Published in final edited form as:

J Exp Psychol Appl. 2009 December ; 15(4): 361–374. doi:10.1037/a0017120.

Human Figure Drawings and Children's Recall of Touching

Maggie Bruck

Division of Child and Adolescent Psychiatry, Johns Hopkins University

Abstract

In 2 studies, children ages 3 to 7 years were asked to recall a series of touches that occurred during a previous staged event. The recall interview took place 1 week after the event in Study 1 and immediately after the event in Study 2. Each recall interview had 2 sections: In 1 section, children were given human figure drawings (HFDs) and were asked to show where the touching took place; in the other section, the same questions were asked without the HFDs (verbal condition). Children were randomly assigned to 2 different conditions: HFD 1st/verbal 2nd or verbal 1st/HFD 2nd. There were 2 major findings. First, HFDs elicited more errors than the verbal condition when used to probe for information that the child had already been asked. Second, regardless of interview method, children had poor recall of the touches even when these occurred minutes before the interview. It is suggested that cognitive mechanisms involving memory and semantics underlie children's poor recall of touching in both verbal and HFD conditions.

Keywords

memory; forensic interviews; symbolic development

Human figure drawings (HFDs) are commonly used by professionals who interview children about suspected sexual abuse. It is assumed that these drawings will decrease children's linguistic and emotional or motivational limitations, as well as memory problems, and thus will result in the elicitation of more complete and accurate details of abuse. There is, however, little scientific information to support claims of their benefits. This article presents the results of two studies that examined young children's ability to use HFDs to report body touches.

Asking young children to provide complete and accurate reports of their past is problematic for several interrelated reasons. When children are asked open-ended questions, (e.g., "Tell me what happened"), their responses tend to be sparse although accurate when there is no source of memory contamination, such as suggestive questions (Baker-Ward, Gordon, Ornstein, Larus, & Clubb, 1993; Peterson & Bell, 1996; Steward & Steward, 1996). To elicit more details, interviewers often resort to a number of different suggestive conversational strategies; although these techniques can elicit more details, they also generate more errors (e.g., Brady, Poole, Warren, & Jones, 1999; Ceci & Bruck, 1995; Peterson & Bell, 1996; Peterson & Biggs, 1997).

This pattern also describes the accuracy of children's reports about touching that occurs during medical examinations or laboratory-staged events. Specifically, in nonsuggestive contexts, children (ages 3 to 8 years) provide accurate but sparse details of touching, but when asked more specific and suggestive questions, errors increase (Baker-Ward et al., 1993; Krackow & Lynn, 2003; Merritt, Ornstein, & Spicker, 1994; Pezdek & Roe, 1997; Saywitz, Goodman,

Nicholas, & Moan, 1991). Although many children err because they falsely acknowledge touching that did not occur (commission errors), they more commonly deny or fail to report actual touches (omission errors). Major explanations for this failure to provide accurate touch information include emotional or motivational factors (embarrassment), lack of appropriate linguistic terms, or memory problems (Quas, Davis, Goodman, & Myers, 2007).

To overcome memory, motivational, and language problems, some interviewers have used anatomically detailed or regular dolls when questioning children about body touches (see American Professional Society on the Abuse of Children, 2002). It has been found, however, that children ages 3 to 6 years produced more errors when they used anatomically detailed or regular dolls to demonstrate past events (Bruck, Ceci, & Francoeur, 2000; Bruck, Ceci, Francoeur, & Renick, 1995; DeLoache & Marzolf, 1995; Gordon, Ornstein, Nida, & Follmer, 1993; Greenhoot, Ornstein, Gordon, & Baker-Ward, 1999; Steward & Steward, 1996). The dolls decrease performance for two major reasons. First, children under the age of 5 may not have achieved the cognitive insight that the doll is a symbol (representation) of their own body (DeLoache & Marzolf, 1995). Second, the novelty of the dolls (especially those with anatomical parts) prompts children to play rather than to use the dolls to demonstrate actual events (e.g., Bruck et al., 1995, 2000).

To avoid some of the problems posed by the dolls, HFDs are often included in interviews with children in cases of suspected sexual abuse (Holmes & Vieth, 2003). Children are presented with line drawings of a front and back view of a same-gender unclothed child, and sometimes they are shown line drawings of an unclothed adult of the same sex as the suspected perpetrator. In some protocols, children are asked to name the body parts (including the genital and anal regions) and then to point to the areas where touching occurred, or the interviewer may point to a region and ask if it was touched.

The advantage of HFDs versus dolls is that HFDs invite less exploration and play, thus possibly reducing errors. Furthermore, from a developmental perspective, 3- and 4-year-old children have mastered many of the skills needed to use the drawings. They can understand a picture as a symbol and use it to solve problems (DeLoache & Burns, 1994; Preissler & Carey, 2004; Suddendorf, 2003). Also, they have linguistic labels for most body parts (Witt, Cermak, & Coster, 1990; MacWhinney, Cermak, & Fisher, 1987), and they have a spatial representation of these parts (Johnson, Perlmutter, & Trabasso, 1979).

Although the existing studies indicate that by 4 years of age children have acquired many of the prerequisite skills to use HFDs, it is possible that children may not be able to use HFDs as symbols of their bodies at the same ages that they can symbolically represent objects. Specifically, HFDs may pose difficulty because they are not iconic with the child's body, making them a more abstract symbol than dolls; drawings are two-dimensional symbols of three-dimensional objects. Because even children younger than 9 years of age have difficulty with the manipulation and integration of images of objects over space and time (Enns & Girgus, 1986; Kail, Pelligrino, & Carter, 1980; Kosslyn, Margolis, & Barrett, 1990), they may not grasp the nature of the drawing as a symbol.

Although HFDs are commonly used in forensic and clinical interviews (Carnes, Nelson-Gardell, Wilson, & Orgassa, 2001; Conte, Sorenson, Fogarty, & Rosa, 1991; Kendall-Tackett & Watson, 1992) and are recommended for use in a number of professional guidelines (American Academy of Child and Adolescent Psychiatry, 1997; American Professional Society on the Abuse of Children, 2002; Holmes & Finnegan, 2002), until recently there have been very few studies to examine their feasibility. The existing studies are now reviewed.

Steward and Steward (1996) interviewed children ages 3 to 6 years about their prior medical procedures. Children interviewed with HFDs provided more accurate details, but also more errors about genital and anal touches than did children who were interviewed without the HFDs.

Aldridge and colleagues (2004) examined the degree to which HFDs increased the number of details that children (4 through 13 years of age) provided about sexual abuse during the formal National Institute of Health and Child Development protocol interview. This protocol emphasizes the use of open-ended questions and prompts and discourages the use of specific and suggestive leading questions to elicit information (Lamb, Orbach, Hershkowitz, & Esplin, 2008). In Aldridge et al. (2004), children who had provided information about abuse (touches) in this (verbal) standard protocol were then presented with HFDs and asked a number of questions about being touched and about touching others. Aldridge et al. concluded that the HFDs were beneficial because children produced a large number of forensically relevant details with the HFDs that had not been provided in the first part of the interview. This was particularly so for the youngest children (ages 4 to 7 years) whose recall increased by 27% with the drawings.

There are two major difficulties with the interpretation of these results. First, because of the nature of the crime (sexual abuse), the accuracy of the children's statements could not be validated. Although the authors favored the position that HFDs encouraged children to provide forensically valuable information, it is also possible that this additional information was false. Thus, it is crucial to undertake studies in which children's statements can be validated. Second, it is possible that the presentation of the HFDs resulted in additional information because the children were given a second opportunity to recall the event rather than because of any inherent characteristics of the HFDs.

There have been two subsequent laboratory studies to address the validity of children's reports about previous touches. Willcock, Morgan, and Hayne (2006, Study 1) had an adult touch children (5 and 6 years of age) in various places while helping them dress for a class trip. One month later, each child was shown a same-gendered clothed cartoon drawing and asked to show where she/he had been touched. Performance was poor; on average, only 38% of the touches were reported, and 47% of all touches reported were incorrect. The procedure was repeated (Study 2) with children either being interviewed immediately after, 1 day after, or 1 month after the touching. Although the immediate group produced the most correct touches, their performance was still poor (49% of all touches were recalled). Error rates did not differ among groups (these ranged from 36% to 55%). The authors concluded that children's poor performance reflected their failure to understand the representational nature of drawings. However, because there was no comparison condition in which children were asked about the touching without drawings, these conclusions are not valid.

Brown, Pipe, Lewis, Lamb, and Orbach (2007) remedied this problem by including a comparison condition. Children (5 through 7 years of age) were visited by a photographer, who dressed the child as a pirate and then took the child's picture. About 1 month later, the children were asked about the touching during the visit; children were either questioned with gender-neutral drawings of an unclothed child and adult or without the drawings. Children in the drawing condition produced more false reports of touching than did children in the no-drawing condition. Condition did not influence reports of correct touches; children correctly reported a minority of the touches (approximately 35%).

The two studies described here also examined benefits (i.e., increases in accurate details) and risks (increases in errors) associated with HFDs; in addition, they contain a number of features not included in the previous studies. First, motivated by related methodological and applied issues, the present studies were designed to compare HFD with no-HFD conditions in a within-

and between-groups design. In forensic interviews, there is no agreement on the introduction of HFDs: Sometimes they are introduced at the beginning of the interview; other times they are introduced later. Risk and benefits associated with HFDs may be a function of their particular sequence. Aldridge et al. (2004) used only one sequence: no-HFD first (full verbal protocol) and HFD second. To determine whether the increase in details in the HFD section of that study was due to the HFDs rather than to repeated or additional interviewing, it is necessary to vary the order of the HFD and no-HFD components of the interview. In the present studies, children were randomized to a verbal 1st/HFD 2nd interview sequence (the one most similar to Aldridge et al.) or to a HFD 1st/verbal 2nd interview sequence. If the added information obtained in the Aldridge et al. study was simply because of an additional opportunity to provide information, then just as much new information should be obtained in the verbal 2nd portion as in the HFD 2nd portion of the protocols. Furthermore, if HFDs increase the number of details, then this advantage should be seen in the HFD 1st/verbal 2nd and the verbal 1st/HFD 2nd interview sequences. Finally, these comparisons will also yield information about the best sequence to use in forensic interviews. Although there are other potential control conditions to examine the effects of repeated interviews (e.g., verbal 1/verbal 2 or HFD 1/HFD 2), their inclusion would not address the major objectives of the new studies, which were designed to address the risks and benefits of using HFDs as supplementary tools to verbal interviews with young children. This point is elaborated in the Discussion section.

A second novel feature of the present studies is the inclusion of 3- and 4-year-old children. This age group is of great interest because they provide so little information in interviews and consequently could benefit most from extra assistance, such as HFDs. On the other hand, because of incomplete understanding of symbols and other cognitive immaturities, this age group might have the greatest difficulty reporting touch with HFDs.

Third, Study 1 was designed to examine the risks and benefits of repeating mildly suggestive questions with HFDs versus no HFDs. In some studies, when children were told that they made mistakes and should try again, they changed their answers on the repeated round of questions (e.g., Cassel, Roebbers, & Bjorkland, 1996; Poole & White, 1991). Study 1 examined the degree to which this type of questioning might produce more errors for HFDs compared with no HFDs.

Fourth, the cognitive characteristics associated with performance in the HFD and no-HFD conditions were examined. One prediction was that children with relatively poor knowledge of body parts would benefit most from the HFDs. In Study 2, it was predicted that children with relatively poor linguistic skills and better visual spatial skills would benefit most from the HFD conditions. A measure of visual spatial skills was included on the assumption that successful use of HFDs requires a mapping of a spatial representation of one's own body onto to a spatial representation of a drawing.

Fifth, Study 2 was designed to examine some factors that might account for children's relatively poor performance in reporting touch. As shown by Willcock et al. (2006) and Brown et al. (2007; and to preview the results of Study 1 of this article), children reported very few touches in all conditions. The effects of memory load (the number of touches to be remembered) and of semantic representation of the word *touch* on children's recall of touching were examined.

Study 1

Method

Design—During a staged magic show, the child touched the magician 3 times and the magician touched the child 4 times. One week later, a female research assistant interviewed the child about the magic show. All children were asked open-ended questions about the magic show. Then, the children were questioned about touching either with the verbal protocol first

and with the HFD protocol second (verbal 1st/HFD 2nd) or with the HFD protocol first and with the verbal protocol second (HFD 1st/verbal 2nd). For all children, a section of the second protocol was repeated. At the end of the interview, children's comprehension of body part names was tested.

Participants—Fifty-eight children, ages 3 to 7 years, were tested in local day-care, after-school, and summer programs. There were 12 children each in the 3-, 4-, and 5-year-old age groups and 11 children each in the 6- and 7-year-old age groups. Gender was equally distributed (45% girls) in this racially diverse sample (62% Caucasian, 24% African American, 14% other). Children came from well-educated families (70% of the mothers had completed college or obtained a graduate degree). English was the primary language of all participants. Parental consent was obtained for all children.

Magic Show—Each child individually participated in a 10-min magic show. A magician dressed the child in a magic cape, checked out the magic wand, performed two tricks, hurt her arm on the table, and gave the child a sticker. During this scripted event, the magician touched the child 4 times (e.g., “To make this trick work, I have to pull your ear”), and the child touched the magician 3 times (e.g., Magician: “My back is itchy, can you scratch it for me?”).

Two magic shows were created, each with a different set of body part touches. These versions were counterbalanced across age groups and interview sequences. The body parts and instruments used in each interview version are presented in Appendix A.

Interview—One week after the magic show, an unfamiliar female research assistant interviewed each child individually. The interview consisted of five main sections.

Event recall of magic show: The child was asked, “Tell me as much as you can about the magic show.” This was followed by five open-ended prompts asking the child to provide more information (e.g., “What happened next?”). For the data analysis, the children's responses were segmented into utterances. An utterance is a statement bound by pauses containing one verb. For example, “The magician waved a wand, and we held hands” contains two utterances. Each utterance was classified as correct or incorrect. Interrater reliability was calculated for 20% of the protocols in both Studies 1 and 2. There was perfect agreement on the number of incorrect utterances and high interrater reliability for the accurate responses, $r(34) = .98$.

Stratifying for age, children were randomly assigned to a verbal 1st/HFD 2nd or to a HFD 1st/verbal 2nd sequence.

Verbal interview about touching: The interviewer asked, “Did the magician touch you?” If the child assented, the interviewer asked, “Where?” The child was asked to provide more information (e.g., “Tell me about it”). This sequence of questions was repeated until the child stopped providing information. Then, a similar line of questions was asked about the child's touching the magician. The total number of accurate and inaccurate details about the body parts and the instruments of touching were counted. In addition, errors of commission (e.g., “He hugged me” when there was no hugging) were added to the inaccurate detail score.

Next, the child was asked 20 yes–no recognition questions. Seven questions included accurate information about the target touches. Seven foil questions included incorrect information about touching. The foil questions were the accurate questions from the alternate magic show. Thus, the interview schedules were identical for children who attended the two magic shows; the only difference was that a touch question could be an actual touch for one child but a foil for another. If a child assented to any of the 14 questions, the interviewer asked, “What did the magician touch you/you touch the magician with?” There were also six filler questions

interspersed among the 14 touch questions about non-touching events (e.g., “Did the magician give you a magic wand?”). Correct responses were “yes” for half the items and “no” for the other half. Total correct recognition scores were calculated by adding correct yes target responses and correct instruments. The number of correct no answers to the seven foil questions was calculated.

HFD interview about touching: The interviewer showed the participant drawings of the front and back view of a same-gendered child (see Appendix B, e.g., of female drawings). The child was asked to point to the mouth, neck, and heel of the drawing. Then, the same procedures as just described for verbal interviews were followed except the child was told, “Point to where the magician touched you with this magic marker.” After each mark, the child was asked to give more information. The child was then shown drawings of the front and back view of a big person and was told that this person looks like the magician. The same procedures were followed as just explained for the child drawings. The scoring was the same as described for the verbal prompt questions.

Next, 20 yes–no recognition questions were asked using the verbal recognition format, except rather than naming body parts, the interviewer pointed to the body part on the HFD and asked, “Did the magician touch you/you touch the magician here?” Children who assented were asked, “What did the magician touch you/you touch the magician with?” The same list of questions and the same scoring methods were used for the verbal and HFD data.

Repeated recognition questions: After the second round of recognition questions, the interviewer said that she had been talking to other kids who saw the magic show, and she thought that the child had missed a few things. She said she was going to ask some questions again, and the child just had to answer yes or no. The interview version for these repeated recognition questions was the one that the child had just been given. For the purpose of these analyses, children were classified as HFD 2/HFD 3 or verbal 2/verbal 3.

Body part comprehension: At the end of the session, the children were asked to touch 14 different parts of their body; these included the body parts that were used in the interviews.

Results

Preliminary Analyses—There were no significant correlations in the predicted direction between parents’ educational levels and each of the major outcome variables after partialing out the effects of age.

Analyses of variance (ANOVAs) were carried out to examine the effect of gender (as a function of interview sequence and age) on each of the major outcome variables. There was only one effect: an Age $\times$ Gender interaction for accurate prompt details in the first round of questioning, $F(1, 40) = 4.54, p < .05, \eta_p^2 = .24$. Five- and 7-year-old girls provided more details in the first round of questioning compared with age-matched boys. It is important to note that, because there was no interaction with interview sequence and because this was the only variable to show a gender effect, gender was not included in the remaining analyses.

The variable of magic show version (A vs. B) was not significant, nor did it interact with the variables of age or interview version in any of the analyses. Thus, this variable was removed from all statistical models.

To ensure that there were no baseline differences in recall of the magic show prior to the randomization of the interviewing sequences, I carried out a 2 (interview sequence: verbal 1st/HFD 2nd vs. HFD 1st/verbal 2nd) $\times$ 5 (age level) $\times$ 2 (utterance type: accurate vs. inaccurate)

ANOVA with repeated measures on the last variable of utterances produced in response to the first open-ended question and subsequent prompts, “Tell me everything you can remember that happened in the magic show.” There were no main effects or interactions involving the variable of interview sequence. Thus, any subsequent significant effects involving interview sequence cannot be attributed to prandomization baseline differences.

Finally, the number of touches that children reported in their free recall was examined. Only 3% ($n = 15$) of all utterances contained touching information; two of the 15 touch utterances were incorrect. Most children reported no touches.

Prompted Recall for Touches—Accurate and inaccurate details were analyzed separately by means of an Age (five levels) $\times$ Interview Sequence (HFD 1st/verbal 2nd vs. verbal 1st/HFD 2nd) $\times$ Round of Questioning (first round vs. second round) ANOVA with repeated measures on the last two variables (see Table 1).

Accurate details: There was a main effect of age, $F(4, 48) = 3.00, p < .05, \eta_p^2 = .20$. Results of planned comparisons revealed that the 7-year-old children produced more accurate details than all younger children who performed similarly. As shown in the last row of Table 1, children provided very little accurate information (percentages are presented in Table 1 to make the data comparable with those presented in Study 2). There were no other main effects or significant interactions.

Inaccurate details: There were significant interactions of Round of Questioning $\times$ Age, $F(4, 48) = 2.82, p < .05, \eta_p^2 = .19$, and Round of Questioning $\times$ Interview Sequence, $F(1, 48) = 13.10, p < .01, \eta_p^2 = .19$. These were modified by a significant Round of Questioning $\times$ Age $\times$ Interview Sequence interaction, $F(4, 48) = 3.12, p < .05, \eta_p^2 = .21$.

To highlight developmental differences and to directly compare HFD with verbal errors in each round of questioning, I conducted two parallel analyses. In the first, I conducted an Interview Version (HFD 1 vs. verbal 1) $\times$ Age (five levels) ANOVA on the errors produced in the first round of questions. This yielded an Age $\times$ Interview Version interaction, $F(4, 48) = 3.36, p < .05$. Planned comparisons revealed that the 5-year-olds made more errors on HFD 1 versus verbal 1, whereas there were no interview differences for the other children. Also, the 5-year-olds made more errors on HFD 1 than all other age groups. There were no age differences for the verbal condition. For the second analysis, a 2 (interview version: HFD 2 vs. verbal 2) $\times$ 5 (age level) ANOVA conducted on Round 2 errors produced an interview version effect, $F(1, 48) = 5.2, p < .05$. When interviews were repeated (i.e., Round 2), HFDs produced more errors than verbal interviews for all age groups.

New details: The number of details produced for the first time in the second round of questioning was examined. I carried out 5 (age level) $\times$ 2 (interview version) ANOVAS on new accurate details and on new inaccurate details. There were no significant effects for new accurate details, which were relatively infrequent (approximately .5 details per child). For the new inaccurate details, however, there was a main effect of interview version, $F(1, 48) = 9.87, p < .01, \eta_p^2 = .17$. In the second round of questioning, HFDs elicited significantly more new errors ($M = 1.69, SD = 2.07$) than did verbal interviews ($M = 0.48, SD = 0.79$). There was a significant main effect of age, $F(4, 48) = 2.77, p < .05, \eta_p^2 = .18$. The most new errors were made by the 3- (2.3) and 5-year-old children (1.3). The error rates were lower and similar among the remaining three age groups (ranges = 0.6–0.7).

Recognition Tests—The first set of analyses of the recognition test compared children's performance on foil and target items in the first and second rounds of questioning; these analyses compared children's performance on verbal and HFD conditions. The responses to the seven foil items (questions that contained incorrect information) were examined by means of an Age $\times$ Interview Sequence $\times$ Round of Questioning ANOVA with repeated measures on the last variable. The dependent variable was the number of correct rejections of the foil questions (see Table 2, where percentages are presented to make the data comparable with Study 2). There was a significant main effect of age, $F(4, 48) = 2.97, p < .05, \eta_p^2 = .20$. The 3-year-old children rejected fewer items than the 5-, 6-, and 7-year-olds. All other comparisons were nonsignificant. With the exception of the 3-year-olds, performance on the foils was very high. The parallel analysis of the number of correct responses to the target items (correct yes responses + correct instruments) did not produce any significant effects or interactions.

Next, to determine whether repeating recognition questions using the same interview version affected children's performance similarly for HFD and verbal conditions, I compared performance on the second and third rounds of questioning by means of a 5 (age level) $\times$ 2 (interview condition: verbal vs. HFD) $\times$ 2 (round of questioning: 2 vs. 3) ANOVA with repeated measures. For the foil items, there was only a main effect of age, $F(4, 48) = 3.48, p < .01, \eta_p^2 = .22$. The 3-year-old children performed the worst, but there were no other age differences. For the target items, there was a significant effect of round of questioning, $F(1, 48) = 3.99, p < .05, \eta_p^2 = .08$. Repeating questions in the third round decreased accuracy from the second round. There were no effects involving interview condition.

Comprehension of Body Parts—There were no age differences in children's identification of 14 body parts. The overall accuracy rate was 92%. There were no significant correlations between body comprehension scores and each of the major dependent variables, probably due to ceiling effects of the body comprehension scores.

Discussion

There were two major findings. First, HFDs had a major effect on errors for prompted recall when they were introduced after a full verbal interview. Consequently, children provided more new false information with the HFDs. Interview version and sequence did not influence the responses for the accurate prompt details or for the recognition data. Second, both the recognition and prompt recall data indicate that regardless of age, interview protocol, or round of questioning, the children performed very poorly. They recalled few touches even though they had good comprehension of the body part names.

In Study 2, three major hypotheses were tested to account for the children's poor performance. First, perhaps they had "forgotten" about the touches during the 1-week delay. In this case, testing immediately after the magic show should increase performance. Second, perhaps there were too many touches to remember, and thus they never were processed or retained. In this case, reducing the number of touches in the magic show should improve performance. Finally, perhaps the children did not associate the action that involved touching with the word *touch* but with another associated action (e.g., rub, tap, scratch).

Study 2

Method

Design—The design was similar to that of Study 1 with the following major changes. Children were interviewed immediately after the magic show. Children were randomly assigned to a four-touch or a six-touch magic show. The recognition questions in the third round of

questioning were omitted because of their limited effect in Study 1. To explore various hypotheses regarding children's failure to acknowledge touching, at the end of the interview, the interviewer prompted children who denied actual touches in various ways to recall touching.

Finally, to examine individual differences in reporting on the HFD and verbal protocols, the interviewer gave children a vocabulary and a spatial perception task. The guiding hypothesis was that children with poor vocabulary but good spatial skills would benefit most from the HFD protocols, whereas children with poorer spatial skills would benefit most from the verbal protocols.

Participants—One hundred children, ages 3 to 7 years, were tested in local day-care, after-school, and summer programs. Gender was equally distributed (49% girls) in this racially diverse sample (51% Caucasian, 38% African American, 5% Asian, 6% other). Children came from well-educated families (75% of the mothers had completed college or obtained a graduate degree). English was the primary language of all participants. Parental consent was obtained for all children.

There were 20 children in each age group. At each age level, children were randomly assigned to a four-touch magic show or a six-touch magic show. Stratifying for age and touch condition, children were randomly assigned to a verbal 1st/HFD 2nd protocol or to a HFD 1st/verbal 2nd protocol.

Magic Show—The Study 1 magic show versions were used, with the following changes: Instead of seven touches, there were either six or four touches, half of the touches were made by the magician and half of the touches were made by the child (see Appendix A). The touches were made more explicit by verbal elaboration (e.g., “First, I need to measure your wrists. I’m measuring your wrists really loosely. Now, I’m touching your wrists really tight. I touched your wrists and you are a big boy”).

Touching Interviews—The same format used in Study 1 was used in Study 2 with the following change. Instead of repeating the recognition questions for a third time, the child was probed about actual touches that they had not reported or that they had denied in both the HFD and verbal sections of the interview. For example, if the child failed to report the detail that the child touched the magician's chin with a wand, the child was asked, “Did the magician do something to make his wand work? Tell me about it?” If the child failed to answer, the second prompt was, “Did you do something with the wand on her face? Tell me about it?” If the child provided the correct information, then he or she was asked, “Before I asked you if you touched the magician's chin and you said you didn't. So when you did something with the wand on her face, did you touch her?” “Why did you not tell me about the touching before?” To investigate whether the children's previous failure to report was due to not understanding the meaning of *touch*, I categorized their responses to these questions as (a) there was no touching (denial), (b) forgot to tell before, (c) the reported action was not the same as touching (e.g., “I rubbed her toes, but that is not the same as touching”), and (d) reported wrong body part (e.g., “I touched her knee, not her cheek”).

Procedure—A female interviewer tested all children immediately after the magician left the room. After the touching interview, the child was asked to touch on his or her own body the 12 body parts that were mentioned in the two versions of the magic show. Correct responses served as a measure of body comprehension.

Finally, the Vocabulary and Block Design subtests of the revised Wechsler Preschool and Primary Scale of Intelligence (Wechsler, 1989) were given to the 3-, 4-, and 5-year-olds and

the same subtests of the fourth edition of the Wechsler Intelligence Scale for Children (Wechsler, 2003) were given to the 6- and 7-year-olds.

Results

Preliminary Analyses—The effects of gender, parent education, and magic show version on the major outcome variables were examined following the procedures used in Study 1. There was only one effect: an Age $\times$ Gender interaction for the percentage of accurate foils in the first round of questioning ($p < .05$). Thus, these variables were excluded from the remaining analyses.

As was the case for Study 1, children in the two interview sequences did not differ in their number of spontaneous utterances they produced in response to the open-ended prompts about the magic show. Further analysis of this data set revealed that only 6% of all the utterances involved a “touching” action. Four of the 40 touch utterances were incorrect.

Prompted Recall

Accurate details: The number of accurate touches and instruments produced in response to the open-ended questions, “Did you touch the magician/the magician touch you?” “Tell/show me where,” and “Was there anywhere else?” were expressed as proportion of the total number of actual touches and instruments; for the four-touch group, the denominator was 8 (4 touches + instruments) and for the six-touch group the denominator was 12 (6 touches + 6 instruments).

A 5 (age level) $\times$ 2 (interview sequence) $\times$ 2 (touch version: four vs. six) $\times$ 2 (round of questioning) ANOVA with repeated measures on the last variable was carried out on the percentage of accurate details (see Table 3). Age was the only main effect, $F(4, 80) = 26.21$, $p < .01$, $\eta_p^2 = .57$. This was modified by a significant interaction of Touch Version $\times$ Age, $F(4, 80) = 2.99$, $p < .05$, $\eta_p^2 = .13$. As shown in Figure 1, only the 7-year-olds showed the predicted effect that lessening the memory load would result in improved performance. The interaction also was obtained because there were fewer developmental differences in the heavy compared with the light memory load condition. There was also a Round of Questioning $\times$ Age interaction, $F(4, 80) = 5.89$, $p < .01$, $\eta_p^2 = .23$. More items were recalled on the second round of questioning by the 6- and 7-year-old children only, whereas all other age groups recalled the same number of items in the first and second rounds of questioning (see Figure 2). Because of this increase by the older children, developmental differences were more attenuated in the first recall as compared with the second recall.

Thus, the touch manipulation influenced the performance of the 7-year-old children alone and only the 6- and 7-year-old children provided more information in the second round of questioning compared with the first round of questioning. Given the fact that the children had just experienced the magic show, their levels of recall were quite low, especially for the 3- and 4-year-old children.

Inaccurate details: The number of inaccurate details was analyzed using the same model as that for the accurate recall except that raw scores were not adjusted for number of touches because there was no constraint on the number of errors that children could make. Analysis of these data produced a Sequence $\times$ Round of Questioning interaction, $F(1, 80) = 5.06$, $p < .05$, $\eta_p^2 = .06$. Error rates in the first round of questioning were similar for HFD and verbal interviews. However, in the second round of questioning, more errors were made in the HFD condition than in the verbal condition (see last column of Table 3). There was also a significant effect of

age, $F(4, 80) = 3.42, p < .01, \eta_p^2 = .15$. The 4-year-olds produced more errors (0.75) than the other age groups, which did not differ from each other (ranges = 0.1–0.4).

New details: A 5 (age level) $\times$ 2 (interview version: verbal 2 vs. HFD 2) ANOVA was conducted on the new accurate details (expressed as a function of the total number of touches and instruments) in the second round of questioning. There was a main effect of age, $F(4, 90) = 9.54, p < .01, \eta_p^2 = .30$. The greatest amount of new accurate information was produced by the 6- (21%) and 7-year-old (27%) children. There were no age differences among the three youngest groups, which produced the fewest new details (ranging from 4% to 7%). It is important to note that interview format did not influence this pattern.

Analyses were also conducted on the number of new errors made in the second round of questioning. There were no significant effects or interactions. New errors were relatively rare ($M = 0.33$).

Recognition Test—Correct responses to the foil items were entered into an Age (5) $\times$ Touch Version (2) $\times$ Sequence (2) $\times$ Round of Questioning (2) ANOVA with repeated measures on the last measure. The dependent variables were the proportions of correct responses as a function of the total number of foil questions (four vs. six). There was a significant effect of age, $F(4, 80) = 2.45, p < .05, \eta_p^2 = .11$. Planned comparisons confirmed the pattern shown in Table 4 that 3- and 4-year-olds made more errors (said yes to absent touches) than did 5-, 6-, and 7-year-olds, who performed similarly.

Next, performance on the target items was analyzed. The dependent measure was the number of yes responses to target present items plus the number of accurate instruments provided to the prompt (e.g., “What did the magician touch you/you touch the magician with?”). The total was divided by the number of possible accurate responses (eight vs. 12). These variables were entered into an Age $\times$ Touch Version $\times$ Sequence $\times$ Round of Questioning ANOVA with repeated measures on the last measure. There was a significant main effect of age, $F(4, 80) = 34.69, p < .01, \eta_p^2 = .63$ (see Table 4). Planned comparisons for the main effect of age showed that 6- and 7-year-olds were most accurate, followed by the 5-year-olds, who performed better than the 3- and 4-year-olds. A significant Touch $\times$ Age interaction, $F(4, 80) = 2.58, p < .05, \eta_p^2 = .11$, modified this relationship only slightly, as shown in Figure 3. The interesting aspect of the interaction was the fact that number of touches affected the performance only of the 3- and 7-year-olds, each in a different direction. The 7-year-olds showed the anticipated pattern that reduced memory load would result in relatively better performance; this same result was found for the prompt data (see Figure 2). The reverse pattern was obtained for the 3-year-olds. Their poorer performance on the low memory load was unexpected and may merely reflect floor effects; the 3-year-olds remembered very few items in both touch conditions.

There were main effects of sequence, $F(1, 80) = 5.73, p < .05, \eta_p^2 = .07$, and a significant Sequence $\times$ Round of Questioning interaction, $F(1, 80) = 4.58, p < .05, \eta_p^2 = .05$, which was modified by a Round of Questioning $\times$ Sequence $\times$ Age interaction, $F(4, 80) = 3.36, p < .01, \eta_p^2 = .14$.

Simple effects of the Age $\times$ Sequence interaction were examined at each round of questioning. For the first round of questioning, there was a main effect of age, $F(4, 90) = 29.88, p < .01$. Planned comparisons revealed the same pattern as reported for the main effect of age above, $3 = 4 > 5 > 6 = 7$. There was also a main effect of interview version, $F(1, 90) = 9.99, p < .01$. Overall children provided more correct information in the verbal 1 (55%) than in the HFD 1

(41%) condition. A similar analysis was carried out on the second round of questioning data in which verbal 2 and HFD 2 scores were compared as a function of age group. There was only a main effect of age, $F(4, 90) = 23.78, p < .01$. Planned comparisons yielded the same developmental patterns as reported above.

In summary, although the children were asked to recall touches immediately after the magic show, their performance was quite poor on both prompt and recognition measures. Only the older children showed the effects of memory load on both recognition and prompt measures. It is possible that there were no effects for the younger children because of floor effects even in the low-load condition. In addition, there were consistent effects of interview version and sequence across measures. HFDs resulted in more prompt errors when introduced second. They also resulted in fewer accurate recognition responses when introduced first. These effects were found across all ages.

Analysis of Prompts to “No” Errors—After all questions had been asked in both the verbal and HFD sections, the interviewer selected all target items that the child had consistently denied, stating that no touching had taken place when in fact there had been touching. As described above, the interviewer prompted the child in various ways to remember the events leading up to the actual touch.

Figure 4 shows for each age group the percentage of prompted responses that were categorized as denial of touch (NO), forgot to tell you before (FORGOT), what happened was not the same as touching (NOT THE SAME), and the touch occurred on another part of the body (DIFFERENT PART). Other responses represented 15% of the data, and these were omitted from the following analysis. Children who did not make any omission errors were excluded from this analysis, resulting in the following sample sizes for each age group: 20 (3 years), 20 (4 years), 15 (5 years), 15 (6 years), and 11 (7 years).

The types of errors were entered as repeated measures into an ANOVA with age as the between-groups effect. There was significant main effect of error type, $F(3, 228) = 8.78, p < .001$, $\eta_p^2 = .10$, which was modified by an Age $\times$ Error Type interaction, $F(12, 228) = 2.83, p < .001$, $\eta_p^2 = .13$. As shown in Figure 4 and confirmed by follow-up tests, the most common response for the 3- and 4-year-old children was to deny touching. The frequency of this response declined across other age groups. In contrast, the most frequent response for the 6- and 7-year-olds was to provide the correct response with prompting and to further say that they did not report it before because they had just “forgotten. The other response of interest was when children used an appropriate verb other than *touch* (e.g., *scratch*) and when asked whether *touch* could also be used, they replied, “That action was not touching.” This response occurred most frequently among 5-year olds but also occurred among the other age groups.

These data indicate that the 3- and 4-year-old children’s poor performance reflects their poor encoding of the touch, and they suggest that even at 7 years of age their semantic system does not have *touch* as an entry for the actions carried out in the magic show.

Predictors of Performance: Body Part Comprehension, Vocabulary, and Block Design—Unlike Study 1, children’s performance on the body part comprehension task was correlated with age, $F(4, 95) = 11.60, p < .01, r = .52$. Accuracy rates were relatively high across all age groups: 85% (3-year-olds), 83% (4-year-olds), 96% (5-year-olds), 97% (6-year-olds), and 99% (7-year-olds). There were no age differences for the standardized Vocabulary and Block Design scores.

Multiple regression analyses were conducted on four different outcome variables: prompt inaccurate details, prompt accurate details (%), target recognition (%), and foils (%). Separate

analyses were conducted on round of questioning 1 and round of questioning 2 measures. Age in months was entered in the first step, and then vocabulary, block design, and body comprehension were entered in the second step. Analyses were carried out for the HFD 1st/verbal 2nd group and next for the verbal 1st/HFD 2nd group. Significant standardized beta weights for the vocabulary, block design, and body comprehension tasks are reported in Table 5.

As can be seen after controlling for age, there were relatively few significant predictors; only 50% of the analyses were significant. For the verbal 1st/HFD 2nd condition, six of the eight analyses produced significant predictors, and these involved receptive and productive vocabulary tests. In all cases, high scores on the vocabulary and body comprehension tests were associated with accurate and errorless performance on both HFD and verbal interviews.

A second set of analyses was conducted in which the Vocabulary $\times$ Block Design interaction was entered into the model after forcing in age. The criterion variable was the change score from round of questioning 1 to round of questioning 2 for each of the four major outcome variables. None of these equations were significant, suggesting that profiles of strengths and weaknesses on language and visual spatial tasks did not predict relative performance on HFD versus verbal conditions.

General Discussion

The initial goal of this research was to explore the risks and benefits of using HFDs to assist young children's recall of touching. Two studies were designed to examine whether adding HFDs to verbal interviews would increase accuracy (benefits) or increase errors (risks). Of related concern was the optimal sequence of presenting HFDs in interviews so as to obtain the most complete and accurate information. Finally, these issues were placed within a developmental framework, with special attention to the performance of the 3- and 4-year-old children who might benefit most from interviews with various nonverbal cues of which HFD is one instance.

Pooling the results from Studies 1 and 2, there were three consistent findings concerning HFDs. First, there were more prompt errors when HFDs were introduced after a verbal interview. This result was not due to a general effect of repeating questions because there was no increase in errors for the verbal interview when it was introduced after the HFD 1 condition. Also, it was not inherent characteristics of the HFDs that produced more errors because questioning with HFD first elicited the same number of errors as questioning with the verbal interview first. Thus, when HFDs were used to give children a second opportunity to recall information, they were associated with risk.

The second consistent finding was that HFDs were not associated with any benefits in terms of eliciting additional accurate details. In fact, in Study 2, the HFD protocol elicited fewer correct responses than the verbal protocol when recognition questions were asked for the first time.

The third finding concerned developmental trends. This was the first study to include a preschool group (3- and 4-year-olds) that generally provides little information in most interviews and thus might stand to gain the most from HFDs; because of problems with symbolic representation, this group might perform particularly poorly in the HFD conditions. Although the 3- and 4-year-old children performed worst on most measures, they showed the same pattern of benefits and risks as the older children in the verbal and HFD conditions. The HFDs did not provide them with any added benefits in recalling complete and accurate information, nor did they reduce the quality of the information.

Study 2 was designed to examine some of the cognitive factors that might be associated HFD performance. Correlational analyses did not provide any evidence that HFDs were particularly helpful for children with poor vocabulary skills or with poor visual spatial skills. Rather, across all ages, children with good vocabulary skills excelled in their performance in both HFD and verbal conditions.

Finally, it should be noted that in the present studies, most of the questioning was carried out in nonsuggestive neutral situations—ones in which there would be minimal error rates. There was one exception in Study 1 where there was a very mild suggestion; children were told to answer recognition questions again because some of their previous answers might have been wrong. Here, there were repeated questions in the same condition for HFDs (HFD 2 vs. HFD 3) and for no-HFDs (verbal 2 vs. verbal 3). Children provided fewer accurate responses to target recognition items equally for both HFD and verbal versions. Although the suggestive manipulation worked as in previous studies (e.g., Cassel et al., 1996; Poole & White, 1991), there was no added risk for one condition over another. The results of the study may have underestimated these repeated effects because performance was so poor in all comparison conditions. Because of these floor effects, there is a need for additional research on the risks and benefits of repeating HFD questions.

The results of the studies presented in this article aid in the interpretation of previous studies and in doing so provide important new data on the feasibility of using HFDs. First, they provide an alternative interpretation of Aldridge et al. (2004). To review, in that study children who had disclosed sexual abuse during a full verbal protocol, which included open-ended and specific follow-up questions, were then asked specific information about touching with HFDs. Aldridge et al. found that HFDs were associated with increased information; however, the accuracy of this additional information is unknown because it could not be verified. In the present studies, where the verbal 1st/HFD 2nd condition best parallels the procedures of Aldridge et al.¹ and where the accuracy of the information could be verified, HFDs elicited increased information after a full verbal protocol but this information was inaccurate (see Brown et al., 2007, for similar results).

Second, the results of the present studies challenge the conclusions of Willcock et al. (2006) that children's poor performance on HFDs reflected problems with symbolic representation. Because there was no control condition of questioning without HFDs, it is difficult to determine the degree to which HFDs compromised children's performance. In the present study when HFDs were used first, as in Willcock et al., they produced the same number of errors as the verbal condition. Furthermore, pilot data for the present studies indicated that even the youngest children (an age group not included in Willcock et al.) had achieved the necessary symbolic representation skills; they had no difficulty using HFDs to show where a sticker had been placed on their own bodies.

Although the risk and benefits of using HFDs were the intended major focus of Study 1, the children's poor recall at all ages for both interview versions became the more important finding. Because children's poor recall of touching has also been reported by a number of researchers (Brown et al., 2007; Pezdek & Roe, 1997; Quas et al., 2007; Willcock et al., 2006), Study 2 was designed to examine some factors that might influence young children's memories of touching.

¹It is not a perfect match, but the studies share the following elements. All the children in Aldridge et al. (2004) had mentioned touch during the free recall with follow-up verbal questions. This is analogous to the present verbal 1 condition where children were asked open-ended questions about the magic show followed by specific questions about touching. In both studies, the HFD 2nd condition is the same with one exception. In the present study, the HFD questions repeated the specific verbal questions. In Aldridge et al., it is not known how many of the HFD questions requested information already presented in the full verbal protocol, although it is clear that the children were asked about touching in both phases of the interview.

To test the hypothesis that the children forgot the touches during the 1-week delay in Study 1, I tested children immediately after the magic show in Study 2. The youngest children (3-, 4-, and 5-year-olds) continued to perform poorly even when there was no delay. The 6- and 7-year-old children performed much better in Study 2, although they still provided few prompt accurate details, and performance on the target recognition items was far from perfect. Thus, forgetting could have accounted for some of the variance in the poor performance of the oldest children in Study 1.

The second hypothesis for children's poor performance in Study 1 was that there were too many touches to remember. Thus, a reduction in the number of touches should decrease memory load and increase memory. However, when the number of touches was manipulated in Study 2 (with no delay), only the 7-year-old children benefited from the lower memory load. The number of touches did not influence performance of the younger age groups.

The third hypothesis was that the children in Study 1 did not remember the touches because these were not salient and thus not noticed by many of the children. So, in Study 2, there was verbal elaboration and repetition after each touch. The slight rise in performance from Study 1 to Study 2 may have reflected this extra emphasis, although it is difficult to tease apart the effects of delay and salience of touch.

Finally, the follow-up questioning after the completion of the HFD and verbal interviews in Study 2 suggests some hypotheses worthy of future study. Specifically, it seemed that many of the 3- and 4-year-old children were unaware that the touches had taken place, which is consistent with their very poor performance on the prompt and recognition questions. Even with specific prompting about the event in which the touch took place, they denied the specific touch. In contrast, the older children (6- and 7-year-olds) were more likely to remember the touch under prompt conditions, claiming that they had just forgotten to report it. Finally, there were a small but significant number who denied or failed to mention a target touch because that action was not considered *touch* but was described as a more active verb such as *rubbing* or *scratching*. This suggests that young children's lexical and semantic system for *touch* is quite narrow.

Some researchers have interpreted children's poor recall of touching in the context of disclosure of child sexual abuse; specifically, they have concluded that it reflects motivational factors—the children are afraid, they are embarrassed, they do not want to tell about the touching (see discussion of Quas et al., 2007). Clearly, this cannot be the case for the present two studies, nor for other laboratory studies where the touches were socially sanctioned and innocuous. Rather, the results of the present studies suggest a cognitive basis for failures to report in the laboratory and even perhaps in the real world where children make false denials of sexual behaviors. Specifically, the results suggest that the children did not process or encode the touches not because of embarrassment, guilt, or fear, but because young children are so used to being touched by adults, that the touches are simply not noticed or segmented into identifiable events. In other words, there is so much going on (in this study, the magic show) that unless the touch is significant for the child, it may not get noticed. Second, children may not disclose because they may not have an adequate representation of the neutral word *touch*. Thus, using more specific terms such as *rub*, *scratch*, and so forth could produce more information.

In fact, it could be children's generally poor memory for touches that could explain in part the increased errors when questions were repeated with HFDs. If children do not have a memory for touch, then asking them a question for a second time is not going to improve that memory. However, the use of HFDs in concert with repeated questions may prime inaccurate pointing responses when asked, "Show me"

These findings offer practitioners some guidelines about the supplementary use of HFDs in interviews with young children (ages 3 through 7 years). The results of the two studies in this article suggest that if HFDs are used to follow up on information already questioned in a full verbal interview, they tend to elicit errors. Therefore, this would indicate that it would be best to use HFDs first in the interview. However, there was some suggestion in one of the analyses (Study 2, recognition questions) that HFDs elicited fewer accurate details than no-HFD interviews when they were presented first. Thus, on the basis of the current data, HFDs did not provide incremental validity as a diagnostic or assessment tool (Wolfner, Faust, & Dawes, 1993); namely, HFDs did not show an advantage over the verbal conditions (increased accuracy or decreased errors) in either of the two studies or in any of the test conditions. The results, therefore, suggest that it might be best to use only a verbal protocol when interviewing children about touching, even if they are young preschoolers. Perhaps HFDs could be used at the very beginning of interviews only to assess children's knowledge of body part names while the interviewer is also assessing other cognitive functions. However, it is first necessary to test this protocol to ensure that this specific use of HFDs actually helps interviewers in understanding children's statements and does not provide any negative effects.

The present data also indicate the importance of taking into account the nature of the allegations that children make or that have been reported. Children may not differentiate and thus not remember appropriate and inappropriate touches even if they have occurred a short while before. Of course, more research is required on children's recall of these types of touches that occur during sanctioned interactions (toileting, rubbing of buttocks). The concern is that children's denial of these nonmemorable experiences may provoke interviewers to use suggestive questioning techniques. Although in some cases suggestions might lead to true assents, there is also the risk that they might lead to false assents, especially for children who do not remember the touches or do not understand the semantics of the questions about touching.

Before concluding, it is necessary to emphasize the limitations of the present research. The current studies examined children's (ages 3 through 7) recall of innocuous nonpainful touches that in some circumstances could have sexual connotations. Different types of paradigms are required to determine how HFDs influence older and younger children's recall of genital or anal touches that are also often painful.

In summary, young children of the ages tested in this study demonstrated difficulty when asked to recall innocuous events that involved touching another person or being touched by another person. Children's difficulties in reporting touches may reflect incomplete encoding of the touch and poor semantic representation of the concept *touch*. HFDs were not designed to overcome these difficulties and, consequently as shown by these and other studies, do not assist children in providing accurate reports about touching.

Acknowledgments

This research was supported by Grant 5R01HD52034 from National Institute of Child Health and Human Development.

I would like to thank Chuck Brainerd for his help on the article and Kate Ollerhead for her help on all aspects of this project.

References

- Aldridge J, Lamb M, Sternberg K, Orbach Y, Esplin P, Bowler L. Using a human figure drawing to elicit information from alleged victims of child sexual abuse. *Journal of Consulting and Clinical Psychology* 2004;72:304–316. [PubMed: 15065963]

- American Academy of Child and Adolescent Psychiatry. Practice parameters for the psychiatric assessment of children and adolescents. *Journal of the American Academy of Child & Adolescent Psychiatry* 1997;36:37S–56S. [PubMed: 9432518]
- American Professional Society on the Abuse of Children. Guidelines for psychological evaluation of suspected sexual abuse in young children. Chicago: Author; 2002.
- Baker-Ward L, Gordon BN, Ornstein PA, Larus DM, Clubb PA. Young children's long-term retention of a pediatric examination. *Child Development* 1993;64:1519–1533. [PubMed: 8222886]
- Brady MS, Poole DA, Warren AR, Jones HHR. Young children's responses to yes–no questions: Patterns and problems. *Applied Developmental Science* 1999;1:47–57.
- Brown DA, Pipe ME, Lewis C, Lamb ME, Orbach Y. Supportive or suggestive: Do human figure drawings help 5- to 7-year-old children to report touch? *Journal of Consulting and Clinical Psychology* 2007;75:33–42. [PubMed: 17295561]
- Bruck M, Ceci SJ, Francoeur E. A comparison of three and four year old children's use of anatomically detailed dolls to report genital touching in a medical examination. *Journal of Experimental Psychology: Applied* 2000;6:74–83. [PubMed: 10937313]
- Bruck M, Ceci SJ, Francoeur E, Renick A. Anatomically detailed dolls do not facilitate preschoolers' reports of a pediatric examination involving genital touch. *Journal of Experimental Psychology: Applied* 1995;1:95–109.
- Carnes CN, Nelson-Gardell D, Wilson C, Orgassa UC. Extended forensic evaluation when sexual abuse is suspected: A multi-site field study. *Child Maltreatment* 2001;6:230–242. [PubMed: 11471630]
- Cassel W, Roebbers C, Bjorklund D. Developmental patterns of eyewitness responses to repeated and increasingly suggestive questions. *Journal of Experimental Child Psychology* 1996;61:116–133.
- Ceci, SJ.; Bruck, M. *Jeopardy in the courtroom: A scientific analysis of children's testimony*. Washington, DC: American Psychological Association; 1995.
- Conte JR, Sorenson E, Fogarty L, Rosa JD. Evaluating children's reports of sexual abuse: Results from a survey of professionals. *American Journal of Orthopsychiatry* 1991;61:428–437. [PubMed: 1951650]
- DeLoache JS, Burns NM. Early understanding of the representational function of pictures. *Cognition* 1994;52:83–110. [PubMed: 7924202]
- DeLoache JS, Marzolf DP. The use of dolls to interview young children: Issues of symbolic representation. *Journal of Experimental Child Psychology* 1995;60:155–173. [PubMed: 7545206]
- Enns JT, Girgus S. A developmental study of shape integration over space and time. *Developmental Psychology* 1986;22:491–499.
- Gordon B, Ornstein P, Nida R, Follmer A. Does the use of dolls facilitate children's memory of visits to the doctor? *Applied Cognitive Psychology* 1993;7:459–474.
- Greenhoot A, Ornstein P, Gordon B, Baker-Ward L. Acting out the details of a pediatric check-up: The impact of interview condition and behavioral style on children's memory reports. *Child Development* 1999;70:363–380. [PubMed: 10218260]
- Holmes L, Finnegan M. The use of anatomical diagrams in child sexual abuse forensic interviews. *American Prosecutors Research Institute Update* 2002;15:1–5.
- Holmes LS, Vieth VI. Finding words/half a nation: The forensic interview training program of CornerHouse and the American Prosecutors Research Institute. *American Professional Society on the Abuse of Children Advisor* 2003;15:4–8.
- Johnson LR, Perlmutter M, Trabasso T. The leg bone is connected to the knee bone: Children's representation of body parts in memory, drawing and language. *Child Development* 1979;50:1192–1202.
- Kail R, Pellegrino J, Carter P. Developmental changes in mental rotation. *Journal of Experimental Child Psychology* 1980;29:102–116. [PubMed: 7354259]
- Kendall-Tackett KA, Watson MW. Use of anatomical dolls by Boston-area professionals. *Child Abuse & Neglect* 1992;16:423–428. [PubMed: 1617476]
- Kosslyn SM, Margolis JA, Barrett AM. Age differences in imagery ability. *Child Development* 1990;61:995–1010. [PubMed: 2209202]

- Krackow E, Lynn SJ. Is there touch in the game of Twister? The effects of innocuous touch and suggestive questions on children's eyewitness memory. *Law and Human Behavior* 2003;27:589–604. [PubMed: 14724958]
- Lamb, ME.; Hershkowitz, I.; Orbach, Y.; Esplin, PW. Tell me what happened: Structured investigative interviews of child victims and witnesses. Chichester, UK and Hoboken, NJ: Wiley; 2008.
- MacWhinney K, Cermak S, Fisher A. Body part identification in 1- to 4-year-old children. *American Journal of Occupational Therapy* 1987;4:454–459. [PubMed: 3688161]
- Merritt KA, Ornstein PA, Spicker B. Children's memory for a salient medical procedure: Implications for testimony. *Pediatrics* 1994;94:17–23. [PubMed: 8008531]
- Peterson C, Bell M. Children's memory for traumatic injury. *Child Development* 1996;67:3045–3070. [PubMed: 9071770]
- Peterson C, Biggs M. Interviewing children about trauma: Problems with “specific” questions. *Journal of Traumatic Stress* 1997;10:279–290. [PubMed: 9136092]
- Pezdek K, Roe C. The suggestibility of children's memory for being touched: Planting, erasing, and changing memories. *Law and Human Behavior* 1997;21:95–106. [PubMed: 9058575]
- Poole D, White L. Effects of question repetition on the eyewitness testimony of children and adults. *Developmental Psychology* 1991;27:975–986.
- Preissler MA, Carey S. Do both pictures and words function as symbols for 18- and 24-month-old children? *Journal of Cognition and Development* 2004;5:185–212.
- Quas J, Davis E, Goodman GS, Myers JEB. Repeated questions, deception, and children's true and false reports of body touch. *Child Maltreatment* 2007;12:60–67. [PubMed: 17218648]
- Saywitz K, Goodman G, Nicholas G, Moan S. Children's memory of a physical examination involving genital touch: Implications for reports of child sexual abuse. *Journal of Consulting and Clinical Psychology* 1991;59:682–691. [PubMed: 1955603]
- Steward MS, Steward DS, Farquahar L, Myers JEB, Reinart M, Welker J, Joye N, Driskill J, Morgan J. Interviewing young children about body touch and handling. *Monographs of the Society for Research in Child Development* 1996;61:1–187. [PubMed: 8972588]
- Suddendorf T. Early representational insight: Twenty-four-month-olds can use a photo to find an object in the world. *Child Development* 2003;74:896–904. [PubMed: 12795396]
- Wechsler, D. Wechsler Preschool and Primary Scale of Intelligence—Revised. New York: Psychological Corporation; 1989.
- Wechsler, D. WISC–IV. San Antonio, TX: Psychological Corporation; 2003.
- Willcock E, Morgan K, Hayne H. Body maps do not facilitate children's reports of touch. *Applied Cognitive Psychology* 2006;20:607–615.
- Witt A, Cermak S, Coster W. Body part identification in 1-to 2-year-old children. *American Journal of Occupational Therapy* 1990;44:147–153. [PubMed: 2178442]
- Wolfner G, Faust D, Dawes RM. The use of anatomically detailed dolls in sexual abuse evaluations: The state of the science. *Applied and Preventive Psychology* 1993;2:1–11.

Figure 1.

Study 2: Prompt questions: Percentage of accurate details (*SE*) as a function of touch condition and age group.

Figure 2.
Study 2: Prompt questions: Percentage of accurate details (*SE*) as a function of round of questions and age group.

Figure 3.

Study 2: Percentage of accurate target recognition details (*SE*) as a function of touch condition and age group.

Figure 4.
Study 2: Percentage of response types (*SE*) by each age group for “Why children did not assent to touching?”

Appendix B.
HFDs of Females

Table 1

Study 1: Percentage Correct Details and Raw Incorrect Details to Prompts

Variable	Correct details (%)					Incorrect details (raw)				
	3-year-olds	4-year-olds	5-year-olds	6-year-olds	7-year-olds	3-year-olds	4-year-olds	5-year-olds	6-year-olds	7-year-olds
Round 1										
HFD 1st	1 (3)	7 (5)	14 (9)	2 (4)	17 (17)	0.50 (0.80)	1.40 (1.1)	3.33 (3.2)	0.83 (.75)	0.83 (1.3)
Verbal 1st	5 (6)	5 (7)	12 (16)	7 (5)	13 (6)	0.67 (0.80)	0.43 (0.50)	0.33 (0.50)	1.40 (1.1)	.40 (.60)
Round 2										
HFD 2nd	11 (6)	11 (14)	10 (14)	10 (8)	14 (11)	3.67 (2.9)	0.57 (1.0)	2.17 (2.1)	1.80 (1.1)	1.20 (1.0)
Verbal 2nd	7 (6)	9 (14)	8 (11)	0 (0)	21 (18)	1.17 (1.5)	1.00 (1.0)	0.83 (0.80)	1.67 (0.80)	0.67 (0.80)
Total	6 (6)	8 (9)	11 (12)	5 (6)	17 (14)	1.5 (2.1)	0.79 (0.90)	1.67 (2.2)	1.14 (0.90)	0.73 (0.90)

Note. HFD = human figure drawings. Standard deviations appear within parentheses.

Table 2

Study 1: Percentage Correct Recognition Assents and Instruments and Foils

Variable	Correct assents and instruments (%)							Fails (%)				
	3-year-olds	4-year-olds	5-year-olds	6-year-olds	7-year-olds	Total assents		3-year-olds	4-year-olds	5-year-olds	6-year-olds	7-year-olds
Round 1						24 (22)						76 (33)
HFD 1st	21 (26)	34 (28)	29 (08)	4 (6)	27 (26)			71 (35)	57 (40)	86 (16)	93 (8)	93 (12)
Verbal 1st	30 (19)	24 (27)	24 (16)	14 (12)	36 (23)			45 (47)	78 (38)	76 (39)	86 (10)	77 (37)
Round 2						21 (19)						80 (29)
HFD 2nd	29 (22)	17 (18)	18 (15)	14 (13)	27 (21)			52 (46)	82 (32)	88 (17)	83 (12)	83 (19)
Verbal 2nd	23 (26)	27 (14)	24 (20)	6 (11)	31 (26)			60 (42)	71 (36)	95 (7)	1 (0)	95 (7)
Round 3						19 (20)						84 (30)
HFD 3rd	23 (25)	15 (22)	13 (17)	16 (12)	27 (21)			67 (52)	86 (27)	98 (06)	86 (10)	86 (20)
Verbal 3rd	19 (24)	26 (23)	17 (20)	4 (6)	30 (26)			60 (44)	63 (51)	95 (7)	98 (06)	10 (0)
Total	26 (22)	25 (22)	24 (17)	9 (11)	30 (23)			59 (41)	74 (36)	90 (19)	91 (10)	90 (19)

Note. HFD = human figure drawings. Standard deviations appear within parentheses.

Table 3
Study 2 Prompts: Percentage Correct Details and Raw Scores of Incorrect Details Collapsed Across Touch Conditions

Variable	Correct details (%)						Incorrect details (raw)					
	3-year-olds	4-year-olds	5-year-olds	6-year-olds	7-year-olds	3-year-olds	4-year-olds	5-year-olds	6-year-olds	7-year-olds	Total errors	
Round 1												
HFD 1st	5 (7)	12 (16)	30 (27)	28 (15)	29 (17)	0.20 (0.42)	0.64 (1.28)	0.0 (0.0)	0.40 (0.97)	0.20 (0.63)	0.29 (0.81)	
Verbal 1st	3 (8)	17 (10)	26 (20)	33 (18)	40 (22)	0.0 (0.0)	1.0 (1.0)	0.20 (0.42)	0.10 (0.32)	0.10 (0.32)	0.27 (0.60)	
Round 2												
HFD 2nd	4 (7)	7 (10)	19 (16)	46 (20)	49 (15)	0.50 (1.3)	1.22 (1.5)	0.30 (0.48)	1.1 (1.9)	0.10 (0.32)	0.63 (1.3)	
Verbal 2nd	5 (10)	8 (10)	22 (18)	35 (20)	43 (15)	0.20 (0.63)	0.27 (0.47)	0.0 (0.0)	0.0 (0.0)	0.50 (0.71)	0.20 (0.49)	

Note. HFD = human figure drawings. Standard deviations appear within parentheses.

Table 4
Study 2: Percentage Correct Recognition Assents and Instruments and Foils

Variable	Correct assents and instruments (%)					Fails (%)				
	3-year-olds	4-year-olds	5-year-olds	6-year-olds	7-year-olds	3-year-olds	4-year-olds	5-year-olds	6-year-olds	7-year-olds
Round 1										
HFD 1st	15 (20)	17 (20)	39 (30)	58 (20)	75 (20)	89 (30)	93 (20)	100 (0)	95 (20)	92 (20)
Verbal 1st	25 (30)	31 (10)	63 (20)	75 (20)	79 (20)	87 (30)	76 (20)	96 (10)	88 (30)	100 (0)
Round 2										
HFD 2nd	18 (21)	17 (15)	48 (35)	62 (18)	67 (23)	87 (30)	97 (10)	98 (10)	100 (0)	98 (10)
Verbal 2nd	18 (25)	32 (24)	42 (19)	66 (16)	78 (12)	88 (30)	62 (30)	98 (10)	91 (20)	98 (10)
Total	19 (20)	23 (20)	48 (30)	65 (20)	75 (20)	88 (30)	84 (30)	98 (10)	94 (20)	97 (10)

Note. HFD = human figure drawings. Standard deviations appear within parentheses.

Table 5

Study 2: Significant Standardized Beta Weights for Regression Analyses

Significant predictors of 1st ROQ measures		
Criterion variables 1	Verbal 1	HFD 1
% Prompts accurate 1	Vocabulary (+.30)	<i>ns</i>
Prompt errors 1	<i>ns</i>	Body comp (−.30)
% Recognition 1	Body comp (+.35)	<i>ns</i>
% Foils 1	<i>ns</i>	<i>ns</i>
Significant predictors of 2nd ROQ measures		
Criterion variables 2	HFD 2	Verbal 2
% Prompts accurate 2	Vocabulary (+.34)	<i>ns</i>
Prompt errors 2	Body comp (−.28)	<i>ns</i>
% Recognition 2	Vocabulary (+.31)	Block (+.32)
% Foils 2	Body comp (+.35)	<i>ns</i>

Note. ROQ = round of questioning; Body comp = body comprehension correct items; Block = block design; % Prompt accurate = percentage of accurate details on prompts; % Recognition = percentage of correct touches and instruments.

Appendix A

Touches and Instruments Used in Studies 1 and 2

Action	Study 1		Study 2	
	Version 1	Version 2	Version 1	Version 2
Adult touches child	Wrist	Ankle	Wrist	Ankle
	Ear (scarf)	Head (wand)	Ear (scarf)	Knee (scarf)
	Leg	Knee (scarf)	Head (wand) ^a	Back (brush) ^a
	Stomach (sticker)	Shoulder		
Child touches adult	Cheek (wand)	Chin (wand)	Elbow	Chin (wand)
	Elbow	Toe	Back (brush)	Toe
	Back	Hand (sticker)	Cheek (sticker) ^a	Stomach (sticker) ^a

Note. The instruments used in the touches are in parentheses next to the corresponding touch.

^aThese items were only given to the children in the six-touch group.