuracy was not better than adults. On the other hand, like recall, true memory exceeded false memory at all age levels when semantically related distractor false alarms were the measure of false memory. This relation held at all age levels for both types of lists and for both types of poststudy activities.

Finally, consider the false-alarm rates for critical distractors for DRM lists that had been previously recalled (Table 2, last column). Earlier, it was mentioned that written recognition tests have been used in previous adult studies and that this may be responsible for high levels of false recognition. Using oral recognition, the present adult sample had an 80% false-alarm rate for high-list critical distractors and a 51% false-alarm rate for low-list critical distractors. These values are quite comparable to those that have been reported for adults using written recognition tests, in which critical distractors from previously recalled high lists have produced an 82% false-alarm rate and critical distractors from previously recalled low lists have produced a 45% false-alarm rate (see Stadler et al., 1999). We therefore conclude that the high levels of false recognition that have been obtained in prior adult studies were not due to the mismatch between presentation and test modalities.

GENERAL DISCUSSION

In the present experiments, young children were quite resistant to the DRM false-memory illusion on recall tests. Lists whose critical unpresented words are recalled 64% of the time according to adult norms (Stadler et al., 1999) were recalled less than 9% of the time by kindergartners (averaging across experiments). False recall did not increase reliably between the ages of 5 and 7. By early adolescence, false recall had increased, but considerable room remained for further development, because lists that produced 53% false recall in an adult sample produced only 27% in young adolescents. Further, the effects of the high–low list manipulation on recall were qualitatively different in children and adults. High lists increased adults' false recall while not affecting their true recall, but had the opposite effect on young children.

The developmental picture for recognition was analogous. In the adult sample of Experiment 3, the mean A' values were .92 for critical distractors from high lists and .81 for critical distractors from low lists. Young children's A' values for critical distractors were much lower, and A' values for high lists were not larger than those for low lists. Young children's A' values for semantically related distractors were also much lower than were adults', and again, A' values for high lists were not larger than those for low lists. Young adolescents' mean A' values were higher than were young children's for both critical and semantically related distractors, and A' values for critical distractors (but not semantically related distractors) were elevated for high lists. Young adolescents' mean A' values were smaller than were adults' for semantically related distractors but not for critical distractors. Thus, as with recall, recognition measures of false memory increased between early childhood and early adolescence, and increased further between early adolescence and early adulthood. Also, as with recall, the effects of the high–low list manipulation were qualitatively different in children than in adults. In children, as just noted, high lists failed to elevate A' values for distractors that preserved some of the meaning content of DRM lists, and reduced A' values for targets. In the adult sample of Experiment 3, as in prior adult research (Brainerd et al., 2001; Stadler et al., 1999), high lists elevated A' values for critical distractors and semantically related distractors, but did not affect A' values for targets. In short, there were age dissociations in the list manipulation's effects on true and false recognition.

The most instructive feature of the present developmental trends is that they were predicted on theoretical grounds—specifically, on the basis of fuzzy-trace theory's analysis of the false-memory illusion (Brainerd & Reyna, 1998b; Reyna & Lloyd, 1997). According to that analysis, two of the illusion's key controlling factors are participants' ability to store gist memories of meaning relations between the words on DRM lists and their tendency to process such gist memories during recall or recognition. In this connection, the extensive developmental literature on children's memory for categorized lists shows that the ability to spontaneously connect gist across individual words is limited in young children (e.g., Bjorklund, 1978) and undergoes a protracted period of development (e.g., Bjorklund & Hock, 1982). Therefore, the DRM illusion has the potential to display large developmental increases, and the data of the present study confirmed this expectation. In addition to predicting this particular result, the gist-failure explanation generates further predictions that can be tested in future research. For instance, it specifies some conditions in which young children's susceptibility to the DRM illusion should increase—namely, when children are induced to store gist memories of meaning relations between list words and to process them on memory tests. Although various methods could be used to ensure such storage and processing, a simple procedure would be to cue DRM list themes at study and at test (i.e., inform children that a to-be-studied list is composed entirely of, e.g., medical words or furniture words and then remind them of this fact just prior to recall).

Beyond the general finding of lowered susceptibility to the DRM illusion in young children (and also in young adolescents), specific results pointed to a gist-failure explanation. Patterns of intrusions on free-recall tests were especially probative results. If participants extract the gist of individual items and connect it across items (e.g., "medical words"), they should not output words that are wildly inconsistent with list themes (e.g., CAKE) immediately after study. Although adults' intrusions rarely violate list themes (Payne et al., 1996), such violations were the most common intrusions in young children. Another finding that bears directly on the gist-failure explanation concerns recall of targets versus critical distractors from high lists. In adults, true and false recall are comparable for these lists, suggesting that the processing of list gist predominates during free recall. In young children, on other hand, true recall greatly exceeded false recall for high lists, suggesting that gist processing did not predominate during free recall.

As mentioned, there are two aspects to the storage of gist memories of DRM lists, either of which may contribute to gist failures in young children: understanding the meanings of individual words and connecting shared meanings across words. Our gist-failure explanation of low DRM susceptibility emphasizes the second factor, an emphasis that follows from children's known limitations in connecting gist across items and also from the fact that DRM lists consist almost entirely of familiar words whose meanings are understood by children. However, the first factor may also have contributed to low DRM susceptibility because there were a few uncommon words in the lists presented in the Appendix (e.g., BILLOWS, HESITANT, IRE), that would not be familiar to most young children. Although the first factor may have contributed, two considerations suggest that it was not a major reason for low DRM susceptibility. First, and most important, inspection of the Appendix reveals that many DRM lists do not contain uncommon words (e.g., the BREAD, COLD, FOOT, AND SLEEP lists), so it is possible to estimate the effects of the first factor by comparing the false-recall and false-recognition levels of lists that did and did not contain uncommon words. When this analysis was conducted, it was found that young children's levels of false recall in Experiments 1 through 3 were the same for the two types of lists. In Experiment 3, young children's levels of false recognition were found to be the same for the two types of lists. Second, using child and adult norms for word frequency (e.g., Paivio, Yuille, & Madigan, 1968; Thorndike & Lorge, 1944), the mean familiarity of the words that comprise individual DRM lists can be computed. When the mean familiarity of the lists in the Appendix was computed, this statistic was not reliability correlated with young children's false recall (Experiments 1–3) or false recognition (Experiment 3).

Although the present developmental patterns are predicted by fuzzy-trace theory and by what is known of age variability in verbatim and gist memory, it might be argued that there are other accounts of the DRM illusion that could explain this study's results. Although many alternative accounts might be possible, the major alternative to fuzzy-trace theory's explanation in the adult literature is Roediger and McDermott's (2000) activation-monitoring theory. This theory posits that the DRM illusion is caused by bidirectional associative relations between targets and critical unpresented words, combined with source-monitoring failures on recall and recognition tests. Specifically, target presentations (BED, REST) are assumed to automatically activate critical unpresented words (SLEEP), perhaps retrieving them to consciousness as DRM lists are studied. Such activation produces false memories of critical unpresented words, in addition to the true memories that result from target presentations. To avoid recalling or recognizing critical unpresented words during a memory test, participants must correctly monitor the source of false memories (internal activation rather than external presentation) versus true memories (external presentation rather than internal activation). This account would seem to predict that the DRM illusion would be more pronounced in children than in adults, with such a prediction following from two empirical considerations. First, classic developmental research on word-association norms showed that "associative relations are an early form of semantic representation that changes minimally in development" (Bjorklund & Jacobs, 1985, p. 599), so that DRM targets should activate critical unpresented words in children as well as adults. Second, source monitoring on memory tests is known to be poorer in children than in adolescents or adults, particularly when it involves discriminating memories that originate from internal events from memories that originate from external presentation (Foley & Johnson, 1985; Foley, Johnson, & Raye, 1983), so that children will be less able to rely on source monitoring to suppress false-memory responses. Because activation-monitoring theory seems to predict the opposite of the developmental trend that was observed in the present experiments, the theory would have to be enriched with further assumptions to reconcile it with the data.

Finally, since this article was written, we have become aware of two unpublished studies, (Ghetti, Goodman, & Qin, 2001; Price, Metzger, Williams, Phelps, & Phelps, 2001) that contain further developmental data on the false-memory illusion. Price et al. reported two experiments whose designs were similar to one of the conditions of Experiment 3 in that participants of different ages (1) listened to 15-item DRM lists, (2) performed immediate free recall, and (3) responded to a yes–no recognition test after all lists had been recalled. The participants were adolescents and young adults in one experiment and 7-year-olds and young adults in the other. The observed developmental trends paralleled those of the present study: False recall and false recognition of critical unpresented words increased between childhood and adolescence, and increased further between adolescence and young adulthood.

In contrast, the findings of Ghetti et al.'s (2001) study were not similar to those of the current study. Ghetti et al. found that false recall of critical unpresented words decreased from 25% in 5-year-olds to 19% in 7-year-olds to 15% in young adults, and that false recognition of these words decreased from 47% in 5-year-olds to 38% in 7-year-olds to 36% in young adults. However, Ghetti et al.'s procedures were different than those of the present experiments and in the adult literature. Their DRM lists were very short (7 words rather than 15) and words were presented visually (as printed letter strings or as pictures) as well as orally. Assuming that Ghetti et al.'s findings prove to be replicable, additional research will be needed to determine why such procedural modifications reverse the developmental trend that was obtained in the present experiments. However, a working hypothesis is suggested by fuzzy-trace theory's analysis of the DRM task. If the processing of gist memories of list themes on memory tests generates high rates of false recall and false recognition, experimental manipulations that shift memory-test performance toward reliance on verbatim memories of targets rather than gist memories will reduce false recall and false recognition (Brainerd & Reyna, 2001). Moreover, such manipulations will have larger effects on participants who are more likely to retain verbatim memories of targets (e.g., adults and older children) than on participants who are less likely to retain them (e.g., younger children). Obviously, verbatim memories of DRM targets are apt to be better if only a few of them are presented and if targets are accompanied by supporting visual stimuli. Thus, the presentation of shorter lists and supporting visual stimuli may have reversed the developmental trend in the false-memory illusion because (1) these variables shift memory-test performance toward reliance on verbatim memory, and (2) the size of the shift increases with age.

ACKNOWLEDGMENTS

This research was supported in part by a National Science Foundation grant (SBR-9730143), by National Institutes of Health grants (NIH31620 and P50AT00008), and a Managed Health Care Forum grant.

ADDRESSES AND AFFILIATIONS

Corresponding author: C. J. Brainerd, Department of Special Education, Rehabilitation, and School Psychology, University of Arizona, Tucson, AZ 85721; e-mail: brainerd@u.arizona.edu. V. F. Reyna and T. J. Forrest are also at the University of Arizona.

APPENDIX

EXPERIMENT 1 LISTS

Anger: mad, fear, hate, rage, temper, fury, ire, wrath, happy, fight, hatred, mean

Black: white, dark, cat, charred, night, funeral, color, grief, blue, death, ink, bottom

Bread: butter, food, eat, sandwich, rye, jam, milk, flour, jelly, dough, crust, slice

Chair: table, sit, legs, seat, couch, desk, recliner, sofa, wood, cushion, swivel, stool

Cold: hot, snow, warm, winter, ice, wet, frigid, chilly, heat, weather, freeze, air

Doctor: nurse, sick, lawyer, medicine, health, hospital, dentist, physician, ill, patient, office, stethoscope

Foot: shoe, hand, toe, kick, sandals, soccer, yard, walk, ankle, arm, boot, inch

Fruit: apple, vegetable, orange, kiwi, citrus, ripe, pear, banana, berry, cherry, basket, juice

Girl: boy, dolls, female, young, dress, pretty, hair, niece, dance, beautiful, cute, date

High: low, clouds, up, tall, tower, jump, above, building, noon, cliff, sky, over

King: queen, England, crown, prince, George, dictator, palace, throne, chess, rule, subjects, monarch

Man: woman, husband, uncle, lady, mouse, male, father, strong, friend, beard, person, handsome

Mountain: hill, valley, climb, summit, top, molehill, peak, plain, glacier, goat, bike, climber

Music: note, sound, piano, sing, radio, band, melody, horn, concert, instrument, symphony, jazz

Needle: thread, pin, eye, sewing, sharp, point, prick, thimble, haystack, thorn, hurt, injection

River: water, stream, lake, Mississippi, boat, tide, swim, flow, run, barge, creek, brook

Rough: smooth, bumpy, road, tough, sandpaper, jagged, ready, coarse, uneven, riders, rugged, sand

Sleep: bed, rest, awake, tired, dream, wake, snooze, blanket, doze, slumber, snore, nap

Slow: fast, lethargic, stop, listless, snail, cautious, delay, traffic, turtle, hesitant, speed, quick

Soft: hard, light, pillow, plush, loud, cotton, fur, touch, fluffy, feather, furry, downy

Spider: web, insect, bug, fright, fly, arachnid, crawl, tarantula, poison, bite, creepy, animal

Sweet: sour, candy, sugar, bitter, good, taste, tooth, nice, honey, soda, chocolate, heart

Thief: steal, robber, crook, burglar, money, cop, bad, rob, jail, gun, villain, crime

Window: door, glass, pane, shade, ledge, sill, house, open, curtain, frame, view, breeze

EXPERIMENTS 2 AND 3 HIGH LISTS

Chair: table, sit, legs, seat, couch, desk, recliner, sofa, wood, cushion, swivel, stool, sitting, rocking, bench

Doctor: nurse, sick, lawyer, medicine, health, hospital, dentist, physician, ill, patient, office, stethoscope, surgeon, clinic, cure

Rough: smooth, bumpy, road, tough, sandpaper, jagged, ready, coarse, uneven, riders, rugged, sand, board, ground, gravel

Sleep: bed, rest, awake, tired, dream, wake, snooze, blanket, doze, slumber, snore, nap, peace, yawn, drowsy

Smell: nose, breathe, sniff, aroma, hear, see, nostril, whiff, scent, reek, stench, fragrance, perfume, salts, rose

Smoke: cigarette, puff, blaze, billows, pollution, ashes, cigar, chimney, fire, tobacco, stink, pipe, lungs, flame, stain

Sweet: sour, candy, sugar, bitter, good, taste, tooth, nice, honey, soda, chocolate, heart, cake, tart, pie

Window: door, glass, pane, shade, ledge, sill, house, open, curtain, frame, view, breeze, sash, screen, shutter

EXPERIMENTS 2 AND 3 LOW LISTS

Army: Navy, soldier, United States, rifle, Air Force, draft, military, Marines, march, infantry, captain, war, uniform, pilot, combat

Fruit: apple, vegetable, orange, kiwi, citrus, ripe, pear, banana, berry, cherry, basket, juice, salad, bowl, cocktail

High: low, clouds, up, tall, tower, jump, above, building, noon, cliff, sky, over, airplane, dive, elevate

King: queen, England, crown, prince, George, dictator, palace, throne, chess, rule, subjects, monarch, royal, leader, reign

Lion: tiger, circus, jungle, tamer, den, cub, Africa, mane, cage, feline, roar, fierce, bears, hunt, pride

Man: woman, husband, uncle, lady, mouse, male, father, strong, friend, beard, person, handsome, muscle, suit, old

Shirt: blouse, sleeves, pants, tie, button, shorts, iron, polo, collar, vest, pocket, jersey, belt, linen, cuffs

Thief: steal, robber, crook, burglar, money, cop, bad, rob, jail, gun, villain, crime, bandit, bank, criminal

REFERENCES

Ackerman, B. P. (1992). The sources of children's errors in judging causal inferences. *Journal of Experimental Child Psychology, 54*, 90–119.

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.

Bjorklund, D. F. (1978). Negative transfer in children's recall of categorized materials. *Journal of Experimental Child Psychology, 26*, 299–307.

Bjorklund, D. F. (1987). How changes in knowledge base contribute to the development of children's memory. *Developmental Review, 7*, 93–130.

Bjorklund, D. F. (1999). *Children's thinking: Developmental function and individual differences* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Bjorklund, D. F., & Hock, H. H. (1982). Age differences in the temporal locus of memory organization in children's recall. *Journal of Experimental Child Psychology, 33*, 347–362.

Bjorklund, D. F., & Jacobs, J. W. (1985). Associative and categorical processes in children's memory: The role of automaticity in the development of organization in free recall. *Journal of Experimental Child Psychology, 39*, 599–617.

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

Brainerd, C. J., & Mojardin, A. H. (1998). Children's spontaneous false memories for narrative statements: Long-term persistence and mere-testing effects. *Child Development, 69*, 1361–1377.

Brainerd, C. J., & Reyna, V. F. (1996). Mere memory testing creates false memories in children. *Developmental Psychology, 32*, 467–476.

Brainerd, C. J., & Reyna, V. F. (1998a). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology, 71*, 81–129.

Brainerd, C. J., & Reyna, V. F. (1998b). When things that never happened are easier to remember than things that did. *Psychological Science, 9*, 484–489.

Brainerd, C. J., & Reyna, V. F. (2001). Fuzzy-trace theory: Dual processes in reasoning, memory, and cognitive neuroscience. *Advances in Child Development and Behavior, 28*, 41–100.

Brainerd, C. J., & Reyna, V. F. (2002) Recollection rejection: How children edit their false memories. *Developmental Psychology, 38*, 156–172.

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, 53*(2–3, 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.

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*, 307–327.

Bruck, M., & Ceci, S. J. (1997). 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.

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom*. Washington, DC: American Psychological Association.

Ceci, S. J., & Bruck, M. (1998). The ontogeny and durability of true and false memories: A fuzzy trace account. *Journal of Experimental Child Psychology, 71*, 165–169.

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

Deese, J. (1959). On the prediction of occurrence of certain verbal intrusions in free recall. *Journal of Experimental Psychology, 58*, 17–22.

Foley, M. A., & Johnson, M. K. (1985). Confusions between memories for performed and imagined actions: A developmental comparison. *Child Development, 56*, 1145–1155.

Foley, M. A., Johnson, M. K., & Raye, C. L. (1983). Age-related changes in confusion between memories for thoughts and memories for speech. *Child Development, 54*, 51–60.

Ghetti, S., Goodman, G. S., & Qin, J. (2001, April). *False memory formation in children and adults: Recalling, recognizing, and attributing source to events that never happened*. Paper presented at the biennial meeting of the Society for Research in Child Development, Minneapolis, MN.

Howe, M. L. (1991). Misleading children's story recall: Forgetting and reminiscence of the facts. *Developmental Psychology, 27*, 746–762.

Howe, M. L. (2000). *The fate of early memories*. Washington, DC: American Psychological Association.

Jacoby, L. L. (1996). Dissociating automatic and consciously controlled effects of study/test compatibility. *Journal of Memory and Language, 35*, 32–52.

Mather, M., Henkel, L. A., & Johnson, M. K. (1997). Evaluating characteristics of false memories: Remember/know judgments and memory characteristics questionnaire compared. *Memory & Cognition, 25*, 826–837.

Paivio, A., Yuille, J. C., & Madigan, S. A. (1968). Concreteness, imagery, and meaningfulness values for 925 nouns. *Journal of Experimental Psychology Monograph, 76*(No. 1, Pt. 2), 1–25.

Payne, D. G., Elie, C. J., Blackwell, J. M., & Neuschatz, J. S. (1996). Memory illusions: Recalling, recognizing, and recollecting events that never occurred. *Journal of Memory and Language, 35*, 261–285.

Pipe, M. E., Gee, S., Wilson, J. G., & Egerton, J. M. (1999). Children's recall 1 or 2 years after an event. *Developmental Psychology, 35*, 781–789.

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. (1995). Tell me again and again: Stability and change in the repeated testimonies of children and adults. In M. S. Zaragoza, J. R. Graham, C. N. Gordon, R. Hall, R. Hirschman, & Y. S. Ben-Porath (Eds.), *Memory and testimony in the child witness* (pp. 24–43). New York: Harper & Row.

Price, J. L., Metzger, R. L., Williams, D., Phelps, N. Z., & Phelps, A. M. (2001, April). *Children produce as many false memories as adults (sometimes!)*. Poster presented at the biennial meeting of the Society for Research in Child Development, Minneapolis, MN.

Reyna, V. F. (1996). Conceptions of memory development, with implications for reasoning and decision making. *Annals of Child Development, 12*, 87–118.

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

Reyna, V. F., & Brainerd, C. J. (1998). Fuzzy-trace theory and false memory: New frontiers. *Journal of Experimental Child Psychology, 71*, 194–209.

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.

Reyna, V. F., & Lloyd, F. (1997). Theories of false memory in children and adults. *Learning and Individual Differences, 9*, 95–123.

Reyna, V. F., & Titcomb, A. L. (1997). Constraints on the suggestibility of eyewitness testimony: A fuzzy-trace theory analysis. In D. G. Payne & F. G. Conrad (Eds.), *A synthesis of basic and applied approaches to human memory* (pp. 157–174). Mahwah, NJ: Erlbaum.

Roediger, H. L., III, & McDermott, K. B. (1995). Creating false memories: Remembering words not presented on lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803–814.

Roediger, H. L., III, & McDermott, K. B. (2000). Tricks of memory. *Current Directions in Psychological Science, 9*, 123–127.

Schacter, D. L., Verfaellie, M., & Pradere, D. (1996). The neuropsychology of memory illusions: False recall and recognition in amnesic patients. *Journal of Memory and Language, 35*, 319–334.

Schneider, W., & Pressley, M. (1989). *Memory development from 2 to 20*. New York: Springer-Verlag.

Seamon, J. G., Luo, C. R., Schwartz, M. A., Jones, K. J., Lee, D. M., & Jones, S. J. (2002). Repetition can have similar and different effects on accurate and false recognition. *Journal of Memory and Language, 46*, 323–340.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34–50.

Stadler, M. A., Roediger, H. L., & McDermott, K. B. (1999). Norms for words that create false memories. *Memory & Cognition, 27*, 494–500.

State v. Michaels, 136 N.J. 299, 642 A.2d 1372 (N.J. 1994).

Thorndike, E. L., & Lorge, I. (1944). *The teacher's book of 30,000 words*. New York: Teachers College, Columbia University, Bureau of Publications.

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory, 7*, 233–256.

Underwood, B. J. (1957). Interference and forgetting. *Psychological Review, 64*, 49–60.