JOURNAL OF EXPERIMENTAL CHILD PSYCHOLOGY 52, 297–318 (1991)

Developmental Changes in Memory Source Monitoring

D. STEPHEN LINDSAY

Williams College

MARCIA K. JOHNSON

Princeton University

AND

PAUL KWON

Williams College

Previous research suggests that children are more likely than adults to confuse memories of actions they imagined themselves performing with memories of actions they actually performed (Realization Judgments), but are not more likely to confuse memories of actions they had imagined performing with memories of actions they saw another person perform (Reality Monitoring). We approach these findings in terms of a theory about the processes by which people identify the sources of their recollections (Source Monitoring). This approach suggests that children may be more likely than adults to confuse memories from different sources whenever the sources are highly similar to one another. Experiments 1 and 2 tested this hypothesis by manipulating the perceptual and semantic similarity of two sources of information and testing 4- and 6-year-old and adult subjects' recollection of the sources of particular pieces of information. Experiment 3 tested the hypothesis that children are more likely than adults to mistakenly identify memories of things they imagined another person doing as memories of things they witnessed that person doing. The findings indicate that (a) people are more

Aspects of this research were supported by Williams College and the Williams College Bronfman Science Center, by a Natural Science and Engineering Research Council of Canada postdoctoral fellowship to Steve Lindsay, by the University of Victoria, and by NSF Grant BNS 8510633 to Marcia Johnson. Experiments 1 and 2 were included in Lindsay's doctoral dissertation, which was supervised by Johnson at Princeton University; Experiment 3 was included in Kwon's undergraduate honor's thesis, which was supervised by Lindsay at Williams College. We thank Larry L. Jacoby for helpful discussions of Experiments 1 and 2, and the reviewers for their comments on a draft of this article. Correspondence should be addressed to Steve Lindsay, who is now at the Department of Psychology, University of Victoria, P.O. Box 3050, Victoria, B.C., Canada V8W 3P5.

0022-0965/91 $3.00

likely to confuse memories from similar than dissimilar sources, (b) source monitoring improves during the preschool and childhood years, and (c) children may be especially vulnerable to the effects of source similarity.

People often remember some aspects of an event while forgetting or misremembering other aspects. Of particular interest in the present context, people are sometimes uncertain or mistaken about when or where a remembered event occurred, about how an event was perceived (in what media, through what modalities), or even about whether an event actually happened or was merely imagined. Previous research suggests that, relative to adults, children more often confuse memories of things they had merely imagined doing with memories of things they had actually done (see Johnson & Foley, 1984, and Lindsay & Johnson, 1987, for reviews). In the current article, this developmental trend in memory for the reality status of past events is examined in the context of the more general problem of how people identify the sources (or conditions of acquisition) of their event memories. We propose that, relative to adults, children are more likely to confuse memories from different sources whenever those sources are highly similar to one another in terms of their characteristic perceptual and semantic content. One important implication of this hypothesis, tested in Experiment 3, is that children may be more likely than adults to mistake a recollection of imagining someone doing something as a memory of witnessing that person doing that thing.

Johnson and Raye (1981) proposed a framework for understanding the processes by which actual and imagined events are discriminated and confused in memory, a process they termed "reality monitoring." In the reality monitoring model, the reality status of recollected events is identified via decision-making processes performed during remembering. Johnson and Raye proposed that most reality monitoring decisions are made quickly, without reflective deliberation, by a fast guess procedure: Recollections that are rich in perceptual, contextual, and semantic detail and that include relatively few indications of internal cognitive operations are identified as memories of actual events, whereas recollections that are not vivid and lack detail but include indications of many cognitive operations are classified as memories of imagined events. More extended reasoning processes may also be performed, especially if a recollection is not classified as actual or imagined by the fast guess procedure. These include reflection on the content of the memory and its relation to other knowledge and biases based on beliefs about how memory works. For example, one might identify a vivid recollection of unaided flight as a memory of a dream, based on one's knowledge that people cannot fly.

Are children more prone than adults to confuse memories of actual

and imagined events? This appears to have been the opinion of Freud (e.g., 1911/1958) and Piaget (e.g., 1945/1962), but recent research suggests a more complex developmental pattern. The studies published to date indicate that children as young as 6 years perform as well as adults when identifying the origins of memories, except when they must discriminate between recollections of actions they imagined themselves performing and recollections of actions they actually performed (Foley, Aman, & Gutch, 1987; Foley, Johnson, & Raye, 1983; Foley & Johnson, 1985; Foley, Durso, Wilder, & Freidman, 1991). For example, Foley et al. (1983) found that 6-year-old children had difficulty discriminating between memories of words they had actually said aloud and memories of words they had imagined saying during an acquisition phase, but that they were as accurate as adults when discriminating between memories of words they had imagined saying and memories of words they heard another person say, or when remembering which of two other people had said particular words. Foley and Johnson (1985) reported a similar pattern of results with tasks involving memories for actions (e.g., "Did you really touch your nose, or did you just imagine yourself touching your nose?"). Relative to adults, young children were more likely to be confused about which things they had actually done and which they had merely imagined, but children were not more likely to mistake memories of what they had done with memories of what another person had done, nor did they more often misidentify which of two other people had performed particular actions. In a recent paper, Foley et al. (1991) noted that although 6- to 9-year-old children are more likely to be confused than adults when asked to discriminate what they had done from what they had imagined doing, "To date, there is no evidence that children are more likely to confuse what they have imagined with what they have seen" (p. 4).

Why is the Actual-Self/Imagined-Self memory discrimination a particularly difficult one for children? Foley et al. (1983) proposed that the cues involved in differentiating self from non-self in memory are well developed by 6 years of age, whereas those involved in differentiating between memories of different types of self-generated acts (Realization Judgments) are less developed. It may be that memories of actual and imagined self-generations are more similar to one another in children than in adults, either because children's imagery is more rich in perceptual detail and hence more similar to perceptual experience (e.g., Kosslyn, 1978) or because their overt actions require more reflective control of the kind involved in purposeful imagery. On the other hand, it might be that memories of actual and imagined self-generated acts are highly similar in adults and children alike, and that discriminating between them demands sophisticated mnemonic skills and reasoning processes with which adults are more adept (Foley et al., 1983; Foley & Johnson, 1985; Johnson & Foley, 1984). By either account, the developmental trend in Realization

Judgments is explained in terms of the high degree of similarity between memories of actual and imagined self-generated actions.

In what ways are memories of actual and imagined self-generations similar? One possibility is that memory records of actual and imagined events are quantitatively similar in terms of the attributes emphasized in the reality monitoring model proposed by Johnson and Raye (1981). For example, actual and imagined self-generated acts might have similar amounts of perceptual detail. It is also possible, however, that it is similarity of the perceptual and semantic *content* of memories of actual and imagined self-generations that makes memories derived from these two sources confusable. That is, actual and imagined self-generations might be particularly confusable not because they are equally vivid, but because they both involve the same actor (the self) and therefore share similar perceptual and semantic content (cf. Foley, Santini, & Sopasakis, 1989).

The initial Johnson and Raye (1981) reality monitoring model emphasized decision processes that evaluate memory attributes quantitatively (e.g., the *amount* of perceptual detail), but a number of studies indicate that qualitative analyses of the content of memories are also involved in reality monitoring decisions. For example, Johnson, Foley, and Leach (1988) reported a study in which subjects heard a confederate say some words and imagined hearing other words. Some subjects were instructed to imagine hearing the words in the confederate's voice, whereas other subjects were to imagine the words in their own or a third person's voice. Later, subjects were asked to indicate whether particular words had been spoken or imagined. As predicted, subjects who had imagined the words in the confederate's voice had more difficulty discriminating between memories of spoken and imagined words than did subjects who had imagined the words in their own or a third person's voice. In a related study, Foley et al. (1989) found that subjects are more likely to confuse words they said with words they imagined saying if they had imagined themselves (rather than somebody else) saying the words.

These results suggest that reality monitoring may be viewed in the context of the more general problem of discriminating between memories from different sources—the process of source monitoring (Hashtroudi, Johnson, & Chrosniak, 1989; Johnson, 1988; Johnson, Hashtroudi, & Lindsay, 1991; Lindsay, 1987; Lindsay & Johnson, 1987). The term "source" here refers to a variety of attributes that, collectively, specify the conditions under which an event memory was acquired (e.g., its spatial and temporal context, medium, and modality). Source monitoring is involved in remembering when and where an event occurred, the people and objects involved in it, the sensory modalities through which it was perceived, etc.[1] In contrast to the notion that the sources of memories are specified by abstract, proposition-like tags or labels (e.g., Anderson & Bower, 1972, 1974), we propose that sources are identified via decision processes performed during remembering. According to the source monitoring model, knowledge about the characteristic properties of particular sources is used to interpret the source-relevant information accessed from an event memory, much as knowledge about the characteristic properties of particular objects is used to identify those objects in ongoing perceptual experience. The accessible memory records of hearing an utterance, for example, might include information about the perceptual qualities of the speaker's voice, and one's knowledge about the characteristic sound of that person's voice might allow one to "recognize," as it were, the speaker of the remembered utterance. These decisions are usually made quickly and without conscious reflection.

[1] We do not mean to suggest that there is a clear distinction between the source and content of an event. On the contrary, the meaning of an event often has a lot to do with its source (e.g., Kolers, 1977). Nonetheless, the distinction between source and content is analytically useful.

The possibility that the perceptual and semantic content of event memories aids identification of their sources has important implications for the interpretation of the literature on children's reality monitoring. For one thing, it reopens the question of the developmental course of reality monitoring. As noted above, previous research suggests that it is only in Realization Judgments (discriminations between memories of actual and imagined acts) that developmental differences are observed between children as young as 6 years of age and adults, and that children do not make other types of reality monitoring errors more often than adults. Interestingly, Realization Judgments are the only case in the developmental literature to date in which subjects are asked to identify the origin of memories of actual and imagined events that can be discriminated only on the basis of their reality status. In the other reality monitoring conditions (e.g., Actual-Other/Imagined-Self), accurate discriminations could also be based on memory information about the actor involved in the event. Thus the only condition in which developmental effects have been observed is also the only condition in which discriminations could only be made on the basis of the reality status of the events, so it may be that children are in fact less able to identify the reality status of their recollections. Alternatively, children might perform more poorly than adults whenever the to-be-discriminated memory sources are highly similar, regardless of whether or not the discrimination involves reality monitoring per se.

The current studies were designed to explore these issues. In Experiments 1 and 2, young children and adults received information from two external sources and were later asked to identify the sources of particular details. We predicted that age-related improvements would be larger in those conditions in which the to-be-discriminated sources were highly similar to one another. In Experiment 3 we tested the hypothesis that, relative to adults, children are more likely to mistakenly identify a memory of something they imagined another person doing as a recollection of something they witnessed that person doing.

EXPERIMENT 1

Method

Overview. Child and adult subjects listened to a tape-recorded list of words, half presented from a speaker on the subject's left and half from a speaker on the subject's right. For half of the subjects in each age group, the same person's voice came from both speakers, whereas for the remaining subjects a different voice (one male and one female) came from each speaker. Subjects were later asked to remember the source (left or right speaker) of particular words.

Subjects. The subjects were 24 4-year-old children (M age of 4.2 years, with a range of 4.1 to 4.3 years, SD = .05) and 24 Princeton University undergraduates. Parents in the Princeton area who had previously expressed interest in child studies were telephoned and invited to include their children in the study. The undergraduates responded to posters advertising the study, and were paid for participating.

Materials. Sixty words, selected from dictionaries and story books written for young children, were randomly divided into three sets (A, B, and C) of 20 words each. These were used to construct the study and test lists. Three study lists of 40 words each (AB, AC, and BC) were constructed by randomly interleaving the words from the three possible pairings of sets, with the constraint that no more than two words from a set occurred in immediate succession. Each study list was used to make three stereo tape recordings. In each tape the words from one set were recorded on one channel and those from the other set were recorded on the other channel. Words were presented at a 3-s rate. For each list, there was one tape in which all of the words were spoken by a woman, one in which all of the words were spoken by a man, and a third in which the words recorded on one channel were spoken by the woman and the words recorded on the other channel were spoken by the man. The right/left location of the speakers was counterbalanced across tapes.

A 48-item memory test was constructed using 16 words randomly selected from each of the three sets (A, B, and C) presented in a random order, with the constraints that no more than two words from a given set occurred in succession and that approximately equal numbers of words in each condition (right, left, and new) occurred in each quarter of the test.

Procedure. Subjects were tested individually in a quiet lab room. The subject sat on a chair between two speakers, one on either side. Assignment of subjects to word list and presentation condition was predetermined by a counterbalanced schedule. Subjects were told that the experiment concerned people's ability to repeat words aloud. For adults, this description was supported by a cover story about hemispheric effects on speech perception, and subjects were led to believe that pronunciation times were being measured. In that way, we hoped to ensure that none of the subjects expected a memory test.

For the children, the right and left speakers were identified by a stuffed toy that sat atop each one—the speaker on the subject's right was the "pig" speaker, that on the left was the "teddy bear" speaker. The experimenter began sessions with children by introducing the two stuffed animals and explaining that each one had its own speaker. To ensure attention to the words, subjects were asked to repeat each word aloud as soon as they heard it. After the acquisition phase, subjects engaged in a brief (approximately 1 min) distractor task (adults counted backward by threes and children sang the ABC song) and then received the test instructions. The experimenter explained that the test included words presented from each speaker and words that had not been presented from either speaker. Special efforts were made to ensure that the children understood that some of the words on the test had not been presented on the tape. The experimenter read each test item, and the subject responded by saying "Right" ("Pig"), "Left" ("Bear"), or "No" (children were also allowed to respond by pointing to the speaker of choice).

Results

The .05 level of significance was used for all statistical tests.

Source monitoring. A source monitoring score was calculated by dividing the number of old words attributed to the correct source by the number of old words recognized as old. Chance on this measure is .50. Adults' scores were significantly greater than children's (M = .67 and .59, respectively, $F(1, 44) = 4.56$, $MSe = 0.15$), but the main effect for acquisition condition fell short of statistical significance (M = .60 and .65 for the Same and Different Voices conditions, respectively, $F(1, 44) = 2.09$, $MSe = 0.15$, $p < .16$). As predicted, age and acquisition condition interacted, $F(1, 44) = 3.02$, $MSe = 0.15$, $p < .05$, directional).[2] Children in the Same Voice condition had significantly lower source monitoring scores (M = .53) than children in the Different Voices condition (M = .64, $F(1, 44) = 4.62$, $MSe = 0.15$), whereas adults' scores were not affected by acquisition condition (M = .67 and .66, respectively, $F < 1$).

[2] The test is "directional" in that we halved the tabled critical value of F (as in a one-tailed t test) under the rationale that a specific form of interaction had been predicted.

The age effect was reliable in the Same Voice condition, $F(1, 44) = 7.49$, $MSe = 0.110$) but not in the Different Voices condition, $F < 1$. These source monitoring scores were reliably greater than chance (.5) for adults in both conditions ($t(11) = 6.72$ and $t(11) = 4.44$) and for children in the Different Voices condition ($t(11) = 3.09$), but not for children in the Same Voices condition ($t(11) = 1.0$, $p > .33$).

Old/new recognition. The old/new recognition score was the proportion of test items correctly identified as old or new. This measure reflects both hits and correct rejections, and has a chance value of .50. Recognition was well above chance for children ($M = .78$) as well as adults ($M = .82$). Neither age nor acquisition condition had a significant effect on recognition, nor did these factors interact (all $Fs < 1.52$, $p > .20$).

Discussion

Although we correctly anticipated that children would be more affected by source similarity than adults, we had expected that adults would also be influenced by the manipulation. Contrary to the latter prediction, acquisition condition had no effect on adults' source monitoring. Perhaps adult's linguistic expertise allowed them to quickly and easily extract the meaning of the words, so that they paid less attention to the sensory qualities of the voices. This would make later source monitoring difficult (consistent with the fact that adults in both conditions made many source monitoring errors) and would reduce the effect of perceptual similarity.

Children made more source monitoring errors than adults, and the effect was due entirely to the especially poor performance of children in the Same Voice condition. Children in the Same Voice condition performed at chance, whereas those in the Different Voices condition performed as well as adults. The finding that children and adults did not differ in old/new recognition suggests that the age difference in source monitoring cannot be attributed to general factors that would also affect recognition performance (paying attention, trying hard on the test, etc.). Thus Experiment 1 supports the hypothesis that children are particularly likely to confuse memories from similar sources. Experiment 2 was designed to explore this hypothesis with more naturalistic and complex materials.

EXPERIMENT 2

Method

Overview. Subjects in three age groups (4- and 6-year-old children and college-age adults) watched two video tapes, in each of which a person told a story about a circus. For some subjects, the two storytellers were very similar to one another, whereas for other subjects the two storytellers were quite dissimilar. Some subtopics were common to both stories, whereas others were unique to one story or the other. Later, subjects were asked to remember which storyteller had said particular things.

Subjects. There were 32 subjects in each age group (4-year-olds ranged from 3.9 to 4.8 years, $M = 4.3$ years, $SD = .22$; 6-year-olds ranged from 6.0 to 6.5 years, $M = 6.3$ years, $SD = .10$; adults were undergraduate students). Subjects were recruited as in Experiment 1.

Materials. A list of 24 circus acts (e.g., the elephant act, the magician's act) was generated and randomly divided into 3 sets (A, B, and C) of 8 acts each. Two details about each act were described (e.g., "The biggest elephant had a big blue and gold blanket on his back" and "The baby elephant carried a purple umbrella in his trunk"). The acts in Sets A and B were used to generate two pairs of stories. Sixteen details about twelve circus acts were mentioned in each pair of stories. Of these, eight described details of circus acts that were mentioned only in one story or the other (four in each) and eight described details of circus acts that were mentioned in both stories of the pair, with a different detail about these "Common" acts mentioned in each story. In one pair of stories, the circus acts in Set A were used as Common items and those in Set B were used as Unique items. This assignment was reversed in the other pair of stories. The acts in Set C were used only as distractor items on the test.

Three amateur actors from a community theater group were video taped individually reading the stories aloud. Two of the storytellers were teenage girls who were quite similar in appearance and voice and who read the stories in similar styles. The third storyteller was a male senior citizen who read the stories in a style quite distinct from that used by the girls.

All subjects were given a 28-item memory test composed of items derived from the detail descriptions (e.g., "Did one of the storytellers say that the biggest elephant had a blanket on his back?"). Four test items were derived from each of the following conditions: Story 1-Common, Story 1-Unique, Story 2-Common, and Story 2-Unique. Assignment of particular items to these conditions was rotated across subjects. Of the remaining 12 items on the test, 8 were Unrelated distractors (from Set C), which concerned circus acts not mentioned in either story, and 4 were Related distractors, which concerned new details about circus acts that were mentioned in one or both of the stories.

Procedure. Subjects were tested individually in a quiet lab room. Subjects were told that they would watch two video tapes, in each of which a person would tell a story about going to the circus, and that they would later be asked questions "about what happened at the circus." Subjects in the Similar Storytellers condition saw the two teen-age girls tell the stories, whereas subjects in the Dissimilar Storytellers condition saw one of the girls tell one story and the man tell the other. The two stories in a pair were always presented in the same order, with a brief (approximately 30 s) pause between them. Assignment of subjects to storytellers and to story pair was predetermined by a counterbalanced schedule.

Subjects were tested immediately after watching the video tapes. The test instructions differed somewhat for adults and children. Adults were told that the test included details mentioned only by the first storyteller, details mentioned only by the second storyteller, and details not mentioned by either storyteller, and that they should respond "First," "Second," or "Neither" to each test item as it was read. Children were told that the test consisted of details mentioned in one or the other of the stories and details not mentioned in either story, and that they would be tested on their memory for details from the stories. Upon the child's first "Yes" response, the experimenter introduced the source monitoring task by asking the child which storyteller had mentioned that detail.

Results

The .05 level of confidence was used for all statistical tests.

Source monitoring. The source monitoring score measured subjects' accuracy at attributing items said by a girl storyteller to that girl. Thus for subjects in the Dissimilar Storytellers condition the source monitoring score was calculated by dividing the number of items correctly attributed to the girl by the number of items said by the girl which the subject recognized as old. For subjects in the Similar Storytellers condition, the source monitoring score was the mean of the source monitoring scores associated with each of the two girls. This measure allows comparison of source memory for things said by the same storytellers (the girls) in two conditions: one in which the alternative source was highly similar (the other girl) and one in which it was highly dissimilar (the man). Chance performance on this measure is .50. Two source monitoring scores were calculated for each subject, one for Common items and one for Unique items. These scores were analyzed in a 3 × 2 × 2 mixed-model analysis of variance, with age and acquisition condition (Similar vs Dissimilar storytellers) as between-subjects factors and item type (Common vs Unique) as a within-subjects factor.

The data are presented in Figure 1. As is clear in the figure, source monitoring accuracy increased reliably with age (M = .61, .72, and .90 for 4-year-olds, 6-year-olds, and adults, respectively, $F(2, 90) = 18.03$, $MSe = 0.76$). As is also clear, source monitoring scores were greater on Unique items (M = .79) than on Common items (M = .69), $F(1, 90) = 9.99$, $MSe = 0.47$, and performance was better in the Dissimilar Storytellers condition (M = .82) than in the Similar Storytellers condition (M = .66), $F(1, 90) = 14.59$, $MSe = 0.47$. There were no two-way interactions (all $Fs < 1$). There was, however, a significant three-way interaction between age, acquisition condition, and item type, $F(2, 90) = 3.60$, $MSe = 0.47$. Floor and ceiling effects appear to have contributed to this interaction, and obscured its interpretation, but there is some slight indication that 4-year-olds were more sensitive to increases in similarity than 6-year-olds. Specifically, in the Dissimilar Storytellers condition 4-year-olds performed as well as 6-year-olds on Unique items (both means .80), but they showed a nonsignificant tendency to perform more poorly than 6-year-olds on Common items (means of .59 and .77 for 4- and 6-year-olds, respectively, $t(30) = 1.67$, $p < .06$, one-tailed).

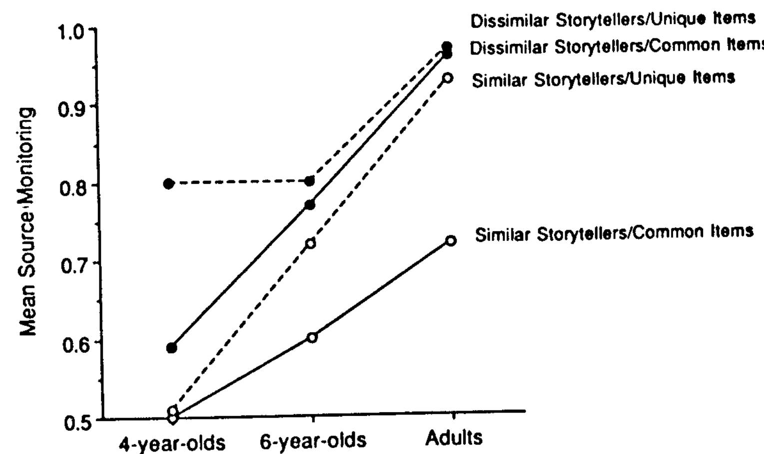

FIG. 1. Mean source monitoring scores as a function of subject's age, similarity of storytellers, and item type, Experiment 2.

The pattern of means also indicates that children in both age groups used memory information about both the perceptual characteristics and the semantic content associated with a source when attempting to remember which storyteller had talked about a particular detail: Although 4-year-olds' performance did not reliably exceed chance (.5) in the other three conditions (all $t(15) < 1.05$, $p > .30$), they performed quite well when the storytellers were dissimilar and only one storyteller had referred to the circus act in question (80% correct in both age groups, which is reliably above chance, $t(15) = 4.03$ for 4-year-olds and $t(15) = 3.96$ for 6-year-olds). On the other hand, 6-year-olds' performance was well above chance in the other conditions (all $ts(15) > 3.95$), but not when similar storytellers had both referred to the circus act in question (M = .60, $t(15) = 1.34$, $p > .19$).

Old/new recognition. The old/new recognition score was the proportion of test items correctly identified as old or new. This measure reflects both correct recognition of old items and correct rejection of new items, and has a chance value of .5. Recognition performance improved with age (M = .71, .84, and .89 for 4-year-olds, 6-year-olds, and adults, respectively, $F(2, 90) = 28.98$, MSe 0.10). Unlike source monitoring, however, old/new recognition was not affected by acquisition condition ($F < 1$), and age and acquisition condition did not interact, $F(2, 90) = 1.91$, $MSe = 0.10$, $p > .20$. Furthermore, an analysis of the source monitoring scores of a subset of 6-year-old and adult subjects matched on old/new recognition revealed a significant improvement with age (means of .79 and .90 for the child and adult subjects, respectively, $F(1, 32) = 7.52$, $MSe = 0.30$).

False alarms. There were 12 distractor items on the test: 8 Unrelated distractors that concerned details of circus acts not mentioned in either story and 4 Related distractors that concerned new details about circus acts mentioned in one or both stories. The false alarm rate (proportion of new items attributed to a storyteller) was much greater among the 4-year-olds ($M = .26$) than among the 6-year-olds ($M = .03$) and adults ($M = .04$), $F(2, 90) = 13.00$, $MSe = 0.86$. False alarms were more frequent to the Related than Unrelated distractors ($M = .12$ and .09, $F(2, 90) = 4.08$, $MSe = 0.13$). Acquisition condition had no effect on false alarm rates, and there were no significant interactions (all $Fs < 1$).

Of the four Related distractors, two referred to circus acts mentioned in both stories and two referred to circus acts mentioned in only one story. When subjects made a false alarm to a Uniquely Related distractor, they could attribute it either to the storyteller who had described the circus act referred to in that distractor or to the other storyteller. False alarms were too infrequent among 6-year-olds and adults to permit statistical analysis. Among 4-year-olds, the Uniquely Related distractors were attributed to the related storyteller more than twice as often as to the unrelated storyteller (totals, across subjects, of 13 and 6, respectively, binomial $p < .06$). Thus 4-year-olds tended to attribute a new detail about a circus act to the storyteller who had talked about that act. This finding can be interpreted as new evidence of inferential remembering in young children (Edwards & Middleton, 1988; Moeser, 1976; Prawat & Cancelli, 1976), although it may reflect similarity-based false recognitions rather than conscious inferences.

Discussion

Children made more source monitoring errors than adults, and subjects in all three age groups were more likely to be confused about which storyteller had said a particular thing if the two storytellers were similar to one another than if they were dissimilar. Furthermore, across age groups, subjects made source monitoring errors more frequently when the memory in question referred to a topic that both storytellers had talked about than when it referred to a topic mentioned by only one of the storytellers. Finally, although floor and ceiling effects may have masked interactions between age and source similiarity, there was some slight support for our prediction that age-differences are more likely to be observed when sources are similar to one another: In the Dissimilar Storytellers condition 4- and 6-year-olds performed equally well on circus acts mentioned by only one storyteller, but 4-year-olds showed a nonsignificant tendency to be less accurate than 6-year-olds on circus acts mentioned by both storytellers. The pattern of means also indicates that 4- and 6-year-old children can use both perceptual and semantic aspects of their memories when identifying the source of a remembered utterance.

EXPERIMENT 3

Taken together, the results of Experiments 1 and 2 demonstrate that people are more likely to confuse memories from similar than dissimilar external sources, and suggest that young children may be especially vulnerable to the effects of similarity. The developmental literature on reality monitoring, reviewed in the introduction, indicates that children may also be especially likely to confuse memories of actual and imagined self-generated acts—another case in which the to-be-discriminated memories would have similar content. Together, these results suggest that people may be relatively likely to confuse memories of things they imagined someone doing and memories of things they witnessed that person doing (because memories of perceived and imagined acts performed by the same actor should share similar content), and that young children may be especially prone to such errors. Experiment 3 tested these hypotheses.

Overview

The study tested children's and adults' source monitoring of memories of everyday actions (e.g., "touch nose," "cross arms"). For all subjects some actions were imagined and some were actually performed. The study used a 2 × 2 between-subjects design, with actor of the imagined actions as one factor (Imagined-Self vs Imagined-Other) and actor of the actual actions as the other factor (Actual-Self vs Actual-Other). Thus the same actor was involved in actual and imagined actions in the Actual-Self/Imagined-Self and Actual-Other/Imagined-Other conditions, whereas different actors were involved in actual and imagined actions in the Actual-Self/Imagined-Other and Actual-Other/Imagined-Self conditions. We predicted that subjects in both age groups would make more source monitoring errors when the same actor was involved in actual and imagined actions, and that only in those conditions would children make more errors than adults.

Method

Subjects. Data from 48 children and 48 adults were analyzed.[3] The children ranged in age from 7.1 to 10.4 years, with a mean age of 8.7 years, $SD = 1.0$ years. Children were recruited via letters sent home

[3] Data from three children and one adult were dropped and replaced. These subjects scored more than two standard deviations below the mean old/new recognition score, and the experimenter's notes indicated that they had been confused about the test instructions.

through their teachers. The adults were undergraduates at Williams College who participated for optional extra credit in an introductory psychology course.

Materials. A pool of 36 actions (e.g., touch nose, cross arms) was generated and randomly divided into three subsets (A, B, and C) of 12 actions. Three study sets were made by randomly interleaving items from each of the possible pairings of subsets (AB, AC, BC) with the constraint that no more than two items from a given subset occurred in immediate succession. Two versions of each study set were prepared, with assignment of subsets of actions to the Imagined and Actual conditions counterbalanced. Thus each action was used equally often as an Imagined, Actual, and New item.

A set of silent color video tapes was prepared. In the Actual-Other/Imagined-Other condition, each Actual trial consisted of a 10-s clip in which an actress seated in a chair performed an action (e.g., touched her nose), and each Imagined trial consisted of a 10-s clip of the actress sitting quietly in the chair with her hands in her lap. After each trial, the picture faded to black for a 2-s pause between trials, and then faded back for the next trial. The videos used in the other conditions were edited copies of the one used in the Actual-Other/Imagined-Other condition. In the Actual-Other/Imagined-Self condition, the picture remained black during the Imagined trials, whereas in the Actual-Self/Imagined-Other condition the picture remained black during the Actual trials. No video was used in the Actual-Self/Imagined-Self condition. The memory test consisted of all 36 actions in a random order, with the constraint that no more than two items in a given condition occurred in immediate succession.

Procedure. Subjects were told that the exeriment concerned visual imagery, and that they would be asked to perform (or watch) some simple actions and to imagine themselves (or another person) performing other actions. Two practice trials (one actual action and one imagined action) were used to clarify the instructions. On each trial, the experimenter asked the subject to watch or to imagine the actress performing an action or to perform or to imagine themselves performing an action (e.g., "Please watch the girl touch her nose" or "Please imagine touching your nose"). At the end of each Imagined action trial, subjects were asked to rate the vividness of the image they had generated. At the end of each Actual action trial, subjects were instructed to image that action and to rate the vividness of that image. Vividness ratings were collected for both types of trials to ensure that subjects could not later discriminate between memories of Imagined and Actual trials on the basis of whether or not a vividness rating had been made. Adults rated vividness on a 4-point scale; children simply indicated whether or not they had formed a clear image.

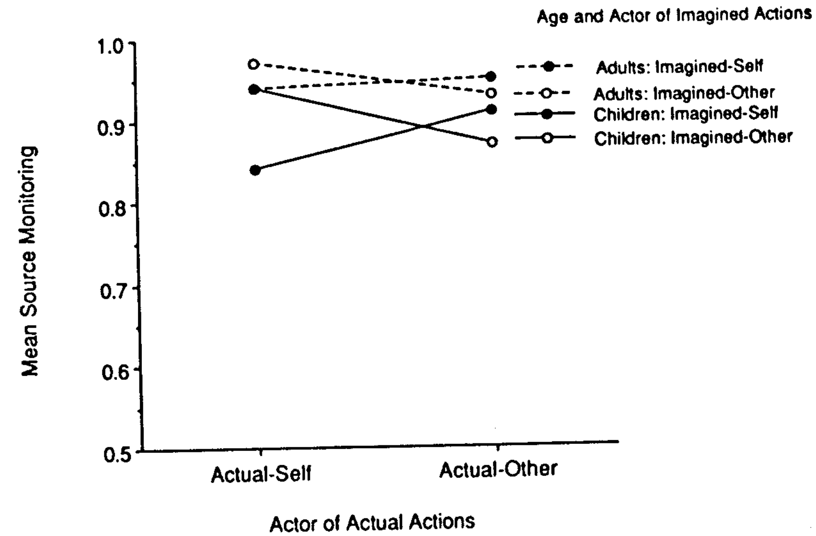

FIG. 2. Mean source monitoring scores as a function of subject's age and actor (self vs other) involved in actual and imagined actions, Experiment 3.

After the acquisition phase, child subjects were engaged in conversation for approximately 5 min and adults participated in an unrelated experiment for 30 min (this difference in retention interval was intended to equate age groups on old/new recognition—see Hirst, Johnson, Kim, Phelps, Risse, & Volpe, 1986). Subjects were then given a surprise memory test. The experimenter read each test item aloud and subjects were asked to indicate whether or not that action had been mentioned during the acquisition phase and, if so, whether it had been performed or imagined. Adults rated their confidence in each judgment on a 7-point scale, with 1 indicating "A completely uninformed guess" and 7 indicating "100% confident." Confidence ratings were not collected from children.

Results

The .05 level of significance was used for all statistical tests.

Source monitoring. The source monitoring score was the proportion of old items correctly identified as old for which the reality status (actual or imagined) was also correctly identified. These data were analyzed in a three-way analysis of variance with age, Actual actor, and Imagined actor as between-subjects factors. The mean source monitoring scores are presented in Fig. 2. Neither Actual actor nor Imagined actor had a significant main effect (both $Fs < 1$), but, as predicted, there was a significant interaction between these two factors, $F(1, 88) = 9.37$, $MSe = 0.006$, such that subjects made more reality monitoring errors when the same actor was involved in both Actual and Imagined actions than when different actors were involved. Adults significantly outperformed children (M = .95 and .89, respectively, $F(1, 88) = 15.04$, $MSe = 0.006$), and there was some indication that, as predicted, age effects were greater in the conditions in which the actual and imagined actions involved the same actor, although the predicted three-way interaction fell short of significance, $F(1, 88) = 2.61$, $MSe = 0.006$, $p < .06$, directional. Planned comparisons revealed that, as expected, children made significantly more errors than adults in the Actual-Other/Imagined-Other condition (M = .87 and .93 for children and adults, $F(1, 88) = 5.07$, $MSe = 0.006$) and in the Actual-Self/Imagined-Self condition (M = .84 and .94, $F(1, 88) = 10.52$, $MSe = 0.006$) but not in the Actual-Other/Imagined-Self condition (M = .91 and .95, $F(1, 88) = 1.81$, $p > .18$) or in the Actual-Self/Imagined-Other condition (M = .94 and .97, $F < 1$).[4] The interpretation of these predicted effects is compromised by the near-ceiling performance of adults in the conditions in which different actors performed the actual and imagined actions: All we can conclude with confidence is that performance on this reality monitoring task improved with age, at least when the same actor (self or other) performed both the actual and the imagined actions.

Old/new recognition. The old/new recognition score was the proportion of items correctly identified as old or new. These data were analyzed in a three-way analysis of variance with age, Actual actor, and Imagined actor as between-subjects factors. Adults performed at a higher level than children (M = .91 and .86, $F(1, 88) = 9.97$, $MSe = 0.005$). None of the other effects was significant, and the pattern of means differed altogether from that of the source monitoring data. Moreover, when children and adults were equated on old/new recognition (by dropping the two children with the lowest and the two adults with the highest old/new recognition scores in each condition), the analysis of source monitoring scores yielded the same results as the overall analysis: A significant effect of age and a significant interaction between Actual and Imagined actor (such that subjects made more errors when the same actor was involved in both Actual and Imagined events than when different actors were involved). Further, as in the overall analysis, planned comparisons indicated that children performed more poorly than adults in the Actual-Self/Imagined-Self condition (means of .87 and .94, $F(1, 72) = 6.00$, $MSe = 0.005$) and in the Actual-Other/Imagined-Other condition (means of .87 and .93, $F(1, 72) = 4.68$, $MSe = 0.005$), but not in the Actual-Other/Imagined-Self condition (mean of .96 for both age groups) or in Actual-Self/Imagined-Other condition (means of .91 and .94, $F < 1.24$). As in the overall analysis, interpretation of these effects is compromised by the near-ceiling performance of adults in the conditions in which different actors performed the actual and imagined actions. We can, however, conclude that source monitoring improved with age even when children and adults were matched on old/new recognition, at least in the conditions in which the same actor performed the actual and imagined actions.

[4] We have made a preliminary analysis of these data using the technique developed by Batchelder and Riefer (1990). The results are identical except that Batchelder and Riefer's technique indicates that in the Actual-Other/Imagined-Self condition children's source monitoring was significantly poorer than adults'.

Confidence ratings. Because confidence data were not collected from children, only the highlights of the analyses of adults' confidence in their reality monitoring decisions are reported here. Mean confidence was reliably greater for correct responses ($M = 6.15$, $SD = 0.61$) than for incorrect responses ($M = 4.86$, $SD = 1.43$), $t(37) = 7.28$.[5] Thus subjects demonstrated insight into their accuracy. There was a strong positive correlation ($r = .70$) between confidence in correct responses and confidence in incorrect responses: Subjects who were confident when correct also tended to be confident when incorrect. Further, subjects were often quite confident in incorrect judgments. For example, subjects were "100% confident" in 22% of their errors.

Discussion

As predicted, subjects were more likely to confuse memories of imagined and actual actions if the same actor (self or other) was involved in both kinds of action than if one actor was involved in the imagined actions and another in the actual actions. Also as predicted, relative to adults, children more often confused memories of actual and imagined actions, at least when the same actor was involved in both. It is clear from the old/new recognition data that this developmental trend cannot be accounted for in terms of general attentional or motivational factors. Finally, as expected, it was only in the conditions in which the same actor was involved in both actual and imagined events that children performed reliably more poorly than adults, although the high level of performance of both age groups may have masked developmental differences in the different-actors conditions. In any case, our results replicate earlier findings of developmental differences in the accuracy of discriminations between memories of actual and imagined self-generations (Realization Judgments), and provide new evidence that (a) people sometimes confuse memories of what they imagined a person doing with memories of what they witnessed that person doing and (b) children may be more likely to make such errors than adults.

These findings do not rule out the possibility that Realization Judgments

[5] Confidence ratings were not collected from the first four subjects (one in each condition).

(i.e., Imagined-Self vs Actual-Self) are particularly difficult for children. Indeed, there was a small and nonsignificant tendency for children in the Actual-Self/Imagined-Self condition to make more errors than those in the Actual-Other/Imagined-Other condition. It is possible that with a more powerful assessment this would prove to be a reliable effect. We would interpret such a finding as evidence that memories of actual and imagined actions are more similar if the self is the actor of both than if another person is the actor of both (e.g., perhaps it is easier to vividly imagine oneself than to vividly imagine another person, or perhaps the cognitive operations involved in doing and imagining oneself doing something are more similar than those involved in watching and imagining another person doing something). It is also possible that the current findings underestimate children's ability to differentiate between memories of actual and imagined events, because in our procedure *all* of the events were imagined—subjects were asked to image each actual action after they performed it (or watched the other person perform it). This was done so that reality monitoring discriminations could not be based on whether or not an image of the action had been rated for vividness. Thus our subjects had to discriminate between memories of actions that were actually performed and then imagined and memories of actions that were only imagined. It might be that children would perform as well as adults if actual and imagined acts could be discriminated on the basis of whether or not the action had been imagined. Our prediction is that the outcome would depend on the overall similarity of the actual and imagined acts; the greater the similarity, the more likely that children would make more source monitoring errors than adults.

GENERAL DISCUSSION

Taken together with the results of previous studies (e.g., Foley et al., 1983, 1987; Foley & Johnson, 1985; Foley et al., 1989, 1991), our findings suggest that children are more likely than adults to confuse memories from different sources whenever those sources are highly similar to one another, but that children can perform just as well as adults provided the sources are relatively discriminable. The evidence that suggested that children's deficit is specific to discriminations between memories of actual and imagined self-generations (e.g., Foley et al., 1989, 1991) is better described as evidence for the more general hypothesis that children have difficulty discriminating between memories from highly similar sources.

What makes two sources "highly similar" to one another, and what is it about discriminating between highly similar sources that is particularly difficult for young children? These are fundamental questions, about which we speculate as follows: A number of factors affect the ease with which the source of a memory is identified, including (a) the amount and nature of source-relevant information accessible from memory records of the

event itself; (b) the amount and nature of knowledge about the characteristic properties of that source; (c) the extent to which the attributes of the memory in question are uniquely specific to memories from its source; and (d) the stringency and appropriateness of the decision-making processes and criteria employed during remembering. These factors may be clarified by analogy to ongoing perception: Perceptual errors are relatively likely when stimuli are degraded or unfamiliar, when targets are similar to distractors, and when decision criteria are low or inappropriate. Likewise, all other things being equal, source monitoring errors are relatively likely when memories are vague or incomplete, when the source is unfamiliar, when more than one source characteristically gives rise to memories with properties similar to the memory in question, or when the attribution is made quickly rather than with careful deliberation. (See Johnson et al., 1991, for a more extended discussion of source monitoring processes.)

We speculate that any or all of these factors may contribute to interactions between age and condition in developmental studies of source monitoring. For example, it may be that children's ongoing experience (and hence the memory records of their experience) differs from adults' in ways that affect some kinds of source monitoring discriminations but not others. That is, children and adults may attend to different aspects of events, or may differently elaborate upon them (but see Foley et al., 1991). Alternatively, it may be that the kinds of memory records that quickly and easily come to mind when remembering an event differ for children and adults. Or, perhaps more likely than either of these, age-related changes in source monitoring may be due to differences in the retrieval strategies children and adults use when they are uncertain of the source of a memory (Ackerman, 1985; Kobasigawa, 1977). Developmental differences in strategic retrieval could contribute to age by condition interactions because the more difficult the discrimination the more likely that accurate performance would require special retrieval strategies. Yet again, age-related changes in metamemory, biases, or decision criteria may contribute to developmental effects in source monitoring. These conjectures about the cognitive mechanisms underlying age-related changes in source monitoring suggest directions for future research.

The results of the current studies counsel against sweeping generalizations about the developmental course of source monitoring. Source monitoring is not a single skill that a child acquires at a particular age. Rather, source monitoring involves decisions about a number of different aspects of event memories (remembering who, remembering where, remembering how, etc.) that we have grouped together because they constitute a useful and interesting category for analysis. Furthermore, accurate source monitoring depends upon a number of kinds of mental activities, such as perceptual analysis during encoding, retrieval of memory records, and decision-making processes. Thus it is likely that developmental changes in source monitoring accuracy will turn out to be gradual rather than sudden and domain specific rather than general. These considerations also suggest that source monitoring development will reflect individual differences along a number of dimensions.

Forgetting source is a common everyday experience (e.g., Linton, 1982). Source monitoring is also involved in a broad range of memory phenomena studied in the laboratory, including interference (Winograd, 1968), amnesia (Huppert & Piercy, 1976; Schacter, Harbluk, & McLachlan, 1984), false recognitions and intrusions (Alba & Hasher, 1983), misinformation effects (Lindsay, 1990; Lindsay & Johnson, 1989), memory misattributions (Jacoby, Kelley, & Dywan, 1989; Kelley & Lindsay, 1991), sleeper effects (Leippe, Greenwald, & Baumgardner, 1982), aging (McIntyre & Craik, 1987; Hashtroudi et al., 1989, 1990), and differences between autobiographical and episodic remembering (Tulving, 1984). Studies of developmental changes in memory for source potentially provide one way of clarifying the basic mechanisms involved in source monitoring. In turn, any understanding of age-related changes in source monitoring may illuminate other aspects of children's memory development (e.g., Ceci, Toglia, & Ross, 1987; Nelson & Gruendel, 1979; Todd & Perlmutter, 1980).

REFERENCES

Ackerman, B. P. (1985). Children's retrieval deficit. In C. J. Brainerd & M. Pressley (Eds.), *Basic processes in memory development: Progress in cognitive development research* (pp. 1–46). New York: Springer-Verlag.

Alba, J. W., & Hasher, L. (1983). Is memory schematic? *Psychological Bulletin,* **93,** 203–231.

Anderson, J. R., & Bower, G. H. (1972). Recognition and retrieval processes in free recall. *Psychological Review,* **79,** 97–123.

Anderson, J. R., & Bower, G. H. (1974). A propositional theory of recognition memory. *Memory and Cognition,* **2,** 406–412.

Batchelder, W. H., & Riefer, D. M. (1990). Multinomial processing models of source monitoring. *Psychological Review,* **97,** 548–564.

Ceci, S. J., Toglia, M. P., & Ross, D. F. (Eds.). (1987). *Children's eyewitness memory.* New York: Springer-Verlag.

Edwards, D., & Middleton, D. (1988). Joint recollection. *Journal of Social,* **5,** 3–25.

Foley, M. A., Aman, C., & Gutch, D. (1987). Discriminating between action memories: Children's use of kinesthetic cues and visible consequences. *Journal of Experimental Child Psychology,* **44,** 335–347.

Foley, M. A., Durso, F. T., Wilder, A., & Friedman, R. (1991). Developmental comparisons of explicit versus implicit imagery and reality monitoring. *Journal of Experimental Child Psychology,* **51,** 1–13.

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.

Foley, M. A., Santini, C., & Sopasakis, M. (1989). Discriminating between memories: Evidence for children's spontaneous elaborations. *Journal of Experimental Child Psychology,* **48,** 146–169.

Freud, S. (1958). Two principles of mental functioning. In J. Strachey (Ed. and Trans.), *The standard edition of the complete psychological works of Sigmund Freud* (Vol. 9, pp. 218–226). London: Hogart Press. (Original work published 1911.)

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1989). Aging and source monitoring. *Psychology and Aging,* **4,** 106–112.

Hashtroudi, S., Johnson, M. K., & Chrosniak, L. D. (1990). Aging and qualitative characteristics of memories for perceived and imagined complex events. *Psychology and Aging,* **5,** 119–126.

Hirst, W., Johnson, M. K., Kim, J. K., Phelps, E. A., Risse, G., & Volpe, B. T. (1986). Recognition and recall in amnesics. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **12,** 445–451.

Huppert, F. A., & Piercy, M. (1976). Recognition memory in amnesic patients: Effects of temporal context and familiarity of material. *Cortex,* **12,** 8.

Jacoby, L. L., Kelley, C. M., & Dywan, J. (1989). Memory attributions. In H. L. Roediger III, and F. I. M. Craik (Eds.), *Varieties of memory and consciousness: Essays in honour of Endel Tulving* (pp. 391–422). Hillsdale, NJ: Erlbaum.

Johnson, M. K. (1988). Discriminating the origin of information. In T. F. Oltmanns & B. A. Maher (Eds.), *Delusional beliefs: Interdisciplinary perspectives* (pp. 34–65). New York: Wiley.

Johnson, M. K., & Foley, M. A. (1984). Differentiating fact from fantasy: The reliability of children's memory. *Journal of Social Issues,* **40,** 33–50.

Johnson, M. K., Foley, M. A., & Leach, K. (1988). The consequences for memory of imagining in another person's voice. *Memory and Cognition,* **16,** 337–342.

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1991). *Source Monitoring.* Manuscript submitted for publication.

Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review,* **88,** 67–85.

Kelley, C. M., & Lindsay, D. S. (1991). *Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to questions.* Manuscript submitted for publication.

Kobasigawa, A. (1977). Retrieval strategies in the development of memory. In R. V. Kail, Jr., and J. W. Hagen (Eds.), *Perspectives on the development of memory and cognition* (pp. 177–201). Hillsdale, NJ: Erlbaum.

Kolers, P. A. (1977). On the representation of experience. In D. Gerver and W. Sinaiko (Eds.), *Language interpretation and communication* (pp. 245–258). New York: Plenum Press.

Kosslyn, S. M. (1978). The representational-development hypothesis. In P. Ornstein (Ed.), *Memory development.* Hillsdale, NJ: Lawrence Erlbaum.

Leippe, M. R., Greenwald, A. G., & Baumgardner, M. H. (1982). Delayed persuasion as a consequence of associative interference: A context effect. *Personality and Social Psychology Bulletin,* **8,** 644–650.

Lindsay, D. S. (1987). *Whence Comes this Memory?* Unpublished doctoral dissertation, Princeton University, Princeton, NJ.

Lindsay, D. S. (1990). Misleading suggestions can impair subjects' ability to remember event details. *Journal of Experimental Psychology: Learning, Memory, and Cognition,* **16,** 1077–1083.

Lindsay, D. S., & Johnson, M. K. (1987). Reality monitoring and suggestibility: Children's ability to discriminate among memories from different sources. In S. J. Ceci, M. P.

Toglia, & D. F. Ross (Eds.), *Children's eyewitness memory* (pp. 92–121). New York: Springer-Verlag.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory and Cognition*, **17**, 349–358.

Linton, M. (1982). Transformations of memory in everyday life. In U. Neisser (Ed.), *Memory observed: Remembering in natural contexts* (pp. 77–91). San Francisco: Freeman.

McIntyre, J. S., & Craik, F. I. M. (1987). Age differences in memory for item and source information. *Canadian J. of Psychology*, **41**, 175–192.

Moeser, S. D. (1976). Inferential reasoning in episodic memory. *Journal of Verbal Learning and Verbal Behavior*, **15**, 193–212.

Nelson, K., & Gruendel, J. (1979). At morning it's lunchtime: A scriptal view of children's dialogues. *Discourse Processes*, **2**, 73–94.

Piaget, J. (1962). *Play, dreams, and imitation in childhood* (C. Gattegno & G. M. Hodgson, Trans.). New York: Norton. (Original work published 1945.)

Prawat, R. S., & Cancelli, A. (1976). Constructive memory in conserving and nonconserving first graders. *Developmental Psychology*, **12**, 47–50.

Schacter, D. L., Harbluk, J. L., & McLachlan, D. R. (1984). Retrieval without recollection: An experimental analysis of source amnesia. *Journal of Verbal Learning and Verbal Behavior*, **23**, 593–611.

Todd, C. M., & Perlmutter, M. (1980). Reality recalled by preschool children. In M. Perlmutter (Ed.), *New directions of child development: Children's memory* (Vol. 10, pp. 69–85). San Francisco: Jossey-Bass.

Tulving, E. (1984). Precis of elements of episodic memory. *The Behavioral and Brain Sciences*, **7**, 223–268.

Winograd, E. (1968). List differentiation as a function of frequency and retention interval. *Journal of Experimental Psychology Monographs*, **76**, 1–18.

RECEIVED: September 10, 1990; REVISED: August 2, 1991