Journal of Experimental Psychology: Applied
2001, Vol. 7, No. 1, 27–50

Copyright 2001 by the American Psychological Association, Inc.
1076-898X/01/$5.00 DOI: 10.1037//1076-898X.7.1.27

Children's Eyewitness Reports After Exposure to Misinformation From Parents

Debra Ann Poole
Central Michigan University

D. Stephen Lindsay
University of Victoria

This study examined how misleading suggestions from parents influenced children's eyewitness reports. Children (3 to 8 years old) participated in science demonstrations, listened to their parents read a story that described experienced and nonexperienced events, and subsequently discussed the science experience in two follow-up interviews. Many children described fictitious events in response to open-ended prompts, and there were no age differences in suggestibility during this phase of the interview. Accuracy declined markedly in response to direct questions, especially for the younger children. Although the older children retracted many of their false reports after receiving source-monitoring instructions, the younger children did not. Path analyses indicated that acquiescence, free recall, and source monitoring all contribute to mediating patterns of suggestibility across age. Results indicate that judgments about the accuracy of children's testimony must consider the possibility of exposure to misinformation prior to formal interviews.

During the past decade, there has been keen interest in young children's performance when interviewed about autobiographical events. This interest was sharpened by broad social changes that led the public to be more concerned about crimes to which child victims are often the sole witnesses (e.g., child sexual abuse), but it also reflected a movement in psychology away from relatively artificial research paradigms (e.g., studies of memory for word lists) and toward more naturalistic and multifaceted approaches to research. Although the field of eyewitness testimony receives considerable attention for its forensic implications, researchers increasingly view eyewitness paradigms as tools for studying basic issues in memory and cognition.

There have been several recent reviews of the literature on interviewing children for forensic purposes (e.g., Ceci & Bruck, 1995; Poole & Lamb, 1998), and we do not offer an exhaustive recapitulation here. In its broadest outlines, the literature can be crudely summarized by saying that even very young children (i.e., 3- to 4-year-olds) can provide quite detailed and accurate accounts of past autobiographical events under some conditions but that even older children and adults often provide impoverished and inaccurate accounts under others. The aim of the current research was to provide additional insight into factors that compromise or enhance the amount and accuracy of the information interviewers gain from young children. Borrowing terms from Wells's (1993) analysis of eyewitness suspect identification, our research examined both "system variables" (i.e., factors that are under the control of investigators in forensic cases, such as the way questions are phrased) and "estimator variables" (i.e., factors that may affect the amount or accuracy of information children report but that cannot be controlled by investigators, such as the age of the child or the nature of the alleged event).

Numerous studies have documented that the accuracy of children's testimonies can be degraded when interviewers ask misleading questions or provide social feedback that favors particular answers (Ceci & Bruck, 1995). Rather than emphasizing suggestive interviewing techniques, the central focus of the current study was on the effects of misleading information presented to young children by their parents prior to and outside of an interviewing situation. It is likely that many children who are involved in forensic investigations were exposed to information from trusted adults, ranging from overheard conversations to deliberate coaching before participating in a formal forensic interview. This raises the worrisome possibility that false reports based on prior exposure to suggestions may intrude into forensic interview responses even when those interviews are conducted in an optimally nonsuggestive manner.

Concern about children's experiences prior to formal interviews stems from evidence that suggestions presented in a variety of contexts sometimes degrade children's testimonies. In Leichtman and Ceci's (1995) "Sam Stone" study, for example, 3- to 6-year-olds were repeatedly told stories that depicted Sam as a clumsy,

Debra Ann Poole, Department of Psychology, Central Michigan University; D. Stephen Lindsay, Department of Psychology, University of Victoria.

This material is based on work supported by National Science Foundation Grant SBR-9409231. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. We thank the members of our research team for assistance in data collection and coding (Sonya Lucke, Scott Kuligoski, Lisa Wolf, Carolyn Miska, Chad Stabler, Rachel Johnson, Erin Henderson, Doug Haseley, Rachel Franceschina, Laura Whitlock, and Susan Williams), and Dan and Linda King (National Center for PTSD, Boston VA Medical Center) for consultations regarding data analyses. We also extend special thanks to the parents and children in Michigan who have volunteered their time to support our work.

Correspondence concerning this article should be addressed to Debra Ann Poole, Department of Psychology, 231 Sloan Hall, Central Michigan University, Mt. Pleasant, Michigan 48859.

bumbling person, followed by an uneventful visit by Sam to their classroom. In a subsequent interview, 37% of these children reported that Sam committed at least one misdeed in response to direct questioning compared with only 10% of the children who had not been exposed to the negative stereotype. Similarly, in a study by Bruck, Ceci, Francoeur, and Barr (1995), a research assistant mischaracterized children's reactions to a routine inoculation experience. This mild suggestion did not influence 5-year-olds' descriptions of their experiences 1 week later, but misleading narratives repeated approximately 1 year after the inoculations did alter children's responses. An important feature of these studies is that findings were based on children's responses during a final interview that did not contain overtly suggestive statements or strongly misleading questions.

We developed a paradigm to assess the impact of misleading suggestions from parents on children's eyewitness reports. In our initial study (Poole & Lindsay, 1995), 17 3-year-old children interacted with an unfamiliar man, named Mr. Science, who showed them four "science demonstrations" (e.g., using two funnels and a rubber tube to make a telephone). We selected these demonstrations because children found them engaging, they were likely to be relatively novel, and we could counterbalance complete mini-events across the various conditions (e.g., experienced versus suggested).[1] Approximately 3 months after the initial session, parents of each child received a story book, titled "A Visit to Mr. Science," that described experienced and nonexperienced events. After parents had read the story to their children three times, each child was interviewed using a protocol that began with rapport building, then open-ended prompts, then direct questions about experienced and suggested events, and, finally, a source-monitoring procedure designed to help children differentiate between experienced and suggested events.

The most striking finding of our 1995 study was that even in the free-report phase of the final interview, a substantial percentage (41%) of the children falsely reported experiencing at least one of the events they had merely heard about in the story (including one child who spontaneously reported having something "yucky" put in her mouth). Leading questions increased false reports, with 94% of the children falsely responding "yes" to direct questions about one or more suggested event and 88% describing these events. A subsequent source-monitoring procedure—in which children were reminded of the story, explicitly told that some things described in the story might not actually have happened to them, and asked to indicate whether they had actually experienced each event—did not reduce the rate of false reports.

The Current Study

Our findings with the Mr. Science procedure (Poole & Lindsay, 1995) were unexpected, but the sample size was small and only 3- to 4-year-old children were tested. There are good reasons to believe that the results might not be representative of older preschoolers or early elementary school children. In the eyewitness literature, young preschoolers often contribute disproportionately to error rates both in free recall (e.g., Goodman, Quas, Batterman-Faunce, Riddlesberger, & Kuhn, 1994; Poole & White, 1993) and in response to direct questions (e.g., Steward et al., 1996). Moreover, 3- and 4-year-old children have proven especially poor at conveying the source of their knowledge, even minutes after that knowledge is imparted by experimenters (e.g., Gopnik & Graf, 1988). Young preschoolers' poor source monitoring, high rates of false recognition, and tendency to confabulate about autobiographical events led Schacter, Kagan, and Leichtman (1995) to propose that immature frontal lobe development might be involved in the heightened suggestibility in this age group.

The foregoing considerations imply that older children would be less vulnerable to the suggestive influences brought to bear on the 3- to 4-year-old children we tested. Nonetheless, it would be naïve to assume that older children would be immune to parental suggestions. The growing literature on memory falsification in adults documents that suggestibility is by no means limited to early childhood (e.g., Hyman & Billings, 1998; Loftus, 1997). Although most prior studies have found reductions in suggestibility from preschool to the school years, the magnitude and even existence of such developmental trends varies considerably across studies (Ceci & Bruck, 1993). Developmental changes in some cognitive processes during the preschool and early childhood years likely lessen susceptibility to suggestions, but developments in other processes may have the opposite effect. For example, improvements in memory source-monitoring skills (which would presumably decrease false reports) might be partly counteracted by developmental increases in conversational skills and motivation to demonstrate relevant knowledge in response to interviewers' questions.

One purpose of the current study, therefore, was to assess the effects of misinformation from parents with a larger sample and age range of children. The current study also improved on our earlier procedure by measuring reports of two different types of body touch (one suggested and one control, counterbalanced across children). We also assessed the efficacy of new source-monitoring instructions, designed to help children differentiate between memories of events they had actually experienced and memories of events they had merely heard described in the story.[2]

An additional aim of the current study was to investigate the relative "survival" of true and false memories over time, an issue that has both practical and theoretical implications. To do so, we conducted a final interview 1 month after exposure to the story. Speculating about the likely time course of true and fictitious reports, two scenarios seem possible. On the one hand, there are reasons to believe that reports of false events mentioned in the story will decline steeply between the second interview (conducted soon after exposure to the suggestions) and the final interview

[1] Interviewers conducted a nonsuggestive interview immediately after exposure to the science demonstrations to assess the efficacy of open-ended prompts designed to elicit additional free-report information from young children without compromising accuracy. As predicted, we found that even the youngest children did well, that prompts to report visual details and auditory details enabled children of all ages to provide additional information about their experiences with Mr. Science, and that almost all of this additional information was accurate. Subsequent research confirmed that children report more new information in response to these focused prompts than when they are repeatedly asked to report everything they remember about an event (Elischberger & Roebers, in press).

[2] Throughout this article, we use the term *source monitoring* to refer to behaviors that have to do with differentiating (or failing to differentiate) between memories from different sources, rather than to refer to any particular theoretical account of the mechanisms involved in such behaviors.

compared with reports of actual events experienced during the interaction with Mr. Science. Both laboratory research (from Ebbinghaus on) and more recent naturalistic studies of autobiographical memory (e.g., Thompson, Skowronski, Larsen, & Betz, 1996) have consistently shown that forgetting over periods of weeks or months follows a negatively accelerated function, such that the more recent story material would be forgotten at a greater rate between the second and third interviews than the older memories of experienced events. There is also evidence that memories of unique, participatory experiences (such as the interactions with Mr. Science) are generally more robust than memories of passively received material (such as the story) (e.g., Tobey & Goodman, 1992). Consistent with these general principles, Brainerd and Poole (1997) reviewed a number of studies in which errors induced by suggestibility procedures dropped out of children's narratives more rapidly than accurate reports did. On the other hand, the parents in our study read the misleading story to their children several times before the second interview, and parental authority and repetition might produce highly stable levels of false reporting of suggestions from the story.

Finally, in recent years there has been increasing interest in the extent to which the reliability of children's reports can be estimated by individual difference measures. The relatively large sample size and age range tested in the current study enabled us to conduct analyses of relationships between descriptive variables (e.g., parental education) and accuracy across the various phases of our interview procedure. We also conducted exploratory analyses of how developmental changes in acquiescence, recall, and memory source monitoring mediate suggestibility.

In summary, the current project addressed nine questions about the impact of misinformation from parents on children's event reports:

1. How does the frequency of false reports of parental misinformation change between 3 and 8 years of age when information is elicited with open-ended questions/invitations?

2. Does continued prompting with open-ended invitations compromise or enhance children's accuracy?

3. How does the frequency of false reports of parental misinformation change between 3 and 8 years of age when information is elicited with direct yes–no questions?

4. When interviewers ask children direct questions and prompt them to describe events, is the true state of affairs "diagnosed" more accurately by basing decisions on children's initial "yes" or "no" answers, or by basing decisions on the presence or absence of narrative descriptions of events?

5. What is the developmental trend in source-monitoring performance between 3 and 8 years of age?

6. Are responses to questions about fictitious touching experiences more accurate than responses to questions about fictitious nontouching events?

7. What is the relative stability of true and false reports across an interval with no further intentional memory contamination?

8. What individual difference factors correlate with suggestibility? More specifically, do descriptive variables (e.g., socioeconomic status) and personality characteristics that are assessed during nonsubstantive portions of the interview (e.g., talkativeness) differentiate children who have high versus low rates of false reports?

9. Do acquiescence, recall, and memory source monitoring mediate the decline in suggestibility that is often observed between the early preschool and elementary school years?

We postpone consideration of prior research on these questions to the discussion section.

Method

Participants

We recruited 128 families by distributing letters at day care centers and posting advertisements in local newspapers. These procedures, with an enrollment limit of one child per family, were intended to recruit children from different neighborhoods and reduce the possibility of children discussing their participation with one another. The final sample included children from 14 small towns, predominantly within a 30 mile (48-km) radius of Mt. Pleasant, Michigan. Families received $10 for each of three sessions.

Data from 11 children were lost (because of experimenter error, a family move, or the inability to schedule the third interview within target dates). In addition, two 3-year-olds made no verbal or nonverbal responses during one of the interviews, and one family of a 4-year-old declined the third interview. Data therefore were based on 114 children; 19 3-year-olds (including one child who was 2 years, 11 months old at Session 1), 19 4-year-olds, 18 5-year-olds, 18 6-year-olds, 18 7-year-olds, and 22 8-year-olds.

There were more girls (58%) than boys (42%), but there was not a significant discrepancy among age groups in the percentage of girls, $\chi^2(5) = 4.18$, $p = .52$. The families represented diverse levels of parent education and family income. For example, 17% of the mothers completed no education past high school, but 22% held degrees beyond the bachelors level. Family income was also varied (median income bracket = $30,000–39,999). There was no significant discrepancy among age groups in mother's educational level or family income, $ps > .54$.

Procedure

Overview. The children participated in three sessions, with individual children assigned to different female interviewers for each of the sessions. In Session 1, each child interacted individually with Mr. Science for 16 min, after which an interviewer asked nonsuggestive questions about this experience. Session 1 thus provided baseline data on the children's accuracy when target events were fresh in their minds and responses were not contaminated by exposure to misinformation. Approximately 3.5 months later, parents read a story to their children that described some experienced and some nonexperienced events. Shortly afterward, Session 2 took place in the children's homes. This interview included open-ended invitations, followed by direct yes–no questions and a source-monitoring procedure. Session 3 was a repetition of the Session 2 interview after another 1-month delay.

Session 1. On arriving at the laboratory, each parent completed an informed consent, consent to approve or disapprove of using audio or videotapes for professional demonstrations, and a family variables questionnaire. Each child then interacted one-on-one with Mr. Science for 16 min. Sessions began with Mr. Science setting a timer, followed by four demonstrations (e.g., spinning tops and reaching for them with and without prism glasses). For each activity, Mr. Science first demonstrated, then encouraged the child to explore the materials.

After participating in the science demonstrations, the child left the science area and was questioned by one of several female interviewers. This immediate interview began with orienting comments (e.g., "I'm going to ask you some questions. . .") and three rapport-building questions about family and the child's favorite activities. The interviewer then introduced the science demonstrations as the topic and proceeded through five open-

ended questions/invitations, hereafter referred to as the *tell, more, looked, heard,* and *think* prompts:

1. I am going to ask now about what happened today in the science room. Start with the first thing that happened and *tell* me everything you can, even things you don't think are very important. But don't guess or make anything up. Just tell me what you saw or heard or did in the science room, so I will know about them too. (If the child said nothing, they were prompted, "You can start talking about the science room now.")

2. Can you tell me *more* so that I will know all about what happened in the science room too?

3. Sometimes we remember a lot about how things *looked.* Think about all of the things that were in the science room. Tell me how everything looked.

4. Sometimes we remember a lot about sounds, or things that people said. Tell me about all of the things you *heard* in the science room.

5. *Think* about what you told me. Is there something you didn't tell me that you can tell me now?

The order of looked and heard prompts was counterbalanced across participants, with individual condition assignments fixed across the three sessions.

Session 2: The misinformation manipulation. We mailed storybooks to families 3 months after Session 1. Each family was instructed to read the four-page story ("A Visit to Mr. Science") to their child three times before Session 2. Parents whose children could read were told to hold the book and read as they normally would to a younger child. Parents were instructed to record the dates on which they read the story in spaces printed on the back of the book. Although we informed parents that it was desirable to read the story on 3 consecutive days, we also explained that this might not be possible for some families and that they could catch up by reading twice in 1 day (e.g., morning and evening).

The story, which used the child's own name throughout, described accurate contextual information (e.g., coming to a building covered with ivy) and four science demonstrations: two that the child had experienced and two nonexperienced. Thus, there were four science-demonstration conditions for each child: two experienced-only demonstrations, two experienced-heard demonstrations (experienced demonstrations that were also reported in the book), two heard-only demonstrations (demonstrations mentioned only in the book), and two control demonstrations (neither experienced nor heard). Specific demonstrations were counterbalanced across conditions so that each demonstration appeared approximately equally often in each of the four conditions. Each story also described a nonexperienced event about body touch: that Mr. Science put something yucky in the child's mouth (a wet wipe that was used to wash the child's face and hands) or that Mr. Science hurt the child's tummy (when he pushed a little hard to apply a reward sticker). In fact, wet wipes and stickers were not part of the actual events. We refer to nonexperienced touch events that were described in the story as touch-heard events, with touch-control events referring to touch events that were neither experienced nor described in the story.

Each parent was informed that not all children had experienced the same demonstrations and that the story described some events their child had experienced and some that other children had experienced. The parents were instructed not to correct their children or remind them of earlier comments about their experiences with Mr. Science.

Sessions 2 and 3: The interviews. Interviews after the misleading suggestions were conducted in the child's home by assistants who did not know which demonstrations or story individual children had experienced. The interview format, which was identical for Sessions 2 and 3, mimicked the "phased" structure recommended for forensic interviews (Poole & Lamb, 1998): rapport building, open-ended prompts, and direct questions, followed by a source-monitoring procedure developed specifically for our study. Interviews began with two rapport questions regarding events of the day. The interviewer then introduced the topic and began the open-ended portion of the interview with the tell prompt:

> All right. We are going to talk about something different now. A while ago you went to visit Mr. Science. Mr. Science played some games with you and then you answered some questions into a tape recorder, just like this one. Do you remember playing with Mr. Science? Good. I want you to tell me everything that happened when you were playing with Mr. Science. I wasn't in the room, so I don't know what happened. Start with the first thing that happened and tell me everything you can, even things you don't think are very important. But don't guess or make anything up—just tell me what you saw or heard or did the time you played in the science room with Mr. Science.

Open-ended questioning continued with the *more, looked, heard,* and *think* prompts, as in Session 1.

Next, the interviewer provided instructions intended to communicate to the child that she or he had permission to say "no" in response to direct questions:

> Now I have some other questions to ask about the time you visited Mr. Science. If I ask you about something that Mr. Science didn't do that time you visited him, I want you to say "no." But if you remember something that I ask about, then I want you to tell me about it.

The interviewer then proceeded through 10 direct question pairs, one pair for each of the eight science demonstrations and the two details about body touch. Each question pair began with a yes–no question (e.g., "Did Mr. Science have a machine with ropes to pull?"), followed by one prompt for a "yes" response (e.g., "Tell me about the machine") or another for a "no" response (e.g., "Can you tell me about the machine?"). The order of question pairs was randomized for each participant with the restriction that one question from each science demonstration condition and one touch event appeared in the first and second block of five questions.

The final portion of the interview was a source-monitoring procedure in which the children were asked to indicate whether particular events appeared in the story and/or actually happened with Mr. Science. We revised the procedure because 3- and 4-year-olds performed poorly in our prior study (Poole & Lindsay, 1995). In that study, the interviewers proceeded through each of the events, first asking if the event had actually occurred and then asking if the event had been mentioned in the story. This procedure required children continually to shift their reference point between the events and the story, and all children were asked a large number of yes–no questions regardless of how much information they had already provided. The goal of the revised procedure was to capitalize on children's memory of the story and their prior responses, to reduce the number of yes–no questions needed to assign events to sources. This was accomplished by first asking each child to reconstruct the story with the following instructions:

> I have been asking you about that time you visited Mr. Science. Did your mom or dad read you a story about Mr. Science too? I don't know what happened in the story. What did Mr. Science do in the story that your mom or dad read?

After the child responded, the interviewer encouraged appropriate "no" responses in preparation for asking a series of yes–no questions:

> Now I am going to ask you whether some things were in the story. For example, I might ask you if Mr. Science floated across the room in the story. If I ask you about something, and you don't remember it, just tell me "no." Did the story say that Mr. Science floated across the room? Good, that's right, he didn't float across the room, so you were right to say "no." Let's do some more.

Next, the interviewer asked specific questions about any events (from the eight science demonstrations and two details about body touch) that the child had not mentioned during free recall of the story (e.g., "Did the story say that Mr. Science made an eyedropper go up and down in a bottle?").

Following this reconstruction of the story, the interviewer shifted the topic to the experienced events:

> Okay. You know, there might have been some things in the story that you really did when you visited Mr. Science. But there might have been things in the story that you didn't really do . . . things that were only in the story. Now I am going to ask you about what really happened when you visited Mr. Science. If you don't remember something, or it was just in the story, say "no."

The interviewer then asked about those events to which the child had responded "yes" during direct questioning (e.g, "Did Mr. Science *really* show you how to make an eyedropper go up and down in a bottle?").

In summary, there were 20 possible yes–no questions during the source-monitoring phase of the interview: 10 "story" questions that addressed whether each event appeared in the story, and 10 "event" questions that addressed whether each event really had happened. Individual children, however, answered story questions about only those items they had not mentioned during free recall of the story, and event questions about only those items they had previously accepted with a "yes" response during the direct questions phase of the interview.

Interviewers retrieved the storybooks at the Session 2 interview. Approximately 1 month later, the interview procedure was repeated for Session 3 by a new interviewer.

Results

The data address a number of topics regarding children's suggestibility, the impact of various interviewing techniques, and the interpretation of children's responses. Because a detailed presentation of each issue is cumbersome, we periodically refer readers to a supplementary report for additional documentation.[3] The general strategy for analyzing children's responses during each phase of the interview (i.e., open-ended prompts, direct questions, and source-monitoring questions) was to conduct separate trend analyses on each of the event conditions, followed by a priori comparisons of selected pairs of conditions. An alpha level of .05 was used for all statistical tests.

Data Preparation and Preliminary Analyses

Data Scoring

Working from typed transcripts that did not indicate children's ages or condition assignments, two assistants read responses to open-ended prompts and recorded which of the 10 target events each child described. Intercoder agreement across all 114 participants was 93%; discrepancies were primarily oversights that were resolved by rereading the transcripts.

The amount of information in narrative responses to open-ended and yes–no questions was coded in three passes. Coders first divided responses into units of information called syntactic units (SUs). SUs included words that described an actor (e.g., "Mr. Science"), an action (e.g., "spun"), or a direct object (e.g., "the tops"), as well as temporal terms (e.g., "then"), prepositional phrases, and quotes, each of which counted as a single SU. Within each response, complete propositions that were redundant or off topic were not coded. Intercoder agreement for Total SUs, based on nine interviews (105 responses) randomly selected with the restriction that they represented the various age groups, was 93%.

After dividing the narratives into SUs, SUs in response to the open-ended prompts were coded as "New SUs" if they had not been previously mentioned. New SUs for the second through fifth open-ended prompts thus represented how much information was gained by continued prompting. Intercoder agreement across the 105 responses in the nine interviews coded by both raters was 86%. For Total and New SUs, data were based on decisions of the primary coder for each transcript.

Two assistants then read through the transcripts and assigned SUs to accuracy categories. Information from the open-ended prompts was divided into four categories: (a) accurate information; (b) detail errors (e.g., saying the lab coat was blue when it was actually black); (c) spontaneous intrusions (i.e., off-topic talk that might not be recognized as such by an individual who did not know what had actually happened in the science room); and (d) suggested information (i.e., nonexperienced events described in the story or the child's spontaneous elaborations based on the story). Assignment of information to categories was largely determined by the child's condition assignment. For example, if spinning tops was a heard-only demonstration, then all statements that the child made about spinning tops were coded as suggested information. Because some error types occurred only rarely, intercoder reliability was based on all transcripts from 60 randomly selected subjects to ensure an adequate sample. Intercoder agreement for categorizing SUs was 99.9% for accurate information, 92.2% for detail errors, 100% for suggested information, and 98.4% for spontaneous intrusions; discrepancies were resolved by discussion.

For direct questions, SUs were divided into accurate information, wrong event (information that was wrong for the event in question, but was an accurate description of another event that the child did not explicitly mention), detail errors, spontaneous intrusions, suggested information, or incorrect rejections (e.g., saying "We didn't do that" to an experienced event). Intercoder agreements, based on transcripts from the 60 reliability subjects, were 99.3%, 100%, 82.5%, 99.6%, 95.2%, and 100%, respectively; discrepancies were resolved by discussion.

Preliminary Analyses

There were no significant differences among age groups in the number of days between Sessions 1 and 2 ($p = .65$, the range for individual children was 99–140 days), the number of days between Sessions 2 and 3 ($p = .71$, the range was 21–57 days), or the number of times parents reported reading the book to their children ($p = .79$, the range for individual children was 2–5 times; range of averages for age groups was 2.9–3.1).

To avoid eliminating data from children who were too shy to enter the science room alone, a parent was allowed to observe the science demonstrations if the child was reluctant to separate. We therefore considered whether parents who observed the events might have volunteered information that improved the accuracy of their children's reports. Sixteen parents chose to observe Session 1, accompanying 9 3-year-olds, 1 4-year-old, 3 5-year-olds, 2 6-year-olds, and 1 7-year-old. For the 3-year-olds and the older

[3] Interview transcripts and data files from this study are archived at the National Data Archive on Child Abuse and Neglect, Cornell University, FLDC-MVR Hall, Ithaca, NY 14853. The supplementary report is available from the archive website at www.ndacan.cornell.edu.

children separately, we compared children whose parents were present versus children whose parents were not present on each of five measures: age, and the proportions who responded "yes" to questions about experienced and heard-only events in Sessions 2 and 3. There were no significant differences on any of these measures ($ps > .05$).

Performance Before Misleading Suggestions (Session 1)

The initial interview gave us an estimate of free recall performance shortly after the event and prior to the suggestibility manipulation. Free recall was measured in two ways: as the proportion of events that the children mentioned during free recall from the various event conditions, and as the amount of accurate and inaccurate information (SUs) that they reported.

Proportion of Events Reported

As expected, the proportion of events that the children mentioned in their initial interview increased steadily with age, $F(1, 108) = 92.30$, $MSE = .07$, $p < .001$. From 3 to 8 years, respectively, the proportions of events mentioned (out of 4 possible, standard deviations in parentheses) were .09 (.21), .25 (.32), .51 (.30), .71 (.18), .68 (.32), and .74 (.21). No child spontaneously described any of the nonexperienced target events prior to exposure to the story.

The Amount of Accurate and Inaccurate Information in Response to Open-Ended Prompts

Because the majority of information reported in response to open-ended prompts was categorized as "new" (91%), and because performance patterns were comparable for Total and New SUs, we report only analyses on New SUs (hereafter, simply referred to as SUs). Table 1 lists the mean number of SUs volunteered by children of various ages and the proportion of information in each of the error categories. As expected, the amount of new information that the children reported increased dramatically with age, as indicated by a significant linear age trend, $F(1, 108) = 52.89$, $MSE = 2576.16$, $p < .001$. The proportion of detail errors was low, averaging only .01 of SUs, and the rate of detail errors did not decrease significantly with age ($ps > .62$). The rate of spontaneous intrusion errors did decrease significantly with age, however, $F(1, 93) = 7.47$, $MSE = .01$, $p < .01$. The following section clarifies why some younger children offered irrelevant but potentially confusing material in their narratives.

Table 1
Mean Number of SUs Mentioned in Response to Open-Ended Prompts in Session 1, and the Proportion of Information That Was Inaccurate (With Standard Deviations)

		Proportion inaccurate	
Age	SUs	Detail errors	Spontaneous intrusion errors
3	5.7 (10.1)	.01 (.03)	.10 (.23)
4	13.3 (13.8)	.00 (.01)	.02 (.07)
5	39.6 (29.2)	.02 (.05)	.05 (.14)
6	65.6 (53.7)	.01 (.02)	.00 (.00)
7	73.2 (53.6)	.02 (.02)	.00 (.00)
8	102.9 (87.5)	.01 (.01)	.00 (.00)
Age linear	$p < .001$	*ns*	$p < .01$

Note. SUs = syntactic units.

The Impact of Continued Prompting

A breakdown of the proportion of children who reported any new accurate information in response to each successive prompt documented that prompting yielded additional information across all ages, but that older children were more likely to add information than were younger children. In response to the fourth prompt, for example, proportions of children who provided some new information, from 3- to 8-year-olds, were .26, .53, .56, .83, .67, and .96, respectively. As in our previous study (Poole & Lindsay, 1995), fewer children continued to add additional information in response to the final think prompt, but a minority did so (.11, .11, .22, .22, .28, .27).

Children's error rates on the first two questions were compared with their error rates on the final three questions by entering overall error rates (detail errors plus spontaneous intrusions) into a question block (proportion of new information that was inaccurate for the first two versus final three questions) by linear age trend analysis. There was no significant impact of continued prompting, $p = .56$, and no Prompting × Linear Age Trend interaction, $p = .95$. Therefore, there was no evidence that continued prompting with these open-ended questions or invitations was a risky practice.

Although error rates were generally low throughout the interview (see supplementary report for a breakdown), there was a disturbing 20% rate of spontaneous intrusion errors for 3- to 4-year-olds on the final prompt (compared with a 6% intrusion rate by the 5- to 6-year olds, and a 0% intrusion rate by the 7- to 8-year-olds). However, this figure resulted from the fact that one child intruded 5 SUs ("*Ghosts, pumpkins*, and *a spider. Didn't crawl on me*") to a question that prompted only four other 3- to 4-year-olds to add some additional accurate information.

Patterns of spontaneous intrusion errors led us to inspect the transcripts to determine whether there might be a problem with the final think prompt. We found that despite four prior questions about the science room, some children did not assume that the final prompt ("Is there something you didn't tell me that you can tell me now?") referred to the science experience. In abuse investigations, children who spontaneously return to earlier questions about family, friends, or school could lead interviewers to suspect that numerous individuals were somehow involved in the alleged abuse. An obvious solution to this problem is to mention the topic redundantly throughout questioning (e.g., "Is there something you didn't tell me about the science room that you can tell me now?").

Summary of Session 1 Performance

Session 1 confirmed that even young children can accurately report recent, complex events when they have not been influenced by misinformation or highly specific questioning. Reports by the youngest two age groups were quite sparse but, as in previous research, the vast majority of the information reported by even the youngest children was accurate. Over all age groups, only 1% of

the children's reports were categorized as detail errors, and the 2% spontaneous intrusion rate might have been reduced by phrasing the final prompt to refer unambiguously to the science experience. Under these optimal interviewing conditions, there was no evidence that accuracy was compromised by continued prompting with our open-ended invitations.

Performance After Misleading Suggestions From Parents (Sessions 2 and 3)

The data from Sessions 2 and 3 address the nine questions posed in the introduction, and in this section we summarize the results pertaining to each of those questions. We first consider responses to open-ended questions (Question 1) and the impact of continued nonleading prompting (Question 2). Next, we summarize the accuracy of children's answers to direct yes–no questions (Question 3) and evaluate their willingness to describe experienced and suggested events when they were encouraged to do so (Question 4). We then explore whether the source-monitoring procedure reduced false reports of suggested events and its effects on accurate reports of experienced events (Question 5). Because we expected that false reports of fictitious touching experiences (which did not fit the overall theme or script of science demonstrations) would be less frequent than false reports of suggested, script-consistent science demonstrations, throughout these subsections we compare children's accuracy for these two types of events (Question 6). The final three subsections explore basic mechanisms that underlie varying rates of suggestibility across time and individuals (Questions 7–9).

Responses to Open-Ended Prompts

Proportion of events reported. Contrary to claims that misleading suggestions rarely affect children's free-recall narratives, a substantial percentage of the events children reported in response to open-ended prompts were nonexperienced events described in the story. In Session 2, they collectively reported 216 experienced science demonstrations in response to open-ended prompts, but they also reported 41 heard-only demonstrations and 17 touch-heard events. Thus, 21% of the reported events were never experienced. In Session 3, they reported 232 experienced demonstrations but only 18 heard-only and 9 touch-heard events, reducing the rate of false reports to 10%.

Table 2 details the infiltration of nonexperienced events into the children's narratives. Statistics across the bottom of the table refer to the results of separate session (2 vs. 3) by linear age trend analyses for each condition, which are described in Table 3. Because there were no significant Session × Linear Age Trend interactions, we describe only the main effects of age and session.

There were significant age trends only for experienced-only and experienced-heard demonstrations. This means that there was a gradual improvement with age in the number of experienced events reported, but no significant tendency for reports of nonexperienced events to decline with age. Across age groups, the children reported, on average, .18 of heard-only demonstrations and .15 of touch-heard events in Session 2, and .08 of heard-only demonstrations and .08 of touch-heard events in Session 3. No child reported a control demonstration in either session, although one 4-year-old did report a touch-control event in Session 3 (possibly because of direct questions about that event in Session 2).

These data run counter to claims that information in children's free recall is highly accurate, and they are also inconsistent with any blanket statement that older children are less suggestible than younger children. We wondered, though, whether the lack of an age trend for nonexperienced events reflected a few older children who made a disproportionate number of errors. This was not the case. For example, the proportion of children (from 3 to 8 years

Table 2
Mean Proportion of Events Reported in Response to Open-Ended Prompts in Sessions 2 and 3 (With Standard Deviations)

	Event condition					
Age	Experienced-only	Experienced-heard	Heard-only	Control	Touch-heard	Touch-control
			Session 2			
3	.08 (.19)	.16 (.29)	.16 (.34)	.00 (.00)	.11 (.32)	.00 (.00)
4	.34 (.37)	.50 (.47)	.29 (.42)	.00 (.00)	.32 (.48)	.00 (.00)
5	.25 (.35)	.67 (.34)	.08 (.19)	.00 (.00)	.00 (.00)	.00 (.00)
6	.56 (.34)	.67 (.38)	.19 (.30)	.00 (.00)	.11 (.32)	.00 (.00)
7	.31 (.30)	.69 (.46)	.17 (.30)	.00 (.00)	.28 (.46)	.00 (.00)
8	.57 (.32)	.84 (.28)	.18 (.33)	.00 (.00)	.09 (.29)	.00 (.00)
			Session 3			
3	.03 (.11)	.21 (.30)	.08 (.19)	.00 (.00)	.00 (.00)	.00 (.00)
4	.37 (.37)	.50 (.44)	.16 (.24)	.00 (.00)	.26 (.45)	.05 (.23)
5	.42 (.35)	.56 (.38)	.06 (.16)	.00 (.00)	.06 (.24)	.00 (.00)
6	.61 (.32)	.64 (.29)	.03 (.12)	.00 (.00)	.17 (.38)	.00 (.00)
7	.58 (.35)	.81 (.35)	.08 (.26)	.00 (.00)	.00 (.00)	.00 (.00)
8	.57 (.32)	.80 (.30)	.07 (.18)	.00 (.00)	.00 (.00)	.00 (.00)
Age linear	$p < .001$	$p < .001$	*ns*	—	*ns*	*ns*
Session	$p < .05$	*ns*	$p < .01$	—	*ns*	*ns*
A × S	*ns*	*ns*	*ns*	—	*ns*	*ns*

Table 3
Mean Proportion of Events Reported in Response to Open-Ended Prompts in Sessions 2 and 3: Statistics List

Source	*df*	*F* values for Linear Age Trend × Session (2 vs. 3) analyses: Experienced-only	Experienced-heard	Heard-only	Control	Touch-heard	Touch-control
Age linear	1	38.28***	49.67***	.29	—	.71	.76
(*MSE*)	(108)	(.14)	(.18)	(.10)		(.12)	(.00)
Session	1	5.95*	.02	12.59**	—	3.90	1.08
A × S	1	1.82	.04	.09	—	.58	.76
(*MSE*)	(108)	(.16)	(.08)	(.05)		(.07)	(.00)

	df	*F* values for condition comparisons: Experienced-only vs. Experienced-heard	Experienced-only vs. Heard-only	Heard-only vs. Touch-heard
Session 2				
Age linear	1	45.49***	9.89***	.08
(*MSE*)	(108)	(.14)	(.09)	(.13)
Condition	1	31.76***	15.38***	.57
A × C	1	2.59	10.36***	.00
(*MSE*)	(108)	(.10)	(.11)	(.09)
Session 3				
Age linear	1	67.45***	25.33***	2.45
(*MSE*)	(108)	(.12)	(.06)	(.05)
Condition	1	13.27***	88.04***	.02
A × C	1	.18	26.63***	.33
(*MSE*)	(108)	(.10)	(.08)	(.05)

* $p < .05$. ** $p < .01$. *** $p < .001$.

old, respectively) who reported one or more suggested events (heard-only or touch-heard) in Session 2 were .26, .53, .17, .39, .39, and .36, $p = .29$. It is possible that the superior verbal skills of the older children, coupled with their desire to explain as much as possible, contributed to unexpected intrusions of suggested events into their answers to open-ended prompts.

Regarding session effects, there were significant changes across sessions for experienced-only and heard-only demonstrations, which were reported less frequently after the 1-month delay. In contrast, experienced-heard demonstrations maintained a high reporting rate across sessions.

To directly compare selected conditions, Table 3 lists the results of three event condition by linear age trend analyses conducted at each session. At both sessions the children were more likely to report experienced events that had also been described in the story (experienced-heard events) than experienced-only events (*M*s = .59 vs. .36 in Session 2 and .59 vs. .43 in Session 3). In Session 2, there was a significant tendency for experienced-only events to be reported more frequently than heard-only events (*M*s = .36 vs. .18). In Session 3, experienced-only demonstrations continued to be reported more frequently than heard-only demonstrations, but there was also a significant Condition × Linear Age Trend interaction. It is evident from the means in Table 2 that the difference in reporting rate between experienced-only and heard-only demonstrations was much greater for the older than for the younger children in Session 3. Our source-monitoring procedure in Session 2 may have promoted better differentiation between actual and fictitious events for the older children after the 1-month delay. Finally, the children reported comparable proportions of heard events regardless of whether or not the events described physical touching. Perhaps suggestions of physical touching are not less likely to be reported than other suggestions when the touching is described as socially appropriate (i.e., in the context of cleaning up or applying a sticker) during initial suggestions.

The amount of accurate and inaccurate information in response to open-ended prompts. Table 4 lists the mean number of SUs mentioned in response to open-ended prompts in Sessions 2 and 3, and the proportion of information that fell into each of the three categories of inaccurate information. Asterisks across the bottom of the table report the results of session (1 vs. 2 vs. 3) by linear age trend analyses on SUs, detail errors, and spontaneous intrusion errors, and a session (2 vs. 3) by linear age trend analysis on suggested information. Statistical results are described in Table 5.

The mean number of SUs increased dramatically across age groups in all three sessions. There also was a significant session effect, with the mean number of SUs increasing from 51.2 in Session 1 to 68.4 in Session 2 and 61.3 in Session 3, and this pattern was similar across age groups. Improvement in the sheer amount of information reported between Sessions 1 and the later interviews probably is because of increased practice with the interview procedure and the fact that the story reinforced memories of the experienced events.

Age trends in error rates varied across the three types of inaccurate information. Average rates of detail errors were low, ranging from 1–2% across age groups and interviews, and detail errors did not vary as a function of age or session. In contrast, there was

Table 4
Mean Number of SUs Mentioned in Response to Open-Ended Prompts in Sessions 2 and 3 and the Proportion of Information That Was Inaccurate (With Standard Deviations)

Age	SUs	Proportion inaccurate: Detail errors	Proportion inaccurate: Spontaneous intrusion errors	Proportion inaccurate: Suggested information
		Session 2		
3	15.4 (22.0)	.01 (.01)	.05 (.11)	.15 (.21)
4	58.2 (69.8)	.01 (.02)	.00 (.00)	.20 (.25)
5	71.2 (60.2)	.01 (.01)	.00 (.00)	.02 (.06)
6	64.9 (49.5)	.01 (.02)	.00 (.00)	.10 (.12)
7	80.3 (79.3)	.02 (.03)	.00 (.00)	.07 (.10)
8	113.7 (73.2)	.01 (.01)	.00 (.00)	.06 (.10)
		Session 3		
3	13.7 (18.9)	.01 (.02)	.00 (.00)	.06 (.12)
4	41.0 (35.2)	.02 (.04)	.01 (.03)	.16 (.19)
5	52.8 (39.4)	.01 (.01)	.01 (.03)	.01 (.03)
6	73.3 (52.8)	.02 (.02)	.00 (.00)	.03 (.08)
7	67.3 (55.2)	.02 (.02)	.00 (.02)	.02 (.06)
8	112.1 (89.7)	.02 (.02)	.00 (.01)	.02 (.06)
	Session 1 vs. 2 vs. 3			Session 2 vs. 3
Age linear	$p < .001$	*ns*	$p < .05$	$p < .01$
Session	$p < .01$	*ns*	*ns*	$p < .01$
A × S	*ns*	*ns*	*ns*	*ns*

Note. SUs = syntactic units.

a significant age trend for spontaneous intrusions. As reported earlier, some 3- to 5-year-olds had difficulty focusing on the target events, particularly during Session 1, but no age group intruded much potentially confusing material during Sessions 2 and 3.

Finally, analysis of the amount of suggested information reported in Session 2 versus Session 3 showed a significant age trend and session effect. As shown in Table 4, the 3- and 4-year-olds reported relatively high rates of suggested information in response to open-ended prompts. In Session 2, .17 of the 3- to 4-year-olds' narratives consisted of suggested information, while the older children averaged only .06 suggested information. The younger groups averaged .11 suggested information in Session 3 compared with only .02 for the older children. Age differences in the proportion of information that was inaccurate reflected the fact that, compared with the younger children, the older children more often reported experienced events that had not been reinforced by the story (i.e., experienced-only).

The impact of continued prompting. As in Session 1, continued prompting was useful at all ages, although more of the older children responded to each prompt and relatively few children volunteered additional information in response to the final think prompt (see supplementary report). A question block (first two vs. final three questions) by linear age trend analysis at each session revealed no significant main effects or interactions involving ques-

Table 5
Mean Number of SUs Mentioned in Response to Open-Ended Prompts in Sessions 2 and 3 and the Proportion of Information That Was Inaccurate: Statistics List

Source	*df*	*F* values for Linear Age Trend × Session (1 vs. 2 vs. 3) analyses: SUs	Detail errors	Spontaneous intrusion errors	*df*	*F* values for Linear Age Trend × Session (2 vs. 3) analyses: Suggested information
Age linear	1	47.88**	2.77	4.03*	1	9.57**
(*MSE*)	(108)	(6787.60)	(.00)	(.00)	(90)	(.02)
Session	1	6.64**	.72	.88	1	12.91**
A × S	1	.98	.04	.96	1	.04
(*MSE*)	(108)	(1332.44)	(.00)	(.00)	(90)	(.01)

Note. SUS = syntactic units.
* $p < .05$. ** $p < .01$.

tion block for detail and spontaneous intrusion errors combined, $ps > .08$.

The proportions of new information that described suggested events were also entered into question block (first two or final three questions) by age trend analyses. In Session 2, there was a significant effect of question block, $F(1, 91) = 5.04$, $MSE = .02$, $p < .05$. Shortly after hearing the misleading information, all age groups tended to report more of the misleading information during the earlier as opposed to the latter portion of the interview. It is possible that the looked and heard prompts promoted better accuracy because the children were at a loss to provide perceptual details for fictitious events. In Session 3, there was also a significant effect of question block, but also a Question Block × Linear Age Trend interaction. Separate t tests comparing rates of suggested information volunteered in response to the first two versus the final three questions indicated that the youngest [$t(21) = 3.97$, $p < .01$] and oldest groups [$t(36) = 2.28$, $p < .05$] were significantly more accurate on the final three questions in Session 3, whereas the 5- to 6-year-olds showed a similar accuracy rate during the two portions of the interview, $p = .10$. This interaction may not be stable across a replication sample; in any case, the overall pattern was that suggested information infiltrated responses more often during the earlier open-ended prompts.

Summary. Even after 3- and 4-month delays, the children reported few detail errors about experienced events in response to open-ended prompts, and rarely included spontaneous intrusions in their reports. They often erred, however, by reporting events described only in the story. In the interview conducted shortly after exposure to the story, 40 of the 114 children (35%) reported a total of 58 nonexperienced events in free recall, including 17 events that dealt with an unpleasant touching experience. Although reports of nonexperienced events tended to decrease from Session 2 to Session 3, 24 of the children (21%) reported 27 nonexperienced events after a 1-month delay with no further misleading suggestions, including 9 suggested events that involved touching. It was surprising that children across this broad age range reported comparable numbers of nonexperienced events in response to open-ended prompts. Because the older children recalled more information about experienced events, however, suggested information comprised a smaller proportion of their narratives compared with those of the younger children. Continued open-ended prompting was not responsible for reports of suggested information; in fact, the children were generally less accurate during initial free recall than they were on subsequent open-ended prompts.

Responses to Direct Questions

The children's accuracy deteriorated markedly when interviews moved on to direct yes–no questions in Sessions 2 and 3. Recall that each direct question pair began with a yes–no question followed by one prompt for a "yes" response ("Tell me about the telephone") and another for a "no" response ("Can you tell me about the telephone?"). We were especially interested in the extent to which the children would describe events when interviewers ignored their "no" answers and suggested that they might, nonetheless, be able to describe the target events.

Yes–no responses. Because only "yes" responses were unambiguously accurate answers to experienced events and clearly false answers to nonexperienced events, we did not distinguish between "no" responses and occasional "I don't know" or "I don't remember" answers. Furthermore, the later responses occurred only rarely (between 0–2% of responses at each age) and were not significantly related to age, $p = .22$.

Table 6 reports, for each age group, the mean proportion of "yes" responses in each of the six event conditions for Sessions 2 and 3. The results of separate trend analyses for each condition are reported at the bottom of the table and are described in Table 7. As shown in Table 6, there were linear age trends across all condi-

Table 6
Mean Proportion of Yes Responses to Direct Questions in Sessions 2 and 3 (With Standard Deviations)

	Event condition					
Age	Experienced-only	Experienced-heard	Heard-only	Control	Touch-heard	Touch-control
			Session 2			
3	.58 (.30)	.76 (.39)	.68 (.45)	.21 (.35)	.37 (.50)	.00 (.00)
4	.84 (.24)	.97 (.11)	.58 (.45)	.26 (.42)	.42 (.51)	.16 (.37)
5	.67 (.38)	1.00 (.00)	.47 (.44)	.11 (.21)	.38 (.50)	.00 (.00)
6	.83 (.24)	1.00 (.00)	.56 (.45)	.14 (.33)	.33 (.49)	.00 (.00)
7	.81 (.30)	.94 (.16)	.36 (.45)	.06 (.24)	.33 (.49)	.00 (.00)
8	.93 (.18)	.95 (.15)	.34 (.47)	.09 (.25)	.36 (.49)	.00 (.00)
			Session 3			
3	.63 (.37)	.82 (.30)	.66 (.37)	.34 (.41)	.53 (.51)	.26 (.45)
4	.89 (.21)	.89 (.21)	.58 (.45)	.29 (.42)	.58 (.51)	.32 (.47)
5	.86 (.23)	.97 (.12)	.33 (.42)	.08 (.26)	.17 (.38)	.00 (.00)
6	.83 (.24)	1.00 (.00)	.28 (.35)	.03 (.12)	.28 (.46)	.00 (.00)
7	.89 (.21)	.97 (.12)	.22 (.35)	.03 (.12)	.11 (.32)	.00 (.00)
8	.86 (.32)	.89 (.21)	.14 (.32)	.07 (.18)	.14 (.35)	.05 (.21)
Age linear	$p < .01$	$p < .05$	$p < .001$	$p < .01$	$p < .08$	$p < .01$
Session	$p < .05$	*ns*	$p < .01$	*ns*	*ns*	$p < .01$
A × S	*ns*	*ns*	*ns*	*ns*	$p < .01$	$p < .01$

Table 7
Mean Proportion of Yes Responses to Direct Questions in Sessions 2 and 3: Statistics List

Source	*df*	Experienced-only	Experienced-heard	Heard-only	Control	Touch-heard	Touch-control
		F values for Linear Age Trend × Session (2 vs. 3) analyses					
Age linear	1	10.65**	6.56*	19.55***	12.15**	5.83*	12.34**
(*MSE*)	(108)	(.12)	(.05)	(.26)	(.13)	(.31)	(.07)
Session	1	4.55*	.78	11.70**	.04	2.16	10.37**
A × S	1	2.15	.56	3.61	3.73	9.94**	10.50**
(*MSE*)	(108)	(.03)	(.02)	(.08)	(.04)	(.12)	(.03)

	df	Heard-only vs. Control	Experienced-heard vs. Heard-only	Touch-heard vs. Touch-control	Heard-only vs. Touch-heard
		F values for condition comparisons			
Session 2					
Age linear	1	8.30**	1.84	.62	3.29
(*MSE*)	(108)	(.21)	(.15)	(.14)	(.29)
Condition	1	75.16***	121.50***	52.04***	6.22*
A × C	1	1.86	15.81	.02	3.54
(*MSE*)	(108)	(.09)	(.09)	(.13)	(.15)
Session 3					
Age linear	1	33.78***	12.33**	24.85**	6.90***
(*MSE*)	(108)	(.15)	(.11)	(.16)	(.21)
Condition	1	40.34***	267.44***	20.36***	2.22
A × C	1	4.62	40.99***	1.93	.32
(*MSE*)	(108)	(.07)	(.07)	(.11)	(.11)

* $p < .05$. ** $p < .01$. *** $p < .001$.

tions. Increased age was associated with an increased number of "yes" responses to direct questions about experienced events, and a decreased number of "yes" responses to direct questions about nonexperienced events.

There were significant session effects for experienced-only demonstrations, heard-only demonstrations, and touch-control events. The children were reliably less likely to say "yes" to experienced-only events in Session 2 than in Session 3, although the difference between sessions (*M*s = .78 vs. .83) was very small. Similarly, they were less likely to say "yes" to touch-control events in Session 2 than in Session 3 (*M*s = .03 vs. .11). In contrast, they were reliably more likely to say "yes" in Session 2 than in Session 3 in response to direct questions about heard-only demonstrations (*M*s = .50 vs. .36).

There were significant Session × Linear Age Trend interactions only for touch-heard and touch-control events. Regarding touch-heard events, the proportion of children in the four oldest age groups who made erroneous "yes" responses decreased between the second (.36) and third sessions (.17), whereas the proportion of children in the youngest two age groups who made false "yes" responses did not reliably change across sessions (.39 vs. .55). Similarly, on touch-control events, the oldest four age groups maintained their zero rate of responding to these questions, whereas .29 of the 3- and 4-year-olds acquiesced to the question about the touch-control event in Session 3. Across all four categories of nonexperienced events, the 3- and 4-year-olds made more inaccurate "yes" responses in Session 3 than they had in Session 2, whereas inaccurate responses tended to drop out of the older children's reports across interviews.

Four pairwise condition comparisons, conducted for each session, show the impact of the story on children's reports (see Table 7). Comparisons of heard-only versus control events and touch-heard versus touch-control events documented that the misleading story elevated children's error rates beyond their acquiescence rates on control demonstrations. Comparisons of experienced-heard with heard-only demonstrations indicated that the children were more likely to report events described by their parents that they had actually experienced during their visit with Mr. Science, but the Condition × Linear age trend interaction was significant at both sessions. As shown in Table 6, the 3- and 4-year-olds were poor at differentiating between experienced and nonexperienced events that had recently been discussed by their parents, whereas the older children were much more likely to say "yes" to direct questions about events that had been experienced and heard than to those that had only been heard. Finally, there was a significant condition effect in Session 2 for heard-only versus touch-heard events, because of more erroneous reports of the neutral events. In contrast, there was no condition effect in Session 3, indicating that the children did not seem to differentiate between neutral and touch events after the 1-month delay.

To summarize the data on yes–no responses to direct questions, increased age was associated with increased accuracy for both experienced and nonexperienced event questions. Although all age groups sometimes acquiesced to questions about control demonstrations that were neither experienced nor heard, the younger children were especially prone to such acquiescence. Exposure to misleading suggestions by parents increased rates of inaccurate "yes" responses beyond mere acquiescence rates for both the

nonexperienced science demonstrations and the descriptions of body touch. Although the children responded "yes" less frequently to questions about suggested touch than they did to questions about suggested science demonstrations shortly after exposure to misinformation, false acquiescence to touch-heard questions were nonetheless common: In Session 2, between .33 and .42 of the children in each age group responded "yes" to questions about whether Mr. Science put something yucky in their mouths or hurt their tummies; in Session 3, these rates of false acquiescence increased for the 3- and 4-year-olds (.53 and .58, respectively), but dropped to between .11 and .28 for the 5- to 8-year-olds. More generally, "yes" responses to nonexperienced but suggested events tended to drop out of the testimonies of the 5-, 6-, 7-, and 8-year-olds, but this was not the case for the 3- and 4-year-olds.

Narrative answers to direct questions and prompts to describe events. The literature on children's eyewitness testimony warns interviewers to mistrust answers to yes–no format questions, and our data certainly support this advice. Adults might therefore make judgments about whether events did or did not occur on the basis of whether children actually describe the events when asked to do so. To evaluate the accuracy of narrative responses, the children's answers to the second question of each direct question pair were divided into SUs, and each SU was assigned to one of six categories: accurate information; wrong event (i.e., information that described another target event but that might be misconstrued as relevant to the current question); detail error; incorrect rejection (e.g., "That didn't happen"); spontaneous intrusion (i.e., an unrelated, but not clearly off topic remark); or suggested information (i.e., nonexperienced information from the misleading story).

We calculated the proportion of children ages 3–4, 5–6, and 7–8 years old who responded with relevant narrative information in each event condition, divided into those who initially responded "no" versus those who initially responded "yes" (see supplementary report). For questions about experienced-only or experienced-heard demonstrations, incorrect rejections were not considered narratives because such responses were not attempts to describe the event. Similarly, for questions in the four nonexperienced categories, accurate information and detail errors were not considered narratives because such responses were digressions into other aspects of the science experiences rather than information that would be taken by an adult as an explanation of the event in question.

We found that the majority of children in all age groups who initially acknowledged experienced events provided at least some relevant narrative information when asked to do so. Unfortunately, many also provided narrative information about nonexperienced events. For example, in Session 2 the 7- to 8-year-olds narrated 91% of the experienced events after a "yes" response, but they also narrated 92% of nonexperienced events after a "yes" response. This rate declined somewhat in Session 3 but was still high, at 79%. Furthermore, although fewer children narrated either experienced or nonexperienced events after a "no" response, some did (see supplementary report).

We then computed the impact of different criteria for deciding whether or not an event occurred from answers to direct yes–no questions. Two strategies we compared were to (a) base judgments solely on whether children answered "yes" or "no" to the initial question in each pair, versus (b) credit children with saying "yes" if they provided narrative information about the event. Contrary to intuition, for both sessions the number of accurate answers to experienced-only and touch-heard events was higher when we based decisions on initial yes–no responses than when we based decisions on the presence or absence of a narrative response. For example, in Session 2 the percentage of children with correct "no" responses to touch-heard events was 61% for the 3- to 4-year-olds, 64% for the 5- to 6-year-olds, and 65% for the 7- to 8-year-olds. The percentages of children who provided no narrative information after prompting were 45%, 53%, and 55% for these three age groups. In other words, when children are indiscriminately prompted to provide narrative information, decision accuracy would decline if the presence of narratives was the basis for deciding whether or not events occurred.

The net effect of exposure to suggested information and prompting on children's narrative answers is described in Tables 8 and 9, which summarize the children's performance in Sessions 2 and 3 across all six event conditions. A Session × Linear Age Trend analysis of SUs confirmed that older children reported more narrative information to direct questions than the younger children did, and that the children as a whole reported more narrative information in Session 2 than in Session 3. The analogous analysis of inaccuracy rates (rightmost column in Table 8) showed that accuracy improved from ages 3 to 8, owing to declines with age in the proportion of wrong event, spontaneous intrusion, and suggested information. (A breakdown of narrative information by event condition is available in the supplementary report.)

These data paint a clear picture. Children who have heard their parents describe nonexperienced events sometimes reported having experienced those events when asked directly, and they were often willing to describe those events. Group averages do not communicate the narrative quality of children's true and false responses across age groups, however, because SUs were averaged across children who volunteered zero SUs and those who did respond. Therefore, we also analyzed the length of narrative responses to determine whether degree of elaboration might distinguish true from false reports. These data showed that children's descriptions of heard-only events in Session 2 were longer, on average, than descriptions they generated when interviewers asked them specific questions about novel events, but individual variability was large. Furthermore, narrative length did not discriminate between experienced and heard events in Session 2, and there were no significant differences between conditions in Session 3 (see supplementary report).

Block effects for direct questions. For each major age group (3–4, 5–6, and 7–8 years old) and each session, the mean proportion of "yes" responses to heard-only and control demonstrations in the first question block (direct questions 1–5) was compared with the number of "yes" responses during the second question block (direct questions 6–10) by a t test. None of these tests reached significance, $ps > .25$. Because question block is a between-subjects comparison for touch-heard and touch-control events, chi-square tests were used to analyze block effects for these variables: There was no block effect for individual age groups in either session, $ps > .71$. Thus there is no evidence that high rates of reporting nonexperienced events in this study were because of an interview protocol that simply wore the children down with numerous questions.

Summary of responses to direct questions. As anticipated, direct yes–no questions increased correct reports of experienced events relative to open-ended invitations, especially for younger

Table 8
Mean Number of SUs in Response to Direct Questions and the Proportion of Information in Various Accuracy Categories in Sessions 2 and 3 (With Standard Deviations)

Age	Total SUs	Proportion inaccurate					
		Wrong event	Detail	Spontaneous intrusion	Incorrect rejection	Suggested	Total inaccurate
				Session 2			
3	22.2 (26.8)	.03 (.07)	.01 (.02)	.21 (.27)	.00 (.01)	.25 (.30)	.48 (.32)
4	77.3 (42.4)	.01 (.05)	.01 (.02)	.14 (.25)	.01 (.02)	.19 (.14)	.34 (.26)
5	82.8 (62.1)	.00 (.00)	.02 (.02)	.06 (.12)	.01 (.03)	.25 (.16)	.33 (.15)
6	65.3 (34.5)	.01 (.03)	.01 (.02)	.04 (.06)	.00 (.02)	.19 (.14)	.25 (.15)
7	57.3 (43.6)	.00 (.00)	.02 (.03)	.05 (.10)	.01 (.02)	.14 (.16)	.22 (.21)
8	70.5 (37.8)	.01 (.03)	.02 (.03)	.01 (.03)	.00 (.00)	.18 (.18)	.20 (.20)
				Session 3			
3	24.4 (29.4)	.03 (.08)	.01 (.02)	.16 (.26)	.00 (.02)	.37 (.37)	.55 (.33)
4	62.5 (32.3)	.00 (.02)	.01 (.03)	.09 (.19)	.01 (.04)	.17 (.15)	.29 (.23)
5	54.7 (40.9)	.03 (.11)	.02 (.03)	.09 (.15)	.00 (.01)	.15 (.23)	.26 (.28)
6	59.3 (46.2)	.00 (.00)	.01 (.02)	.06 (.10)	.00 (.00)	.17 (.16)	.24 (.19)
7	53.2 (51.8)	.00 (.00)	.03 (.03)	.02 (.06)	.00 (.00)	.07 (.09)	.12 (.15)
8	52.9 (24.7)	.00 (.00)	.02 (.03)	.04 (.10)	.02 (.08)	.08 (.12)	.16 (.19)
Age linear	$p < .05$	$p < .05$	$p < .05$	$p < .001$	*ns*	$p < .001$	$p < .001$
Session	$p < .01$	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*
A × S	*ns*	*ns*	*ns*	*ns*	*ns*	$p < .05$	*ns*

Note. SUs = syntactic units.

children and especially for experienced-only (as opposed to experienced-heard) events. Also as expected, direct questions increased false reports of nonexperienced events, and this was especially striking for events that parents had described. It is important to emphasize that even the youngest children's responses were not entirely driven by a general tendency to acquiesce to direct questions per se. For example, in Session 2 the percentage of 3-year-olds who falsely reported the touch-heard event rose from 11% in response to open-ended prompts to 37% in response to direct questions, but 3-year-olds' false reports of the touch-control events remained at zero even with direct questioning.

Our results indicate that children's willingness to provide narrative details to back up a "yes" response cannot be taken as evidence of the accuracy of the "yes" response. Regardless of whether or not an asked-about event had actually occurred, most children who responded "yes" subsequently provided at least some narrative details. Unfortunately, there was no evidence in this study that the amount of information children generated about experienced versus suggested events was diagnostic of event condition, because some children across all six age groups gave substantive narrative elaborations in response to direct questions about heard-only or touch-heard events. Furthermore, decisions based on the presence or absence of narrative descriptions were less accurate than were decisions based on children's "yes" or "no" responses.

Source-Monitoring Performance

Prior to formal source-monitoring questions, few children spontaneously mentioned the misleading story in a way that would alert

Table 9
Mean Number of SUs in Response to Direct Questions and the Proportion of Information in Various Accuracy Categories in Sessions 2 and 3: Statistics List

Source	*df*	Total SUs	Proportion inaccurate					
			df	Wrong event	Detail	Spontaneous intrusion	Incorrect rejection	Suggested	Total inaccurate
Age linear	1	5.11*	1	4.23*	4.60*	13.88***	.01	15.01***	29.27***
(*MSE*)	(108)	(2,462.31)	(99)	(.00)	(.00)	(.03)	(.00)	(.04)	(.07)
Session	1	9.77**	1	.04	.34	.05	.08	3.52	2.05
A × S	1	.50	1	.02	.04	.49	.15	6.62*	3.07
(*MSE*)	(108)	(786.29)	(99)	(.00)	(.00)	(.02)	(.00)	(.02)	(.03)

Note. Data are *F* values for Linear Age Trend × Session (2 vs. 3) analyses. SUs = syntactic units.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Table 10
Source Monitoring: Proportion Accurate on Both Story and Event Questions in Sessions 2 and 3 (With Standard Deviations)

	Event condition						
Age	Experienced-only	Experienced-heard	Heard-only	Control	Touch-heard	Touch-control	Overall
			Session 2				
3	.05 (.16)	.68 (.42)	.13 (.28)	.68 (.45)	.21 (.42)	.79 (.42)	.41 (.11)
4	.24 (.39)	.79 (.35)	.32 (.45)	.66 (.47)	.42 (.51)	.84 (.37)	.53 (.21)
5	.31 (.39)	.83 (.34)	.61 (.47)	.94 (.24)	.56 (.51)	1.00 (.00)	.69 (.20)
6	.33 (.42)	.83 (.34)	.67 (.38)	.86 (.33)	.56 (.51)	1.00 (.00)	.69 (.17)
7	.33 (.42)	.86 (.29)	.75 (.39)	.94 (.24)	.56 (.51)	1.00 (.00)	.73 (.18)
8	.61 (.41)	.89 (.21)	.80 (.37)	.91 (.20)	.86 (.35)	1.00 (.00)	.83 (.15)
			Session 3				
3	.00 (.00)	.66 (.41)	.08 (.19)	.63 (.40)	.05 (.23)	.57 (.51)	.34 (.13)
4	.05 (.23)	.74 (.35)	.26 (.39)	.61 (.45)	.42 (.51)	.63 (.50)	.44 (.15)
5	.19 (.35)	.81 (.35)	.47 (.40)	.72 (.39)	.44 (.51)	.67 (.49)	.55 (.14)
6	.31 (.35)	.81 (.39)	.64 (.41)	.61 (.40)	.61 (.50)	.56 (.51)	.59 (.12)
7	.25 (.35)	.75 (.31)	.58 (.43)	.78 (.35)	.72 (.46)	.56 (.51)	.60 (.16)
8	.36 (.41)	.75 (.37)	.77 (.37)	.66 (.39)	.91 (.29)	.59 (.50)	.66 (.17)
Age linear	$p < .001$	*ns*	$p < .001$	$p < .05$	$p < .001$	*ns*	$p < .001$
Session	$p < .001$	*ns*	$p < .05$	$p < .001$	*ns*	$p < .001$	$p < .001$
A × S	*ns*	*ns*	*ns*	*ns*	*ns*	*ns*	$p < .05$

interviewers to a contaminating influence (see supplementary report). Of the 23 (20%) who spontaneously mentioned the story, 20 did so in Session 2, shortly after their parents had read the story. Unexpectedly, the older children were not more likely to mention the story: The 20 children who referred to the story in Session 2 included five 3-year-olds, three 4-year-olds, six 5-year-olds, one 6-year-old, three 7-year-olds, and two 8-year-olds.

The interviewers guided each child through the source-monitoring procedure to assess the ability to differentiate between memories of events experienced with Mr. Science versus memories of events that were only mentioned in the story. We summarized source-monitoring performance in three ways: (a) mean accuracy rates on story and event questions separately; (b) mean accuracy rates for assigning events to sources (i.e., the proportion of events that were accurately assigned to source by correct responses to both the story and event questions); and (c) the proportion of events in each event condition that were accepted during direct questioning that were also accepted during the source-monitoring procedure. The last of these analyses measured whether the source-monitoring procedure helped the children reject their earlier false reports without leading them to reject events that they had actually experienced.

Accuracy on Source-Monitoring Questions About the Story and the Events

Analyses of story and event questions separately clarified the circumstances under which younger children have difficulty specifying the sources of their knowledge (see supplementary report). Regarding the story questions, accuracy increased significantly with age for experienced-only, heard-only, and touch-heard questions, $ps < .05$. Because the correct response was "no" to experienced-only demonstrations but "yes" to heard-only demonstrations and touch-heard events, age differences in yeah saying cannot account for these patterns. There was no significant age trend for touch-control questions (with performance close to ceiling), or for experienced-heard questions (with even the youngest children responding correctly about 74% of the time). The overall pattern is that, compared with the older children, the younger children had more difficulty specifying the source of events that they encountered through only one of the two possible sources. In contrast to the story questions, there were significant age trends across all conditions for the event questions, with the older children performing better, $ps < .01$. This finding is a predictable consequence of the fact that the actual events took place 3–4 months previously, whereas the story questions asked about recently presented information.

Identification of the Sources of Individual Events

Table 10 reports mean accuracy rates for assigning events to sources in each condition. To compile these values, a child was scored as having responded accurately to an event if responses to both the story and the event question were accurate. Overall accuracy rates across all 10 events, shown in the rightmost column, show that accuracy rates increased from 41% at 3 years old to 83% at 8 years old in Session 2, and from 34% at 3 years old to 66% at 8 years old in Session 3 (chance performance is 25%). The Session × Linear Age Trend interaction reflects the attenuated age difference at Session 3 compared with Session 2. Although performance was above chance, the children obviously had some difficulty making source discriminations.

Separate analyses for the six conditions (see Table 11) address whether the children's difficulty in identifying source was because of general problems with the task or the problems specifying the

Table 11
Source Monitoring: Proportion Accurate on Both Story and Event Questions in Sessions 2 and 3: Statistics List

		F values for Linear Age Trend × Session (2 vs. 3) analyses						
Source	*df*	Experienced-only	Experienced-heard	Heard-only	Control	Touch-heard	Touch-control	Overall
Age linear	1	26.53***	2.59	57.08***	4.07*	47.10***	1.43	83.37***
MSE	108	(.18)	(.17)	(.22)	(.20)	(.25)	(.16)	(.04)
Session	1	12.57***	3.20	4.20*	21.80***	.00	46.76***	75.39***
A × S	1	.91	1.02	.01	3.77	2.50	3.19	4.39*
MSE	108	(.07)	(.07)	(.08)	(.07)	(.15)	(.14)	(.01)

* $p < .05$. *** $p < .001$.

sources of specific types of information. Looking first at the Session 2 data, the children generally were accurate on touch-control events, reflecting accurate rejections of both source questions. Accuracy was also above .80 for ages 5 to 8 years for control and experienced-heard demonstrations. The children had more difficulty, however, answering source questions about events that occurred months earlier but were not mentioned in the book (i.e., experienced-only), and events that were only mentioned in the book (i.e., heard-only and touch-heard). Children ages 5 to 8 years therefore had little difficulty with events that did not happen at all, or events with which they became familiar through both sources; however, they had difficulty correctly identifying information from a single source. Trend analyses on these six conditions (Table 13) confirmed significant age trends on experienced-only, heard-only, control, and touch-heard events. Once again, nonsignificant age trends for experienced-heard demonstrations (correctly answered "yes" to both questions) and touch-control events (correctly answered "no" to both questions) indicate that response biases at younger ages were not solely responsible for the age trends in source-monitoring performance.

The children's source-monitoring performance decreased from Session 2 to 3, with significant session effects for experienced-only, heard-only, control, and touch-control events, and declines were statistically comparable across ages for individual event conditions. However, there was a significant interaction for the overall mean as a result of attenuated age differences in Session 3 for control and touch-control events.

Is it possible that accuracy rates for assigning events to sources were low because the children were sometimes asked to identify the source of events that they did not remember at all? To address this issue, we recomputed accuracy rates using only items that received a "yes" response to either a direct question or a source-monitoring story question. These data showed that the children's performance was not improved by focusing only on events they had claimed to be familiar with at some point in the interview (see supplementary report).

Table 12
Source Monitoring: Proportion of Items Accepted During Direct Questions That Were Accepted After Source-Monitoring Instructions in Sessions 2 and 3 (With Standard Deviations)

	Event condition					
Age	Experienced-only	Experienced-heard	Heard-only	Control	Touch-heard	Touch-control
			Session 2			
3	.82 (.39)	.87 (.29)	.86 (.31)	.67 (.52)	1.00 (.00)	—
4	.95 (.16)	.92 (.19)	.81 (.38)	1.00 (.00)	1.00 (.00)	1.00 (.00)
5	.97 (.13)	.86 (.29)	.55 (.47)	.25 (.50)	.29 (.49)	—
6	.92 (.26)	1.00 (.00)	.42 (.42)	.00 (.00)	.83 (.41)	—
7	.91 (.26)	1.00 (.00)	.19 (.37)	.50 (.00)	.33 (.52)	—
8	.93 (.23)	.98 (.11)	.56 (.42)	.33 (.58)	.37 (.52)	—
Age linear	*ns*	$p < .05$	$p < .01$	$p < .05$	$p < .001$	
			Session 3			
3	.87 (.29)	.86 (.29)	.84 (.35)	.72 (.44)	.90 (.32)	1.00 (.00)
4	.92 (.25)	.84 (.33)	.81 (.33)	.71 (.39)	.73 (.47)	.50 (.55)
5	.86 (.29)	1.00 (.00)	.75 (.38)	.50 (.71)	.67 (.58)	—
6	1.00 (.00)	1.00 (.00)	.69 (.46)	.00 (.00)	.40 (.55)	—
7	.94 (.16)	.94 (.16)	1.00 (.00)	1.00 (.00)	.00 (.00)	—
8	1.00 (.00)	1.00 (.00)	.50 (.58)	.67 (.58)	.33 (.58)	1.00 (.00)
Age linear	$p < .05$	$p < .01$	*ns*	*ns*	$p < .01$	

Table 13
Source Monitoring: Proportion of Items Accepted During Direct Questions That Were Accepted After Source-Monitoring Instructions in Sessions 2 and 3: Statistics List

Event condition	Values for linear age trend analyses			
	df	*F*	*df*	*MSE*
Session 2				
Experienced-only	1	0.67	102	.06
Experienced-heard	1	5.97*	105	.03
Heard-only	1	11.04**	60	.16
Control	1	5.62*	17	.16
Touch-heard	1	14.94***	36	.15
Touch-control	—	—	—	—
Session 3				
Experienced-only	1	4.31*	103	.04
Experienced-heard	1	8.07**	107	.04
Heard-only	1	0.74	49	.13
Control	1	0.05	17	.21
Touch-heard	1	7.53*	28	.20
Touch-control	—	—	—	—

Note. Dashes indicate that there were insufficient "yes" responses during direct questions to analyze touch-control events.
* $p < .05$. ** $p < .01$. *** $p < .001$.

Selectively Rejecting Nonexperienced Events With a Source-Monitoring Procedure

Did the source-monitoring procedure help the children reject nonexperienced events without increasing false rejections of experienced events? The data in Tables 12 and 13 address this question. Table 12 lists the mean proportion of events accepted during direct questioning that were also accepted as having "really" happened during the source-monitoring procedure. For example, in Session 2 the 3-year-olds accepted 100% of the touch-heard events that they had previously accepted during direct questioning, whereas the 8-year-olds continued to accept only 37% of these nonexperienced events. This pattern is evident throughout Table 12: The younger children accepted the majority of events during source monitoring that they had accepted during direct questioning, whereas the older children were more successful in "gating out" nonexperienced events. Moreover, the older children did so with only a small risk of rejecting experienced events.

The results of separate session by linear age trend analyses for each event condition are listed by session in Table 12, with a corresponding statistics list in Table 13.[4] In Session 2, there was no significant linear age trend for experienced-only demonstrations, which all ages frequently continued to accept during source monitoring. The older children were more likely to correctly accept experienced-heard demonstrations than were the younger children, however, and less likely to erroneously continue to accept heard-only, control, or touch-heard events. Thus, source-monitoring instructions were effective in helping the older children reject nonexperienced events without a large increase in rejection rates for experienced events. Inspection of the means for Session 2 shows that our source-monitoring procedure was most valuable for the 5- to 8-year-olds.

In Session 3, there were significant age trends on experienced-only and experienced-heard demonstrations, with older children accurately accepting more of these experienced events, and on touch-heard events, with older children accurately rejecting more of these nonexperienced events. Once again, the older children clearly benefited more from the source-monitoring procedure. Unlike Session 2, however, there was no evidence that the older children performed better than the younger children on heard-only or control demonstrations after the 1-month delay, consistent with prior results that show that age trends diminish over time for less memorable material.

Figure 1 recasts these data in a format that permits quick comparisons between conditions and sessions. This figure depicts the mean proportion of experienced and nonexperienced events that children at each age reported they had experienced by "yes" responses to both direct questions and the subsequent source-monitoring event questions. Thus, these data represent final performance at the end of each interview: the proportion of accurately reported experienced events, and inaccurately reported nonexperienced events, for each age group. To provide a visual picture of what was gained or lost by adding the source-monitoring procedure, Figure 1 also plots the proportions of "yes" responses at each age to direct questions only. This figure therefore compares children's performance before and after the source-monitoring procedure.

First, the fact that the older children said "yes" to fewer nonexperienced events during direct questioning, and were more likely to correctly reject those items when asked explicit source-monitoring questions, translated into much lower rates of false positive responses compared with the younger children in both Sessions. The 3- and 4-year-old children did not appear to benefit from the source-monitoring procedure. At most, the data pattern for these younger children suggests that adding the source-monitoring procedure had a modest and nondifferential effect on reports of experienced and nonexperienced events alike. In contrast, the older children appeared to benefit substantially from the source-monitoring procedure, particularly in Session 2, without suffering a cost in terms of accurate reports of experienced events.

As reported earlier, the rate of false reports of an instance of unpleasant touching in response to direct questions did not reliably vary as a function of age in Session 2. As shown in Figure 1, however, age effects were quite dramatic following the source-monitoring procedure. In Session 2, the 3- and 4-year-olds failed to benefit at all from the source-monitoring training: False reports of suggested touching remained as high after the source-monitoring procedure as they had been in response to direct questions. In contrast, the rate of false reporting by older children was substantially reduced by the addition of the source-monitoring procedure. (See the supplementary report for statistical comparisons of data reported in Figure 1, which are largely redundant with previously reported analyses.)

[4] Analyses were conducted separately at each session because the data are based only on participants who had at least one "yes" response in each question category during direct questioning, and thus between-session comparisons would restrict sample sizes.

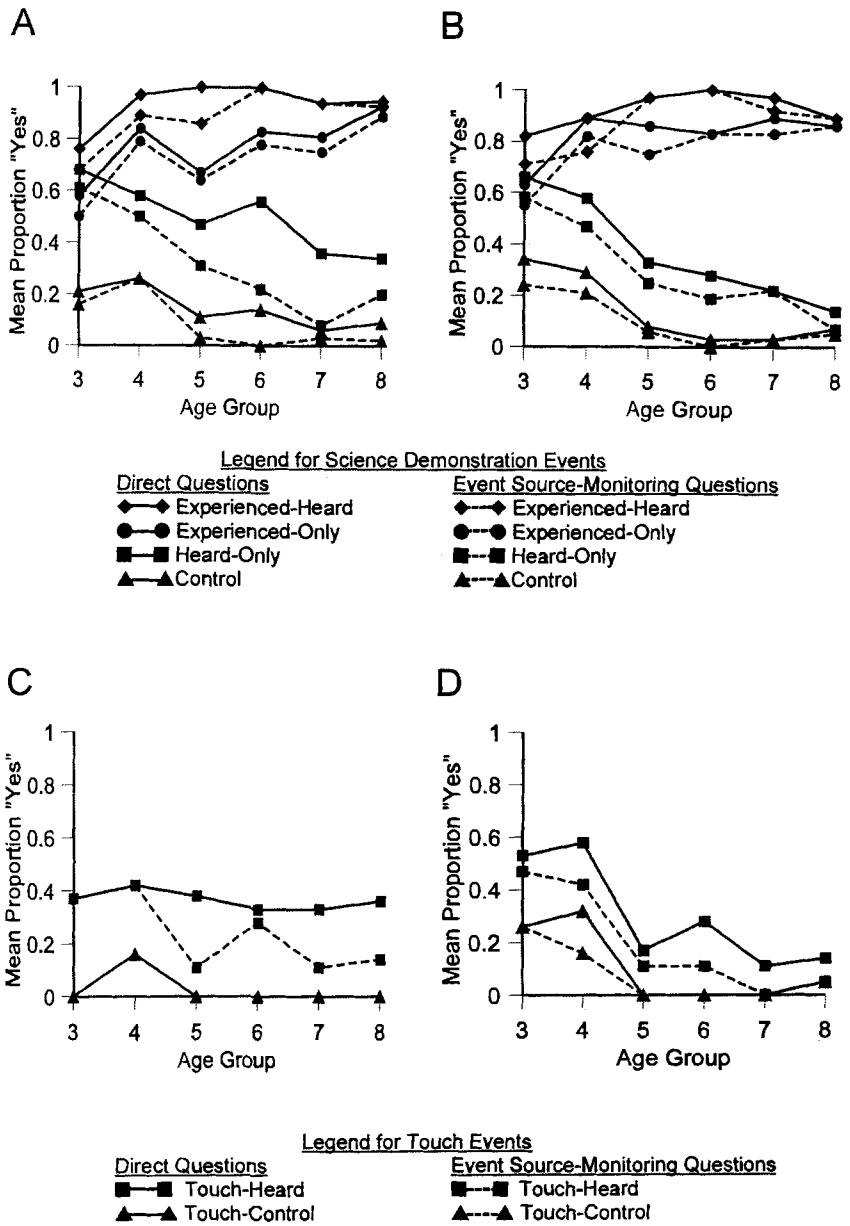


Figure 1. Mean proportion of "yes" responses to direct yes–no questions and event source-monitoring questions. (A) Session 2, science demonstrations. (B) Session 3, science demonstrations. (C) Session 2, touching, (D) Session 3, touching.

Summary of Source-Monitoring Performance

Three aspects of the source-monitoring data are particularly noteworthy. First, on the negative side, our source-monitoring procedure failed to be of any benefit to 3- and 4-year-old children. Importantly, this does not appear to have been because these young children did not understand the source-monitoring questions and consequently answered them in a random or systematically biased way. Even the 3-year-old children, for example, were quite accurate at correctly reporting that experienced-heard events had been mentioned in the story (74% correct) and that touch-control events had not been mentioned in the story (79% correct), and they also were above chance in reporting that experienced-heard events had actually happened (68% correct) and that touch-control events had not (100% correct). These findings suggest that the younger children understood the source-monitoring task and were capable of differentiating between events whose sources they were able to remember. This did not, unfortunately, enable them to selectively reduce their rate of falsely reporting events suggested in the story.

Second, on the positive side, our source-monitoring procedure did decrease false reports by the older children, and did so at relatively little cost to correct reports of experienced events. This pattern was particularly striking in Session 2, in which it would be

relatively easy to discriminate between memories of the interaction with Mr. Science and memories of the story.

Third, and finally, we return to a less encouraging aspect of the results. Even after the source-monitoring procedure, and even among the older children, a nontrivial minority continued to report that they had experienced suggested events. These errors persisted despite the fact that interviewers told the children it was appropriate to say "no," warned them that the story may have described events that did not really happen, gave them the opportunity to demonstrate their familiarity with the story by asking them to describe its contents, and mildly challenged them with direct source-monitoring questions that asked if the event "really" had happened.

Exploring the Mechanisms That Underlie Suggestibility and False Memories

What mechanisms underlie false reports of suggested events? One approach to this question draws on findings from traditional laboratory memory tasks to explain patterns of eyewitness suggestibility. These efforts subsume event reporting under a larger explanatory umbrella that may enable us to describe the characteristics of true and false reports (e.g., Brainerd & Poole, 1997). Another approach searches for clues about the mechanisms that underlie false reporting by analyzing individual differences variables (e.g., Eisen, Goodman, Qin & Davis, 1998; Pipe & Salmon, in press). Finally, other research considers the extent to which errors reflect memory limitations versus errors arising from other mechanisms, such as social compliance, that operate during the target interviews (e.g., Coxon & Valentine, 1997; Lindsay, Gonzales, & Eso, 1995; Newcombe & Siegal, 1997; Pezdek & Roe, 1995). In this section, we present data that speak to each of these issues.

Stability of True and False Reports

Investigators are interested in reporting patterns for individual events because consistency is often touted as an indicator of the veracity of a report (Fisher & Cutler, 1995). Indeed, Brewer, Potter, Fisher, Bond, and Luszcz (1999) found that potential jurors identified inconsistency across statements as the strongest indicator of inaccuracy of a witness's final statements. On the other hand, reports that do not vary may be viewed as artificial because reporting with some variation is seen as a characteristic of normal recall (e.g., Pence & Wilson, 1994).

We expected that memories of the experienced events (which were already 3 or 4 months old during Session 2) would be more stable than memories of the recently presented story. This was the case: Across all age groups, the probability that experienced-only events accepted in Session 2 were also accepted in Session 3 was .94, which was higher than the unconditional probability of acceptance in Session 3 (.78). Reports of suggested events showed stability across sessions (probability of a "yes" in Session 3 given a "yes" in Session 2 was .61 across all ages, compared with an unconditional probability of .30), but were not as consistent as reports of actual events. Unfortunately, there were not sufficient "yes" responses to control events to provide reliable stability estimates at all ages, and therefore we cannot adequately test the prediction that endogenous errors will be more stable than errors based on recently suggested information (e.g., Brainerd & Poole, 1997). (See the supplementary report for details.)

The data on session effects and these analyses converge on some interesting conclusions. First, the greater persistence of true versus implanted reports lends some support for the practice of viewing allegations that drop out of children's reports with greater suspicion than stable reports. This suggests that it could be informative to conduct a second interview to test for the consistency of reports over short time periods that normally are not associated with forgetting of actual events. Even false reports showed some stability over time, however, and thus the reiteration of a report does not by itself diagnose accuracy, and even accurate reports are not perfectly stable.

Individual Differences as Predictors of Suggestibility

We collected information on three descriptive variables: sex, parental education (the highest education level achieved by the child's parents from six specified brackets), and annual family income (from six income brackets). We also created three performance variables from children's behavior during nonsubstantive portions of the interview: Off Topic (0 *stayed on topic during rapport building and when asked the final open-ended prompt*, 1 *strayed off topic*), Talkativeness Session 1 (standardized residual from regressing number of words spoken in response to rapport questions in Session 1 on age), and Talkativeness All Sessions (standardized residual based on length of responses to rapport questions across all three sessions). We reasoned that children who strayed off topic and those who were unusually talkative might be at higher risk of intruding information about suggested events. We computed correlations between these predictors and suggestibility (i.e., the number of suggested events reported)[5] during free recall, direct questioning, and source monitoring at Sessions 2 and 3, for a total of six correlations per predictor.

Because sex, parental education, and family income were not perfectly balanced across ages, we computed partial correlations among these variables and suggestibility with age controlled. We found no consistent relationship between demographic characteristics and suggestibility. The only significant finding was a tendency for children from families with higher incomes to accept fewer nonexperienced events after source-monitoring instructions in Session 3, but the correlation was low ($r = -.19, p < .05$) and the number of tests performed was large. Several performance variables, however, did correlate significantly with suggestibility. In Session 2, straying off topic was associated with the infiltration of suggested events into free recall (partial r with age controlled = .25, $p < .01$), as were both measures of talkativeness (Talkativeness Session 1, $r = .33$, Talkativeness All Sessions, $r = .20$, $ps < .05$). Also in Session 2, suggestibility to direct questions was significantly associated with talkativeness in Session 1, $r = .18$, $p < .05$. The lack of relationships in Session 3 may be because of

[5] We used the number of false reports on suggested items, rather than the difference between false reports on suggested and control items, as our measure of suggestibility for these analyses because in real-world cases investigators would be interested in using individual difference measures to estimate the probability of false reports given suggestions rather than in using them to estimate the effect of suggestions per se.

the low level of suggestibility among the older children in that session. None of our individual difference variables predicted final source-monitoring performance.

These results reveal a simple pattern. Despite the fact that age has accurately been called a "rather crude variable" (Ceci, Huffman, Smith & Loftus, 1994, p. 389), it predicted false acquiescence to questions about suggested events better than the individual difference variables we explored. During free recall, however, when age did not predict the infiltration of suggested events into the children's narratives, talkativeness and straying off topic did predict inaccuracies. In Session 3, even these variables failed to predict suggestibility. This pattern of results underscores the fact that multiple mechanisms underlie reports of nonexperienced events, and therefore experts attempting to identify predictors of suggestibility must grapple not only with characteristics of the child, but also with characteristics of the child's interviewing history and the structure of the current interview (e.g., Lindberg, Keiffer, & Thomas, in press).

Table 14
Zero-Order Correlations Between Age in Months and Acquiescence, Recall, Source Monitoring, and Suggestibility

Variable	Age	Acquiescence	Recall	Source monitoring
		Session 2		
Acquiescence	−.23*			
Recall	.48**	−.08		
Source monitoring	.59**	−.46**	.25**	
Suggestibility	−.21*	.40**	.14	−.48**
		Session 3		
Acquiescence	−.39**			
Recall	.48*	−.23*		
Source monitoring	.59**	−.38**	.45**	
Suggestibility	−.45**	.44**	−.02	−.44**

* $p < .05$. ** $p < .01$.

Mechanisms That Mediate the Age-Suggestibility Relationship

We explored how three mechanisms assessed during the target interviews contributed to the developmental decline in suggestibility. The first is acquiescence. We expected that children who tended to say "yes" to novel (i.e., control) events also would show higher rates of false positive responses to suggested events. A second process is recall. As children develop, their ability to recall events improves while suggestibility typically declines. Recall is not only a marker of memory for experienced events, however. Rather, better recall for experienced events could correlate with better memory for suggested events, which might increase suggestibility (Brainerd & Reyna, 1995; Reyna & Kiernan, 1994), and children who are verbal might have a tendency to narrate events indiscriminately, thereby showing improvements in recall with concomitant increases in suggestibility. Because forensic evaluators often assume that freely recalled information is highly reliable, we were interested in exploring how the quantity of information that children recalled about experienced events related to their suggestibility. Furthermore, we were interested in whether developmental improvements in source monitoring made a unique contribution to declines in suggestibility after factoring out developmental changes in acquiescence and recall. Finally, we wanted to measure not only whether recall had a direct influence on suggestibility, but whether improved recall affected suggestibility indirectly by lowering acquiescence rates and improving source-monitoring scores.

Our phased interview procedure provided the following measures: recall (the number of experienced events narrated during open-ended or direct questions); acquiescence (the number of "yes" responses to direct control questions); source monitoring (the proportion of events accepted as familiar by a "yes" response during direct or source-monitoring questions that were accurately assigned to source); and suggestibility (the number of suggested events accepted either by a "yes" response or a narrative response to direct questions).

Zero-order correlations among these variables and age in months appear in Table 14. Path analyses were conducted to examine multivariate relationships among age, acquiescence, recall, source monitoring, and suggestibility. Figure 2 presents the standardized path coefficients for a model in which acquiescence, recall, and source monitoring all have direct effects on suggestibility, but recall is also an intervening variable that mediates the relationship between acquiescence and suggestibility, and between source monitoring and suggestibility. Specifically, we thought that better recall might lower suggestibility by reducing acquiescence, and also lower suggestibility by improving memory source monitoring. We linked acquiescence and source monitoring in our model because our source-monitoring procedure was based on yes–no questions, and therefore acquiescence would tend to lower source-monitoring scores.

In Session 2, acquiescence, recall, and source monitoring all mediated the age–suggestibility relationship. Note, however, that the direction of effect for recall is positive: As recall improved, suggestibility increased. Furthermore, soon after presentation of the misinformation, recall of experienced events did not contribute to the decline in suggestibility by mediating the effects of acquiescence or source monitoring. To evaluate the adequacy of the model, we also tested the direct path between age and suggestibility and found that it was not significant ($\beta = -.08$, $p = .44$), indicating that the variables we included were adequate to explain developmental changes in suggestibility.

The pattern of relationships between variables was similar in Session 3, with two exceptions. One month after exposure to misinformation, recall of experienced events had both direct and indirect effects on suggestibility. As in the previous session, recall of experienced events related positively to suggestibility, but recall was also associated with better source monitoring, which reduced suggestibility. Also unlike Session 2, the direct path between age and suggestibility was significant ($\beta = -.33$, $p < .01$), indicating that we failed to account fully for developmental trends after a delay. This finding is not surprising because our measures were crude (e.g., ranging from 0–3 for acquiescence and suggestibility) and there were floor effects for some age groups in Session 3, which would exacerbate reliability problems. Alternatively, it may be that factors in addition to recall, source monitoring, and acquiescence are involved in developmental declines in suggestibility with delayed recall.

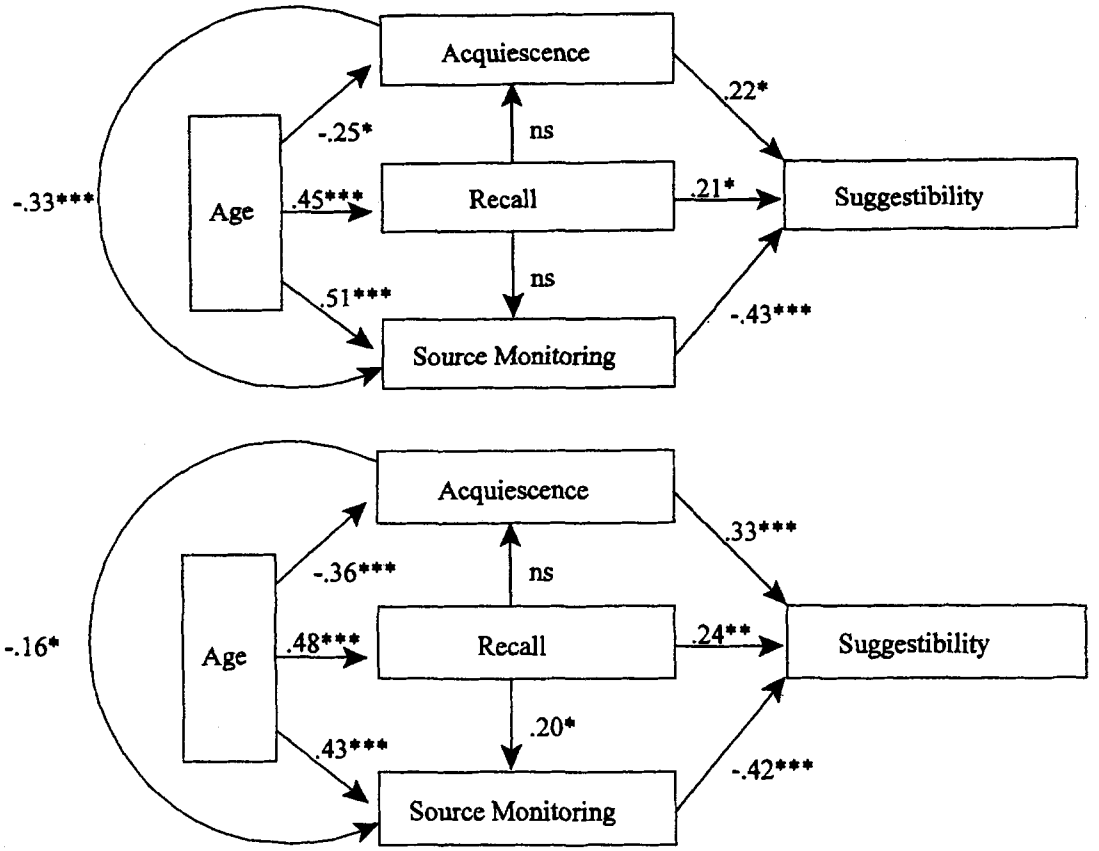


Figure 2. Standardized path coefficients for a model in which acquiescence, recall, and memory source monitoring mediate developmental declines in suggestibility. Top panel = Session 2; bottom panel = Session 3. $^{*}p < .05$. $^{**}p < .01$. $^{***}p < .001$.

Discussion

Here we revisit the nine issues posed in the introduction, highlighting ways in which findings from this and other recent investigations challenge frequently promulgated generalizations regarding children's eyewitness reports.

The Accuracy of Children's Responses to Open-Ended Prompts and the Impact of Continued Prompting

One striking mismatch between common generalizations regarding children's eyewitness testimony and laboratory findings concerns the accuracy of children's responses to open-ended questions or invitations (e.g., "Tell me everything about that"). As an example of the typical conclusion, Douglas (1996) wrote, "In response to free recall [questions], young children. . . rarely recall incorrect information." Widespread confidence in the accuracy of children's free-recall reports is challenged by numerous studies that have found that responses to open-ended invitations are not insulated from memory contamination (e.g., Leichtman & Ceci, 1995; Warren & Lane, 1995). Our current finding that reports of fictitious events often infiltrated children's narratives even during early, nonsuggestive stages of interviews reinforces such results and indicates that the use of open-ended prompts does not guarantee accuracy when child witnesses have previously been exposed to misinformation.

There were two surprising features of our results. First was the lack of a developmental trend across this rather broad age range, in that the 8-year-olds were as likely to mention fictitious events in free recall as were the youngest children. Although we know from the source-monitoring phase of the interview that 8-year-olds more often realized that they were describing events that did not really happen, their superior verbal skills, possibly combined with a desire to show the interviewer what they knew, produced narratives in which experienced and nonexperienced events were commingled.

A second surprising finding was that continued prompting with open-ended invitations did not increase the proportion of false information reported by the children. In fact, some of the youngest children zeroed in on the topic of conversation rather slowly, focusing on the science experience only after one or two prompts that mentioned the event. These findings support the recommendation that interviewers prolong the free narrative section of an interview by using a variety of open-ended prompts rather than shifting to specific questions after the witness provides an initial description (as is often the case in analyses of actual forensic interviews; Davies, Wilson, Mitchell, & Milsom, 1996; Warren, Woodall, Hunt, & Perry, 1996).

The Accuracy of Children's Initial Responses to Direct Questions and Their Narrative Responses After Prompting

Analyses of real-world investigative interviews have found that interviewers frequently ask yes–no format questions (e.g., Warren et al., 1996). This practice contrasts sharply with recommended procedures because current protocols discourage reliance on questions that limit the range of expected responses (e.g., Home Office, 1992; Poole & Lamb, 1998; Sorenson, Bottoms, & Perona, 1997). Brady, Poole, Warren, and Jones (1999) reviewed three of the concerns about yes–no questions: (a) children's initial answers are often wrong, presumably because they feel pressured to answer even when they do not know the target information; (b) children may sometimes interpret question repetition as a tacit request for a different answer; and (c) initial yes–no questioning may increase the rate of false reporting during subsequent questioning because it encourages children to respond on the basis of general feelings of familiarity for the events.

Despite supporting data for each of these concerns, there is little empirical support for banning yes–no questions, because even young children often respond accurately when interviewers ask questions about salient events (e.g., Saywitz, Goodman, Nicholas, & Moan, 1991). High error rates, such as have been reported for questions about touching, can sometimes be attributed to children's lack of understanding of the scope of specific words (Goodman, Rudy, Bottoms, & Aman, 1990; Pezdek & Roe, 1997). Further fueling the controversy is the fact that, in practice, interviewers often believe that false "yes" responses are not problematic as long as children are asked to elaborate on their responses, because they assume that children will not offer coherent narratives about events that did not occur.

We found that children's responses to yes–no questions were alarmingly inaccurate, with one important exception: In response to questions about touch events that were neither experienced nor suggested, in the first postsuggestion interview most children of all ages accurately responded "no." Unfortunately, children's responses, when prompted to describe events, did not clarify the truth status of their initial yes or no answers. These results, together with analyses of the length of narratives about true and false events, reinforce the concern expressed by other investigators about the difficulty of distinguishing true from false reports—at least at a level of accuracy that is comfortable for forensic decisions—when children have been misinformed or encouraged to narrate nonexperienced events (Bruck, Hembrooke, & Ceci, 1997; Lamb et al., 1997).

Developmental Trends in Source-Monitoring Performance

Studies of source monitoring have consistently confirmed two developmental trends. First, preschool children often (although not always) find it difficult to report the sources of specific pieces of information (e.g., Lindsay, in press; Roberts & Blades, 2000; Taylor, Esbensen, & Bennett, 1994; Welch-Ross, Diecidue, & Miller, 1997). Second, although source-monitoring ability improves dramatically after 4 years of age, age-related improvements in performance are found throughout childhood (e.g., Ackil & Zaragoza, 1995; Brainerd & Reyna, 1996; Foley & Johnson, 1985; Foley, Johnson, & Raye, 1983). The current study replicated both of these findings. For information that was presented in only one of two possible sources, there was a marked discontinuity in the ability of 3- and 4-year-old children, compared with older children, to report the source of information. Furthermore, the youngest children rarely rejected fictitious events that they had previously reported even after explicit source-monitoring instructions, whereas the older children often did. Yeah-saying biases do not seem to have produced these results, because even the youngest children were highly accurate in rejecting control events (and other studies that did not use a yes–no format have documented similarly low levels of performance by preschoolers when source discriminations were difficult; Lindsay, Johnson, & Kwon, 1991; Taylor et al., 1994).

Because the credibility of child witnesses often hinges on whether they are reporting memories of their own experiences or memories from other sources, one of the major needs in this area is for forensically practical methods of helping child witnesses differentiate among memories from different sources. The current data suggest some optimism that future research will identify reliable source-monitoring procedures suitable for applied purposes, at least for children ages 7 years and older.

Accuracy in Response to Questions About Touching Versus Nontouching Events

One criticism of suggestibility research is that false reports have been restricted to emotionally neutral or mundane events, not salient events such as inappropriate touching. This was an appropriate criticism of work done in the 1980s, but in the 1990s numerous studies demonstrated effects of suggestions on children's reports of events that involved touching, including genital and anal touching (e.g., Bruck, Ceci, Francoeur, & Renick, 1995; Steward et al., 1996). In the current project, we asked whether children would be less suggestible about touching compared with nontouching events for a context in which touching was likely not part of children's script about what should occur (contrary to what might be the case when target events are medical procedures; e.g., Eisen et al., 1998). We phrased the suggestions regarding touch in ways that made them socially appropriate and plausible, partly to avoid ethical difficulties associated with suggestions of abuse.

We were surprised that reports of fictitious touch experiences appeared in children's free-recall narratives as often as reports of fictitious science demonstrations did. This may have been partly because of the socially appropriate context in which we phrased these events. When asked direct questions in Session 2, the children less often reported suggested touching compared with suggested neutral events, but in the final interview a month later they failed to differentiate between these event types in responses to direct questions. These results suggest that touching is not a special class of event for children but, instead, that the frequency of reporting touching will vary as a function of children's expectations and the clarity of their memories.

The Relative Stability of True and False Reports

Our stability data showed the pattern we expected from previous research with different suggestibility procedures (Brainerd & Poole, 1997; Huffman, Crossman, & Ceci, 1997): Reports of suggested events were less stable than reports of experienced events. It is likely that several mechanisms contributed to the

greater decline in reports of suggested rather than experienced events. First, memories of unique, participatory events are generally more robust than memories of common, passive events (e.g., hearing a story). Second, Session 2 reports of experienced events were based on 3- to 4-month-old memories, whereas reports of suggested events were based on hearing the story mere hours or days before; consequently, between Sessions 2 and 3, memories of suggestions would be in a steeper portion of the forgetting curve than memories of experienced events. (See Brainerd & Reyna, 1995, and Brainerd & Mojardin, 1998, for discussions of spontaneous false memories, which may be especially persistent.)

It is important that a nontrivial proportion of the false reports did persist across the 1-month interval. Relatedly, children were sometimes inconsistent across the two interviews by correctly reporting experienced events in the final interview that they had not reported a month earlier. These data illustrate why there is heated debate about the diagnostic value of consistency (Fisher & Cutler, 1995).

Individual Differences in Suggestibility

There are several reasons why articles written in the 1990s contained analyses of individual difference variables more often than those written in the preceding decade. One is that psychologists interested in law saw these data as relevant to debates about the role of competency requirements for children. Requirements such as the "truth-lie ceremony" were largely abolished during the past decade, partly in light of evidence that the ability to pass such questioning does not adequately predict eyewitness accuracy (e.g., Huffman, Warren, & Larson, 1999; Pipe & Wilson, 1994). Abandoning the *voir dire* left many legal and helping professionals uncomfortable, however (McGough, 1994), and experts were aware that identifying meaningful predictors of suggestibility could shape future policy.

Although the search for predictors of eyewitness accuracy has provided converging evidence that suggestibility involves both cognitive and social factors (Bruck, Ceci, & Melnyk, 1997; Pipe & Salmon, in press; Quas, Qin, Schaaf, & Goodman 1997), the enterprise suffers from a number of practical limitations. Data are often inconsistent across studies, so one must be extremely cautious about generalizing from one or two data sets. A related point is that it is tempting to report differences when they are significant but not mention exploratory analyses when they are not. Therefore, it is likely that the strength of relationships in the literature exaggerates the true state-of-affairs. Furthermore, oftentimes the strength of relationships was weak, and the correlational nature of such analyses leaves substantial doubt about the mechanisms responsible for the observed findings.

All in all, the literature on individual differences has as yet failed to illuminate the mechanisms that underlie suggestibility or contribute definitive information that is useful in forensic settings. Results from the current study did not deviate from this generalization. The significant relationships we did find—involving the tendency to stray off topic and to be highly talkative—showed only modest relationships with suggestibility during the free recall portion of the interview, and only in the first postsuggestion session. The now-you-see-it-now-you-don't character of these results may prove to be the rule rather than the exception. One explanation of this state-of-affairs is that suggestibility is a multifaceted process that involves a variety of different mechanisms, and the mechanisms underlying false reports in particular studies depend on factors such as the type of target event, type of suggestions, and the timing and type of test. Therefore, individual differences variables that are associated with a particular mechanism (e.g., low socioeconomic status leading to greater compliance to suggestions from a high-status interviewer) may not predict false responses generated by some other mechanism (e.g., loss of source information at a delayed interview). This problem, together with the fact that it is difficult to estimate language-memory variables reliably in a single interview, leads to findings such as ours that age remains the best overall predictor of accuracy.

Acquiescence, Recall, and Memory Source Monitoring as Mediators of Developmental Declines in Suggestibility

Although it makes intuitive sense to say that young children report more suggested events because they tend to acquiesce, have poorer memories, and are deficient in monitoring source, data that bear on each of these mechanisms are equivocal. For example, a recent study of young children's patterns of responses to yes–no questions did not find prevalent yeah-saying biases (Brady et al., 1999). Similarly, Peterson and Biggs (1997) found that children who were interviewed about a medical experience were more often wrong when they said "no" than when they said "yes," contrary to expectations that children err predominantly by committing false positives. Regarding memory, there is substantial evidence that individuals often become more susceptible to suggestions as memories for target events fade (e.g., Loftus, Miller, & Burns, 1978; Reyna, 1995). However, memory abilities per se may not always be negatively associated with suggestibility because children with better retention may also be more likely to remember suggestions. Similarly, age trends in the ability to differentiate among memories from different sources may not always drive age trends in suggestibility. For one thing, it is clear that false reports of suggestions do not necessarily entail source-monitoring failures: False reports are sometimes knowingly based on postevent information (e.g., Lindsay et al., 1995). Furthermore, acquisition conditions that encourage children to generate nonpresented information spontaneously (inferences, associations, fantasies, etc.), combined with test conditions that emphasize gist-based memory reports (e.g., long delays), could produce elevated suggestibility scores among older children (Brainerd & Poole, 1997). Thus, the relationships among age, acquiescence, recall, memory source monitoring, and suggestibility are likely to vary widely depending on details of the encoding and testing conditions.

In the current study, results of the path analysis confirmed that, for our design, acquiescence, recall, and memory source monitoring made independent contributions to suggestibility. There were two surprising findings. First, recall was positively associated with suggestibility. This result probably reflects the fact that highly verbal children reported more suggested events, but possibly also the fact that children who better remembered the target events also better remembered events from the story. Second, Session 2 recall did not have indirect effects, meaning that higher recall scores were not associated with lower acquiescence or better source monitoring (although recall was associated with better source monitoring in Session 3). Because recall measures verbal enthusiasm as much as memory, these results indicate that apparent maturity and memory, as measured by verbal output in response to open-ended invitations, may not be an accurate indicator of resistance to suggestibility for young children who have been exposed to false information prior to the target interview (cf. Warren, Hulse-Trotter, & Tubbs, 1991).

In conclusion, our suggestibility procedure, in which parents described fictitious events to their children, replicated a number of well-known trends in children's eyewitness testimony. For example, older children were generally less suggestible than younger children and better able to identify the source of information when they were explicitly asked to do so. Furthermore, as in numerous previous studies, children showed a variety of strengths as witnesses, including impressively low rates of false reporting of novel touch events, even in response to direct questions. There were also many surprising findings, however, including the fact that (a) older children were not less suggestible than younger children in response to open-ended prompts; (b) false reports of unpleasant touch were as common as false reports of innocuous science demonstrations; (c) whether or not children described events did not identify experienced events as accurately as yes–no answers did; (d) even 8-year-olds sometimes reported fictitious events despite source-monitoring instructions; and (e) our recall measure yielded counterintuitive relationships with suggestibility and other mechanisms (i.e., acquiescence and source monitoring) that contribute to false reports.

The sheer volume of eyewitness research in the past decade could lull investigators and legal experts into resting comfortably on the existing knowledge base. The findings presented here, however, encourage us to question the wisdom of doing so. Results from the current project, together with findings of the numerous other studies of children's eyewitness suggestibility reported in the past decade, force a reconsideration of several frequently expressed assumptions about children's eyewitness testimony (e.g., that children's free-recall reports are invariably accurate, or that their responses to yes–no questions are peculiarly dubious). In turn, this reconsideration suggests several goals for future research. The most immediate need is to better align experts' statements regarding children's eyewitness reports with the current database. A more far-reaching goal is to forge a better theoretical understanding of the mechanisms that underlie accurate and inaccurate reports by child witnesses. A transition to more theoretically guided research designs will be especially helpful for memory researchers who offer opinions in forensic settings, because without an ability to predict how the unique features of individual cases likely combined to affect memory performance, the frequent conclusion that children's memory is both resistant to suggestion and highly malleable leaves experts unable to argue that they contribute unique insights to the trier of fact. Finally, there is a clear need to widen our understanding of suggestibility to include children between the ages of 9 and 18 years, an age group that has been virtually ignored in this literature to date. Tempting as simplistic dichotomies between "child witnesses" and "adult witnesses" may be, existing data suggest that the strengths and weaknesses of humans as reporters of past experience continually shift and change throughout childhood (and likely throughout the life span), and that age-related differences interact with a host of situational factors to determine the accuracy of eyewitness reports.

References

Ackil, J. K., & Zaragoza, M. S. (1995). Developmental differences in eyewitness suggestibility and memory for source. *Journal of Experimental Child Psychology, 60*, 57–83.

Brady, M. S., Poole, D. A., Warren, A. R., & Jones, H. R. (1999). Young children's responses to yes-no questions: Patterns and problems. *Applied Developmental Science, 3*, 47–57.

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

Brainerd, C. J., & Poole, D. A. (1997). Long-term survival of children's false memories: A review. *Learning and Individual Differences, 9*, 125–151.

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

Brainerd, C. J., & Reyna, V. F. (1996). Mere memory testing creates false memories in children. *Developmental Psychology, 32*, 467–478.

Brewer, N., Potter, B., Fisher, R. P., Bond, N., & Luszcz, M. A. (1999). Beliefs and data on the relationships between consistency and accuracy of eyewitness testimony. *Applied Cognitive Psychology, 13*, 297–313.

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

Bruck, M., Ceci, S. J., Francoeur, E., & Renick, A. (1995). Anatomically detailed dolls do not facilitate preschoolers' reports of a pediatric examination involving genital touching. *Journal of Experimental Psychology: Applied, 1*, 95–109.

Bruck, M., Ceci, S. J., & Melnyk, L. (1997). External and internal sources of variation in the creation of false reports in children. *Learning and Individual Differences, 9*, 289–316.

Bruck, M., Hembrooke, H., & Ceci, S. (1997). Children's reports of pleasant and unpleasant events. In J. D. Read & D. S. Lindsay (Eds.), *Recollections of trauma: Scientific evidence and clinical practice* (pp. 199–213). New York: Plenum.

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

Ceci, S. J., & Bruck, M. (1995). Jeopardy in the courtroom: A scientific analysis of children's testimony. Washington, DC: American Psychological Association.

Ceci, S. J., Huffman, M. L. C., Smith, E., & Loftus, E. F. (1994). Repeatedly thinking about a non-event: Source misattributions among preschoolers. *Consciousness and Cognition, 3*, 388–407.

Coxon, P., & Valentine, T. (1997). The effects of the age of eyewitness on the accuracy and suggestibility of their testimony. *Applied Cognitive Psychology, 11*, 415–430.

Davies, G., Wilson, C., Mitchell, R., & Milsom, J. (1996). *Videotaping children's evidence: An evaluation.* London: Home Office.

Douglas, R. N. (1996). Considering the witness in interviews [8 paragraphs]. *Psycholoquy* [On-line serial], 7 (21). Available FTP: Hostname: princeton.edu Directory: pub/harnard/Psycholoquy/1996.volume.7 File: psyc.96.7.21.witness-memory.7.douglas

Eisen, M. L., Goodman, G. S., Qin, J., & Davis, S. (1998). Memory and suggestibility in maltreated children: New research relevant to evaluating allegations of abuse. In S. L. Lynn & K. McConkey (Eds.), *Truth in memory* (pp. 163–189). New York: Guilford.

Elischberger, H. B., & Roebers, C. M. (in press). Improving young children's free narratives about an observed event: The effects of unspecific verbal prompts. *International Journal of Behavioral Development.*

Fisher, R. P., & Cutler, B. L. (1995). The relation between consistency and accuracy of eyewitness testimony. In G. Davies, S. Lloyd-Bostock, M. McMurran, & C. Wilson (Eds.), *Psychology, law and criminal justice: International developments in research and practice* (pp. 21–28). Berlin: DeGruyter.

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.

Goodman, G. S., Quas, J. A., Batterman-Faunce, J. M., Riddlesberger, M. M., & Kuhn, J. (1994). Predictors of accurate and inaccurate memories of traumatic events experienced in childhood. *Consciousness and Cognition, 3*, 269–294.

Goodman, G. S., Rudy, L., Bottoms, B. L., & Aman, C. (1990). Children's concerns and memory: Issues of ecological validity in the study of children's eyewitness testimony. In R. Fivush & J. A. Hudson (Eds.), *Knowing and remembering in young children* (pp. 249–284). New York: Cambridge University Press.

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development, 59*, 1366–1371.

Home Office. (1992). *Memorandum of good practice on video recorded interviews with child witnesses for criminal proceedings*. London: Home Office with Department of Health.

Huffman, M. L., Crossman, A., & Ceci, S. (1997). "Are false memories permanent?": An investigation of the long-term effects of source misattributions. *Consciousness & Cognition, 6*, 482–490.

Huffman, M. L., Warren, A. R., & Larson, S. M. (1999). Discussing truth and lies in interviews with children: Whether, why, and how? *Applied Developmental Science, 3*, 6–15.

Hyman, I. E., Jr., & Billings, F. J. (1998). Individual differences and the creation of false childhood memories. *Memory, 6*, 1–20.

Lamb, M. E., Sternberg, K. J., Esplin, P. W., Hershkowitz, I., Orbach, Y., & Hovav, M. (1997). Criterion-based content analysis: A field validation study. *Child Abuse and Neglect, 21*, 255–264.

Leichtman, M. D., & Ceci, S. J. (1995). The effects of stereotypes and suggestions on preschoolers' reports. *Developmental Psychology, 31*, 568–578.

Lindberg, M. A., Keiffer, J., & Thomas, S. W. (in press). Eyewitness testimony for physical abuse as a function of personal experience, development, and focus of study. *Journal of Applied Developmental Psychology*.

Lindsay, D. S. (in press). Children's source monitoring. In H. L. Westcott, G. Davies, & R. H. C. Bull (Eds.), *Children's testimony: Psychological research and forensic practice*. Sussex, England: Wiley.

Lindsay, D. S., Gonzales, V., & Eso, K. (1995). Aware and unaware uses of memories of postevent suggestions. In M. S. Zaragoza, J. R. Graham, G. C. N. Hall, R. Hirschman, & Y. S. Ben-Porath (Eds.), *Memory and testimony in the child witness* (pp. 86–108). Thousand Oaks, CA: Sage.

Lindsay, D. S., Johnson, M. K., & Kwon, P. (1991). Developmental changes in memory source monitoring. *Journal of Experimental Child Psychology, 52*, 297–318.

Loftus, E. (1997). Memory for a past that never was. *Current Directions in Psychological Science, 6*, 60–65.

Loftus, E. F., Miller, D. G., & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory, 4*, 19–31.

McGough, L. S. (1994). *Child witnesses: Fragile voices in the American legal system*. New Haven, CT: Yale University Press.

Newcombe, P. A., & Siegal, M. (1997). Explicitly questioning the nature of suggestibility in preschoolers' memory and retention. *Journal of Experimental Child Psychology, 67*, 185–203.

Pence, D., & Wilson, C. A. (1994). *Team investigation of child sexual abuse: The uneasy alliance*. Thousand Oaks, CA: Sage.

Peterson, C., & Biggs, M. (1997). Interviewing children about trauma: Problems with "specific" questions. *Journal of Traumatic Stress, 10*, 279–290.

Pezdek, K., & Roe, C. (1995). The effect of memory trace strength on suggestibility. *Journal of Experimental Child Psychology, 60*, 116–128.

Pezdek, K., & Roe, C. (1997). The suggestibility of children's memory for being touched: Planting, erasing, and changing memories. *Law and Human Behavior, 21*, 95–106.

Pipe, M.-E., & Salmon, K. (in press). What children bring to the interview context: Individual differences in children's event reports. In G. S. Goodman, M. L. Eisen, & J. A. Quas (Eds.) *Memory and suggestibility in the forensic interview*. New York: Erlbaum.

Pipe, M.-E., & Wilson, J. C. (1994). Cues and secrets: Influences on children's event reports. *Developmental Psychology, 30*, 515–525.

Poole, D. A., & Lamb, M. E. (1998). *Investigative interviews of children: A guide for helping professionals*. Washington, DC: American Psychological Association.

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

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

Quas, J. A., Qin, J., Schaaf, J., & Goodman, G. S. (1997). Individual differences in children's and adults' suggestibility and false event memory. *Learning and Individual Differences, 9*, 359–390.

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

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.

Roberts, K., & Blades, M. (2000). *Children's source monitoring*. Mahwah, NJ: Erlbaum.

Saywitz, K. J., Goodman, G. S., Nicholas, E., & Moan, S. F. (1991). Children's memories of a physical examination involving genital touch: Implications for reports of child sexual abuse. *Journal of Consulting and Clinical Psychology, 59*, 682–691.

Schacter, D. L., Kagan, J., & Leichtman, M. D. (1995). True and false memories in children and adults: A cognitive neuroscience perspective. *Psychology, Public Policy, and Law, 1*, 411–428.

Sorenson, E., Bottoms, B., & Perona, A. (1997). *Intake and forensic interviewing in the children's advocacy center setting: A handbook*. Washington, DC: National Network of Children's Advocacy Centers.

Steward, M. S., Steward, D. S., Farquhar, L., Myers, J.E.B., Reinhart, M., Welker, J., Joye, N., Driskill, J., & Morgan, J. (1996). Interviewing young children about body touch and handling. *Monograph of the Society for Research in Child Development, 61* (4–5, Serial No. 248).

Taylor, M., Esbensen, B. M., & Bennett, R. T. (1994). Children's understanding of knowledge acquisition: The tendency for children to report that they have always known what they have just learned. *Child Development, 65*, 1581–1604.

Thompson, C. P., Skowronski, J. J., Larsen, S. F., & Betz, A. (1996). *Autobiographical memory: Remembering what and remembering when*. Mahwah, NJ: Erlbaum.

Tobey, A. E., & Goodman, G. S. (1992). Children's eyewitness memory: Effects of participation and forensic context. *Child Abuse & Neglect, 16*, 779–796.

Warren, A., Hulse-Trotter, K., & Tubbs, E. C. (1991). Inducing resistance to suggestibility in children. *Law and Human Behavior, 15*, 273–285.

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

Warren, A. R., Woodall, C. E., Hunt, J. S., & Perry, N. W. (1996). "It sounds good in theory, but . . .": Do investigative interviewers follow guidelines based on memory research? *Child Maltreatment, 1*, 231–245.

Welch-Ross, M. K., Diecidue, K., & Miller, S. A. (1997). Young children's understanding of conflicting mental representation predicts suggestibility. *Developmental Psychology, 3*, 43–53.

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist, 48*, 553–571.

Received December 27, 1999
Revision received August 31, 2000
Accepted September 6, 2000 ■