SOCIAL INFLUENCE, 2006, 1 (1), 16–47

Psychology Press
Taylor & Francis Group

Suggestive interviewing in the McMartin Preschool and Kelly Michaels daycare abuse cases: A case study

Nadja Schreiber
Florida International University, USA

Lisa D. Bellah, Yolanda Martinez, Kristin A. McLaurin, Renata Strok, Sena Garven and James M. Wood
University of Texas at El Paso, USA

In the 1980s and early 1990s the United States witnessed an outbreak of bizarre "daycare abuse" cases in which groups of young children levelled allegations of sexual and Satanic abuse against their teachers. In the present study, quantitative analyses were performed on a total of 54 interview transcripts from two highly publicised daycare cases (McMartin Preschool and Kelly Michaels) and a comparison group of child sexual abuse cases from a Child Protection Service (CPS). Confirming the impression of prior commentators, systematic analyses showed that interviews from the two daycare cases were highly suggestive. Compared with the CPS interviews, the McMartin and/or Michaels interviewers were significantly more likely to (a) introduce new suggestive information into the interview, (b) provide praise, promises, and positive reinforcement, (c) express disapproval, disbelief, or disagreement with children, (d) exert conformity pressure, and (e) invite children to pretend or speculate about supposed events.

In the 1980s and early 1990s the United States witnessed an epidemic of what some commentators have called "Satanic panic" (Nathan & Snedeker,

Address correspondence to: James Wood, Department of Psychology, University of Texas at El Paso, El Paso, Texas 79968, USA. Email: jawood@utep.edu

Renata Strok is now at the Pacific School of Psychology, Palo Alto, CA. Lisa D. Bellah (formerly Lisa D. Velarde) is at the Federal Medical Center, Federal Bureau of Prisons, in Fort Worth, Texas. Yolanda Martinez is at Rio Vista Rehabilitation Center in El Paso, TX. Kristin McLaurin is at Porter and Hedges law firm in Houston, TX. Sena Garven is at the US Army Leadership School in Fort Leavenworth, KS.

We wish to thank the following members of Por Ninos y Familias for transcribing and scoring the interviews in this study: Damian Aguilar, Gabriela Bernal, Eunice Buendia, Jennifer Carrasco, Gustavo Correa, Diana Martinez, Linda Mollinary, Gabriel Nevarez, Lucia Ramirez, Regina White, LeeAnne Williams, Betty Urquidi, and Zandra Zamora.

© 2006 Psychology Press Ltd

http://www.psypress.com/socinf DOI: 10.1080/15534510500361739

1995; Victor, 1993, 1998). Exposés in popular books, magazine articles, and television programmes described a supposed underground network of devil worshippers who engaged in ritualistic murder, baby breeding, and cannibalism (Hicks, 1991; Loftus & Ketcham, 1994; Ofshe & Watters, 1994; Smith & Pazder, 1980; Spanos, 1996; Victor, 1993; Warnke, 1972; but see Hertenstein & Trott, 1993). Contributing to the panic was a national outbreak of so-called "daycare abuse" cases, in which groups of young children alleged that they had been sexually abused by their caretakers and forced to participate in bizarre ceremonies with Satanic overtones (Nathan & Snedeker, 1995; Rabinowitz, 2003; see also Kelley, Brant, & Waterman, 1993; Waterman, Kelly, Oliveri, & McCord, 1993).

Many social scientists, scholars, and legal authorities now view the stories of Satanic conspiracy that circulated in the 1980s as urban legends, and the daycare abuse cases as historical aberrations (Acocella, 1999; Bottoms & Davis, 1997; Lanning, 1991; Loftus & Ketcham, 1994; Nathan & Snedeker, 1995; Ofshe & Watters, 1994; Victor, 1993, 1998; but see Noblitt & Perskin, 2000; Sakheim & Devine, 1994; Sinason, 1996). Psychological researchers have taken a special interest in the interviewing techniques in these cases, which apparently induced children to make false accusations against their teachers (Ceci & Bruck, 1995; Garven, Wood, & Malpass, 2000; Garven, Wood, Malpass, & Shaw, 1998; Schreiber, Wentura, & Bilsky, 2001). Experimental exploration of these techniques has led to important insights regarding child suggestibility and child forensic interviewing (Ceci & Bruck, 1993, 1995; Poole & Lamb, 1998).

The present study analysed interviewing techniques from two of the most notorious daycare abuse episodes of the 1980s, the McMartin Preschool and Kelly Michaels cases. Not only did these two cases stimulate widespread interest among psychological researchers, but their legal outcomes also affected the fate of similar prosecutions throughout the United States. Thus the McMartin and Michaels cases are significant from the perspective of both psychological science and the law.

OVERVIEW OF THE MCMARTIN PRESCHOOL AND KELLY MICHAELS CASES

McMartin Preschool

The McMartin Preschool case was the first daycare abuse case in the United States to receive national media attention (for a detailed history, see Butler, Fukurai, Dimitrius, & Krooth, 2001). In 1983, seven teachers at the McMartin Preschool in the well-to-do Los Angeles suburb of Manhattan Beach were accused of kidnapping children and flying them to an isolated farm, where the children saw animals tortured and were forced to engage in group sex. All charges were eventually dropped against five of the teachers,

including several elderly women. The remaining two defendants, Peggy McMartin Buckey and her son Raymond Buckey, were tried in one of the longest and most expensive criminal cases in California history. Peggy Buckey was acquitted on all charges and Raymond on most charges. After juries in two separate trials failed to reach a decision on the remaining counts against Raymond, prosecutors dropped all charges against him in 1990.

Prosecutors in the McMartin case relied heavily on videotaped interviews of children. However, these very interviews eventually undermined the prosecution's case: After the trial, jurors publicly criticised them as highly leading (Reinhold, 1990; Timnick & McGraw, 1990; Wilkerson & Rainey, 1990). The interviewers were also criticised in popular-press books and articles (Eberle & Eberle, 1993; Hicks, 1991; Nathan & Snedeker, 1995; Tavris, 1997), academic articles (Butler et al., 2001; Ceci & Bruck, 1993, 1995; Garven et al., 1998, 2000; Green, 1992; Wyatt, 2002; but see Bernet & Chang, 1997; Faller, 1996; Summit, 1994), and an Emmy-award-winning television movie (*Indictment*).

Kelly Michaels

In 1988, Kelly Michaels, a 26-year-old daycare worker in Maplewood, New Jersey, was convicted and sentenced to 47 years in prison for sexually abusing 20 preschool children (for a detailed history, see Nathan, 1988; Rabinowitz, 1990). Children alleged that over a period of 7 months Michaels raped them with spoons, forks, and Lego blocks, compelled them to swallow her urine and faeces, and forced them to lie naked in the shape of a Satanic pentagram.

Michaels' trial attracted little media attention outside the region where it occurred. However, following her conviction, articles by sceptical journalists appeared in the *Village Voice* (Nathan, 1988) and *Harper's Magazine* (Rabinowitz, 1990). An appellate lawyer took up her case and in 1993 Michaels' conviction was reversed by the Appeals Court of New Jersey, which ruled that the children in the case were interviewed in a manner so suggestive as to render their statements unreliable. As with the McMartin case, the interviewing techniques used in the Michaels case were criticised by journalists (Nathan, 1988; Rabinowitz, 1990, 2003) and academics (Bruck & Ceci, 1993; Ceci & Bruck, 1995; Schreiber, 2000; see also Lamb, Sternberg, & Esplin, 1995; but see Lyons, 1995). Following the collapse of the McMartin prosecution and the reversal of Michaels' conviction, similar "daycare abuse" cases came to be viewed with widespread scepticism, so that legal prosecutions of such cases became rare (but see Nathan & Snedeker, 1995; Rabinowitz, 2003).

AIMS OF THE PRESENT STUDY

The present study quantitatively analysed child sexual abuse interview transcripts from the McMartin Preschool and Kelly Michaels cases. A case study using fine-grained, rigorous analysis was deemed important for three reasons. First, psychologists, legal scholars, and journalists who have criticised the interviews in the McMartin and Michaels cases have typically taken an *impressionistic* approach, supporting their conclusions with isolated quotes from interview transcripts (e.g., Bruck & Ceci, 1993). A quantitative analysis of the interviews can provide a more objective basis for opinion, and confirm whether earlier critiques accurately reflected the content of the interviews.

Second, a case study using quantitative analysis may reveal features of the interviews that are not immediately apparent to impressionistic observation. For example, which leading interviewing techniques were used most often in the McMartin and Michaels cases? Were the techniques used in both cases highly similar? If not, what were the differences? The answers to such questions may help to illuminate the nature of suggestive interviewing as it occurs in ''real-world'' cases. Third, analysis of the McMartin and Michaels transcripts provides an opportunity to develop and validate scientific measures of interviewer suggestiveness. Future research on real child forensic interviews can progress more quickly if measures with demonstrated reliability and validity are available.

CONSTRUCTS ANALYSED IN THE PRESENT STUDY

All transcripts in the present study were analysed using scores that fell into three general categories: (1) interview length, (2) form of questions, and (3) suggestive techniques. These three categories were selected because they are linked with relevant theory and have been a focus of prior research (see Ceci & Bruck, 1993, 1995; Myers, Saywitz, & Goodman, 1996; Poole & Lamb, 1998). The remainder of this section provides an overview of the constructs measured in the study, with a summary of relevant research and theory. Because the number of constructs is relatively large, the discussion of individual constructs is necessarily condensed.

Interview length

Four aspects of interview length were analysed for each interview: (1) Total interview length, (2) Number of words spoken by the interviewer, (3) Number of words spoken by the child, and (4) Ratio of interviewer words to child words.

Because studies on child interviewing practices in forensic settings usually analyse interview transcripts rather than audio or video recordings,

interview length has typically been measured not in seconds or minutes, but as the number of exchanges or "utterances" per interview (e.g., Hershkowitz, Lamb, Sternberg, & Esplin, 1997; Sternberg, Lamb, Hershkowitz, Esplin, Redlich, & Subshine, 1996). Some guidelines for child interviewing suggest that interview length can become a matter of concern if the child becomes fatigued or shows signs of wandering attention (e.g., Home Office, 2002).

Published guidelines for child forensic interviews frequently emphasise the importance of allowing children to talk at length and describe their experiences in their own words (e.g., Home Office, 2002; Lamb, Sternberg, & Esplin, 1998; Poole & Lamb, 1998; Warren & McGough, 1996; Yuille, Hunter, Joffe, & Zarparniuk, 1993). However, some observers have reported that law enforcement and child protection interviewers often do considerably more talking than the children who are being questioned (Warren, Woodall, Hunt, & Perry, 1996; Wood, McClure, & Birch, 1996). It has been suggested that a high *ratio of interviewer words to child words* may serve as a rough indicator of unskilful or suggestive interviewing (Underwager & Wakefield, 1990).

Form of questions

Four aspects of the form of questions were analysed in each interview: (1) Open/Narrative questions, (2) Yes/No questions, (3) Choice questions, and (4) Focused/Specific questions. These four categories partially reflect the way that an interviewer has exerted control and influence during the interview by "agenda setting" (limiting the conversation to certain topics) and "limiting and controlling the number of choices and options" (constraining responses to questions) (Pratkanis, in press).

Most published guidelines recommend that child sexual abuse interviewers begin the substantive part of the interview with *open-ended or free-narrative* questions ("Tell me what happened") and employ such questions as much as possible in the remaining parts of the interview (American Professional Society on the Abuse of Children, 2002; Home Office, 2002; Lamb et al., 1998; Poole & Lamb, 1998; Reed, 1996; Warren et al., 1996; Wood & Garven, 2000). Open-ended questions are deemed desirable because they are less likely to be suggestive than other forms of questions and are more likely to be answered accurately by children (Dent & Stephenson, 1979; Hutcheson, Baxter, Telfer, & Warden, 1996; Memon & Vartoukian, 1996). Furthermore, research on child sexual abuse interviews has found that an open-ended question is likely to elicit more information on average than a question of another form (Orbach, Hershkowitz, Lamb, Sternberg, Esplin, & Horowitz, 2000; Sternberg et al., 1996; Sternberg, Lamb, Orbach, Esplin, & Mitchell, 2001).

Most guidelines also indicate that *yes/no* questions (“Did it happen more than once?”), *choice* questions (“Were your clothes on or off?”), and *focused or specific* questions (“Where did that happen?”, “How many times did that happen?”) are appropriate in child interviews if they are generally non-suggestive and used sparingly (e.g., Home Office, 2002; Lamb et al., 1998; Reed, 1996; Wood et al., 1996). Research indicates that such questions can elicit useful information, but often at the cost of reducing children’s accuracy or introducing elements of suggestiveness (Brady, Poole, Warren, & Jones, 1999; Hutcheson et al., 1996; Memon & Vartoukian, 1996). Studies have shown that law enforcement and social service personnel tend to rely heavily on yes-no, choice, and focused questions in child sexual abuse interviews, even when open-ended questions are likely to be more effective (Davies, Wilson, Mitchell, & Milsom, 1995; Hershkowitz et al., 1997; Lamb et al., 1996; Sternberg et al., 1996, 1997; Warren et al., 1995, 1996).

Suggestive techniques

Five types of suggestive techniques were analysed in each interview: (1) Reinforcement, (2) Repetition of Questions, (3) Co-witness Information, (4) Inviting Speculation, and (5) Introducing New Information.

Reinforcement by an interviewer can take several forms, including (a) praising or otherwise rewarding the child for saying what the interviewer wants (“Thanks for telling me! You’re so smart!”), (b) giving the child negative feedback for failing to say what the interviewer wants (“Are you sure? Positive?”), or (c) indicating that praise, rewards, or negative consequences are forthcoming, depending on what the child says (“Let’s see if you’re smart enough to remember what happened!”). Psychologists have long recognised that reinforcement strongly shapes children’s behaviour (Ettinger, Crooks, & Stein, 1994; Tharp & Wetzel, 1969), and recent research has shown that it is a powerful and swift-acting social influence technique when used in child interviews. For example, in a study by Garven et al. (2000; see also Garven et al., 1998), 120 children aged 5 to 7 were visited in their classroom by a young man known as Paco Perez. A week later they were questioned about his visit. All children were questioned using mundane leading questions (“Did Paco break a toy while he was visiting?”) and fantastic leading questions based on the McMartin case (“Did Paco take you somewhere in a helicopter?”). Half of the children were also reinforced with praise for answers that were accusatory towards Paco and mild negative feedback for non-accusatory answers. In interviews that lasted only 3 to 4 minutes, reinforced children were induced to make 35% false accusations against Paco, compared with 12% for non-reinforced children. For fantastic questions, the false accusation rate was 52% for reinforced children versus 5% for non-reinforced children.

When re-interviewed a week later without reinforcement, children reinforced at the previous interview continued to make accusations at about the same rate as previously.

Several authors have identified r*epetition of questions* within and between interviews as a potentially suggestive interviewing technique, on the grounds that repetition can sometimes constitute a form of negative feedback, indicating to a child that previous answers to a question were unacceptable (Garven et al., 1998; Siegal, Waters, & Dinwiddy, 1988). Research indicates that repetition of choice questions (but not open-ended questions) during an interview can reduce the accuracy of children's reports regarding their experiences (Cassel, Roebers, & Bjorklund, 1996; Memon & Vartoukian, 1996; Poole & White, 1991, 1993).

The suggestive technique of *co-witness information* involves telling a child or adult witness what other witnesses have supposedly already said or observed. It is well established that such "social consensus" (Pratkanis, in press) or "social proof" (Cialdini, 2001) can be a powerful influence technique. Recent research has shown that co-witness information in child interviews can create conformity pressure to "go along" with other witnesses and induce stereotypes that influence responses to other questions (Garven et al., 1998, 2000; Leichtman & Ceci, 1995).

The interviewing technique of *inviting speculation* involves asking a child to speculate whether a particular event may have or could have happened, or to pretend that it has happened. Several studies have shown that this technique can reduce the accuracy of children's memory reports, probably by inducing source-monitoring errors (Ceci, Huffman, Smith, & Loftus, 1994; Ceci & Loftus, 1994; Schreiber & Parker, 2004; Schreiber et al., 2001; see also Garry, Manning, Loftus, and Sherman, 1996; Hyman, Husband, & Billings, 1995; Hyman & Pentland, 1996).

The technique of *introducing new information* involves introducing new post-event information (either accurate or inaccurate) into an interview via a question or a statement, even though that information was not previously mentioned by the child (for example, asking the child "Did he touch you on your hiney?" when the child has not previously mentioned sexual touching). The definition for this category is intentionally broad and overlaps with the definitions of other interviewing techniques already discussed here. What Pratkanis (in press) calls "(mis)leading questions" have been studied by social psychologists, who approach it as an influence tactic, and by cognitive researchers, who approach it as a form of post-event misinformation (Dale, Loftus, & Rathburn, 1978; Ceci, Ross, & Toglia, 1987; Leichtman & Ceci, 1995; Loftus & Davies, 1984). Research has shown that children's reports regarding their experiences become less accurate if interviewers ask misleading questions and introduce misinformation, although children generally grow less susceptible to the effect of misleading

questions as they grow older (see reviews by Ceci & Bruck, 1993; Poole & Lamb, 1998).

As can be seen, the present study analysed the McMartin and Michaels interviews on a variety of dimensions related to interview quality and suggestiveness. We also analysed a set of Child Protective Services (CPS) interviews to serve as a comparison group. We expected that by examining the data from many different perspectives the study could (a) reveal any differences between "normal" CPS interviewing and daycare interviewing styles, and (b) clearly delineate distinctive investigative interviewing features in these two famous cases.

METHOD

Sources of interview transcripts

The present study analysed 54 transcripts of child sexual abuse interviews: 14 were from the McMartin Preschool case, 20 from the Kelly Michaels case, and 20 from the Child Protective Service of a city in the western United States.

McMartin transcripts. The transcripts of more than 50 interviews from the McMartin case have been archived in the library of Brown University in Providence, Rhode Island. The present study analysed the subgroup of 14 interview transcripts that were introduced into evidence at the preliminary hearing or trial of Peggy and Raymond Buckey in Los Angeles (see Butler et al., 2001). The interviewing techniques used in these 14 transcripts appear to be virtually the same as the techniques used in the other transcripts in the archive. These 14 transcripts were selected for use in the present study because there was particularly strong evidence of their reliability: In preparation for the trial, these 14 were reviewed for accuracy by both prosecution and defence attorneys. The transcript versions used in the present study included some minor and generally non-substantive hand-written alterations in the margins. It appears, therefore, that these transcripts represented the penultimate rather than final form of the transcripts as they were finally introduced into evidence at the hearing or trial. The ages of the 14 children in the McMartin sample ranged from 4 to 9.5 years. The mean age was 6.89. A total of 36% (5) of the sample was male and 64% (9) was female.

Kelly Michaels transcripts. The 20 Kelly Michaels interviews were selected from an extensive collection of interview transcripts that were filed as evidence in the legal appeal of Kelly Michaels. Many transcripts in this collection were of poor quality or obviously incomplete. Therefore, 20 transcripts were selected that met the following criteria: (1) five or more

pages in length; (2) beginning and end of the interview included in the transcript; (3) clear indications of which statements were made by the interviewer and which by the children; (4) generally correct punctuation; (5) absence of hand-written notes that changed the meaning of the type-written words. The transcripts of the Kelly Michaels interviews used in the present study are archived, like the McMartin transcripts, at the library of Brown University. The ages of the children in the Michaels sample are unknown, although apparently all or nearly all were less than 7 years old. A total of 50% (10) of the sample was male and 35% (7) was female. The gender of the remaining three children was not clear from the interview transcripts.

Child Protective Service (CPS) transcripts. As part of a larger study on child interviewing, the present researchers transcribed audiotapes of over 100 sexual abuse interviews from the CPS of a medium-sized city in the western United States. All tapes had been recorded in an interviewing room at CPS headquarters as part of the standard investigative process. It was our impression, and that of CPS administrators, that the tapes were representative of all interviews conducted at the agency.

The tapes were transcribed by one member of our research team and then checked and corrected by a second member. To protect confidentiality, all potentially identifying information was deleted. From the larger pool of transcripts, 20 were selected as a comparison group in the present study, according to the following three criteria. First, the original tape recordings had to be clear and audible, with no more than four inaudible statements by the interviewer or the child. Second, the interview had to include an allegation of sexual abuse by the child. Although the research team did not have access to the final CPS determinations for these cases, all the allegations of sexual abuse appeared credible and most were strongly compelling. Third, from among the transcripts that met the first and second criteria, the transcripts for the 20 youngest children were selected, to increase the resemblance to the children in the McMartin and Michaels cases.

The mean age of children represented in the final sample of 20 CPS transcripts was 8.12 years, with a range from 4 to 11 years. Thus the children in the CPS sample were somewhat older than those in the McMartin and Michaels samples. Of the CPS sample, 85% (17) were female and 15% (3) were male. A total of 19 of the 20 interviews were conducted during the years 1993–1996 and the remaining interview was conducted in 1987.

The CPS and Michaels transcripts in the present study were previously analysed in a pilot study published by Schreiber (2000). However, the data in the present study did not overlap with Schreiber's data because the two studies used different scorers, different versions of most scoring rules, and only partially overlapping scoring categories.

Preliminary division of transcripts into "exchanges"

As a preliminary step, all interview transcripts were divided into numbered "exchanges". In virtually all cases, each exchange consisted of one "turn" by the interviewer and one "turn" by the child. In a few instances, when more than one interviewer took part in an interview, slightly more complex rules were used to divide transcripts into "exchanges".

Measurement of interview length

Several aspects of interview length were measured in the present study.

Number of exchanges per interview. As already noted, all exchanges in the transcripts were numbered in all transcripts.

Estimated number of words per interview. For each interview, the number of words in a subset of exchanges was counted by hand. For interviews with 1–100 exchanges, all exchanges were counted. For interviews with 101–200 exchanges, every other exchange was counted. For interviews with 201–300 exchanges, every third exchange was counted, and so forth. The total number of words in an interview was then estimated by multiplying the total number of exchanges times the average number of words per exchange. The total number of words was estimated separately for the interviewer and the child.

Estimated length in seconds of each interview. Interview length in seconds was estimated using a formula developed by Velarde (1997) based on transcripts and audiotapes of 79 child sexual abuse interviews:

$$\text{Total Length of Interview (in seconds)} =$$

$$(.375 * \text{Estimated Total Number of Words}) + 221.99$$

In a cross validation study, Velarde (1997) found that this formula generally provided accurate estimates of interview length ($r = .84$), but could underestimate interview length by approximately 20% if (a) the interview involved a great deal of play (e.g., dressing and undressing dolls), or (b) the transcript or tape included a large number of gaps (e.g., interviewer or child leaves the interview room for more than a few seconds).

The four scoring categories for form of question

Four scoring categories were used to identify the form of question used by interviewers to elicit information from children. These categories were (1)

Yes/No questions, (2) Choice questions, (3) Focused/Specific questions, and (4) Open/Narrative questions. The scoring rules for each form of question are briefly described here. More detailed versions of the scoring rules are provided by Martinez (1999) or can be obtained on request from the seventh author of this article.

Yes/No questions. A Yes/No question was defined as an interrogative statement by the interviewer that could be answered easily and naturally with a simple "yes" or "no" response. Example: "Did he touch you?"

Choice questions. A Choice question was defined as an interrogative statement that asked the child to choose from options that were explicitly named. Example: "Did Chester go into the closet or into the bathroom?" Also included in this category were questions or commands that implicitly asked the child to point to or choose an object. Example (while viewing a photograph of several people): "Who was there?"

Focused/Specific questions. A Focused/Specific question was defined as an interrogative statement that did not meet the definitions for Yes/No or Choice questions, and could be answered easily and naturally with an adjective, noun, preposition phrase, or a list of nouns or adjectives. Focused/Specific questions included most questions that begin with "who", "when", and "where". Examples: "Who else was there?", "What colour was the hat?"

Open/Narrative questions. An Open/Narrative question was defined as an interrogative statement that did not meet the definitions for Yes/No, Choice, or Focused questions, but required (a) a verb in the response, or (b) a narrative description of an action or series of actions. Example: "How did you play the naked movie star game?" Also included in the Open/Narrative category were questions that requested a lengthy description, such as "What did the naked movie star game look like?"

Each exchange in each interview was scored for the presence or absence of each of the four forms of question. Some exchanges included several questions or statements by the interviewer, and therefore could be scored as having more than one "form of question". The following excerpts from McMartin Interview 107 illustrate how the exchanges were scored for form of question (interview numbers are recorded with the transcripts at Brown University):

Q231: Is that like a play shot or a real shot? [Scored as Choice]
A231: It's a real poisonous shot.
Q232: From a needle like a doctor gives? [Scored as Yes/No]
A232: Yeah.

Q233: Oh, where does the shot go? [Scored as Focused/Specific]
A233: I''m not sure.
…
Q280: Oh, that's where the music came from. And then what did the kids do? [Scored as Open/Narrative]
A280: Probably they try to get out.
Q281: Oh, did the kids try to get out? [Scored as Yes/No]

The six scoring categories for suggestive techniques

Six scoring categories were used to identify specific suggestive techniques used by interviewers to elicit information from children. These six categories were (1) Positive Consequences, (2) Negative Consequences, (3) Other People, (4) Asked and Answered, (5) Inviting Speculation, and (6) Introducing Information. These scoring categories were selected because they appeared problematic according to theory or research (see Introduction of the present article) and were present at least occasionally in a separate sample of McMartin transcripts examined during a pilot study. Scoring rules for each of the six techniques are briefly described here. More detailed versions of the scoring rules are provided by McLaurin (2000) and Strok (1997), or can be obtained on request from the seventh author of this article.

Positive Consequences. Positive Consequences was defined as giving, promising, or implying praise, approval, agreement, or other rewards to a child, or indicating that the child could demonstrate desirable qualities (e.g., helpfulness, intelligence) by making a statement to the interviewer. Examples (McMartin Interview No. 107, pages 32–33, 38):

Interviewer: Oh, you're so smart. I knew you'd remember
….
Interviewer: So I bet if you guys put on your thinking caps, you can help remember it. Now let's make a test of your brain and see how good your memories are.

Negative Consequences. Negative Consequences was defined as criticising or disagreeing with a child's statement, or otherwise indicating that the statement was incomplete, inadequate, unbelievable, dubious, or disappointing. Example (Michaels Interview No. 19C, pp. 170–171):

Interviewer: Were you ever afraid of Kelly?
Child: No.

Interviewer: No?
Child: No
Interviewer: Would you tell me if you were afraid of her?

The scoring categories of Positive Consequences and Negative Consequences were intended to operationalise the construct of reinforcement, as discussed in the Introduction of this article.

Asked and Answered. Asked and Answered was defined as asking the child a question that she or he has already unambiguously answered in the immediately preceding portion of the interview. Simple repetition of a question was not considered Asked and Answered if the interviewer was simply reflecting back the child's statement, without trying to elicit a new answer. Example (McMartin Interview Number 111, p. 29):

Interviewer: Can you remember the naked pictures?
Child: (Shakes head "no")
Interviewer: Can't remember that part?
Child: (Shakes head "no")
Interviewer: Why don't you think about that for a while, okay? Your memory might come back to you.

As can be seen from this example, repetitions scored as Asked and Answered could also often be scored as Negative Consequences. The scoring category of Asked and Answered was intended to operationalise inappropriate use of repetition by the interviewer, as discussed in the Introduction of this article.

Other People. Other People was defined as telling the child that the interviewer had already received information from another person regarding the topics of the interview. Example (McMartin Interview No. 107, pp. 16–17):

Interviewer: You see all the kids in this picture? Every single kid in this picture has come here and talked to us. Isn't that amazing? … These kids came to visit us and we found out they know a lot of yucky old secrets from that old school. And they all came and told us the secrets. And they're helping us figure out this whole puzzle of what used to go on in that place …

The scoring category of Other People was intended to operationalise the construct of co-witness information, as discussed in the Introduction of this article.

Inviting Speculation. Inviting Speculation was defined as asking the child to offer opinions or speculations about past events (e.g., "What could she have done?") or framing the child's task during the interview as using imagination (e.g., "pretending") or solving a mystery (e.g., "figuring something out"). Example (McMartin Interview No. 101, pp. 60–61):

> Interviewer: Now, I think this is another one of those tricky games. What do you *think*, Rags?
> Child: Yep.
> Interviewer: Yes. Do you *think* some of that yucky touching happened, Rags, when she was tied up and she couldn't get away? Do you *think* some of that touching that – Mr. Ray *might* have done some of that touching? Do you think that's *possible*? Where do you *think* he *would have* touched her? Can you use your pointer and show us where he *would have* touched her? [Emphasis added]

Introducing Information. Introducing Information was defined as introducing new information into an interview that was not previously mentioned by the child. The new information, included in either an interviewer's statement or question, had to represent a substantial addition to or discontinuity with the child's prior statements. An interviewer question or statement only received a rating for introducing information if it (a) introduced new material that was sexual, violent, or negative in content, (b) was contradictory or substantially inconsistent with the child's previous statements, or (c) referred to unusual and highly specific events or ideas (e.g., being flown away from school in a helicopter) not previously mentioned by the child. Example (McMartin Interview 107, p. 32):

> Interviewer: How about Naked Movie Star? You guys remember that game?
> Child: No.
> Interviewer: Everybody remembered that game. Let's see if we can figure it out.

This example received a score for Introducing Information because it introduced information ("Naked Movie Star" game) that was (a) new, (b) sexual, and (c) highly specific, but (d) hadn't previously been mentioned by the child in this interview. In addition, this example received a score for Other People (because the child was told that "everybody" remembered the supposed game), and for Inviting Speculation (because the child was invited to "figure it out"). Because Introducing Information was a broad scoring category, it often overlapped with the other five suggestive categories. The

scoring category of Introducing Information was intended to operationalise "leading the witness" as this term is used in legal settings (Garner, 2004).

Scoring procedures

Teams of trained raters scored each of the 54 interviews for the presence or absence of the various scoring categories described in the previous two sections. Detailed procedures were developed to ensure that scoring was independent and reliable. For example, the procedures used for scoring the four forms of questions were as follows:

(1) *Training.* Six undergraduate and graduate students received detailed written descriptions of the scoring rules for the four forms of questions. For the following 4 weeks, the students studied these rules, practised scoring of interview transcripts, and received feedback and instruction from a teacher who was expert in the rules. To minimise bias, scorers were told that the study was intended to provide an accurate measure of the interviewing techniques used in the McMartin and Michaels cases and the CPS, that there were no *a priori* hypotheses about which techniques would be more or less common in each case, and that the scorers should therefore strive above all for accuracy in scoring.

(2) *Testing*. At the end of the 4-week training period, the six students were given a test to determine their proficiency. To receive a passing grade on the test, a student had to independently achieve adequate agreement (kappa of .50 or higher) with the teacher for all four forms of question (Yes/No, Choice, Focused/Specific, Open/Narrative). Four of six students passed this test. The two with the highest grades were designated Primary Scorers, whereas the two with lower passing grades were designated as Checkers.

(3) *Creation of scoring teams.* Two scoring teams were created. Each team consisted of one Primary Scorer and one Checker.

(4) *Random assignment of interview transcripts to scoring teams*. Each of the 54 interview transcripts was randomly assigned to one of the two scoring teams, so that each team received the same proportion of McMartin, Kelly Michaels, and CPS transcripts.

(5) *Independent scoring of transcripts.* The Primary Scorer and Checker who were assigned a transcript each scored it independently for the four categories of "form of question". For example, each exchange in the transcript was scored independently for Yes/No questions by both the Primary Scorer and the Checker.

(6) *Assessment of agreement between team members.* After the Primary Scorer and Checker had independently scored a transcript, their level of agreement for the scoring of individual exchanges was assessed

by calculating four separate kappa statistics, one for each form of question. If the Primary Scorer and Checker achieved a minimum kappa of .50 for each of the four forms of question, then their scoring was deemed acceptable. If they failed to achieve a kappa of .50 for a particular scoring category, they were asked to *independently* re-score the transcript for that category, taking more care and consulting the scoring rules.

(7) *Completion of scoring by teams.* A team's scoring of a particular transcript for a particular scoring category was considered complete if (a) the team achieved a kappa of .50 during the first or second independent scoring (see prior step), or (b) the team achieved a kappa below .50 on the first independent scoring, but the base rate of the scoring category for both the Primary Scorer and the Checker was very low (i.e., below .05), or (c) the Primary Scorer and the Checker were unable to achieve a kappa of .50 even after independently scoring a protocol twice.

(8) *Computation of final scores for each protocol.* After a team's independent scoring of a transcript was accepted (see step 7), the Primary Scorer was given a detailed list that described all instances in which the Scorer and the Checker had disagreed in their scoring of individual exchanges. The Primary Scorer then re-read the transcript, reviewed all disagreements, and changed any scores in which the Checker appeared to be correct. This procedure was intended to allow the Primary Scorer to detect and correct any obvious scoring errors that he/she had made during independent scoring. The Primary Scorer's final scores of the transcript—after reviewing the disagreements from the Checker—constituted the ''final score'' of the transcript.

Similar procedures were followed for all scoring categories, to ensure an acceptable level of inter-rater reliability. Additional details about these procedures are provided by Martinez (1999), McLaurin (2000), and Strok (1997).

Calculation of dependent variables for forms of question and leading techniques

For each of the four forms of question, the dependent variable was the *proportion of questions* in a particular interview transcript in which a particular form of question was scored as ''present'', as determined by the Primary Scorer's ''final score'' (see Step 8 under ''Scoring procedures''). For example, for Yes/No questions, the main dependent variable was the proportion of all questions in an interview transcript that were Yes/No. Thus, if a transcript contained 100 questions, and 20 of these questions were

scored as Yes/No by the Primary Scorer, then the transcript would receive a score of 0.20 for Yes/No.

A similar procedure was used for all scores that concerned interview length and suggestive techniques. However, for scores in these categories, the calculations were performed for each *exchange* in the interview (rather than for each question). For example, if a transcript contained 100 exchanges and 10 of these were scored as Introducing Information in the Primary Scorer's final score, then the transcript would receive a score of 0.10 for Introducing Information.

RESULTS

Reliability of scoring

Inter-rater reliability for the various scoring categories is reported in Table 1. As an example, inter-rater reliability for Yes/No questions was calculated as follows: (1) The proportion of questions scored as Yes/No by the Primary Scorer during the first independent scoring (before any feedback) was calculated for each of the 54 interviews. (2) The proportion of questions independently scored as Yes/No by the Checker was calculated for the same 54 interviews. (3) The correlation (Pearson's *r*) was then calculated between the 54 scores assigned by the Primary Scorer and the 54 assigned by the Checker. This correlation coefficient is reported in Table 1. A similar approach was used to calculated inter-rater reliability for all other forms of question and suggestive techniques in Table 1.

Inter-rater reliability for total number of words was calculated somewhat differently. (1) Three raters independently scored a total of 203 exchanges in

TABLE 1
Interrater reliability (Pearson's *r* or Spearman's rho) of scoring categories, based on 54 transcripts

Scoring Category	*r* or rho (descriptor)
Number of Words Per Exchange (Interviewer)	.96 (substantial)
Number of Words Per Exchange (Child)	.99 (substantial)
Yes/No	.96 (substantial)
Choice	.74 (moderate)
Focused/Specific	.94 (substantial)
Open/Narrative	.52 (fair)
Positive Consequences	.91 (substantial)
Negative Consequences	.92 (substantial)
Asked and Answered	.80 (moderate)
Other People	.93 (substantial)
Inviting Speculation	.95 (substantial)
General Suggestiveness	.89 (substantial)

Note: Descriptors are based on standards suggested by Shrout (1998).

three interviews (one interview randomly selected from each source). (2) The number of words scored by each rater for each of the 203 exchanges was calculated. (3) The correlation (Spearman's rho) was then calculated between the 203 scores assigned by each rater, and the 203 scores assigned by each of the other two raters. The mean correlations among raters are reported in Table 1.

Standards suggested by Landis and Koch (1977) have often been used to evaluate the quality of inter-rater reliability. However, Shrout (1998) has criticised the Landis and Koch standards as too lax and recommended that the following descriptors be used instead to evaluate the quality of inter-rater reliability in research: virtually none (0.00–0.10), slight (0.11–0.40), fair (0.41–0.60), moderate (0.61–0.80), substantial (0.81–1.00). As may be seen from Table 1, inter-rater reliability for 9 of the 12 scoring categories was greater than .80 and would be regarded as "substantial" for research purposes according to Shrout's descriptors. Inter-rater reliability for two scoring categories (Choice and Asked and Answered) was between .61 and .80 and would be regarded as "moderate". Scoring for one category (Open/Narrative) was .52 and would be regarded as only "fair".

Interviews

Results regarding the length of the McMartin, Kelly Michaels, and CPS interviews are reported in Table 2. To limit Type I error here and

TABLE 2
Length of the McMartin, Kelly Michaels, and CPS interviews: Means and standard deviations per interview

	Interview Source					
	McMartin	*Michaels*	*CPS*			
Measure	*Mean(SD)*	*Mean(SD)*	*Mean(SD)*	*F*	*df*	*p*
No. of exchanges	576.5[a] (189.6)	189.6[b] (100.4)	163.9[b] (67.8)	46.3	(2,51)	<.0001
Est. no. of words spoken by child	2,337.8[a] (1,128.6)	745.5[b] (470.3)	1,092.5[b] (769.3)	17.6	(2,51)	<.0001
Est. no. of words spoken by interviewer	9,631.4[a] (3,635.0)	2,478.3[b] (1,169.4)	1,800.3[b] (737.4)	71.9	(2,51)	<.0001
Est. ratio of interviewer words to child words	4.60[a] (2.04)	4.67[a] (3.79)	2.31[a] (1.19)	4.9	(2,51)	.011
Est. length in time	1h, 14m 16s[a] (1,482s)	23m,51s[b] (582s)	21m,47s[b] (455s)	64.8	(2,51)	<.0001

$N=14$ for McMartin, 20 for Michaels, 20 for CPS.
[a]Means with the same superscript are not significantly different, using a post hoc Bonferroni test ($p < .005$).

throughout the Results, only p values of .005 or less were regarded as statistically significant for main effects and post hoc comparisons. As can be seen, even by this conservative standard the McMartin interviews were significantly longer than interviews from the other two sources. Specifically, the McMartin interviews lasted approximately 1 hour and 14 minutes, whereas the Michaels interviews lasted about 23 minutes and the CPS interviews about 21 minutes. The McMartin interviews were also found to have significantly more exchanges, interviewer words, and child words than the Michaels and CPS interviews.

The ratio of interviewer words to child words was approximately twice as high for the McMartin (ratio = 4.60) and Michaels (ratio = 4.67) as for the CPS interviews (ratio = 2.31). However, the probability value for the overall between-groups difference ($p < .011$) only approached significance according to the Bonferroni test ($p < .005$) used in the present study. It may be that use of the Bonferroni test, which is known to be conservative, here resulted in a Type II error (i.e., failure to detect a genuine effect).

Form of question

Results regarding the four forms of question are reported in the top part of Table 3. As can be seen, the McMartin interviewers used a significantly higher proportion of Yes/No questions, and a lower proportion of Focused/Specific questions, than the Kelly Michaels and CPS interviewers. Other between-groups differences for form of question were not statistically significant.

Leading techniques

Results regarding the six suggestive interviewing techniques are reported in the bottom part of Table 3 and in Figure 1. As can be seen, the McMartin and Kelly Michaels interviewers both used several leading techniques more frequently than did CPS interviewers. Specifically, the McMartin interviewers were significantly more likely than the CPS interviewers to use Positive Consequences, Other People, Inviting Speculation, and Introducing Information. The Kelly Michaels interviewers were more likely than CPS interviewers to use Negative Consequences and Introducing Information.

In addition, the McMartin interviewers used two techniques (Other People and Inviting Speculation) significantly more often than the Kelly Michaels interviewers. No significant between-group differences were found for Asked and Answered.

Correlations with age in the CPS sample

Because children in the CPS sample were somewhat older than children in the McMartin and Michaels samples, an exploratory analysis was performed to

TABLE 3
Proportion of questions or exchanges in which a form of question or suggestive technique was used: McMartin, Kelly Michaels, and CPS interviews

	Interview Source					
	McMartin	*Michaels*	*CPS*			
Form of Question	*Mean (SD)*	*Mean (SD)*	*Mean (SD)*	*F*	*df*	*p*
Yes/No	0.78[a] (0.06)	0.67[b] (0.10)	0.60[b] (0.10)	14.6	(2,51)	<.0001
Choice	0.04[a] (0.03)	0.04[a] (0.02)	0.05[a] (0.02)	1.4	(2,51)	*ns*
Focused/Specific	0.11[a] (0.03)	0.18[b] (0.06)	0.23[b] (0.06)	21.4	(2,51)	<.0001
Open/Narrative	0.07[a] (0.04)	0.11[a] (0.07)	0.12[a] (0.06)	3.3	(2,51)	.0450
Leading Technique	Mean (SD)	Mean (SD)	Mean (SD)			
Positive Consequences	0.18[a] (0.09)	0.10[ab] (0.06)	0.07[b] (0.05)	21.5	(2,51)	<.0001
Negative Consequences	0.08[ab] (0.05)	0.15[a] (0.14)	0.04[b] (0.04)	8.1	(2,51)	.001
Asked and Answered	0.05[a] (0.02)	0.08[a] (0.04)	0.05[a] (0.04)	5.0	(2,51)	.011
Other People	0.07[a] (0.04)	0.02[b] (0.03)	0.00[b] (0.00)	27.9	(2,51)	<.0001
Inviting Speculation	0.08[a] (0.04)	0.03[b] (0.02)	0.01[b] (0.01)	39.4	(2,51)	<.0001
Introducing Information	0.18[a] (0.07)	0.18[a] (0.10)	0.03[b] (0.03)	26.9	(2,51)	<.0001

N=14 for McMartin, 20 for Michaels, 20 for CPS.
[a,b]Means with the same superscript are not significantly different, using a post hoc Bonferroni test ($p < .005$).

determine whether child's age was related to the measures of interview length, form of question, and suggestive techniques shown in Tables 2 and 3. Significant correlations with age, using a relaxed p value of .05, were found for 4 of the 15 measures. Specifically, age of child correlated −.61 with Yes/No questions, .56 with focused Questions, −.59 with Asked and Answered, −.57 with Negative Consequences. That is, when questioning younger children, CPS interviewers used more Yes/No questions and fewer Focused questions, repeated questions more often, and were more likely to indicate disagreement, disbelief, or dissatisfaction with children's answers.

DISCUSSION

Two findings of the present study are particularly notable. First, quantitative analyses confirmed that the interviews in the McMartin and Kelly Michaels daycare abuse cases were characterised by highly suggestive techniques that research has shown can elicit misleading statements and false accusations from children. Overall, substantially more problematic interviewing techniques were found in the daycare abuse cases than in

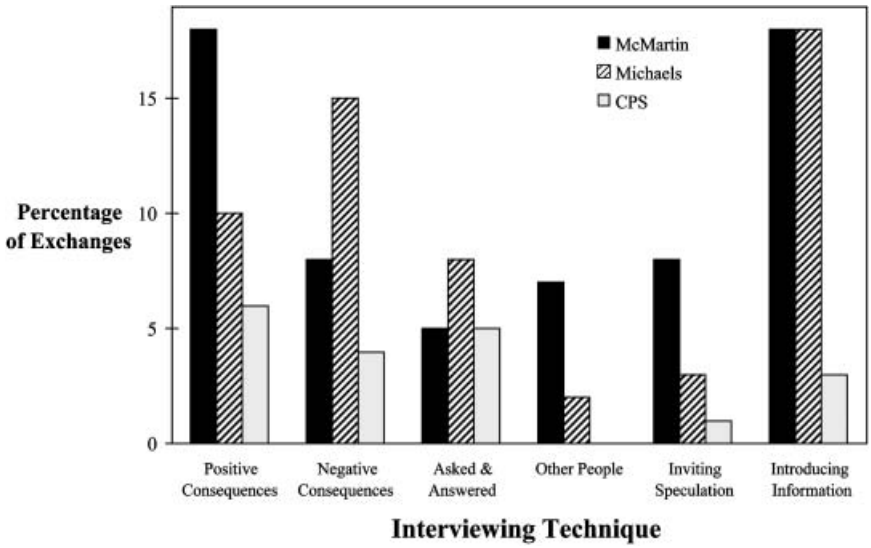

Figure 1. Mean percentage of exchanges in which suggestive interviewing techniques were used. McMartin, Michaels, and CPS cases.

ordinary CPS interviews. Second, analyses also indicated that the style of suggestive interviewing used in the McMartin case differed somewhat from that in the Michaels case.

Suggestiveness of the McMartin Preschool and Kelly Michaels interviews

In this article the scientific scrutiny of the McMartin and Kelly Michaels interviews has come full circle. Based on impressionistic observations, commentators in the late 1980s and early 1990s identified certain child interviewing techniques in these cases as problematic (e.g., Ceci & Bruck, 1993; Nathan, 1988; Rabinowitz, 1990). Laboratory experiments later confirmed that these problematic techniques were suggestive; that is, they could mislead children into making inaccurate statements or false accusations (for reviews, see Ceci & Bruck, 1995; Poole & Lamb, 1998). Finally, the quantitative analyses of the present study have confirmed the impression of the original commentators that the McMartin and Michaels interviews were characterised by intensive and atypical use of these suggestive techniques.

Specifically, the analyses in the present study showed that, in comparison with CPS interviews, the McMartin and Kelly Michaels interviews were characterised by high levels of three suggestive techniques: reinforcement, use of co-witness information, and introducing new information. The

McMartin interviews, but not the Michaels interviews, were also characterised by a fourth suggestive technique, inviting children to speculate. Experiments have established that all four of these techniques—reinforcement (e.g., Garven et al., 1998, 2000), co-witness information (Garven et al., 2000; Leichtman & Ceci, 1995), introducing new information (Ceci & Bruck, 1993), and inviting speculation (Ceci et al., 1994; Schreiber et al., 2001)—can reduce the accuracy of children's statements and even induce them to make false allegations of wrongdoing against other people.

The present findings thus lead to the same conclusion that was reached by the jurors in the McMartin case and by the New Jersey Court of Appeals that overturned Kelly Michaels' conviction—which is to say, that the children's allegations in these cases were elicited by methods that rendered them unreliable. By themselves, these findings cannot establish that the allegations of sexual abuse and Satanic conspiracy in the McMartin and Michaels cases were necessarily false. However, following the Bayesian principle that "extraordinary claims require extraordinary evidence" (Sagan, 1997, p. 60), one can say that the allegations of Satanic abuse in these cases were implausible to begin with (i.e., had a low prior probability) and, in the absence of good evidence to the contrary, are still implausible (i.e., have a low posterior probability of being true). The present findings thus tend to support a sociological analysis by Victor (1998; see also Butler et al., 2001; Cohn, 2000; McGrath, 2001; Sebald, 1995), which concluded that the daycare abuse cases of the 1980s were manifestations of a nationwide "moral panic" regarding Satanism, and not genuine instances of mass sexual abuse.

The present findings cast new light on two well-known studies of the early 1990s in which the allegations in the McMartin case were assumed to be true. First, Gonzalez, Waterman, Kelly, McCord, and Oliveri (1993) reported that in a sample of children in therapy for sexual and ritualistic abuse, 24% recanted their allegations of abuse, a figure two to three times higher than would be expected based on other studies of sexual abuse victims (see review by London, Bruck, Ceci, & Shuman, 2005). However, many or most of the children in the Gonzalez et al. study were from the McMartin case. Based on the present findings, it seems probable that when these children recanted, they were being truthful. That is, they were retracting false allegations that had been improperly extracted from them by suggestive interviews. In the future, therefore, it is questionable whether the results of Gonzalez et al. should be cited as relevant to the disclosure process among genuine sexual abuse victims.

Similarly, in *Behind the Playground Walls*, a book on daycare abuse cases, Waterman et al. (1993; see also Kelley et al., 1993) reported on the substantial psychological problems of the McMartin children, which were supposedly the after-effects of horrendous ritual abuse. However, if the

allegations of most or all of the McMartin children were false, then the psychological problems identified by Waterman et al. cannot be attributed to the effects of ritual sexual abuse. Instead, the children's psychological problems may have been induced by the experience of being falsely diagnosed and treated as sexual abuse victims, accepting these suggested memories as true, or by the stress of repeatedly making false accusations against innocent people who were their friends.

Distinctive styles of suggestiveness in the McMartin and Michaels interviews

As already discussed, the McMartin and Michaels interviews had several features in common. However, the interviewing styles in the two cases also differed in important ways.

McMartin interviews. Compared with the Michaels and CPS interviewers, the interviewers in the McMartin case were much more likely to give children positive reinforcement and invite them to speculate or "pretend". For example, the McMartin interviewers often praised children profusely ("Oh, you're so smart!") for accusing their teachers of sexual or violent wrongdoing. In addition, they often urged children to speculate about what "might" have happened. In most interviews, children were also given puppets and encouraged to "pretend" what might have occurred. Compared with the Michaels and CPS interviewers, the McMartin interviewers were also significantly more likely to ask yes/no questions and less likely to ask "who, what, where" questions (Focused/Specific).

According to the impressions of the present researchers and other observers (e.g., Nathan & Snedeker, 1995), the use of puppets and the steady stream of cheerful praise lent the McMartin interviews a playful "let's have fun" atmosphere that is unusual in child sexual abuse interviews. On average, the McMartin interviews were more than three times as long (1 hour, 14 minutes) as the Michaels (24 minutes) and CPS (22 minutes) interviews. Apparently the McMartin interviewers could sustain children's interest during these long sessions because the children enjoyed playing with the puppets and receiving constant praise.

The McMartin interviewers' use of reinforcement and other suggestive techniques appears to have been deliberate and carefully planned. For example, some interview transcripts show the voice of an experienced interviewer coaching a novice interviewer about how to apply these techniques. In addition, the chief McMartin interviewer (MacFarlane, 1990) published a piece in which she defended the use of reinforcement in child sexual abuse interviews. She contended that there was no good evidence that reinforcement could adversely affect children's accuracy, and

that criticisms of this technique originated from attorneys who were defending accused molestors.

Michaels interviews. Compared with the CPS interviewers, the interviewers in the Kelly Michaels case were significantly more likely to use the technique of Negative Consequences. Specifically, the Michaels interviewers were more likely to express doubt or disbelief when children made a statement that did not fit with the interviewers' preconceptions regarding the supposed abuse. For example, when one small girl denied that she had seen Kelly Michaels make naked children lie on top of each other, the interviewer doubtfully asked "You sure? Positive? If you did see it would you tell me the truth?" (Michaels Interview No. 23C, p. 239).

As the contrast between the "fun" McMartin interviews and the "doubting" Michaels interviews indicates, interviewers may follow multiple routes (different social influence tactics) to reach the same outcome (false allegations). Put another way, there seems to be a toolbox of different techniques from which suggestive interviewers of children can draw, although the present findings suggest that introducing information and reinforcement are especially likely to be used.

Suggestions for practice

Professionals such as police and CPS supervisors, prosecutors, defence attorneys, judges, and expert witnesses sometimes need a straightforward way to determine whether a child interview is unusually suggestive. The findings of the present study suggest four potentially helpful "red flags".

First, some suggestive interviewing techniques virtually *never* occurred in the CPS interviews in the present study. Specifically, CPS interviewers almost never told children what other people had already said (i.e., Other People), asked children to "pretend", or invited them to speculate about what "might" have happened (i.e., Inviting Speculation). Thus, any use of these techniques can be regarded as a "red flag" that something unusual, if not improper, has transpired in a child sexual abuse interview.

Second, the CPS interviewers seldom used positive reinforcement (called "positive consequences" in the present study) except at the very beginning of interviews to build rapport ("My, what pretty eyes you have!") or at the very end ("Thanks for talking with me"). In addition, Poole and Lamb (1998) have recommended the use of reinforcement to encourage children to make narrative statements during the early parts of an interview, before the topic of abuse has been introduced. However, other uses of praise or

promises during a child sexual abuse interview are unusual and may be an indication of suggestiveness.

Third, the CPS interviewers in the present study rarely introduced new outside information into interviews and virtually never gave children information about the alleged abuse or perpetrator. Thus, interviewer statements or questions that provide a child with information regarding a supposed perpetrator or the circumstances of abuse may also be indicators of suggestiveness. Investigative interview guidelines aimed at providing practitioners with non-suggestive interviewing techniques are available (APSAC, 2002; Poole & Lamb, 1998).

Fourth, although CPS interviewers routinely gave mild negative feedback when a child committed a slip of the tongue or made a minor factual error (''Didn't you say before that it was your grandmother's house, not your aunt's?''), they virtually never expressed doubt that a child was telling the truth. Thus, such expressions of doubt or disbelief by an interviewer may also be indicators of suggestiveness.

Limitations of the study

There are several limitations to the present study. First, the interviewing techniques examined here do not represent an exhaustive set of all possible social influence techniques that may have been used in these or other interviews. Future research may well detect additional techniques that were used in the McMartin and Michaels interviews, or that have been used by child interviewers in other cases.

Second, this study focused on only one source of social influence—interviewers' questions—that may have affected children's statements in the McMartin and Michaels cases. However, the children were subject to a broad range of powerful influences, such as the social contagion that spread among families in the McMartin case (''cross-germination'' of information), which were not examined in this study. The social context outside the interview is likely to have directly influenced children in these cases and perhaps also increased their susceptibility to suggestive interviewing techniques.

A third limitation of the study is its reliance on interview transcripts rather than audiotapes or videotapes. As indicated in the Method section, the CPS transcripts, which were independently checked by two members of our research team, and the McMartin transcripts, which were examined for errors by both both prosecutors and defence teams, were probably very accurate. However, the Michaels transcripts, which were apparently transcribed by a New Jersey police department and not independently checked, may well be of lower quality. Furthermore, even assuming complete accuracy, the information available from these transcripts is

incomplete because they fail to record nonverbal behaviour, such as head nods by the child or (in the McMartin case) use of puppets and anatomically detailed dolls by the interviewer. An examination of nonverbal behaviours would probably reveal nuances that the present study failed to detect.

Fourth, because the CPS transcripts in this study represented a cross-section of interviews from a single agency in a single city, it might well be wondered whether they were representative of those from other child protection and police agencies across the United States. Fortunately a study by Warren, Garven, Walker, and Woodall (2000) has partially addressed this issue. Using the same scoring system as the present study, Warren et al. examined 42 transcripts of videotaped CPS sexual abuse interviews collected elsewhere in the US from children 2 to 13 years old (mean age=6). In their sample, suggestive techniques were used by CPS interviewers in the following proportion of exchanges: Positive Consequences, .03; Negative Consequences, .04; Asked and Answered, .07; Other People, .00; Inviting Speculation, .00. Comparison of these numbers with Table 3 and Figure 1 reveal that the rates of suggestive tactics reported by Warren et al. were strikingly similar to the rates for CPS transcripts in the present study, and dissimilar from those for the McMartin and Michaels interviews. These findings provide reassurance that the CPS interviews used in the present study are probably typical of those conducted elsewhere in the country. To say that these interviews were typical, however, is not to say that they were exemplary. To the contrary, a few of the CPS interviews contained clearly inappropriate use of suggestive techniques, especially with younger children. Research has shown that suggestive questioning can have serious adverse effects on children's accuracy, and that the negative impact is greatest among preschoolers (Ceci & Bruck, 1995; Poole & Lamb, 1998).

A final limitation of the study concerns the age of the children in the CPS sample, who were somewhat older than the McMartin and Michaels children. Correlational analyses reported in the Results showed that children's age was unrelated to most interview characteristics in the CPS sample. However, these same analyses indicated that when CPS interviewers questioned very young children (particularly those aged 5 years or younger), they tended to ask more Yes/No and fewer Focused questions, and to make substantially more use of Negative Consequences and Asked and Answered. These findings suggest that the high rate of Yes/No questions and low rate of Focused questions in the McMartin interviews, as well as the high rate of Negative Consequences in the Michaels interviews, may have reflected the younger age of children in these samples. However, it remains doubtful whether age is the explanatory factor for the between-group differences. For

example, in the study by Warren et al. (2000) described earlier, children were on average 2 years younger than in the present CPS sample, but the rate of Negative Consequences was the same, and the rate of Asked and Answered was lower, than the rates reported here. Additional research is necessary before the relationship between children's age and interview characteristics can be clarified.

Contributions to a science of social influence

In closing, the present study contributes to our understanding of social influence in several ways. First, it provides a window into how social influence techniques were used in a high-stakes real-world situation. The McMartin and Kelly Michaels cases commanded national attention and led to the prosecution and imprisonment of innocent citizens. Ordinary children were drawn into a maelstrom of lurid accusations and community-wide panic. The present study clarifies how the social influence tactics of interviewers helped to set these events in motion.

Second, this study provides a partial taxonomy for the social influence tactics used in forensic interviews of children. Interestingly, similar taxonomies for police interrogations of criminal suspects, as proposed by Leo (1996) and Kassin and Gudjonsson (2004), share much in common with the one proposed here. For example, Kassin and Gudjonsson's ''External Pressure'' is similar to our ''Negative and Positive Consequences'', because both involve the use of rewards, punishment, and promises to elicit statements. Similarly, their ''Perception of Proof'' resembles our ''Other People'', since both tactics involve the use of accurate or inaccurate outside information to induce belief and compliance from the person being questioned. If social influence tactics have the potential to elicit confessions (either true or false) from adults, they can certainly induce children to make false sexual abuse accusations.

Third, the present study contributes to the understanding of social influence by introducing a set of scales and a method of analysis that can be used in future research on child interviews. Most of these scales exhibited a high level of inter-rater reliability, and their construct validity was generally supported by findings of significant between-group differences. As the study by Warren et al. (2000) has shown, these scales can be fruitfully applied by new researchers in new samples of children.

Finally, this study has shown how different social influence techniques may be used to arrive at the same outcome. The present findings suggest that while some techniques, such as reinforcement and introducing information, may be common to all or most suggestive child interviews, other idiosyncratic tactics, such as Other People and Inviting Speculation, may be adopted by specific interviewers. Future research is needed to

identify and understand the full range of social influence tactics that individual interviewers have applied in child forensic interviews.

Manuscript received 5 May 2005
Manuscript accepted 16 September 2005

REFERENCES

Acocella, J. (1999). *Creating hysteria: Women and multiple personality disorder*. San Francisco: Jossey Bass.

American Professional Society on the Abuse of Children (APSAC) (2002). *Guidelines for psychosocial evaluation of suspected sexual abuse in young children* (Rev. ed.). Chicago, IL: APSAC.

Bernet, W., & Chang, D. K. (1997). The differential diagnosis of ritual abuse allegations. *Journal of Forensic Science*, *42*, 32–38.

Bottoms, B. L., & Davis, S. L. (1997). The creation of satanic ritual abuse. *Journal of Social and Clinical Psychology*, *16*, 112–132.

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.

Bruck, M., & Ceci, S. J. (1993). *Amicus brief for the case of* State of New Jersey v. Michaels *presented by Committee of Concerned Social Scientists*. Supreme Court of New Jersey, Docket # 36,633. [Reprinted in *Psychology, Public Policy, and Law*, *1*, 1995, 272–322.]

Butler, E. W., Fukurai, H., Dimitrius, J-E., & Krooth, R. (2001). *Anatomy of the McMartin child molestation case*. Lanham, MD: University Press of America.

Cassel, W. S., Roebers, C. E. M., & Bjorklund, D. F. (1996). Developmental patterns of eyewitness responses to repeated and increasingly suggestive questions. *Journal of Experimental Child Psychology*, *61*, 116–133.

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

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom*. Washington, DC: American Psychological Association.

Ceci, S. J., Huffman, M. L. C., Smith, E., & Loftus, E. F. (1994). Repeatedly thinking about a non-event: Source misattributions among preschoolers. In K. Pezdek & W. P. Banks (Eds.), *The recovered memory–false memory debate* (pp. 225–244). San Diego, CA: Academic Press.

Ceci, S. J., & Loftus, E. F. (1994). "Memory work": A royal road to false memories? *Applied Cognitive Psychology*, *8*, 351–364.

Ceci, S. J., Ross, D. F., & Toglia, M. P. (1987). Suggestibility of children's memory: Psycholegal implications. *Journal of Experimental Psychology: General*, *116*, 38–49.

Cialdini, R. B. (2001). *Influence*. Boston, MA: Allyn & Bacon.

Cohn, N. (2000). *Europe's inner demons* (Rev. ed.). Chicago: University of Chicago Press.

Dale, P. S., Loftus, E. F., & Rathburn, L. (1978). The influence of the form of the question on the eyewitness testimony of pre-school children. *Journal of Psycholinguistic Research*, *7*, 269–277.

Davies, G. M., Wilson, C., Mitchell, R., & Milsom, J. (1995). *Videotaping children's evidence: An evaluation*. London: Home Office.

Dent, H., & Stephenson, G. (1979). An experimental study of the effectiveness of different techniques of questioning child witnesses. *British Journal of Social and Clinical Psychology*, *18*, 41–51.

Eberle, P., & Eberle, S. (1993). *The abuse of innocence*. Buffalo, NY: Prometheus Books.

Ettinger, R. H., Crooks, R. L., & Stein, J. (1994). *Psychology: Science, behavior, and life*. Fort Worth, TX: Harcourt Brace.

Faller, K. C. (1996). Interviewing children who may have been abused: A historical perspective and overview of controversies. *Child Maltreatment*, *1*, 83–95.

Garner, B. A. (2004). *Black's law dictionary* (8th ed.). St. Paul, MN: West.

Garry, M., Manning, C. G., Loftus, E. F., & Sherman, S. J. (1996). Imagination inflation: Imagining a childhood event inflates confidence that it occurred. *Psychonomic Bulletin & Review*, *3*, 208–214.

Garven, S., Wood, J. M., & Malpass, R. S. (2000). Allegations of wrongdoing: The effects of reinforcement on children's mundane and fantastic claims. *Journal of Applied Psychology*, *85*, 38–49.

Garven, S., Wood, J. M., Malpass, R. S., & Shaw, J. S. (1998). More than suggestion: The effect of interviewing techniques from the McMartin Preschool case. *Journal of Applied Psychology*, *83*, 347–359.

Gonzalez, L. S., Waterman, J., Kelly, R. J., McCord, J., & Oliveri, M. K. (1993). Children's patterns of disclosures and recantations of sexual and ritualistic abuse allegations in psychotherapy. *Child Abuse and Neglect*, *17*, 281–289.

Green, R. (1992). *Sexual science and the law*. Cambridge, MA: Harvard University Press.

Hershkowitz, I., Lamb, M. E., Sternberg, K. J., & Esplin, P. W. (1997). The relationships among interviewer utterance type, CBCA scores, and the richness of children's responses. *Legal and Criminological Psychology*, *2*, 169–176.

Hertenstein, M., & Trott, J. (1993). *Selling Satan: The tragic history of Mike Warnke*. Chicago: Cornerstone Press.

Hicks, R. D. (1991). *In pursuit of Satan*. Buffalo, NY: Prometheus.

Home Office (2002). *Achieving best evidence in criminal proceedings: Guidance for vulnerable or ntimidated witnesses, including children*. London: Her Majesty's Stationery Office. [Also available on the World Wide Web at http://www.cps.gov.uk/publications/prosecution/bestevidencevol1.html Access date 14 July 2004.]

Hutcheson, G., Baxter, J. S., Telfer, K., & Warden, D. (1996). Child witness statement quality: Question type and errors of omission. *Law and Human Behavior*, *19*, 631–648.

Hyman, I. E., Husband, T. H., & Billings, F. J. (1995). False memories of childhood experiences. *Applied Cognitive Psychology*, *9*, 181–197.

Hyman, I. E., & Pentland, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of Memory and Language*, *35*, 101–117.

Kassin, S. M., & Gudjonsson, G. H. (2004). The psychology of confessions: A review of the literature and issues. *Psychological Science in the Public Interest*, *5*, 33–67.

Kelley, S. J., Brant, R., & Waterman, J. (1993). Sexual abuse of children in day care centers. *Child Abuse & Neglect*, *17*, 71–89.

Lamb, M. E., Hershkowitz, I., Sternberg, K. J., Esplin, P. W., Hovav, M., Manor, T., et al. (1996). Effects of investigative utterance types on Israeli children's responses. *International Journal of Behavioral Development*, *19*, 627–637.

Lamb, M. E., Sternberg, K. J., & Esplin, P. W. (1995). Making children into competent witnesses: Reactions to the Amicus Brief *In re Michaels*. *Psychology, Public Policy, and Law*, *1*, 438–449.

Lamb, M. E., Sternberg, K. J., & Esplin, P. W. (1998). Conducting investigative interviews of alleged sexual abuse victims. *Child Abuse & Neglect*, *22*, 813–823.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Lanning, K. (1991). Ritual abuse: A law enforcement perspective. *Child Abuse & Neglect*, *15*, 171–173.

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

Leo, R. A. (1996). Inside the interrogation room. *The Journal of Criminal Law and Criminology*, *86*, 266–303.

Loftus, E. F., & Davies, G. M. (1984). Distortions in the memory of children. *Journal of Social Issues*, *40*, 51–67.

Loftus, E., & Ketcham, K. (1994). *The myth of repressed memory: False memories and allegations of sexual abuse*. New York: St. Martin's Press.

London, K., Bruck, M., Ceci, S. J., & Shuman, D. W. (2005). Disclosure of child sexual abuse: What does the research tell us about the ways that children tell? *Psychology, Public Policy, & Law*, *11*, 194–226.

Lyons, T. D. (1995). False allegations and false denials in child sexual abuse. *Psychology, Public Policy, and Law*, *1*, 429–437.

MacFarlane, K. (1990). Should alleged child sexual abuse victims be reinforced in forensic interviews? *Violence Update*, *1*(4), 6–7.

Martinez, Y. C. (1999). *Forms of question in the McMartin Preschool and Kelly Michaels cases.* Unpublished MA Thesis, University of Texas at El Paso.

McGrath, M. (2001). *Demons of the modern world*. Buffalo, NY: Prometheus.

McLaurin, K. A. (2000). *Suggestiveness of child interviews in the McMartin Preschool and Kelly Michaels cases.* Unpublished MA thesis, University of Texas at El Paso.

Memon, A., & Vartoukian, A. (1996). The effects of repeated questioning on young children's eyewitness testimony. *British Journal of Psychology*, *87*, 403–415.

Myers, J. E. B., Saywitz, K. J., & Goodman, G. S. (1996). Psychological research on children as witnesses: Practical implications for forensic interviews and courtroom testimony. *Pacific Law Journal*, *28*, 3–91.

Nathan, D. (1988). *Village Voice*, 2nd August, pp. 31–39.

Nathan, D., & Snedeker, M. (1995). *Satan's silence*. New York: Basic Books.

Noblitt, J. R., & Perskin, P. S. (2000). *Cult and ritual abuse: Its history, anthropology, and recent discovery in contemporary America* (Rev. Ed.). Westport, CT: Praeger.

Orbach, Y., Hershkowitz, I., Lamb, M. E., Sternberg, K. J., Esplin, P. W., & Horowitz, D. (2000). Assessing the value of structured protocols for forensic interviews of alleged child abuse victims. *Child Abuse & Neglect*, *24*, 733–752.

Ofshe, R., & Watters, E. (1994). *Making monsters: False memories, psychotherapy, and sexual hysteria*. New York: Scribners.

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

Poole, D. A., & White, L. (1991). Effects of question repetition on the eyewitness testimony of children and adults. *Developmental Psychology*, *27*, 975–986.

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

Pratkanis, A. R. (in press). Social influence analysis: An index of tactics. In A. R. Pratkanis (Ed.), *The science of social influence: Advances and future progress*. Philadelphia: Psychology Press.

Rabinowitz, D. (1990). Out of the mouths of babes and into a jail cell. *Harper's Magazine*, *May*, 52–63.

Rabinowitz, D. (2003). *No crueler tyrannies*. New York: Free Press.

Reed, L. D. (1996). Findings from research on children's suggestibility and implications for conducting child interviews. *Child Maltreatment*, *1*, 105–120.

Reinhold, R. (1990). How lawyers and media turned the McMartin case into a tragic media circus. *New York Times*, 25 January, p. 1D.

Sagan, C. (1997). *Billions and billions: Thoughts on life and death at the brink of the millenium*. New York: Random House.

Sakheim, D. K., & Devine, S. E. (1994). *Out of darkness: Exploring satanism and ritual abuse*. New York: Wiley.

Saywitz, K. J., & Geiselman, R. E. (1998). Interviewing the child witness. Maximizing completeness and minimizing error. In S. J. Lynn & K. M. McConkey (Eds.), *Truth in memory*. New York: Guilford Press.

Schreiber, N. (2000). Interviewing techniques in sexual abuse cases – a comparison of a day-care abuse case with normal abuse cases. *Swiss Journal of Psychology*, *59*, 196–206.

Schreiber, N., & Parker, J. F. (2004). Inviting witnesses to speculate: Effects of age and interaction on children's recall. *Journal of Experimental Child Psychology*, *89*, 31–52.

Schreiber, N., Wentura, D., & Bilsky, W. (2001). "What else could he have done?" Creating false answers in child witnesses by inviting speculation. *Journal of Applied Psychology*, *86*, 525–532.

Sebald, H. (1995). *Witch-children: From Salem witch-hunts to modern courtrooms*. Amherst, NY: Prometheus.

Shrout, P. E. (1998). Measurement reliability and agreement in psychiatry. *Statistical Methods in Medical Research*, *7*, 301–317.

Siegal, M., Waters, L. J., & Dinwiddy, L. S. (1988). Misleading children: Causal attributions for inconsistency under repeated questioning. *Journal of Experimental Child Psychology*, *45*, 438–456.

Sinason, V. (1996). *Treating survivors of satanist abuse*. New York: Routledge.

Smith, M., & Pazder, L. (1980). *Michelle remembers*. New York: Congdon & Lattes.

Spanos, N. (1996). *Multiple identities and false memories: A sociocognitive perspective*. Washington, DC: American Psychological Association.

Sternberg, K. J., Lamb, M. E., Hershkowitz, I., Esplin, P. W., Redlich, A., & Sunshine, N. (1996). The relation between investigative utterance types and the informativeness of child witnesses. *Journal of Applied Developemental Psychology*, *17*, 439–451.

Sternberg, K. J., Lamb, M. E., Hershkowitz, I., Yudilevitch, L., Orbach, Y.,Esplin, P. W., et al. (1997). Effects of introductory style on children's abilities to describe experiences of sexual abuse. *Child Abuse & Neglect*, *21*, 1133–1146.

Sternberg, K. J., Lamb, M. E., Orbach, Y., Esplin, P. W., & Mitchell, S. (2001). Use of a structured investigative protocol enhances young children's responses to free recall prompts in the course of forensic interviews. *Journal of Applied Psychology*, *85*, 997–1005.

Strok, R. (1997). *Social influence and reinforcement techniques in the McMartin and Kelly Michaels cases.* Unpublished MA thesis, University of Texas at El Paso.

Summit, R. C. (1994). The dark tunnels of McMartin. *The Journal of Psychohistory*, *21*, 397–416.

Tavris, C. (1997). A day-care witch hunt tests justice in Massachusetts. *Los Angeles Times*, 11 April, p. B9.

Tharp, R. G., & Wetzel, R. J. (1969). *Behavior modification in the natural environment*. New York: Academic Press.

Timnick, L., & McGraw, C. (1990). McMartin verdict: Not guilty. *Los Angeles Times*, 19 January, p. A1.

Underwager, R., & Wakefield, H. (1990). *The real world of child interrogations*. Springfield, IL: Charles C. Thomas.

Velarde, L. (1997). *Estimating the length of a child sexual abuse interview from the transcript alone.* Unpublished MA thesis, University of Texas at El Paso.

Victor, J. S. (1993). *Satanic panic: Creation of a contemporary legend*. Chicago: Open Court.

Victor, J. S. (1998). Moral panics and the social construction of deviant behavior: A theory and application to the case of ritual child abuse. *Sociological Perspectives*, *41*, 541–565.

Warnke, M. (1972). *The satan-seller*. South Plainfield, NJ: Logos International.

Warren, A., Garven, S., Walker, N., & Woodall, C. (2000, March). *Setting the record straight: How problematic are "typical" child sexual abuse interviews?* Paper presented at the biennial conference of the American Psychology-Law Society. New Orleans, LA.

Warren, A. R., & McGough, L. S. (1996). Research on children's suggestibility: Implications for the investigative interview. *Criminal Justice & Behaviour*, *23*, 269–303.

Warren, A. R., Perry, N. W., Nelson, D. H., Porter, C., Elliott, K., Komori, L., et al. (1995). *Interviewing children: Questions of structure and style.* Poster presented at the Third National Colloquium of the American Professional Society on the Abuse of Children, Tucson, AZ.

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

Waterman, J., Kelly, R. J., Oliveri, M. K., & McCord, J. (1993). *Behind the playground walls: Sexual abuse in preschools*. New York: Guilford Press.

Wilkerson, T., & Rainey, J. (1990). Tapes of children decided the case for most jurors. *Los Angeles Times*, 19 January, p. A1.

Wood, J. M., & Garven, S. (2000). How sexual abuse interviews go astray: Implications for prosecutors, police, and child protection services. *Child Maltreatment*, *5*, 109–118.

Wood, J. M., McClure, K. A., & Birch, R. A. (1996). Suggestions for improving interviews in child protection agencies. *Child Maltreatment*, *1*, 223–230.

Wyatt, W. J. (2002). What was under the McMartin Preschool? A review and behavioural analysis of the "tunnels" find. *Behaviour and Social Issues*, *12*, 29–39.

Yuille, J. C., Hunter, R., Joffe, R., & Zaparnuik, J. (1993). Interviewing children in sexual abuse cases. In G. S. Goodman & B. L. Bottoms (Eds.), *Child victims, child witnesses. Understanding and improving testimony* (pp. 95–115). New York: Guilford Press.