Post-Event Information Affects Children’s Autobiographical Memory After One Year

Kamala London • Maggie Brück • Laura Melnyk

Abstract In two experiments, we examined whether post-event information (PEI) about true and false events persisted in children’s reports after approximately 1 year. In Experiment 1, 4- to 6-year-olds were given PEI and then were given memory tests 2 weeks and 15 months later. Although PEI appeared in free recall at the initial testing, it decreased substantially by the long-term test. In contrast, on recognition tasks the children showed facilitation and misinformation effects at initial and follow-up tests. Experiment 2 replicated lasting misinformation and facilitation effects in recognition memory among 4- to 9-year-olds who were tested after 1-week and 10-month delays. We conclude that true and false reminders about an experienced event continue to affect children’s memory approximately 1 year later.

Keywords Autobiographical memory • Delayed memory report • Suggestibility • Forensic interviews

Over the past 25 years, a great deal of research has been conducted to examine suggestibility in children. In one of the more well-known paradigms, the misinformation paradigm (e.g., Loftus, 1997), a common finding is that children provided with misinformation about an experienced event often later incorporate that information into false reports of the event (e.g., Ackil & Zaragoza, 1995, 1998; Brück, Melnyk, & Ceci, 2000; Marche, 1999; Poole & Lindsay, 1995, 2001; Roberts & Powell, 2006). In most of these studies, the delay between the delivery of misinformation and the memory test is relatively short (several minutes to several weeks) compared to conditions in actual forensic contexts where there may be delays of several months between a suggestive forensic interview and testifying at trial (Bjorklund, Bjorklund, Brown, & Cassel, 1998; Ceci & Brück, 1995; Flin, 1995; Goodman et al., 1992). In other cases, children may be subjected to sources of memory taint (e.g., by parents and therapists) and then, following a delay, interviewed by forensic interviewers. The major purpose of this paper is to examine whether providing children with false information about a previously experienced event affects their autobiographic reports following a delay of approximately a year.

The creation and persistence of false memories should be evaluated in the context of the longevity of children’s accurate memories of events, a strategy that is used in the present paper. Clearly, a significant finding from the literature on children’s autobiographical memories is that preschool and school-aged children can recall some details about personally experienced events with delays of months and even years when interviewed with non-leading techniques (e.g., Butler, Gross, & Hayne, 1995; Fivush & Hamond, 1990; Hamond & Fivush, 1991; Hudson & Fivush, 1991; Pillemer; 1993; Salmon, Bidrose, & Pipe, 1995). For example, Peterson and colleagues (e.g., Peterson, 1996, 1999; Peterson & Bell, 1996; Peterson, Moores, & White, 2001; Peterson & Whalen, 2001) examined long-term event memory for an emergency room visit in children from 2 to 13 years old, with delays of 6 months to 5 years.
Their data revealed that although preschoolers reported less than older children, even the 3-year-olds recalled some central information about highly salient events.

The effects of giving children true reminders about experienced events have also been examined in some studies of autobiographical recall, using methods similar to the misinformation paradigm. In these studies, however, children were provided with true rather than false information about the experienced event and later (usually several days or weeks later) asked to remember the details of the actual event. The general finding is that presentation of true reminders aids accurate recall (Bruck et al., 2000; Harley & Reese, 1999; Melnyk & Bruck, 2004; Nelson, 1995). Most of these studies, though, have only examined short-term effects of exposure to true reminders, with delays ranging from a few minutes to several months. Thus it is not known if such reminders continue to assist autobiographical memory once the reminder sessions have stopped. The second purpose of this paper is to present data to address this issue.

In our review of the literature, we do not cover the few studies on the stability of children’s non-suggested errors over time (e.g., Salmon & Pipe, 1997). Nor do we include studies where there is a significant delay between the event and later misleading questions/information (e.g., Goodman, Hirschman, Hepps, & Rudy, 1991; Goodman, Battersman-Faunce, Schaaf, & Kenney, 2002; Quas, Goodman, Bidrose, Pipe, Craw, & Ablin, 1999) because the focus of these studies is whether children are suggestible when suggestive questioning happens for the first time after a long delay. In contrast, the present studies focus on the persistence of false reminders, with delays ranging from a few minutes to several months. Below, we summarize the few studies that have examined the long-term effects of providing false reminders (misinformation) and true reminders on children’s memory of an event.

Peterson, Parsons, and Dean (2004) interviewed children (ages 2–12) who had been treated in an emergency room for injuries. The children were asked non-leading questions about these events 1 week and 6 months after the emergency room visit. One year later, they were given misleading information about four details. Then, 1 week after misleading information, their memory of the emergency room visits and injuries were obtained through neutral interviews. At the interview 1-week after misinformation (baseline), only 2% of children’s responses reflected the incorporation of misleading information; 1 year later, only 1% of the responses reflected misinformation. Neither of these findings is surprising given the fact that the children were initially given misinformation about well-rehearsed events for which they had strong memory representations. Because the children never showed an initial misinformation effect, one would not expect to find a misinformation effect 1 year after the initial suggestion. The important issue at hand is whether reports made as a result of suggestions persist after a long period of no suggestions.

Huffman, Crossman, and Ceci (1997) examined the persistence of preschool children’s true and false reports that had been elicited 2 years previously when the children were repeatedly asked to “think hard” about two true events and about eight false events (Ceci, Crotteau-Huffman, Smith, & Loftus, 1994). At that initial testing, children assented to 80% of the true events and 22% of the false events. After initial testing, the children were extensively debriefed that the false events had not occurred. Two years later, when asked whether these events had occurred, children assented to 77% of the true events (a drop of seven percentage points) and to 13% of the false events (a drop of nine percentage points). Although this study provides some data on the longevity of false reports generated from “suggestive” interviewing techniques (i.e., to think hard about the event), it should be noted that these very mild suggestive techniques did not produce high rates of false reports in the initial testing, which could have contributed to a low rate of false reports in the follow-up. Perhaps if the suggestions had been more explicit (e.g., telling the child rather than asking the child to imagine certain events), then there may have been higher rates of false reports at baseline, allowing a better picture of the longevity of the false event reports. Furthermore, after the original study, Ceci et al. (1994) extensively debriefed subjects that the false events had not occurred (which could deflate reports of non-experienced events). Additionally, for the experienced events, memory might have been facilitated due to continued conversations with family members about these experienced events.

Poole and Lindsay (2001) provided data on the longevity of true and false reports following misinformation. Children aged 3 to 8 years participated in a 16-min staged interaction with “Mr. Science.” After a 3-month delay, parents read children stories that contained true events and false details about the staged event. These stories were repeatedly read over several days. Children were interviewed soon after exposure to this misinformation and then 1 month later. Misinformation was reported by 35% of children at the initial testing and 21% of children at the 1-month follow-up session. Thus, although there was a reduction in the misinformation effect after 1 month, the effect still persisted.

Melnyk and Bruck (2004) examined long-term misinformation and facilitation effects in preschool- and kindergarten-aged children who participated in a magic show. After a 2-week delay, the children were provided with true reminders and false reminders about the event. The first memory test was given 6 weeks after the event,
and the second memory test was given 5 months later. Melnyk and Bruck (2004) found that reminders had the same effects on children’s memory over a 5-month period. Specifically, the rates of facilitation and misinformation effects were the same at 6 weeks and at 5 months after the information had been provided.

We present two studies that expand on the previous research in several ways. First, we examined whether misinformation and facilitation effects that were previously found to persist for up to 5 months would persist even longer. Given the long delays that may occur between interviews in some forensic contexts (e.g., delay between a forensic interview and testimony; Flin, 1995), this is an important issue. The delay between the first memory and follow-up memory interviews was 15 months in Experiment 1 and 10 months in Experiment 2. Second, these two studies included a wide age range of children: In Experiment 1 the initial sample included preschool- and kindergarten-aged children (4- to 6-year-olds); in Experiment 2, the initial sample included 4- to 9-year-olds, providing an opportunity to examine age differences in long-term stability of misinformation and facilitation effects. Finally, most of the earlier studies on long-term misinformation effects exposed children to multiple suggestive interviews. In Experiment 2, only a single suggestive interview was used. Hence, these data examine whether misinformation and facilitation effects induced by a single interview session can persist without intervening “booster” sessions.

Given prior work revealing lasting reminder and suggestibility effects after a 5-month delay, we hypothesized that these effects would be seen following delays of approximately a year. This hypothesis was tested in 4- to 6-year-olds in Experiment 1. Given recent work revealing strong suggestibility effects in school-aged children, we hypothesized reminder and suggestibility effects would also persist among this older group of children. This hypothesis was tested in Experiment 2.

**EXPERIMENT 1: MISINFORMATION AND FACILITATION EFFECTS AFTER A 15-MONTH DELAY**

**Method**

**Participants**

Bruck et al. (2000) tested 87 4- to 6-year-olds (52 to 75 months old, \( M = 67 \) months, \( SD = 4.6 \) months) from middle-class suburban neighborhoods in the Montreal area. Most of the children were ages 4 and 5 years; five of the 45 children had recently turned 6 years (ages 73 to 75 months). They were attending preschool or kindergarten classes. Thirteen to 16 months later (\( M = 15.02 \) months), 45 of these children (23 girls) were located and received parental consent to participate in a follow-up interview. At the time of follow-up, these children were between the ages of 68 and 91 months (\( M = 82, \ SD = 5 \)). There were no differences on demographic or memory test measures between the follow-up sample and the group of children who had dropped out of the study.

**Procedure**

**Original Bruck et al. (2000) Study**

**Session 1: Magic Show.** As described in Bruck et al. (2000), children individually participated in a scripted magic show containing 16 target events. Events included the following: The magician demonstrated two magic tricks, gave the child various magical paraphernalia to help with the tricks, tripped on her shoelaces, fell, and asked the child to help.

**Sessions 2 & 3: Suggestive Interviews.** Sixteen days later, children were given eight reminders about the magic show: four reminders described actual events (true reminders) and four reminders contained false information (false reminders, or false suggested items). Each reminder was in the form of a statement (e.g., “I heard that you wore red magic helper boots…”) that was followed by a forced choice question (e.g., “Were the boots shiny or dirty?”). An example of a false reminder is that the magician wore black gloves when she did not (see Bruck et al., 2000, Appendix, for a list of the true and false reminders). Half the children answered the questions verbally and half the children answered the questions verbally and then drew their answers (e.g., “Draw me a picture of the red boots”). At the end of the session, the experimenter repeated all the reminders, prefacing them with, “You told me that…” Twelve days later, the same suggestive procedures were repeated.

**Session 4: Initial Memory Test.** The initial memory test was conducted approximately 12 days after the last suggestive interview. Children were first asked for free recall of the magic show. Then, they were asked 16 yes/no recognition questions probing target details about the magic show: four about the true reminders, four about false reminders (e.g., “Did you wear red magic helper boots?”), four true controls (true details that were not rehearsed in the suggestive interviews), and four false controls (false details that were not rehearsed in the suggestive interviews).

**Session 5: Follow-up Memory Test.** Approximately 15 months following the staged event, the children were individually re-tested. After a brief warm-up, each child
was asked about the name of the class and the teacher of the previous year. The experimenter then said she was there to talk about a magic show that happened at that time. She showed the child a photograph of the magician who had visited and asked if the child remembered the magician.

The formal memory test that followed was the same test administered in the previous year. During an initial free recall section, the child was asked to tell everything he or she remembered about the magic show and was prompted for additional details until nothing further was said. Next, the child answered the 16 yes/no recognition questions that were used in the initial memory test.

The reminder status of each item was counterbalanced across children, making the experimenter unaware of which items for a given child were reminded during the suggestive interviews. That is, an item that served as a reminder for one child was not a reminder for another child.

Scoring

The free recall responses were first segmented into utterances. An utterance was defined as a statement bound by pauses containing one verb. For example, “The magician waved a magic wand, and we held hands,” contains two utterances. The number of correct and incorrect utterances as a function of whether the utterance was reminded during the suggestive interviews was counted. Utterances were coded by two separate research assistants. Coding was compared, and rare cases of discrepancy were resolved via discussion.

Results

Because there were no main effects or interactions involving the initial interviewing condition (drawing vs. no-drawing), the data were collapsed across this variable. Data were also collapsed across age since there was no relationship between the age and any of the memory measures at the original interview or follow-up interview, rs between -.25 and +.25, all ps > .08. There was no consistent pattern among age and the dependent measures, and most rs were less than .10. Also, the patterns in the statistical tests remained identical when age was controlled. Finally, because there were no significant correlations between length of delay and any of the recognition or recall measures, length of delay was not included as a covariate.

Free Recall Data

Three children produced no utterances in the initial interview and 14 children produced no utterances in the follow-up interview. Due to different base rates at the initial interview, correct and incorrect utterances are analyzed separately.

Correct Spontaneous Utterances

A repeated measures t-test revealed that the number of correct utterances decreased between the initial and follow-up interviews, t(44) = 3.46, p < .005 (see Table 1). In order to further examine the specific effects of the reminded information on free recall, we examined the proportion of utterances that included true reminders. At the initial interview, 29% of all correct utterances (69/234) included true reminders; however, by follow-up the rate had dropped to 4% (5/133). This pattern suggests that true reminders did not have a long-term influence on free recall. Furthermore, the disappearance of true reminders from the follow-up recall data accounts for the decrease in total correct recall between initial and follow-up interviews. That is, once all utterances that mentioned true reminders were excluded from the free recall data, there were no longer differences between spontaneous recall at the first interview (M = 3.67) and the follow-up interview (M = 2.84), t(44) = 1.36, p = .18.

Incorrect Spontaneous Utterances

As shown in Table 1, there were no significant changes in the number of incorrect utterances between the initial and follow-up interviews, t(44) = -1.27, p = .20. As was true for the correct utterances, the effects of reminders did not persist over a 15-month period. Specifically, at the initial interview, 75% (51/68) of all incorrect statements reflected

<table>
<thead>
<tr>
<th>Age (years) at 1st interview (n)</th>
<th>Correct utterances</th>
<th>Incorrect utterances</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial test</td>
<td>Follow-up test</td>
</tr>
<tr>
<td></td>
<td>Initial test</td>
<td>Follow-up test</td>
</tr>
<tr>
<td>Experiment 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4-6 years (n = 45)</td>
<td>5.20 (3.15)</td>
<td>1.49 (1.20)</td>
</tr>
<tr>
<td></td>
<td>2.96 (3.25)</td>
<td>2.20 (3.84)</td>
</tr>
<tr>
<td>Experiment 2</td>
<td>3.36 (2.59)</td>
<td>1.57 (1.65)</td>
</tr>
<tr>
<td>4-5 years (n = 14)</td>
<td>1.14 (1.51)</td>
<td>1.29 (1.39)</td>
</tr>
<tr>
<td>6-7 years (n = 18)</td>
<td>7.00 (6.09)</td>
<td>3.06 (4.21)</td>
</tr>
<tr>
<td>8-9 years (n = 19)</td>
<td>10.16 (2.93)</td>
<td>2.42 (1.92)</td>
</tr>
<tr>
<td>Total 4-9 years (n = 51)</td>
<td>7.18 (5.00)</td>
<td>1.78 (1.81)</td>
</tr>
</tbody>
</table>
false reminded information, whereas at follow-up there were no false reminders (0/99) in the spontaneous statements. Thus, although most of the incorrect spontaneous statements at the initial interview reflected the effects of prior misinformation, the spontaneous errors at follow-up reflected other types of memory errors (probably based upon forgetting the magic show).

**Recognition Data**

The following analyses of the recognition data not only allow an examination of children’s memory of items when explicitly asked about them, but they also provide a method to examine the long lasting effects of providing true and false reminders. Specifically, if true reminded items are remembered better than true control items (i.e., children answer “yes” to these questions), then this supports the hypothesis that the reminders facilitate children’s memory. Similarly, if children provide more “yes” responses to false reminder than to false control questions, then this indicates there was a misinformation effect (i.e., that the presentation of false information interferes with memory beyond what would be expected from the normal effects of other types of memory distortion). In Brück et al. (2000), we found significant facilitation and misinformation effects; the present analyses examine the stability of these effects following delay. Given the different base-rates at the initial interview between true and false items, the true and false items are analyzed separately.

**True Recognition Items**

The number of accurate assents for true items was entered into a 2 (time: initial interview vs. follow-up interview) x 2 (item type: true reminder vs. true control) repeated measures Analysis of Variance (ANOVA). There was a main effect of item type, \( F(1, 44) = 60.92, p < .001, \eta^2_p = .58 \), as well as a Time x Item Type interaction, \( F(1, 44) = 19.94, p < .001, \eta^2_p = .31 \). As shown in the top third of Table 2, and confirmed by planned comparisons, accuracy on true reminder items decreased with time, whereas there were no changes in accuracy on true control items. Further analysis of the interaction showed a significant facilitation effect at follow-up (i.e., there were significantly more accurate assents for reminded compared to control items), although its size had decreased over time, directly due to the decrease in the recognition of the true reminded items.

1 In Tables 2 and 3 the recognition data are presented in proportions even though the raw scores were entered into the statistical analyses. We do this to facilitate general comparisons between Experiments 1 and 2.

**False Recognition Items**

The number of inaccurate assents for false items was entered into a 2 (time: initial interview vs. follow-up interview) x 2 (item type: true reminder vs. true control) repeated measures ANOVA. There was a significant main effect of item type, \( F(1, 44) = 118.93, p < .001, \eta^2_p = .73 \), as well as a Time x Item Type interaction, \( F(1, 44) = 27.25, p < .001, \eta^2_p = .38 \). The interaction (see top third of Table 3) reflected the fact that although there were significant misinformation effects at both time periods (i.e., there were more false assents to false reminder items than to false control items, all \( ps < .001 \)), the size of the

<table>
<thead>
<tr>
<th>Table 2 Mean proportion correct (with standard deviations) by time and reminder status for true recognition questions for Experiments 1–2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years) at 1st interview (n)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td><strong>Experiment 1</strong></td>
</tr>
<tr>
<td>4–5 years (n = 45)</td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
</tr>
<tr>
<td>4–5 years (n = 14)</td>
</tr>
<tr>
<td>6–7 years (n = 18)</td>
</tr>
<tr>
<td>8–9 years (n = 19)</td>
</tr>
<tr>
<td><strong>Total 4–9 years</strong> (n = 51)</td>
</tr>
</tbody>
</table>

**Note:** \( d = \) Cohen’s effect size

<table>
<thead>
<tr>
<th>Table 3 Mean proportion incorrect (with standard deviations) by time and reminder status for false recognition questions for Experiments 1–2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years) at 1st interview (n)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td><strong>Experiment 1</strong></td>
</tr>
<tr>
<td>4–6 (n = 45)</td>
</tr>
<tr>
<td><strong>Experiment 2</strong></td>
</tr>
<tr>
<td>4–5 years (n = 14)</td>
</tr>
<tr>
<td>6–7 years (n = 18)</td>
</tr>
<tr>
<td>8–9 years (n = 19)</td>
</tr>
<tr>
<td><strong>Total 4–9 years</strong> (n = 51)</td>
</tr>
</tbody>
</table>

**Note:** \( d = \) Cohen’s effect size
misinformation effect was smaller at follow-up than at baseline. The major reason for this decrease was the fact that assents to false control items significantly increased by 20% whereas there was no significant change of false assents to the false reminder items. These data suggest that a decrease in the misinformation effect was not due to an increase in correctly rejecting formerly provided misinformation but instead reflected the relative increase in other types of memory errors.

Summary of Results from Experiment 1

Although children incorporated a substantial amount of suggested information (true and false reminders) into their initial spontaneous recall of the magic show, 15 months later spontaneous recall was almost devoid of these reminders. Overall, though, their free recall reports at the 15-month interview were sparse. Thus, suggestibility effects produced by two interviews did not persist in their free recall 15 months later. In contrast, suggestions in the form of true and false reminders resulted in significant facilitation effects and misinformation effects in the initial and follow-up recognition memory tests; as shown in Tables 2 and 3, the effect sizes were very large for the initial session and moderate to large for the follow-up period.

EXPERIMENT 2: MISINFORMATION AND FACILITATION EFFECTS IN 4- TO 9-YEAR-OLDS FOLLOWING A 10-MONTH DELAY

Experiment 2 was both a replication and extension of Experiment 1. First, a broader age group of children (between 4 and 9 years of age) was included at the initial interview to examine developmental changes on the major measures. Second, in Experiment 2, each interview session was conducted by an unfamiliar person. This procedure allowed us to determine if memory performance in the follow-up interview of Experiment 1 was inflated due to the fact that most of the children (74%) were interviewed by the same interviewer who conducted their initial suggestive interview 15 months previously. The interviewer’s presence may have provided extra cues for those children to retrieve misinformation as well as accurate information. Third, there was only one suggestive interview in Experiment 2 rather than two (as was the case for Experiment 1). Some work indicates that two suggestive interviews can produce larger misinformation and facilitation effects than one suggestive interview (Melnik & Bruck, 2004); thus, one might predict less enduring long-term misinformation and facilitation effects in Experiment 2. However, Melnik and Bruck (2004) also reported that the timing of the first suggestive interview is of critical importance in augmenting facilitation and misinformation effects, such that an initial suggestive interview that occurs more than several days after the event produces facilitation and misinformation effects of the same magnitude as two suggestive interviews. Therefore, it might be the case that the robustness of the facilitation and misinformation effects is similar in Experiment 1 and Experiment 2.

Method

Participants

Bruck, London, Landa, and Goodman (2007) tested 68 normal functioning children who ranged in age from 4 to 9 years with a mean of 6.71 years (SD = 1.68). These children were recruited through Baltimore-area daycares and schools. All children had average IQs.

Between 7 and 13 months later, we located, received parental consent, and retested 51 of these children. The children’s ages at the time of follow-up ranged from 4 to 10 years (M = 8 years, SD = 2.15 months). There were 30 girls (59%).

The 17 children who were not re-interviewed could not be contacted because they had moved or did not respond to our calls. No differences were found in memory performance at the original memory test or demographic variables between these 17 children and the remaining sample.

Procedure

Original Study

The procedures were highly similar to those followed in Experiment 1. Children saw a magic show and were questioned about 1 week later in a session designed to provide true and false reminders about the magic show (with many of the same reminders as used in Experiment 1). Examples of the false reminded items were that the magician brushed her hair, blew up a balloon, and hugged the child (see Bruck et al., 2007, Appendix, for a complete list of the true and false reminded items). In contrast to Experiment 1, the suggestive interviewer in Experiment 2 wore a large yellow stovepipe hat and called herself the “Yellow Lady.” Several days after the suggestive interview, the children were given a memory test that included many of the same questions as in Experiment 1. Its structure was identical to that of Experiment 1.

Follow-up Memory Test

Following an approximate 10-month delay since witnessing the magic show (M = 303 days, SD = 46 days), a
fourth person visited all children at their school, home, or daycare. The entire follow-up session lasted about 30 min. After a brief warm-up, they were reminded that a magician had visited them the previous year. The interviewer asked them if they remembered the magic show. Children who reported that they did not remember the magic show were told, “Well, I am going to go ahead and ask you some questions about the magic show. You think really hard and maybe you will remember some things about it.” Children were administered the same memory test that was given in the original study. This consisted of free recall and a 20-item recognition task.

Results

Children were categorized into three age groups: 4- to 5-year-olds (48 to 72 months old, \( M = 60.71 \) months, \( n = 14 \)), 6- to 7-year-olds (73 to 94 months old, \( M = 82.11 \) months, \( n = 18 \)), and 8- to 9-year-olds (100 to 119 months old, \( M = 109.37 \), \( n = 19 \)). All analyses in Study 2 were ANOVAs with planned contrasts for the age variable (age in years) to extract the linear trend. Because there were no significant correlations between length of delay and any of the recognition or recall measures, length of delay was not included as a covariate.

Free Recall Data

Four children (two 4- to 5-year-olds, two 6- to 7-year-olds) recalled no details about the magic show at the initial interview and this number increased to eight children (all 4- to 5-year-olds) at the follow-up interview.

Table 1 shows the mean number of correct and incorrect utterances at initial and follow-up recall by age group. Separate analyses were carried out on the true and false utterances.

Correct Spontaneous Utterances

A mixed factor 2 (time: initial vs. follow-up interview) x 3 (age group: 4- to 5-year-olds, 6- to 7-year-olds, 8- to 9-year-olds) repeated measures ANOVA on number of correct utterances yielded main effects of time, \( F(1, 49) = 49.94, p < .001, \eta_p^2 = .50 \), and age, \( F(1, 49) = 27.39, p < .001, \eta_p^2 = .36 \). There was also a significant Time x Age interaction \( F(1, 49) = 7.73, p < .01, \eta_p^2 = .14 \). As shown in Table 1 and confirmed by planned comparisons, there was a smooth developmental trend for the initial recall data; however, by follow-up the 6- to 7-year-olds and 8- to 9-year-olds performed similarly but better than the 4- to 5-year-olds (ps < .04). All age groups provided less information at follow-up than at the initial interview.

Incorrect Spontaneous Utterances

A mixed factor 2 (time: initial vs. follow-up interview) x 3 (age group: 4- to 5-year-olds, 6- to 7-year-olds, 8- to 9-year-olds) repeated measures ANOVA on number of incorrect utterances yielded an effect of age, \( F(1, 49) = 3.98, p = .05, \eta_p^2 = .08 \). There was no main effect or interaction involving the factor of time.

As shown in Table 1, the older two age groups gave more incorrect utterances compared to the youngest age group. This finding, however, must be qualified by the fact that the two oldest age groups also produced more correct spontaneous utterances. From this viewpoint, it turns out that the narratives of the youngest age group contained proportionately more incorrect information than those of the older groups. Collapsed across the initial and follow-up memory tests, 39%, 29%, and 27% of utterances were incorrect for the youngest, middle, and oldest age groups, respectively, \( F(2, 45) = 3.62, p < .05, \eta_p^2 = .14 \).

Although this analysis indicates that the children made a substantial number of errors, these errors did not reflect the incorporation of the false reminders into free recall either at the initial (9% of all errors reflected the false reminders) nor follow-up interviews (3% of all errors reflected the false reminders). There were no age differences on these measures.

Recognition Data

Consistent with Experiment 1, separate analyses were conducted for the true and false items. As with the free recall data, trend analyses were conducted with planned contrasts to extract the linear trend for the age variable.

True Recognition Items

The number of accurate assents for true items was entered into a mixed factor 2 (time: initial interview vs. follow-up interview) x 2 (item type: true reminder vs. true control) x 3 (age group: 4- to 5-year-olds, 6- to 7-year-olds, 8- to 9-year-olds) repeated measures ANOVA. There was a main effect of item type, \( F(1, 49) = 39.73, p < .001, \eta_p^2 = .45 \), age, \( F(1, 49) = 9.34, p < .005, \eta_p^2 = .16 \), as well as an Age x Item Type interaction, \( F(1, 49) = 6.07, p = .02, \eta_p^2 = .11 \). The main effect of time was not significant.
(p = .14). Although all age groups showed a significant facilitation effect (true reminder minus true control items), the oldest children showed the smallest effect because of ceiling effects. The interaction also reflects the fact that there were greater age differences on control compared to reminded items. For the reminded items, 4- to 5-year-olds (81%) and 8- to 9-year-olds (92%) significantly differed from one another, but neither group differed from the 6- to 7-year-olds (89%). However, each age group differed on the non-reminded items (at 61%, 74% and 86% for 4-5-, 6-7-, and 8-9-year-olds, respectively). There are two possible explanations for this. First, perhaps the reminders are particularly helpful to the younger children who have poorer memory for the event. Second, perhaps ceiling effects on the reminded items did not allow for a more sensitive test of age differences.

**False Recognition Items**

The number of inaccurate assents for false items was entered into a mixed factor 2 (time: initial interview vs. follow-up interview) x 2 (item type: false reminder vs. false control) x 3 (age group: 4- to 5-year olds, 6- to 7-year-olds, 8- to 9-year-olds) repeated measures ANOVA. There were main effects of item type, F(1, 49) = 82.88, p < .001, n^2 = 63, and time, F(1, 49) = 6.47, p < .02, n^2 = .12, that must be viewed in light of the significant interaction between these variables F(1, 49) = 19.00, p < .001, n^2 = .28.

Misinformation effects were found at both testing times (i.e., there were more false assents for false reminded vs. false control items). However, as was the case in Experiment 1, the size of the misinformation effects decreased from the Time 1 interview to Time 2 due to increases in false assents to the control items from Time 1 (19%) to Time 2 (40%). In contrast, there was no change in the rate of false assents to the false reminded items from Time 1 to Time 2 (57% assents at both sessions).

**Summary of Results of Experiment 2**

Analyses of the free recall data for both correct and incorrect items revealed that the children incorporated few reminders into their narratives at either the initial or follow-up interviews. The recognition data however presented a different view. Children of all ages showed significant facilitation and misinformation effects at the initial interview (with more assents to reminded than to control items) and at the follow-up interview 10 months later.

There were a few developmental trends. For the free recall analyses, the youngest children produced fewer correct utterances than the oldest children at both time periods. The youngest children also made relatively more false statements at both time periods compared to the two older groups. The oldest group of children also showed the steepest forgetting curve which may simply reflect the fact that they had the most items to forget from baseline. Importantly, there were no age trends associated with misinformation effects.

**GENERAL DISCUSSION**

The overall aim of the present studies was to examine the long-term effects of suggesting true and false information soon after children have experienced an event. With a few minor exceptions, the pattern of results and effects sizes were very similar for Experiments 1 and 2, despite between-study differences in the number of suggestive interviews, the familiarity of the follow-up memory interviewer, and the ages of the children. The major finding of both studies was that there were significant misinformation and facilitation effects on the follow-up recognition tasks.

Such effects were not observed for the free recall data. The results of Experiment 1 showed that, although the suggested true and false reminders infiltrated children’s initial reports, by follow-up very few of these appeared in their free recall. In Experiment 2, there were also very few suggestions in the follow-up free narratives, but there were also very few suggested utterances in the initial free narratives. Thus, in Experiment 2, floor effects at baseline could have accounted for the lack of suggestions in the follow-up free recall data. The major explanation for these discordant findings appears to be the fact that there were two suggestive interviews in Experiment 1 and only one suggestive interview in Experiment 2. Perhaps multiple suggestive interviews are required for children to incorporate reminders into their free recall (but see Melnyk & Bruck, 2004, for failure to find free recall differences as a function of number of interviews). Another possible explanation for the lack of reminders in children’s long-term recall is that, in both studies, children said very little at the follow-up free recall test. Perhaps a more salient event or more pressured suggestions would lead to increased free recall. Regardless of the explanation for the differences in free recall rates across studies, it is nonetheless clear that children simply provided very little information in response to free recall questions at the long-term follow-up, either suggested or not.

Due to floor effects in children’s free recall at the delayed interview sessions, the recognition questions provided our primary means of examining the long-term effects of post-event information (PEI). The presentation of true PEI continued to bolster children’s recognition memory performance after short delays (approximately 1 week after the suggestive interview) and longer delays (10–
15 months post-event). Similarly, false PEI continued to detrimentally affect children’s performance on the false recognition questions.

For the false items, a consistent pattern of results was obtained across studies. Specifically, the size of the misinformation effect (assent rates to false reminder items compared to false control items) decreased from the initial to the follow-up tests. From a practical standpoint, it is important to note that children’s overall rates of assenting to the false suggested items did not change from the initial to the follow-up interviews. Instead, children showed increased assent rates to the false items that were not reminded during the suggested interview. This same pattern was also reported in Experiment 1 of Melnyk and Bruck (2004).

There are several hypotheses to explain this pattern. The first is that children’s memory was poorer at the delayed interview test and thus they made more errors. But if this were the case, then accuracy on the true control items should have also decreased; this did not occur. The second hypothesis is that the presentation of the false control questions in the initial memory test was itself a form of misinformation that increased children’s false assents to these questions in the follow-up test. In a pilot study with 10 8- to 10-year-olds, we provide preliminary data to address this issue. The procedures were the same as those in Experiment 2 except that the delay period was 4 months, and at the delayed memory test, we included a new set of false control items that were not previously used during the initial memory test. The results revealed that the new false control items (that were presented at the follow-up period only) produced much lower false assent rates, similar to the old false control assents in the initial interview (i.e., both sets of false control items had the same assent rate upon first presentation). Thus repetition of false controls, even over a very long delay period, appears to increase the risk that they will be accepted as true. This preliminary finding is consistent with work by Brainerd and Reyna (1996) who found increased false alarm rates on delayed memory tests when participants received earlier memory tests, particularly for distracter items that overlapped in meaning with experienced events. These “mere testing effects” were found in memory tests using unrelated words and sentences from short narratives about common objects and events (also see Roediger & Marsh, 2005). The delay periods in these studies however were quite short compared to the many months in our studies.

**Developmental Trends**

By using a broader age in the sample, Experiment 2 allowed us to explore developmental trends in event memory and suggestibility. There were some, but not as many as might be predicted. Consistent with many past studies (e.g., Bjorklund et al., 2000; Lamb et al., 2003; Price & Goodman, 1990), we found a positive association between age and amount of correct free recall. At the initial memory test, children produced more correct utterances with each increasing age group. At the delayed memory test, the trend was more protracted because prior differences between the two older age groups disappeared. There were also differences in Experiment 2 in the relative amount of incorrect information in the children’s narratives. The youngest group’s narratives had a greater density of false information compared to those of the two older groups.

Importantly, we found no age differences on the suggestibility measures in both free recall and recognition. The rate of incorporation of suggested true and false reminders was the same for all age groups at both the initial and follow-up test periods. Although an inverse relationship between suggestibility and age has been reported in numerous studies (e.g., Bright-Paul, Jarrold, & Wright, 2005; Ceci, Ross, & Toglia, 1987; Chae & Ceci, 2005; Holliday, 2003), our findings are not unique. A number of other researchers have also failed to report age differences in suggestibility (e.g., Finnilä, Mahlberg, Santtilä, & Niemi, 2003). In some situations older children are more suggestible than younger children (Ceci, Papierno, & Kulikofska, 2007; Connolly & Price, 2006; Finnilä et al., 2003; Schreiber & Parker, 2004). One common factor among studies not finding age-suggestibility correlations is the mode of delivering the initial suggestions. In these studies, children were forced to provide an incorrect answer in the suggestive interviews rather than just hearing the suggestion. Thus this technique may be very powerful in eliciting false reports (as noted in Table 3, over 50% of the falsely suggested items were reported in the initial and final interviews) and thus it might serve to level age differences that are observed when less suggestive techniques are employed.

Before moving on to discuss the forensic implications of our findings, a caveat regarding the ecological validity of our study procedures is necessary. In our experiments, children were interviewed about a relatively brief positive interaction with a stranger. Though some of the reminded details involved central actions (e.g., that the child banged on a table when they did not, that the magician hugged them when she did not), some of the reminded details did involve peripheral information. One might argue our effects might be limited to non-central details about positive interactions. However, other research findings have revealed the effects of suggestive questioning are not limited to irrelevant and peripheral details of unemotional events. Children’s erroneous reports as a result of suggestive techniques include central details to negative and
painful events, such as doctor’s office and emergency room visits (Bruck et al., 2000; Burgwyn-Bales, Baker-Ward, Gordon, & Ornstein, 2001) and other forms of bodily touching (Poole & Lindsay, 1995; White, Leichtman, & Ceci, 1997). Furthermore, in our experiments, children were interviewed during short sessions with very mild suggestive techniques and, nonetheless, facilitation and misinformation effects were apparent in children’s recognition memory approximately one year following the event.

Forensic Implications

The results of our study address the issue of whether suggestive interviewing effects ever completely fade when suggestions have ceased for a long period of time. We conclude cautiously that they do in free recall if the suggestions were initially mild and if free recall was sparse in the first interview. Under these circumstances, even though there may be initial suggestibility effects in free recall, these will disappear months later and generally narratives will be very sketchy. According to Experiment 1, although the follow-up narratives will be devoid of initial false suggestions, these will be replaced by other errors that were not suggested. There are a number of actual cases where children’s allegations of abuse were associated with highly suggestive interview technique and resulted in guilty convictions for the alleged perpetrators. Years later, when questioned as young adults, some of these witnesses reported that they “knew” they were abused but could not “remember” any of the details (e.g., see Jones, 2004). Although this does not demonstrate that the initial allegations were false, the inability to recall traumatic childhood events is inconsistent with results of studies that have re-interviewed young adults who as children gave courtroom testimony about abuse. In these cases, the adults remembered significant details about the abuse (Goodman et al., 2003).

The problem is that when children provide little free recall at any interview, this may prompt interviewers to use more suggestive techniques that can either taint the child’s testimony or that can regenerate previously tainted testimony which may have been given many months previously, as we found in the two studies reported in this paper. Thus, the general guideline to primarily use open-ended questions in interviews with children (see Lamb et al., 2003; Memorandum of Good Practice, 1992; Poole & Lamb, 1998; Sterberg, Lamb, Esplin, Orbach, & Hershkowitz, 2002) pertains to all interviews regardless of the delay from prior interviews.

When evaluating the reliability of children’s previous reports, interviewers may feel less constricted in their strategies if they conclude there were no previous suggestive interviews with the child. However, this is a dangerous belief because any conversation that involves the alleged crime might be suggestive and without actual verbatim transcripts one cannot determine this fact (see Bruck, Ceci & Principe, 2006, for a review of memory for conversations). As such, open-ended questions should be relied upon as much as possible and recognition questions only asked to clarify the child’s own utterances. The best course for interviewers is to conduct forensically “safe” interviews whereby they not only rely on open-ended questions, but continually keep in mind that there may have been prior suggestive interviews. Our results clearly show the long-term detrimental effects of ignoring this practice.

REFERENCES


