Canadian Journal of Behavioural Science / Revue canadienne des sciences du comportement
2013, Vol. 45, No. 1, 1–11

0008-400X/13/$12.00 DOI: 10.1037/a0028507

The Effect of Forced-Choice Questions on Children's Suggestibility: A Comparison of Multiple-Choice and Yes/No Questions

E. M. Rocha
University of Northern British Columbia

T. A. Marche and J. L. Briere
University of Saskatchewan

Previous research has suggested that, when interviewing young children, responses to yes/no questions are less reliable than responses to multiple-choice questions (Peterson & Grant, 2001). However, according to fuzzy trace theory, some forms of multiple-choice questions should elicit higher error rates than yes/no questions. Fuzzy trace theory is a theory of cognitive development that suggests there are two types of memory traces: Verbatim traces include exact details of an experience, whereas gist traces represent the patterns and meanings extracted from that experience. Based on the assumptions of this theory, we explored the effect of question format (yes/no vs. multiple-choice), temporal delay (short delay vs. long delay) and age (4- to 6-year-olds, 7- to 9-year-olds, and 10- to 12-year-olds) on children's suggestibility for a naturalistic, potentially stressful event; namely, a dental procedure. Following the dental procedure, and again after a 6- to 8-week delay, children ($N = 68$) were given 24 forced-choice questions regarding the dental event. Consistent with fuzzy trace theory, the findings suggest that (a) multiple-choice questions can be more problematic than yes/no questions, especially after a delay, and (b) younger children are more suggestible than older children, particularly when asked "no" and "absent feature" questions. The findings are discussed with respect to implications for interviewing children.

Keywords: suggestibility, children, interviewing, memory, forced-choice questions

When children are interviewed about personally experienced events, their free recall tends to be highly accurate, even with younger children (e.g., Quas, Goodman, Ghetti, & Redlich, 2000). However, children's accounts of such events tend to be less complete than those of adults (e.g., Goodman & Reed, 1986), which often leads to the use of specific questions (e.g., yes/no questions, multiple-choice questions). Determining the most appropriate questioning techniques continues to be a challenge for researchers (e.g., Ceci & Bruck, 2006), as does understanding the effect of delay on children's recall of meaningful events. Given the lengthy delays that are common in criminal court (e.g., up to 2 years; Sas, 2002), children may be asked to answer questions about events that happened years earlier. Little is known regarding the impact of delay on children's susceptibility to specific questions. The present study used a theoretical framework, fuzzy trace theory, to make predictions regarding differences in susceptibility to suggestion across age, delay, and question type.

True and False Memory From a Fuzzy Trace Theory Perspective

One of the prominent theoretical accounts of true and false memory is fuzzy trace theory (e.g., Brainerd & Reyna, 2002; Brainerd & Reyna, 2005). According to the *fuzzy trace theory*, there are two types of memories, *verbatim* and *gist* memory traces. Verbatim traces are vivid, realistic representations of to-be-remembered information that support quick and accurate recall, whereas gist traces capture meanings and patterns that are instantiated by to-be-remembered materials. Fuzzy trace theory offers a dual-process explanation of true and false memory (Brainerd & Reyna, 2005; Lampinen, Neuschatz & Payne, 1998) that relies on five principles (Brainerd & Reyna, 1998): (a) parallel storage of verbatim and gist traces; (b) different time courses of verbatim and gist memory; (c) dissociated retrieval of verbatim and gist traces; (d) opponent judgments about false-memory; and (e) developmental variability.

In terms of the first principle, the processing and storage of verbatim and gist traces occurs concurrently, which is supported by research indicating that there is statistical and experimental dissociation between memory for presented material and false memory for material that preserves the meaning of the to-be-remembered information (e.g., Brainerd & Reyna, 2005; Reyna & Kiernan, 1994, 1995). Verbatim memories become more inaccessible with time than gist memories (e.g., Murphy & Shapiro, 1994), which means that reliance on gist retrieval, and conse-

This article was published Online First June 11, 2012.
E. M. Rocha, Department of Psychology, University of Northern British Columbia; T. A. Marche and J. L. Briere, Department of Psychology, University of Saskatchewan.
We thank Dr M. Tynan and the late Dr M. McKim for comments on the research. M. Hrabok is thanked for assistance with data collection. Appreciation is also expressed to the dentists who provided access to their patients and to the parents and children who participated. E. M. Rocha was funded by a PhD Fellowship from the Social Sciences and Humanities Research Council of Canada, and was a trainee member in Pain in Child Health, a Strategic Training Initiative in Health Research of the Canadian Institutes of Health Research. This project was part of E. M. Rocha's doctoral dissertation in clinical psychology at the University of Saskatchewan under the supervision of T. A. Marche.
Correspondence concerning this article should be addressed to Elizabete M. Rocha, Psychology Department, University of Northern British Columbia, 3333 University Way, Prince George, B.C., V2N 4Z9. E-mail: rochae@unbc.ca

quently false memory, increases with time. Consistent with these theoretical assumptions, true- and false-memories are statistically independent and experimentally dissociated when they are initially recalled but are replaced with dependency and association as time passes (e.g., Brainerd & Reyna, 1996).

In terms of opposing judgments about false memory, whereas true memory is initially based on verbatim traces and then shifts toward gist traces over time, false memory is because of gist traces (Reyna & Kiernan, 1994). True memories are typically recalled via direct access of verbatim traces, at least initially, and produce vivid, item-specific phenomenology, known as *recollection*. However, memories based on gist retrieval are reconstructed and usually produce the vague, global phenomenology that is typically called *familiarity* (e.g., Jacoby, 1991). During recollection, both retrieval operations support true memory such that events can be recalled via direct access of verbatim traces or reconstruction from gist (Brainerd & Reyna, 2002). However, with false memory, direct access and reconstruction are opponent processes, with reconstruction prompting false memory and direct access suppressing it through a process called *recollection rejection*. The theory predicts that phenomenological discrimination of true and false memories becomes more difficult as false memories become more consistent with the gist of participants' experience. It is interesting to note that in certain situations, gist-based reconstruction supports illusory vivid phenomenology that imitates verbatim-based recollective phenomenology, a phenomenon known as *phantom recollection* (Brainerd, Wright, Reyna, & Mojardin, 2001). According to the theory, when experiences repeatedly cue the same meaning, strong gist memories are formed, which can give rise to phantom recollection.

Regarding the last principle, developmental variability, both verbatim and gist memory improve during early childhood to young adulthood. Gist memory of meaningful material improves during childhood because the ability to process the meaning of experiences, as well as connect the meaning across different experiences, improves (Brainerd & Reyna, 2002; Reyna & Kiernan, 1995). Children, especially young children, are typically more susceptible to false memories than are older children and adults (see Ceci & Bruck, 1993, for a review). Given that gist traces are the basis for false memories (Reyna & Titcomb, 1997), the reason we see developmental declines in false memories, according to fuzzy trace theory, is because children's ability to store and retrieve verbatim traces, which are used to support the correct rejection of false information (Brainerd, Reyna, & Forrest, 2002; Brainerd, Stein, & Reyna, 1998), increases with age (Reyna & Brainerd, 1995; Brainerd et al., 2002; Brainerd, Reyna, & Kneer, 1995). However, in situations where children are not capable of storing and retrieving the pertinent gist traces (understanding the meaning of their experiences), false memories can increase with age, a finding known as developmental reversals (e.g., Brainerd, Reyna, & Ceci, 2008; Holliday, Reyna, & Hayes, 2002; Reyna, Holliday, & Marche, 2002).

Age and Suggestibility

Generally, younger children have been considered more susceptible than older children to misleading information (e.g., Brady, Poole, Warren & Jones, 1999; Quas et al., 2007). However, recent research shows that suggestibility effects are also present throughout the elementary school years. For example Finnilä, Mahlberga, Santtilaa, and Niemib (2003) compared children's (aged 4–8) response to a low- versus high-pressure interview that included misleading questions. They found no significant age differences in yes responses to suggestive questions; the high-pressure interview, irrespective of age, induced the most false reports. Similarly, London, Bruck, and Melnyk (2009) found no age differences in suggestibility on recognition memory tests in children aged 4 to 9 years. For example, the 9-year-old children were as likely as the younger children to not correct the interviewer's false statements. Furthermore, when these children were interviewed after a delay, they were just as likely to acquiesce to experiencing misinformation. Thus, school-aged children often are suggestible, especially with social pressure and over time, when original traces may have faded.

Reyna et al. (2002) noted that studies examining suggestibility effects across the early school years can differ considerably, leading to different results. In particular, a lack of control for children's initial level of memory may mask subtle developmental differences in memory. When techniques that control for differences in initial learning are employed, younger children are found to retain less information than older children (Titcomb & Reyna, 1995). Brainerd and Reyna (1998) argued that the inconsistencies among studies examining false recognition effects (i.e., false recognition effects sometimes decrease with age, sometimes increase with age and are sometimes age invariant) can be explained using tenets of fuzzy trace theory. For example, age differences in false recognition effects are affected by the level of specificity contained in the questions, which can have differential effects on suggestibility. Such differences are not often accounted for in suggestibility research.

Forced-Choice Questions and Suggestibility

According to Peterson and Biggs (1997) and Peterson and Grant (2001), *forced-choice* (or closed) questions refer to specific questions that have limited response alternatives and include multiple-choice and yes/no questions. Most forced choice questions are considered suggestible, in that they introduce false information and thus incorrect answers are suggested to the child. Despite recommendations otherwise (e.g., Lamb, Orbach, Hershkowitz, Esplin, & Horowitz, 2007), forced-choice questions are frequently used in forensic settings (e.g., Lamb et al., 2003; McGough & Warren, 1994), yet there has been relatively little work comparing the pros and cons of different forced-choice questions. There is a good deal of research that indicates that yes/no questions can be problematic, especially for young children. Peterson and Grant (2001) suggest that conversational dictates encourage children to answer questions and be cooperative, which leads some children to frequently say "yes" to yes/no questions (Poole & Lamb, 1998), and to change their answers to yes/no questions (Poole & White, 1991). This is especially true when children are younger (Okanda & Itakura, 2010) and the questions are incomprehensible (Fritzley & Lee, 2003). Peterson, Dowden, and Tobin (1999) reported that when children aged 3 to 5 years answered yes/no questions about traumatic injuries, they were biased toward making a "yes" response. In contrast, Peterson and Biggs (1997) found that 2- to 5-year-old children were biased toward saying "no." Fritzley and Lee (2003) evaluated the response bias for yes/no questions in children aged 2 to 5 years and found key developmental differences. When asked comprehensible questions (i.e., questions they understood), 2-year-olds displayed a consistent "yes" bias, 4- to 5-year-olds did not exhibit a response bias, and the results for the 3-year-olds were mixed. They argued that yes/no questions are appropriate for use with children aged 4, 5, and older. However some caveats should be considered. For example, 4- to 5-year-old children showed a nay-saying bias when faced with incomprehensible questions (questions they did not understand). Also, the questions asked of children concerned familiar and unfamiliar *objects*, rather than people or experiences, and, as such, the results may or may not be generalizable to questions regarding personally experienced events.

Aside from yes/no questions, there are other common question types used to interview children, although they have garnered little empirical examination. For example, two-alternative or multiple-choice questions are often used in developmental research (Fritzley & Lee, 2003), as well as in forensic child interviews (e.g., McGough & Warren, 1994; Peterson & Grant, 2001). Despite their frequent use, the only published study to compare yes/no and multiple-choice questions has been that of Peterson and Grant (2001). They questioned preschoolers about a scenario the children participated in using yes/no and multiple-choice questions. They found some difficulties with responses to multiple-choice questions because of children's tendency to use a response set (e.g., typically select the second option; Peterson & Grant, 2001). However, they argued that the overall results indicated that yes/no questions were more problematic than multiple-choice questions because of children's "yes" bias. The authors cautioned that the findings need to be replicated and conducted in more forensically relevant contexts. In addition, they did not use a theoretical foundation upon which to base hypotheses, nor was suggestibility examined. Additional studies are needed, therefore, to examine whether some forced-choice questions are more reliable than others and whether forced-choice questions affect accuracy and suggestibility similarly.

Delay and Suggestibility

The length of the delay before questioning affects suggestibility for previous experiences, sometimes increasing it (e.g., Powell & Roberts, 2002) and sometimes decreasing it (e.g., Akehurst, Burden, & Buckle, 2009). Research on what information children retain after a delay can inform our understanding of children's abilities to provide accurate testimony in legal settings. In general, research has shown age-related differences in rates of forgetting, with older children exhibiting less forgetting (e.g., Baker-Ward, Gordon, Ornstein, Larus, & Clubb, 1993; Flin, Book, Knox, & Bull, 1992).

The type of question asked has been found to influence recall after a delay (Goodman, Hirschman, Hepps, & Rudy, 1991; Poole & White, 1993). Both Goodman et al. (1991) and Poole and White's (1993) studies showed that over time (1 and 2 years), children were particularly inaccurate when responding to specific questions. Thus, children's memories decay over time, but specific questions may be particularly problematic when children's verbatim memories are no longer readily accessible.

The timing of the interview can also influence suggestibility. When 5-year-old children were exposed to misinformation both shortly after the to-be-remembered event and shortly before the memory test (6–8 weeks later), suggestibility effects were heightened (Melnyk & Bruck, 2004). Children forget verbatim information quickly (Brainerd & Reyna, 2005).

At the delayed memory test, when verbatim memory has faded, children are forced to rely on gist memory for the event (Brainerd & Reyna, 1990, 1996) and show heightened suggestibility. Some events are reportedly less affected by delays (Howe, 2000). Events that are more "distinctive" (i.e., are unique, emotionally charged, or stand out, from one's general experience and knowledge) are retained in memory for longer periods (Howe, 2000). For example, reports for medical procedures (e.g., emergency room visits) can be quite accurate 1 year after the event, at least for the gist of the experience (e.g., Goodman, Quas, Batterman-Faunce, Riddlesberger, & Kuhn, 1994; Howe, Courage, & Peterson, 1995; Merritt, Ornstein, & Spicker, 1994). Also, children who rated an event as more painful, which is presumably more salient, showed better recall for a medical procedure than did children who recalled less pain, but only after a lengthy delay (Steward & Steward, 1996). Children's long-term memory for personally experienced events has been found to be good, even after a number of years (Peterson, 2002). Thus, although free recall of event details shows low levels of suggestibility over time (e.g., London et al., 2009), children's suggestibility over time appears to be a function of a number of factors such as the child's age, type of questions asked, timing of interviews, and the salience of the event.

Current Study

Peterson and Grant (2001) argued that multiple-choice questions might be preferable to yes/no questions when interviewing children, because of children's response bias with the latter. We sought to extend the work of Peterson and Grant by including examination of delay and suggestibility and by utilizing a popular theory of children's false memory, fuzzy trace theory to guide predictions. Chae, Ogle, and Goodman (2009) argued that given the decades of scientific study on children's memory and the extensive literature that we have on this topic, the time has come to turn to theory testing, which was one of the main goals of the current study.

The current study examined the influence of three factors on children's accuracy and suggestibility for dental procedures: (a) *Age:* 4- to 6-year-olds, 7- to 9-year-olds, and 10- to 12-year-olds; and (b) *Delay*: 1 day versus 6- to 8-week delay, and (c) *Question Type*: *Yes* questions (correct answer is yes), *No* questions (correct answer is no), *Absent Feature* questions (correct answer is no and the detail probed is not related to a dental visit or is not consistent with dental gist), *Choice 1* questions (2-option multiple-choice questions in which the first choice is the correct answer), *Choice 2* questions (2-option multiple-choice questions in which the second choice is the correct answer), and *Neither Choice* questions (2-option multiple choice questions where neither of the two presented choices is correct). See Table 1 for examples of question types.

First, we predicted that older children should be more accurate, especially at initial testing, because the storage of verbatim traces increases with age. Older children should also be less suggestible, especially at initial testing, because the stronger verbatim memories can be used to reject the misleading information. Second,

Table 1
Question Type Examples

Question format	Example questions	Example response
Yes/no		
Yes	Did someone rinse your mouth with water?	"Yes"
No	Did the dentist rub toothpaste on your gum?	"No"
Absent feature	Did the dentist tickle your eyes when s/he put the eye drops in?	"No"
Multiple choice		
Choice 1	Did someone floss your teeth, or did s/he count your teeth?	"Floss teeth"
Choice 2	Did someone count your teeth, or did s/he floss your teeth?	"Floss teeth"
Neither choice	Did the dentist rub toothpaste or Vaseline on your gum?	"I don't know" or "None"

suggestibility scores should be higher at delayed testing than at immediate testing, given the greater tendency to rely on gist memories with time as verbatim memories fade. If asked about a verbatim detail, children may be likely to acquiesce to the interviewer's suggestion, if the suggestion is consonant with their gist memories.

Third, regarding question type and suggestibility effects, it was predicted that suggestibility rates would increase by question type in the following order: *Absent Feature* questions will elicit the lowest suggestibility scores, followed by *Choice 1* and *Choice 2* questions, followed by *No* questions, and then followed by *Neither Choice* questions. In *Absent Feature* questions, the information presented in the question is not related to dental gist and the misinformation is, therefore, easier to reject. *Choice 1* and *Choice 2* introduce target information and are therefore good cues for verbatim memories. If the verbatim memories are cued by the question, it should be easier for the child to reject the false choice presented. The *No* questions present a semantically related distracter, thus children would have to correctly reject information that is consistent with dental gist. *Neither Choice* questions contain two distracters that are consistent with the gist of the dental experience, and therefore require active rejection of both details, which may be a cognitively more difficult task.

Fourth, regarding question type and accuracy effects, it was predicted that accuracy rates would increase by question type in the following order: *Neither Choice* questions should elicit the lowest accuracy rates, followed by *No* questions, then by *Absent Feature* questions, then by *Choice 1 and 2* questions and, lastly, by *Yes* questions. *No* and especially *Neither Choice* questions (two gist-consistent misleading details to reject) would elicit the lowest recall rates as these questions prompt retrieval of gist traces, which do not contradict misleading details. *Choice 1* and *2* questions contain target information that is present in the question, thus verbatim memory should be cued. With *Absent Feature* questions, the false information is not similar to gist information and is easier to reject. *Yes* questions present only target information in the question, which should cue verbatim memory.

Method

Participants and Setting

Sixty-eight children and their parents (82% mothers) participated. Mean child age was 8.07 years ($SD = 2.19$, range = 4–12 years); 60% were boys. Children who participated were visiting the dentist for a range of procedures: cleanings, check-ups and diagnostic exams, fillings, extractions, and sealants. Preliminary analyses revealed that the intensity of the type of procedure experienced did not influence accuracy or suggestibility scores (i.e., cleanings and check-ups vs. restorative procedures vs. surgical procedures). The study was conducted in six dental practices serving families from both urban and rural settings and covering areas of varied socioeconomic status. Ninety-five percent of families approached agreed to participate.

Materials and Procedure

After obtaining ethics approval, parents of children who were arriving for appointments were asked whether they would be interested in participating in a study of children's responses to dental visits. Children who received medications other than local anaesthetics were not recruited to participate. Parents had the option of accompanying the child or remaining in the waiting room, according to the typical procedure of that practice. Forty percent of parents were present during the entire event; 10% were present for portions of the procedure, and 44% of parents waited in the waiting room for the entire procedure. Parents of younger children were more likely to be present during the procedure ($r = .35$; $p = .008$). No relationship between parental presence during the procedure and children's suggestibility ($r = .02$, $p = .868$) or accuracy scores ($r = -.04$, $p = .772$) was found. Data on parental presence of seven children (7%) was missing (not recorded). The dentist was asked to conduct the treatment session as usual. During the procedure, a researcher recorded the main details that occurred. Following the procedure, children were given a package of sugarless gum for participating in the study and their names were put into a draw for a prize.

Either later that day or the following day ($M = 14.00$ hr, $SD = 21.98$ hr; range = .25–96 hrs), children's memory for the procedure was examined via free recall (collected to keep the interview similar to what occurs in actual interview settings; no analysis was conducted) and 24 specific recognition questions. A list of 24 event details was generated from previous research (Vandermaas, Hess, & Baker-Ward, 1993) and after consultation with a local dentist. From these 24 event details, four questions for each of the six question types were created (e.g., 4 *no* questions; 4 *Choice 1* questions, etc.). The specific event detail was counterbalanced across question types, resulting in four versions of the 24 questions. The 24 event details involved dental staff's actions during the procedure, the persons involved, children's emotional responses, and contextual details. In each version of the 24 recognition questions, all children were presented with four questions of each of the six question formats: *Yes*, *No*, *Absent Feature, Choice 1*, *Choice 2*, and *Neither Choice.* See Table 1 for examples of the yes/no and multiple-choice questions. Children were informed, once at the beginning of each interview, that if they were unsure or did not know an answer, they were to report this.

One of four versions of interview questions was randomly assigned to each participant. Before the first interview, the specific questions were modified based on the information that the researcher recorded during the procedure. For example, if Question 1 (*Yes question*) was worded "Was the dentist's name Dr. ______," then the researcher placed the correct dentist's name in the blank. The interview was conducted as soon after the procedure as possible and, because of the broad geographic area of the participants, was conducted by telephone, with the child only. After a 6- to 8-week delay (M = 49.40 days; SD = 5.96 days; range = 40–71 days), children were telephoned at home and were asked again to freely recall the event and to answer the same 24 recognition questions (parents did not hear the questions being asked). In real-life forensic contexts, a few days, a few weeks or even a few years may pass between the time children experience an event and the time that they are questioned about that experience (Sas, 2002). We aimed to assess children's suggestibility following a 1.5- to 2-month delay to be more comparable to forensic interview situations, while keeping in mind the practical goal of minimising attrition rates. Note that the length of the delay (in days) was not significantly associated with Time 2 suggestibility scores nor with Time 2 accuracy scores. Complete data for the long delay interview are missing for six children in the sample.

Results

This study was part of a larger study on individual differences and children's memory for pain. Preliminary analyses were conducted to rule out effects of individual differences on suggestibility or accuracy. Analyses indicated that children's suggestibility and accuracy scores were not associated with their: self-reported levels of pain, temperament, or anxiety (state and trait).

The main analyses assessed the effect of question type, age, and delay on accuracy and suggestibility. Three age groups were created to achieve relatively balanced sample sizes: Group 1 (n = 24; 4–6 years; M_{age} = 5.84, SD = .73); Group 2 (n = 21; 7–9 years; M_{age} = 7.84, SD = .65) and Group 3 (n = 23; 10–12 years; M_{age} = 10.71, SD = 1.13). Responses to the 24 questions were coded as follows: (a) correct, (b) incorrect—suggestible, and (c) incorrect—not suggestible (e.g., "I don't know," "not sure," "don't remember"). To calculate *suggestibility scores*, the number of questions to which the child was led to respond incorrectly by the interviewer's suggestion (e.g., a "yes" response to "Did the dentist take your blood pressure?") was summed for each of the question types, except *Yes* questions (thus, five question types with a maximum possible score of 4 for each type). *Yes* questions were not included in the *suggestibility* scores because these questions do not introduce any incorrect information to which the child might be susceptible. For example, an incorrect response of "no" or "I don't know" to: "Did the dentist put in a filling?" does not imply a tendency to "go along with" interviewer suggestion. *Accuracy* scores were the sum of correct responses for each of the six question types (thus, 6 question types with a maximum possible score of 4 for each type). The essential difference between scoring of the suggestibility and accuracy data is that "don't know" responses are considered inaccurate, but not suggestible, as they do not imply a tendency to "go along." Also, given that responses to *Yes* questions are not considered suggestible, the suggestibility scores total a possible 20 and accuracy scores total a possible 24. Suggestibility and accuracy scores for each question type were calculated separately for age and time of interview; their means are displayed in Table 2.

The main analyses were conducted via repeated measures analyses of variance (ANOVAs). To control for inflated Type I error rates, all post hoc pairwise comparisons used the Bonferroni correction.

To determine whether children's likelihood of making suggestible responses was related to age, format of the question, or to timing of the interview, a 3 (Age Group [younger, middle, older]) × 5 (Question Type [*No, Absent Feature, Choice 1, Choice 2, Neither Choice*]) × 2 (Time [short delay, long delay]) repeated-measures ANOVA was performed on the suggestibility scores. Age was the between-subjects factor and time and question type were within-subjects factors. The main effect of age was significant, $F(2, 56) = 11.56$, $p < .001$, $\eta_p^2 = .29$. Posthoc comparisons indicated that the youngest group (4- to 6-year-olds) gave more suggestible responses ($p < .001$) than did the older two groups (7- to 9-year-olds and 10- to 12-year-olds), which were not significantly different from one another. Question format was also significant, $F(4, 224) = 68.42$, $p < .001$, $\eta_p^2 = .55$. Follow-up comparisons showed that *Neither Choice* questions received the highest suggestible response rate and were significantly different from all other question types. *No* questions were found to be significantly more susceptible to suggestion than *Choice 2* and *Absent Feature* questions (see Table 2 for means and *SD*s). There was a main effect of time, $F(1, 58) = 24.09$, $p < .001$, $\eta_p^2 = .29$. Children demonstrated significantly more suggestibility following a delay ($M = 3.98$, $SD = 3.03$) than immediately after the event ($M = .51$, $SD = 2.87$), $t(58) = -4.91$, $p < .001$.

There was also a Question Type × Age Group interaction (Figure 1), $F(8, 224) = 3.31$, $p < .001$, $\eta_p^2 = .11$. Follow-up tests indicated that 4- to 6-year-old children had more suggestible responses to *No* ($M = 1.69$, $SE = .18$) and *Absent Feature* ($M = 1.08$, $SE = .15$) questions compared with middle-aged ($M = .68$, $SE = .18$, and $M = .16$, $SE = .15$, respectively) and older-aged ($M = .46$, $SE = .16$, and $M = .32$, $SE = .14$, respectively) children ($ps < .001$). The number of suggestible responses given by 4- to 6-year-old children in response to *Neither Choice* questions ($M = 2.72$, $SE = .24$) approached significance when compared with middle-aged children ($M = 1.76$, $SE = .22$, $p = .068$) but not older-aged children ($M = 1.93$, $SE = .22$, $p = .132$). There was also a Question Type × Time interaction, $F(4, 179.89) = 3.23$, $p = .02$, $\eta_p^2 = .55$ (Figure 2). Follow-up analyses showed that, over time, children responded with more suggestibility to *Neither Choice* questions, $t(61) = 3.83$, $p < .001$.

An ANOVA was also conducted with children's accuracy scores. This analysis was conducted to compare our results to those of Peterson and Grant (2001) who used accuracy, rather than suggestibility, data. A 3 (Age Group [younger, middle, older]) × 6 (Question Type *No, Absent Feature, Choice 1, Choice 2, Neither Choice, Yes*]) × 2 (Time [short delay, long delay]) repeated-measures ANOVA was performed on the accuracy scores. All

Table 2
Means (SD) for Suggestibility Scores, Accuracy Scores, and "I Don't Know" Responses Across Question Type and Time

Variable	Age	Question type	Time 1	Time 2	Total
Suggestibility	4–6 years	No	1.56 (1.10)	1.83 (0.99)	3.60 (1.90)
		Choice 1	0.78 (0.94)	1.00 (1.08)	1.88 (1.73)
		Choice 2	0.28 (0.57)	0.61 (0.61)	0.88 (0.78)
		Neither choice	2.50 (1.25)	2.94 (0.87)	5.25 (1.94)
		Absent feature	1.11 (1.13)	1.06 (1.11)	2.21 (1.96)
		Total (max = 20)	5.71 (3.56)	7.44 (3.09)	
	7–9 years	No	0.74 (0.93)	0.63 (1.01)	1.45 (1.76)
		Choice 1	0.37 (0.68)	0.79 (0.85)	1.16 (1.34)
		Choice 2	0.42 (0.84)	0.53 (0.70)	0.95 (1.31)
		Neither choice	1.68 (1.00)	1.84 (1.07)	3.70 (1.98)
		Absent feature	0.16 (0.50)	0.16 (0.50)	0.40 (0.99)
		Total (max = 20)	3.81 (2.73)	3.95 (2.04)	
	10–12 years	No	0.41 (0.59)	0.50 (0.51)	0.91 (0.87)
		Choice 1	0.23 (0.43)	0.59 (0.96)	0.82 (1.14)
		Choice 2	0.32 (0.48)	0.45 (0.67)	0.80 (1.01)
		Neither choice	1.45 (1.22)	2.40 (1.30)	3.90 (2.32)
		Absent feature	0.27 (0.46)	0.36 (0.58)	0.64 (0.85)
		Total (max = 20)	4.09 (2.95)	5.15 (2.87)	
Accuracy	4–6 years	No	2.29 (0.99)	1.86 (1.17)	3.85 (1.76)
		Choice 1	2.71 (1.20)	2.86 (1.23)	5.41 (2.12)
		Choice 2	3.36 (0.63)	3.14 (0.77)	6.65 (1.06)
		Neither choice	1.29 (1.07)	0.50 (0.76)	1.89 (1.59)
		Absent feature	3.00 (1.11)	2.86 (1.23)	5.58 (1.89)
		Yes	3.57 (0.65)	3.36 (0.74)	6.78 (1.00)
		Total (max = 24)	16.67 (3.38)	14.57 (3.25)	
	7–9 years	No	2.76 (0.97)	3.00 (1.06)	5.70 (1.78)
		Choice 1	3.26 (0.85)	2.82 (0.88)	6.32 (1.53)
		Choice 2	3.35 (0.86)	3.24 (0.75)	6.47 (1.47)
		Neither choice	1.88 (0.86)	1.53 (1.07)	3.40 (1.79)
		Absent feature	3.71 (0.69)	3.70 (0.59)	7.37 (1.16)
		Yes	3.52 (0.62)	3.35 (0.61)	6.79 (1.08)
		Total (max = 24)	18.16 (3.10)	17.95 (2.66)	
	10–12 years	No	3.19 (0.75)	2.90 (0.83)	6.05 (1.33)
		Choice 1	3.43 (0.87)	2.90 (1.22)	6.40 (1.79)
		Choice 2	3.43 (0.98)	3.38 (0.82)	6.80 (1.70)
		Neither choice	2.19 (1.29)	1.05 (1.09)	3.33 (2.29)
		Absent feature	3.71 (0.46)	3.57 (0.68)	7.23 (0.97)
		Yes	3.62 (0.59)	3.48 (0.68)	7.14 (0.89)
		Total (max = 24)	18.17 (3.15)	16.81 (3.08)	
Don't Know	4–6 years	No	0.21 (0.41)	0.24 (0.44)	0.47 (0.62)
		Choice 1	0.29 (0.69)	0.24 (0.44)	0.59 (1.06)
		Choice 2	0.21 (0.42)	0.24 (0.44)	0.47 (0.72)
		Neither choice	0.25 (0.68)	0.53 (0.80)	0.82 (1.33)
		Absent feature	0.08 (0.28)	0.20 (0.41)	0.30 (0.57)
		Yes	0.17 (0.48)	0.00 (0.00)	0.18 (0.53)
		Total (max = 24)	1.21 (1.69)	1.35 (1.54)	
	7–9 years	No	0.43 (0.60)	0.44 (0.78)	0.94 (1.21)
		Choice 1	0.14 (0.48)	0.28 (0.58)	0.44 (1.04)
		Choice 2	0.29 (0.64)	0.18 (0.53)	0.47 (1.18)
		Neither choice	0.33 (0.58)	0.56 (0.86)	0.94 (1.16)
		Absent feature	0.10 (0.24)	0.10 (0.31)	0.20 (0.70)
		Yes	0.14 (0.36)	0.06 (0.24)	0.22 (0.43)
		Total (max = 24)	1.43 (2.06)	1.71 (2.54)	
	10–12 years	No	0.43 (0.66)	0.50 (0.83)	1.00 (1.38)
		Choice 1	0.26 (0.62)	0.30 (0.57)	0.55 (1.15)
		Choice 2	0.17 (0.65)	0.15 (0.49)	0.35 (0.99)
		Neither choice	0.43 (0.79)	0.40 (0.75)	0.80 (1.36)
		Absent feature	0.00 (0.00)	0.09 (0.29)	0.09 (0.29)
		Yes	0.13 (0.34)	0.05 (0.22)	0.05 (0.22)
		Total (max = 24)	1.35 (1.88)	1.52 (2.10)	

Note. Maximum possible score for each question type = 4.

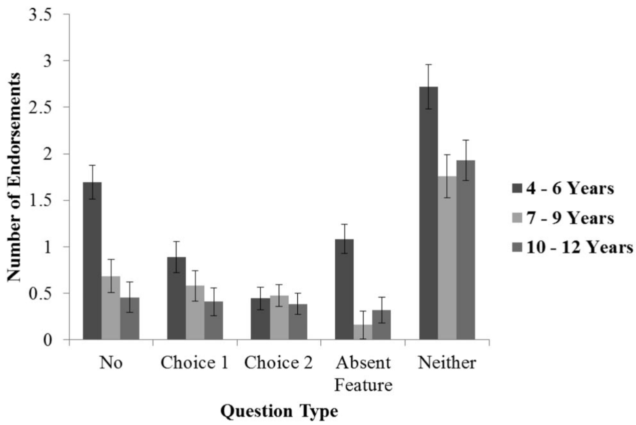

Figure 1. Mean number of suggestible endorsements for each question type across the three age groups. Error bars represent the *SEM*.

three main effects were significant: Age, $F(2, 49) = 5.86$, $p < .001$, $\eta_p^2 = .98$, Question Type, $F(5, 245) = 64.05$ $p < .001$, $\eta_p^2 = .57$, and Time, $F(1, 49) = 38.58$, $p < .001$, $\eta_p^2 = .44$, thus, the pattern of results was similar to what is reported above for suggestibility. Question Type made a difference with *Neither Choice* questions having less accurate responses than all other types (all ps $< .001$). *No* questions were responded to with less accuracy than *Yes, Choice 2*, and *Absent Feature* questions (all ps $< .001$). The youngest group of children evidenced lower accuracy scores than the older two groups of children who were not significantly different from one another (all ps $< .001$). A Question Type × Time interaction was also significant, $F(5, 245) = 3.56$, $p = .004$, $\eta_p^2 = .087$. The interaction showed that the accuracy of *Neither Choice* questions declined significantly from Time 1 to Time 2, $t(61) = 5.30$, $p < .001$.

Discussion

The present study compared children's suggestibility and accuracy rates with yes/no and multiple-choice questions for a dental procedure. We sought to extend the findings of Peterson and Grant (2001) to a personally relevant, potentially distressful event that was expected to elicit varying degrees of distress, thereby making it more comparable to forensic settings. We also investigated the impact of age and length of interview delay on suggestibility and accuracy and applied a theoretical framework, fuzzy trace theory, to guide hypotheses.

As predicted, *Neither Choice* questions elicited the most suggestible and least accurate responses. *Neither Choice* questions contain two distracters that are consistent with the gist of the dental experience; accurate responding requires active rejection of both details. It would be expected that rejecting two pieces of misinformation is cognitively difficult and would result in high error rates. However, Peterson and Grant found that *Neither Choice* questions were just as likely to elicit errors as *Choice 1* and *Choice 2* questions, and they found no response bias for answering multiple-choice questions. Given that no response bias was found, they concluded that including multiple-choice questions where neither of the stated options was correct does not seriously jeopardise children's responses. The results of the present study do not support this conclusion. *Neither Choice* questions were the most problematic, with children incorrectly responding to 58% of them. In addition, *Neither Choice* questions were significantly more problematic after a delay. Including a delay is important in applied suggestibility studies, as long delays between the experienced event and interviewing are common (Sas, 2002).

If it were possible for interviewers to confirm that the target detail, or true event, is included as one of the multiple-choice options (i.e., *Choice 1* or *Choice 2* questions), then multiple-choice questions are less problematic. However, interviewers seldom know the target detail. Thus, if a child is asked, "Did it happen in the living room or the bedroom?" and, in fact, it happened in the kitchen, then children are likely to incorrectly accept one of the provided options. While it may be tempting to offer choices to children about how things happened, the obvious risk is that they might accept the suggestion even if it is not true. They might do this because they tend to defer to adults, or because of their reliance on gist-based processing. An alternate interview approach may be to have a third response option—"Was it C*hoice 1*, C*hoice 2*, or something else?" However, this question format with children has not been empirically tested to date.

The pattern of suggestibility and accuracy results regarding *No* questions was also consistent with predictions—*No* questions were expected to elicit the next highest levels of suggestibility and lowest levels of accuracy, after *Neither* questions. Reviewing the accuracy data, a response bias was also found for yes/no questions, with children more likely to incorrectly respond to *No* questions at both interviews. Peterson and Grant (2001) argued that yes/no questions are more problematic than multiple-choice questions because of children's bias toward responding yes to these questions. They found that when the correct answer was *no*, almost half of children's responses were wrong. Although we also found a response bias, our error rates are lower. In the present study, only 26% of children's responses to *No* questions were incorrect. The higher accuracy rate in our sample may be because of our older-aged sample, the larger age range studied, and/or the salience of the dental event for the children.

Although *No* questions were more problematic than *Choice 2* and *Absent Feature* questions, they were not significantly different from *Choice 1* questions. The finding of differential error rates for

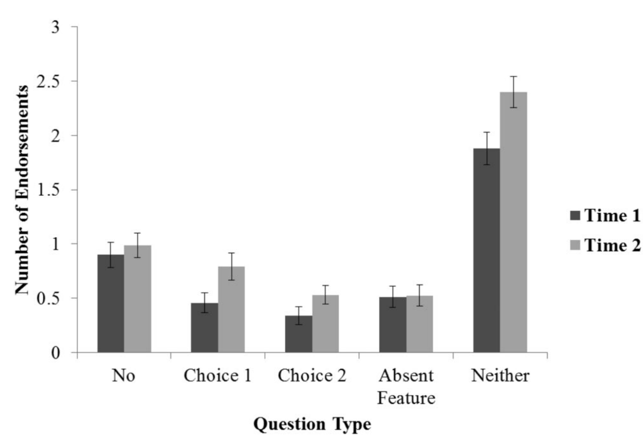

Figure 2. Mean number of suggestible endorsements for each question type at both Time 1 and Time 2.

Choice 1 and *2* questions was not consistent with predictions. Both question types contain target (or true) details and thus, according to fuzzy trace theory, their presentations were expected to trigger a verbatim memory of that detail. Our finding that children were more likely to erroneously choose the second option when the first option was correct (e.g., "Did the dentist brush your teeth or did she floss your teeth?") is, nonetheless, consistent with past research (Walker, Lunning, & Eilts, 1996). It is possible that children did not have access to the verbatim memory trace and were guessing. However we found the effect at both Time 1 and Time 2; at Time 1 children would presumably have stronger verbatim memories. Perhaps the different error rates for *Choice 1* and *Choice 2* questions are not memory-based, but are instead because of differences in response bias. For example, in a study of phenomenological differences between true and false memories, Marche, Brainerd, and Reyna (2010) investigated the effects of question format on suggestibility in adults. They found that participants were more likely to accept suggestions when they were presented in two-alternative forced-choice questions as opposed to yes/no questions. It is important to note that their results revealed a lack of consistent phenomenological differences between forced-choice and yes/no questions. The authors suggested that differences in error rates for these types of questions may not be memory-based and, instead, may be because of differences in response bias (also see Belli, 1989).

Other question-type predictions included the expectation that *Absent Feature* questions would be easy to reject for most children as they do not contain information that is consistent with dental "gist." We found that the youngest children (4–6 years) were significantly more likely to be suggestible in response to *Absent Feature* questions. Our findings are consistent with research showing that younger children have particular difficulty rejecting falsely presented information (Poole & Lindsay, 2001; Quas et al., 2007). We know that younger children tend to rely more on verbatim memories and sometime during the early elementary school years a verbatim-gist shift occurs, with children then demonstrating a gist bias (Brainerd & Gordon, 1994). Thus, if the younger children were relying on verbatim memory, which fades quickly, they may have simply guessed when exposed to questions with misinformation. The present findings speak to the importance of avoiding suggestive questioning with young children who appear more likely to accept false statements with low plausibility. Indeed, Quas et al. (2007) found that when 3- to 5 year-old children were interviewed in a neutral/nonsuggestive manner false reporting was not a problem.

Consistent with predictions and previous research (e.g., Peterson & Bell, 1996; Poole & Lindsay, 1995; Quas et al., 2007), children's overall suggestibility scores increased, while accuracy decreased, after a delay. This finding is consistent with fuzzy trace theory which predicts that children's reliance on gist memories would increase with delay because verbatim memories have faded. Thus, when questioned about a verbatim detail, children will likely acquiesce to the interviewer's suggestion if the suggestion is consistent with the gist of their memories.

We also predicted that age differences would be stronger at Time 1. Storage of verbatim traces increase with age; thus, we expected older children would retain stronger verbatim traces to reject false information. However, we did not find a significant interaction between age and time of interview. It is possible that verbatim and gist memories did not vary across the age group studied. Also, as described above, there are several recent studies that have not found age differences in suggestibility among school-aged children (e.g., London et al., 2009).

As for study limitations, for 12 of the children, the first (short delay) telephone interview was conducted more than 24 hr following the procedure, which may have increased the variability in the short delay accuracy and suggestibility scores. Second, it is possible that some children did not know the meaning of some words used in questioning (e.g., "Did the dentist take your blood pressure?"). Participants' level of language development may have rendered at least some of the questions difficult to comprehend, particularly if they were more structurally complex. For this reason, it is important for interviewers to inform children that if they are unsure or do not know an answer, they are to report this; this was done in the present study.

It is also important to acknowledge that mere-testing effects may have increased suggestibility scores from Time 1 to 2. Brainerd and Mojardin (1998) described the mere-testing effect in which true memory and spontaneous false memory are higher on a delayed memory test if participants received a prior memory test. We know that suggestive questioning reduces the accuracy of children's reports (see Ceci & Bruck, 1993, for a review). The long-held belief has been that nonsuggestive interviewing does not implant false memories and inoculates true memories from forgetting (see Brainerd & Reyna, 2002). However, Brainerd and Mojardin's results indicate that nonsuggestive questioning can substantially increase false memories. Thus, specific questions, like yes/no and multiple-choice questions, even when not deliberately suggestive, can taint children's subsequent testimony. Mere testing effects are consistent with the assumptions of fuzzy trace theory regarding developmental trends in true and false memory (e.g., Reyna & Brainerd, 1995). Whereas memory tests for to-be-remembered information improve the accessibility of verbatim traces on subsequent tests, memory tests that inquire about false items provide practice at retrieving gist traces, consequently increasing false memory reporting. When interview questions ask about experiences that did not happen but that overlap in meaning with what did, false memory creation can exceed true-memory inoculation with repeated testing (e.g., Brainerd & Reyna, 1996). Based on these findings, Brainerd and Mojardin argued that repeated interviews should be avoided given that the overall effect is to increase false memory at the expense of true memory (however, see Goodman & Quas, 2008, for a different perspective on repeated interviewing). Given that some of the questions used in the current study overlapped in meaning with what the children actually experienced, mere testing effects may have contributed to the increase in suggestibility observed in the current study from Time 1 to Time 2.

Despite such limitations, several strengths can be identified. The study was conducted in a naturalistic setting, with a delayed interview, which increases ecological validity. The children were personally affected by the details asked about in the present study (however, results may not necessarily be generalizable to forensic settings as we cannot be certain that being questioned about going to the dentist is as distressful as being questioned about child abuse). Although the interviews were conducted by telephone, research has found that accuracy levels in children's reports are comparable for in-person and telephone data collection (e.g., Baxter et al., 2003). The study allowed for both yes/no and multiple-choice questions to be systematically compared. As forced-choice questions are prevalent in investigative and courtroom interviews, the findings may add valuable information to the literature. As well, researchers have also noted that the child development research relies on interviews, often using forced-choice methodologies (Fritzley & Lee, 2003).

In sum, the present findings do not support a switch to the use of multiple-choice questions when interviewing children. The present data suggest that following a delay, which is typical in forensic interviewing procedures, *Neither Choice* questions were more problematic than *No* questions. These findings should lead us to be more careful in our approach to interviewing children, such that questions are modified to minimise the potential for suggestibility. Many trained professionals who interview alleged child abuse victims use forced-choice questions (e.g., McGough & Warren, 1994; Peterson & Biggs, 1997; Peterson & Grant, 2001), particularly with younger children. Although Brady et al. (1999) did not find question type to influence accuracy, children in their study answered questions regarding a viewed video event, which may not be comparable to a personally experienced event. Our findings, which are based on a personally experienced event, speak to the obvious difficulties with using these types of questions and they underscore the importance of avoiding them in interrogative interviewing.

Given the difficulty of eliciting the required forensically relevant information from children, especially younger children, with open-ended questions, and the susceptibility to bias and suggestibility that can arise when using closed questions, one might wonder what types of questions can be used when interviewing children. There is a wealth of research on various aspects of the interview process as well as numerous documents describing recommended guidelines and best standards. Interview protocols that emphasise the use of open-ended questions while minimising the use of specific questions, like multiple-choice and yes/no questions, can help interviewers obtain complete, forensically relevant accounts of what children experienced. To do so, Poole and Dickinson (in press) argue that investigative interviewers need to be properly trained in conversation habits and need to conduct interviews with a style that involves hypothesis-testing and a child-centered approach.

Résumé

Les recherches antérieures ont suggéré que, lorsque l'on interroge de jeunes enfants, les réponses aux questions oui/non sont moins fiables que les réponses aux questions à choix multiples (Peterson & Grant, 2001). Cependant, selon la théorie de la trace floue, certaines formes de questions à choix multiples devraient engendrer des taux d'erreurs supérieurs à ceux des questions oui/non. La théorie de la trace floue est une théorie du développement cognitif qui suggére l'existence de deux traces en mémoire : des traces verbatim incluant les détails exacts d'une expérience et des traces générales représentant les patrons et les significations extraits de cette expérience. En se fondant sur les postulats de cette théorie, nous avons exploré l'effet du format de la question (oui/non vs choix multiples), du délai temporel (délai court vs délai long) et de l'âge (4-6 ans, 7-9 ans et 10-12 ans) sur la suggestibilité des enfants au cours d'une expérience réelle, potentiellement stressante, c'est-à-dire, une procédure dentaire. Après la procédure dentaire, et ensuite après un délai de 6-8 semaines, les enfants (N = 68) devaient répondre à 24 questions à choix forcé par rapport à cette expérience chez le dentiste. Conformément à la théorie de la trace floue, les résultats suggèrent que (a) les questions à choix multiples peuvent être plus problématiques que les questions oui/non, surtout après un délai, et que (b) les enfants plus jeunes sont plus suggestibles que les enfants plus âgés, particulièrement en présence des choix « non » ou « absence de la caractéristique ». Les résultats sont discutés en fonction de leurs répercussions sur la façon d'interroger les enfants.

Mots-clés : suggestibilité, enfants, interroger, mémoire, questions â choix forcé.

References

Akehurst, L., Burden, N., & Buckle, J. (2009). Effect of socially encountered misinformation and delay on children's eyewitness testimony. *Psychiatry, Psychology and Law, 16*(Suppl. 1), S125–s136. doi:10.1080/13218710802620406

Baker-Ward, L., Gordon, B. N., Ornstein, P. A., Larus, D. M., & Clubb, P. A. (1993). Young children's long-term retention of a pediatric examination. *Child Development, 64,* 1519–1533. doi:10.2307/1131550

Baxter, S. D., Thompson, W. O., Litaker, M. S., Guinn, C. H., Frye, F. H. A., Baglio, M. L., & Shaffer, N. M. (2003). Accuracy of fourth-graders' dietary recalls of school breakfast and school lunch validated with observations: In-person versus telephone interviews. *Journal of Nutrition Education and Behavior, 35,* 124–134. doi:10.1016/S1499-4046(06)60196-6

Belli, R. F. (1989). Influences of misleading postevent information: Misinformation interference and acceptance. *Journal of Experimental Psychology: General, 118,* 72–85. doi:10.1037/0096-3445.118.1.72

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. doi:10.1207/s1532480xads0301_6

Brainerd, C. J., & Gordon, L. L. (1994). Development of verbatim and gist memory for numbers. *Developmental Psychology, 30,* 163–177. doi:10.1037/0012-1649.30.2.163

Brainerd, C. J., & Mojardin, A. H. (1998). Children's spontaneous false memories for narrative statements: Long-term persistence and mere testing effects. *Child Development, 69,* 1361–1377. doi:10.2307/1132271

Brainerd, C. J., & Reyna, V. F. (1990). Can age × learnability interactions explain the development of forgetting? *Developmental Psychology, 26,* 194–203. doi:10.1037/0012-1649.26.2.194

Brainerd, C. J., & Reyna, V. F. (1996). Mere memory testing creates false memories in children. *Developmental Psychology, 32,* 467–478. doi:10.1037/0012-1649.32.3.467

Brainerd, C. J., & Reyna, V. F. (1998). Fuzzy-trace theory and children's false memories. *Journal of Experimental Child Psychology, 71,* 81–129. doi:10.1006/jecp.1998.2464

Brainerd, C. J., & Reyna, V. F. (2002). Recollection rejection: How children edit their false memories. *Developmental Psychology, 38,* 156–172. doi:10.1037/0012-1649.38.1.156

Brainerd, C. J., & Reyna, V. F. (2005). *The science of false memory.* New York, NY: Oxford University Press. doi:10.1093/acprof:oso/9780195154054.001.0001

Brainerd, C. J., Reyna, V. F., & Ceci, S. J. (2008). Developmental reversals in false memory: A review of data and theory. *Psychological Bulletin, 134,* 343–382. doi:10.1037/0033-2909.134.3.343

Brainerd, C. J., Reyna, V. F., & Forrest, T. J. (2002). Are young children susceptible to the false-memory illusion? *Child Development, 73,* 1363–1377. doi:10.1111/1467-8624.00477

Brainerd, C. J., Reyna, V. F., & Kneer, R. (1995). False-recognition reversal: When is similarity distinctive? *Journal of Memory and Language, 34,* 157–185. doi:10.1006/jmla.1995.1008

Brainerd, C. J., Stein, L., & Reyna, V. F. (1998). On the development of conscious and unconscious memory. *Developmental Psychology, 34,* 342–357. doi:10.1037/0012-1649.34.2.342

Brainerd, C. J., Wright, R., Reyna, V. F., & Mojardin, A. H. (2001). Conjoint recognition and phantom recollection. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27,* 307–327. doi:10.1037/0278-7393.27.2.307

Ceci, S. J., & Bruck, M. (1993). The suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin, 113,* 403–439. doi:10.1037/0033-2909.113.3.403

Ceci, S. J., & Bruck, M. (2006). Children's suggestibility: Characteristics and mechanisms. *Advances in Child Development and Behavior, 34,* 247–281. doi:10.1016/S0065-2407(06)80009-1

Chae, Y., Ogle, C. M., & Goodman, G. S. (2009). Remembering negative childhood experiences: An attachment theory perspective. In J. A. Quas & R. Fivush (Eds.), *Emotion and memory in development: Biological, cognitive and social considerations* (pp. 3–27). New York, NY: Oxford University Press.

Finnilä, K., Mahlberga, N., Santtilaa, P., & Niemib, P. (2003). Validity of a test of children's suggestibility for predicting responses to two interview situations differing in their degree of suggestiveness. *Journal of Experimental Child Psychology, 85,* 32–49. doi:10.1016/S0022-0965(03)00025-0

Flin, R., Book, J., Knox, A., & Bull, R. (1992). The effect of a five month delay on children's and adults' eyewitness memory. *British Journal of Psychology, 83,* 323–336. doi:10.1111/j.2044-8295.1992.tb02444.x

Fritzley, H. V., & Lee, K. (2003). Do young children always say yes to yes–no questions? A metadevelopmental study of the affirmation bias. *Child Development, 74,* 1297–1313. doi:10.1111/1467-8624.00608

Goodman, G. S., & Quas, J. A. (2008). Repeated interviews and children's memory: It's more than just how many. *Current Directions in Psychological Science, 17,* 386–390. doi:10.1111/j.1467-8721.2008.00611.x

Goodman, G. S., Hirschman, J. E., Hepps, D., & Rudy, L. (1991). Children's memory for stressful events. *Merrill-Palmer Quarterly, 37,* 109–158. Retrieved from http://www.asu.edu/clas/ssfd/mpq/

Goodman, G. S., Quas, J. A., Batterman-Faunce, J. M., Riddlesberger, M. M., & Kuhn, J. (1994). Predictors of accurate and inaccurate memories of traumatic experiences in childhood. *Consciousness and Cognition, 3,* 269–294. doi:10.1006/ccog.1994.1016

Goodman, G. S., & Reed, R. S. (1986). Age differences in eyewitness testimony. *Law and Human Behaviour, 10,* 317–332. doi:10.1007/BF01047344

Holliday, R. E., Reyna, V. F., & Hayes, B. K. (2002). Memory processes underlying misinformation effects in child witnesses. *Developmental Review, 22,* 37–77. doi:10.1006/drev.2001.0534

Howe, M. (2000). *The fate of early memories: Developmental science and the retention of childhood experiences.* Washington, DC: American Psychological Association. doi:10.1037/10369-000

Howe, M. L., Courage, M. L., & Peterson, C. (1995). Intrusions in preschoolers' recall of traumatic childhood events. *Psychonomic Bulletin and Review, 2,* 130–134. doi:10.3758/BF03214419

Jacoby, L. L. (1991). A process dissociation framework: Separating automatic from intentional uses of memory. *Journal of Memory and Language, 30,* 513–541. doi:10.1016/0749-596X(91)90025-F

Lamb, M. E., Orbach, Y., Hershkowitz, I., Esplin, P., & Horowitz, D. (2007). A structured forensic interview protocol improves the quality and informativeness of investigative interviews with children: A review of research using the NICHD Investigative Interview Protocol. *Child Abuse & Neglect, 31,* 1201–1231. doi:10.1016/j.chiabu.2007.03.021

Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P., Stewart, H., & Mitchell, S. (2003). Age differences in young children's responses to open-ended invitations in the course of forensic interviews. *Journal of Consulting and Clinical Psychology, 71,* 926–931. doi:10.1037/0022-006X.71.5.926

Lampinen, J. M., Neuschatz, J. S., & Payne, D. G. (1998). Memory illusions and consciousness: Examining the phenomenology of true and false memories. *Current Psychology, 16,* 181–224. doi:10.1007/s12144-997-1000-5

London, K., Bruck, M., & Melnyk, L. (2009). Post-event information affects children's autobiographical memory after one year. *Law & Human Behavior, 33,* 344–355. doi:10.1007/s10979-008-9147-7

Marche, T. A., Brainerd, C. J., & Reyna, V. F. (2010). Distinguishing true from false memories in forensic contexts: Can phenomenology tell us what is real? *Applied Cognitive Psychology, 24,* 1168–1182. doi:10.1002/acp.1629

McGough, L. S., & Warren, A. R. (1994). The all-important investigative interview. *Juvenile and Family Court Journal, 45,* 13–29. doi:10.1111/j.1755-6988.1994.tb00810.x

Melnyk, L., & Bruck, M. (2004). Timing moderates the effects of repeated suggestive interviewing on children's eyewitness memory. *Applied Cognitive Psychology, 18,* 613–631. doi:10.1002/acp.1013

Merritt, K. A., Ornstein, P. A., & Spicker, B. (1994). Children's memory for a salient medical procedure: Implications for testimony. *Pediatrics, 94,* 17–23. Retrieved from http://pediatrics.aappublications.org/

Murphy, G. L., & Shapiro, A. M. (1994). Forgetting of verbatim information in discourse. *Memory & Cognition, 22,* 85–94. doi:10.3758/BF03202764

Okanda, M., & Itakura, S. (2010). When do children exhibit a yes bias? *Child Development, 81,* 568–580. doi:10.1111/j.1467-8624.2009.01416.x

Peterson, C., & Bell, M. (1996). Children's memory for a traumatic injury. *Child Development, 67,* 3045–3070. doi:10.2307/1131766

Peterson, C., & Biggs, M. (1997). Interviewing children about trauma: Problems with "specific" questions. *Journal of Traumatic Stress, 10,* 279–290. doi:10.1002/jts.2490100208

Peterson, C., Dowden, C., & Tobin, J. (1999). Interviewing preschoolers: Comparisons of yes/no and wh- questions. *Journal of Law and Human Behavior, 23,* 539–556. doi:10.1023/A:1022396112719

Peterson, C., & Grant, M. (2001). Forced-choice: Are forensic interviewers asking the right questions? *Canadian Journal of Behavioural Science, 33,* 118–127. doi:10.1037/h0087134

Peterson, C. (2002). Children's long-term memory for autobiographical events. *Developmental Review, 22,* 370–402. doi:10.1016/S0273-2297(02)00007-2

Poole, D. A., & Dickinson, J. J. (in press). Evidence supporting restrictions on uses of body diagrams in forensic interviews. *Child Abuse & Neglect.* doi:10.1016/j.chiabu.2011.05.004

Poole, D. A., & Lamb, M. E. (1998). *Investigative interviews of children: A guide for the helping professionals.* Washington, DC: American Psychological Association. doi:10.1037/10301-000

Poole, D. A., & Lindsay, D. S. (1995). Interviewing preschoolers: Effects of nonsuggestive techniques, parental coaching and leading questions of reports of nonexperienced events. *Journal of Experimental Child Psychology, 60,* 129–154. doi:10.1006/jecp.1995.1035

Poole, D. A., & Lindsay, D. S. (2001). Children's eyewitness reports after exposure to misinformation from parents. *Journal of Experimental Psychology: Applied, 7,* 27–50. doi:10.1037/1076-898X.7.1.27

Poole, D. A., & White, L. T. (1991). Effects of question repetition on the eyewitness testimony of children and adults. *Developmental Psychology, 27,* 975–986. doi:10.1037/0012-1649.27.6.975

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. doi:10.1037/0012-1649.29.5.844

Powell, M. B., & Roberts, K. P. (2002). The effect of repeated experience on children's suggestibility across two question types. *Applied Cognitive Psychology, 16,* 367–386. doi:10.1002/acp.801

Quas, J. A., Goodman, G. S., Ghetti, S., & Redlich, A. (2000). Questioning the child witness. What can we conclude from the research this far? *Trauma, Violence and Abuse, 1,* 223–249. doi:10.1177/1524838000001003002

Quas, J. A., Malloy, L. C., Melinder, A., Goodman, G. S., D'Mello, M., & Schaaf, J. (2007). Developmental differences in the effects of repeated interviews and interviewer bias on young children's event memory and false reports. *Developmental Psychology, 43,* 823–837. doi:10.1037/0012-1649.43.4.823

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7,* 1–75. doi:10.1016/1041-6080(95)90031-4

Reyna, V. F., Holliday, R., & Marche, T. (2002). Explaining the development of false memories. *Developmental Review, 22,* 436–489. doi:10.1016/S0273-2297(02)00003-5

Reyna, V. F., & Kiernan, B. (1994). The development of gist versus verbatim memory in sentence recognition: Effects of lexical familiarity, semantic context, encoding instructions and retention interval. *Developmental Psychology, 30,* 178–191. doi:10.1037/0012-1649.30.2.178

Reyna, V. F., & Kiernan, B. (1995). Children's memory and interpretation of psychological metaphors. *Metaphor and Symbolic Activity, 10,* 309–331. doi:10.1207/s15327868ms1004_5

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

Sas, L. (2002). The interaction between children's developmental capabilities and the courtroom environment: The impact on testimonial competency. Department of Justice Canada: Policy Centre for Victims Issues. Retrieved from http://publications.gc.ca/collections/Collection/J3-2-2002-6-1-1E.pdf

Steward, M. S., & Steward, D. S. (1996). Interviewing young children about body touch and handling. *Monographs of the Society for Research in Child Development, 61,* 4–5. doi:10.2307/1166205

Titcomb, A. L., & Reyna, V. F. (1995). Memory interference and misinformation effects. In F. N. Dempster & C. J. Brainerd (Eds.), *Interference and inhibition in cognition* (pp. 40–46). San Diego, CA: Academic Press. doi:10.1016/B978-012208930-5/50009-X

Vandermaas, M. O., Hess, T. M., & Baker-Ward, L. (1993). Does anxiety affect children's reports of memory for a stressful event? *Applied Cognitive Psychology, 7,* 109–127. doi:10.1002/acp.2350070204

Walker, N. E., Lunning, S. M., & Eilts, J. L. (1996, June). *Do children respond accurately to forced-choice questions? Yes or no.* Paper presented at the NATO Advanced Study Institute on Recollections of Trauma: Scientific Research and Clinical Practice, Talmont Sain Hilaire, France.

Received July 20, 2011
Revision received March 26, 2012
Accepted March 26, 2012 ■