

The Accuracy of Mothers' Memories of Conversations With Their Preschool Children

Maggie Bruck
McGill University

Stephen J. Ceci
Cornell University

Emmett Francoeur
Montreal Children's Hospital

Mothers interviewed their 4-year-old children about a structured play activity that occurred minutes earlier while the mothers had been absent. Half of the mothers were forewarned that this was a memory experiment and that they should try to remember the meaning and the exact words used when they interviewed their children about the play activity; the other mothers were not given any forewarning. Approximately 3 days later, mothers' recall and recognition memories for aspects of the interview were assessed. Forewarnings about the memory test did not improve mothers' performances on any measures. As anticipated, mothers' memories for meaning was better than their memories for the exact wording or structure of the conversation. Importantly, they had difficulty recalling how the information was elicited from their children, whether their children's statements were spontaneous or prompted, or whether specific utterances were spoken by themselves or by their children. The authors discuss these results in terms of their practical ramifications for hearsay testimony.

In many situations, adults provide hearsay testimony about their conversations with children. This can occur in informal situations, for example, when one parent relates elements of a conversation with their child to another parent. Hearsay testimony is also given in more formal

situations: For example, when a child is a suspected witness or victim of a crime, an adult may make the first report to the authorities on the basis of an interview or conversation with the child. In courts of law, hearsay exceptions are sometimes made so that adults (e.g., parents, mental health professionals, medical professionals) are allowed to testify about statements made to them by a child. McGough (1994) has provided a historical and philosophical rationale for accepting such hearsay statements made by a child to an adult.

Maggie Bruck, Department of Psychology, McGill University, Montreal, Quebec, Canada; Stephen J. Ceci, Department of Family Studies and Human Development, Cornell University; Emmett Francoeur, Department of Pediatrics, Montreal Children's Hospital, Montreal, Quebec, Canada.

This research was supported by grants from the Natural Sciences and Engineering Research Council and the National Science Foundation. We thank Laila Brown, Kerri Ritchie, Lynn Herscovitch, and Gayle DeDe for their assistance in data collection and analysis.

Correspondence concerning this article should be addressed to Maggie Bruck, Department of Psychology, McGill University, 1205 Docteur Penfield Avenue, Montreal, Quebec, Canada H3A 1B1. Electronic mail may be sent to bruck@hebb.psych.mcgill.ca.

The untested presumption underlying the acceptance of all hearsay testimony is that in the absence of direct malice or motives to lie, adults can accurately recall earlier conversations with children. However, despite the absence of empirical data on this issue, there are good reasons for suspecting that adults cannot accurately recall the contents or structure of conversations with their children. Unlike conversations with other adults, conversations with children require that the adult invest considerable attention to pacing and formatting the exchanges while simultaneously trying to

keep the child on task. This attentional investment may diminish memory for certain aspects of a conversation. It may be considerably less demanding to converse with another adult because that person is more capable of holding up their own end of the conversation.

The present study represents a first step in empirically addressing the question of whether adults can accurately report their conversations with young children. We examined two aspects of mothers' memories of their conversations with their preschool children—gist (content) memory and verbatim (also termed *structure of conversation*) memory. As we argue below, although it is important to accurately recall the content of a conversation, accurate memory of its structure is necessary for evaluating the statements of young children.

Adult Recall of Statements Made by Other Adults

There are two related areas of research that are relevant to the design of studies of adult recall of statements made by children. The first concerns adult memory for prose. The major issue addressed in these studies is whether the actual words of sentences are remembered (verbatim memory) or whether only meaning (gist) is preserved. In these studies, participants first hear a target sentence such as, "A wealthy manufacturer sought out the young inventor." Next, they are given a recognition test that requires identification of a test item that is identical (in terms of words and meaning) to the target. Three different types of test items are usually presented: a sentence that is identical to the target ("A wealthy manufacturer sought out the young inventor."), a paraphrase that preserves the meaning but not the words of the target ("The young inventor was sought out by a wealthy manufacturer."), and a sentence with a different meaning from the target ("The young inventor sought out the wealthy manufacturer.").

Within seconds of the original presentation, adults have poor verbatim but good gist recall of the target sentence. They cannot differentiate the original target sentence from its paraphrase (Garrod & Trabasso, 1973; Sachs, 1967); that is, they are as likely to select the original item as the gist-preserving item. In contrast to poor verbatim recall, adults have good recall of the original

meaning of the sentence; that is, they rarely select the test item with a different meaning. It has thus been argued that very shortly after verbal material is encoded and semantically represented, the surface details fade from the memory, leaving only gistified traces (Bransford & Franks, 1971; Gernsbacher, 1985) that can be accurately accessed long after the original conversation has ended.

The second area of research focuses on adult memory for sentences within connected discourse, such as conversations, classroom lectures, or soap operas. The results of this line of study demonstrate that under certain circumstances, verbatim memory for sentences in discourse may be more robust than memory for these same sentences presented in isolation. For example, Kintsch and Bates (1977) found that 2 and 5 days after listening to a lecture, students were able to discriminate sentences used in that lecture from paraphrases of these sentences. Extraneous statements (e.g., jokes, comments) were best recognized. Thus, in this study, verbatim memory was at least as good as gist memory. Similarly, Keenan, MacWhinney, & Mayhew (1977) found that participants' verbatim memories of a luncheon discussion were best for figures of speech, jokes, and insults—utterances that were deemed to have high pragmatic value or what the authors referred to as *high interactional content*. Keenan and colleagues (MacWhinney, Keenan, & Reinke, 1982) argued that these sentences were memorable because "these surface forms carry information that is relevant to the subsequent dynamics of the speaker-listener relationship" (p. 309).

Although these results suggest that long-term verbatim memory may be quite robust for pragmatically important utterances, it is important to point out that in these same studies, verbatim memory was quite poor for sentences of low pragmatic value. Also, it is important to point out that the conclusions about the robustness of verbatim and gist memories are based on recognition task performance rather than on recall accuracy. When participants are asked for free recall of prior conversations, performance is quite poor. For example, Stafford, Burggraf, & Sharkey (1987) found that immediately after a conversation, participants could recall only about 10% of its content, and 1 month later, this figure had dropped to 4%. Kintsch and Bates (1977) asked

students to write down everything they could possibly remember from a lecture given 5 days previously; they were told to include jokes as well as seemingly irrelevant material. The students' protocols were sparse, containing an average of three to four sentences.

Problems Generalizing From Adult Research on Sentence Recognition

There are several problems in attempting to generalize the results of studies of adults' memories for lectures and television shows to the case of parents' memories of conversations with their children. A primary difficulty is that the existing literature does not explicitly examine memory for the structure of conversations. That is, for the most part, it seems that the (recognition) tasks only required participants to make judgments about the accuracy of individual sentences presented in isolation. Although individual sentences are a component of conversations, their isolated presentation does not fully exploit some of the important parameters of conversations. Namely, conversations are fluid interactions between two or more participants in which the crucial units include not individual utterances but the verbal exchanges in which these utterances are couched.

The present article focuses on memory for a specific type of conversation: an interview. An *interview* is a verbal interaction between at least two people in which one of the participants (the interviewer) has the goal of obtaining specific information from the other participant (the interviewee). As such, interviews can be carried out in highly structured situations (e.g., job interviews, mental health assessments, forensic investigations) as well as in less structured settings (e.g., when a parent attempts to find out what his or her child did in school).

Memory of an interview therefore involves more than recall of the meaning or wording of single sentences; it also requires recall of its structure—how the information was obtained and, particularly, memory for the types of devices that were used to obtain that information. In the case of adults' conversations with children, some important elements of structure include the types and sequences of elicitation techniques that were used to obtain information (e.g., repeated questions, probes, suggestions) as well as children's responses to these techniques. Accurate memory

for these conversational features are important because past research has revealed that children's reports are most reliable when they are elicited by open-ended questions that are not repeated; conversely, children's reports are least reliable when they are elicited by specific leading questions, especially when these questions have been repeated within and across sessions (Ceci & Bruck, 1995; Poole & White, 1991).

If adults who provide hearsay testimony cannot accurately recall the interrogative context in which children's statements were elicited, then this might render the reliability of the child's statement both unknown and unknowable. In short, adults must be able to reconstruct the context of the interview that led to a child's statement, not merely what that statement was. Otherwise, there is no basis for evaluating the validity of the child's statements because it is unknown if these were highly prompted or coached in some way.

For example, a parent or therapist may testify or write in a report that a child said that he had been touched inappropriately by a defendant, but the evaluation of the reliability of this statement differs depending on whether it was elicited in Context A or Context B.

Context A

Adult: What happened at school today?

Child: The man with a moustache touched me.

Context B

Adult: What happened at school today?

Child: Nothing.

Adult: Tell me.

Child: Nothing happened.

Adult: Did you see the man?

Child: No.

Adult: Did you see the man with the moustache?

Child: Yes.

Adult: Did he touch you?

Child: What?

Adult: He touched you didn't he?

Child: Yeah, the man with a moustache touched me.

In this example, it might be possible that the adult accurately remembers that the child said, "The man with a moustache touched me," but it is also important to recall how this utterance actually fit into the structure of the conversation: Was it a spontaneous utterance? Was the statement made in response to several questions? How were the questions worded?

Another important feature of the structure of

conversations involves the determination of role assignment—specifically, who said what. (Role assignment is a variant of the ability to monitor external sources; see Johnson, Hashtroudi, & Lindsay, 1993.) In other words, one could have excellent memory for the gist of a conversation, and even for the exact words used in a sentence, but nevertheless be inaccurate as to whether Person A or Person B was the speaker.

There are a few studies on role assignment accuracy in adult–adult conversations. Studies of cryptomnesia (wherein the participant inadvertently claims to have generated an idea or sentence that was generated by another) are the most pertinent. Cryptomnesia data suggest that participants in an interaction may come to falsely claim a response as their own, when in fact it was made by someone else (e.g., Brown, Jones, & Davis, 1995; Brown & Murphy, 1989; Marsh & Bower, 1993). One might hypothesize that it is easier to make accurate judgments about speakers and listeners in adult–child conversations because both the language styles and contents of the two participants are so different, making the utterances less confusable. In general, the more that two sources can be perceptually differentiated (differing gender, age), the easier it will be to separate their respective sources (Johnson et al., 1993). As reasonable as this hypothesis sounds, there are no available data to test it.

However, memory for adult–child conversations may be poorer than memory for adult–adult conversations for a number of interrelated reasons. Specifically, the structure of adult–child conversations differs from that of adult–adult conversations in that the former often requires the adult participant to provide much of the interactional structure, to flexibly use a number of strategies in order to elicit information, and to keep the child participating in the interaction (see Ceci & Bruck, 1995 for a review). Specifically, because young children provide little informative information in response to open-ended questions, such as “Tell me what happened at school today,” adults must use a variety of strategies to sustain the conversation until they obtain the desired information. Therefore, adults often ask many questions that serve as probes or even prompts to keep the child on task (e.g., Fivush, 1993). Importantly, some of the questions may be leading, incorporating information that adults think might be important or reflecting adults’

notions of what happened. If the child does not respond satisfactorily to the questions, the questions may be repeated, or at times, the adult might actually provide the response (for examples of these interviewing techniques in legal cases, see Ceci & Bruck, 1995).

At the same time, the adult participant must carefully guide the interview with children, keeping it on track and not allowing the child’s attention to be diverted for too long. Because of this added effort and complexity, it may be more difficult to remember the gist of the conversation as well as the exact wording of the utterances than would be the case when college students listen to a lecture or try to recall dialogue from a soap opera. In these latter cases, there is no need to compose on-line strategies for eliciting statements from the lecturer or actor; therefore their attentional resources can be devoted entirely to the task of processing the speaker’s message.

Although we have suggested a number of reasons why adult recall of conversations with young children might be poor, there is also an alternative hypothesis: Some adults, especially caretakers who are very familiar with the idiosyncrasies and nuances of their children’s language, may have particularly good memories of conversations with their children. Thus, adults’ memories of their conversations with children might be good despite the added attentional resources required to elicit them. Because empirical evidence is lacking, either hypothesis could be true.

One important consideration in designing a study on memory for conversation is the possibility that verbatim or gist memory might vary as a function of the importance of remembering the conversation. If the conversation is deemed to be unimportant, there may be little effort to process verbal material so that it is well represented in memory for future recall. There are times, however, when a participant may make an effort to try to remember the exact words because of their potential importance. Intuitively, it seems that the latter strategy might increase verbatim as well as gist memory. Surprisingly, there are few data in the literature on memories of conversations to address this issue. Stafford and Daly (1984) found that forewarning participants to remember an upcoming conversation improved their short-term free recall of that conversation. One month later, when they were again asked to recall their conversations, there was still an advantage of

forewarning (Stafford et al., 1987). However, in another study, Kausler and Hakami (1983) found no advantage of forewarning when participants' short-term memories of a conversation were evaluated. In light of these inconsistent results, we examined whether forewarning mothers to remember their conversations with their children would increase their recall and recognition of that conversation.

In the present study, mothers interviewed their 4-year-old children about a play activity. Some mothers were forewarned that this was a memory experiment and that we would later ask them to recall exactly what was said during their interview. Other mothers were not given this forewarning; they were merely told that this was a study of mother-child interaction. Approximately 3 days later, the mothers' memories of their conversations during the interview were assessed. As in previous studies, we assessed mothers' gist memories relative to their verbatim memories. The particular focus on verbatim memory was the degree to which mothers could remember conversational exchanges that highlighted the spontaneous versus close-ended questioning of their children and the degree to which they were able to remember the speaker of utterances (i.e., their children or themselves). We expected that mothers would have good recognition memory of the utterances within the conversation—after all, we reasoned, parents are very familiar with their children's conversational style and might readily be able to detect typical from atypical interactional exchanges. On the other hand, perhaps these elements of organization do not become firmly represented in adult memory, and only the gist remains—particularly under conditions when much of an adult's attentional effort goes into keeping the conversation with a child going and focused on the topic at hand. Finally, we predicted that the forewarned mothers would have better memory for the gist and perhaps for verbatim aspects of the conversations.

Method

Participants

Twenty-four mothers with preschool children between the ages of 3 and 5 years were recruited for the study. Most of the children were clients in a pediatric practice of one of the authors (Emmett

Francoeur) and were from middle-class backgrounds.

Procedures

The mothers were asked to participate in a study on mother-child conversation. They were told that participation would require a 1-hr visit with their child to our university laboratory and a follow-up session several days later with the mother only.

Child play. Mothers and their children arrived in the laboratory. As soon as the children felt comfortable, they were separated from their mothers to play with a research assistant in another room. The play period was structured, lasting approximately 20 min. There were six major activities and one surprise event: (a) The assistant and the child placed stickers in a book and the assistant placed a sticker on the child's cheek, (b) the child put on a smock, (c) the child and the assistant used a plastic template to color a picture of an animal, (d) the child made and cut playdough hair with Sesame Street characters, (e) the assistant performed a magic trick, (f) the assistant sang a song, and (g) there was also a surprise event: During the coloring activity, a male confederate came into the room looking for his firehat, which he found. He then noticed the assistant using his favorite crayon and asked for it back. The assistant asked if she could keep it until the child left. The confederate acquiesced, blew the child a kiss, and left the room. As soon as the play session was finished, the child rejoined his or her mother.

Instructions to mother. While the child was playing with the research assistant in an adjoining room, all mothers were provided with the following information and instructions. They were told that the purpose of the study was to examine how mothers obtain information from their children about events that children have knowledge about but mothers do not. The mothers were informed that their children would be playing in the other room for approximately 20 min and that when the child returned, they were to try to find out in as much detail as possible what had happened during the play period. They were told to continue questioning their child until they thought they could give an accurate description of what had happened or until the child could provide no additional information. Mothers were instructed

to keep the conversation as natural as possible, but if the conversation went off topic, they should bring it back on topic, unless they felt that the off-topic exchanges might elicit relevant information.

The mother was then given a list of activities that might help her interview her child. She was told that these were the types of activities that the child might have done, but not all of the ones on the list might in fact have occurred. (In fact, only two of the six activities had occurred, but the other four were reasonable activities that potentially could have occurred.¹) Finally, the mother was told that a surprising or unusual event had occurred during the play session and that she should try to find out about this event. These procedures were instituted to help ensure that the mothers would use a variety of elicitation techniques and cues when interviewing their children.

Mothers were randomly assigned either to a memory or to a control condition. Mothers in the memory condition were told that in addition to studying the nature of mother-child conversations, the study was designed to examine mothers' memories of this conversation. These mothers were told that in the follow-up interview, we would ask them to remember details of their conversation with their children, including their memories for the exact words used in the conversation. Mothers in the control group were told that at the follow-up visit, they would be asked a few questions about the experiment.

Mother-child conversation. When the children had finished the play session, they rejoined their mother. The assistant, who had given the instructions to the mother, was present for this session. The mother was instructed to begin the interview and to continue until she felt that she had obtained all or as much information about the activities as possible. If, at the end of the interview, the mother had not found out about the surprising event, she was prompted to question her child about the surprising event. If she still had not obtained the required information, the firehat was shown as a retrieval cue. The session was videotaped and was later transcribed.

Recognition test construction. The transcript of the interview was used to construct a recognition test that was administered to mothers 3 or 4 days later. The recognition test contained 20 separate passages from the interview: 5 of the passages were exact wordings of the original mother-child conversation (verbatim items), 5

passages were modified to contain gist changes (gist change items), and 10 passages were modified to change the surface structure but to retain the gist of the conversation (surface structure change items). Each of the 20 passages contained utterances from both the mother and the child (i.e., there were at least two speaker turns per passage).

For the gist-change items, the meaning of a selected part of the conversation was changed. For these items, the gist changes did not make the passage incongruent with the rest of the conversation (i.e., it was not bizarre or antithetical to the content of the interview); however, the changes were salient and were not about small details. Thus, we did not make gist changes that included changes to the color of the crayon or the number of stickers but rather to central aspects of the interview. The following is an example of a gist change that is quite salient.

Original Text	Gist-Altered Text
Mom: Did you watch a movie?	Mom: Did you watch a movie?
Child: No.	Child: No.
Mom: No, no movie?	Mom: No, no movie?
Child: No.	Child: No.
Mom: Snow White?	Mom: Snow White?
Child: (shakes head)	Child: (shakes head)
	Santa Claus Movie.
Mom: No.	Mom: Santa Claus Movie?
Child: No, and seven doors and I want it.	Child: (nods) Ya.

For the 10 surface structure change items, the structure of the conversation was changed but the gist was left intact. There were three different types of surface structure change items. The first involved changing the structure of the conversation so that the children's original utterances, which had been highly prompted by the mothers,

¹ The use of the list resulted in only five errors of commission (children assenting to false questions) and seven errors of omission (children denying true questions). There are several reasons for these low rates. First, mothers often did not ask questions if the child had already told them about an activity that was similar in content to the suggested activity. Second, mothers often rephrased the suggested activity in a way that it was no longer false. For example, mothers were given the false detail "broken crayon," but most only asked their children if they did something with a crayon.

were made to appear spontaneous, or, conversely, children's original spontaneous utterances were made to appear highly prompted. This type of change, called *spontaneous-prompted*, is illustrated in the following example.

Original Text	Altered Text
Mom: What did you do? Tell me one of what you did.	Mom: What did you do? Tell me one of what you did.
Child: I know some of what I did.	Child: <i>We drew a horsy. With Lynne's crayons. One with sparkles.</i>
Mom: What was it?	
Child: Making a horsy drawing.	
Mom: A horsy?	
Child: Ya.	
Mom: You made a horsy?	
Child: Ya, we drew a horsy.	
Mom: You drew it? What did you draw it with?	
Child: My crayon.	
Mom: But were they yours or were they Lynne's?	
Child: Lynne.	
Mom: They were Lynne's crayons. Were they colored?	
Child: They have sparkles on them.	
Mom: Sparkle crayons, That's our favorite. What color?	Mom: Sparkle crayons, That's our favorite. What color?
Child: Purple.	Child: Purple.
Mom: And any other color.	Mom: And any other color.
Child: Nope just the one.	Child: Nope just the one.
Mom: You make a purple horse?	Mom: You make a purple horse?

As can be seen, although the altered text appears spontaneous, it was actually prompted by specific, leading questions asked by the mother during the original interview. The changes in the altered text entailed omission of repeated prompts and also the concatenation of text, leaving out intervening text.

The second type of surface structure change item involved a change of speakers. Thus, within

a selected passage, we put the mother's original words in the child's mouth or put the child's original words in the mother's mouth. In the following example, the passage was altered to have the child say what the mother had actually said.

Original Text	Altered Text
Mom: There's the firehat.	Mom: There's the firehat.
Child: Yah.	Child: Yah.
Mom: Is that the firehat they were wearing?	Mom: Is that the firehat they were wearing?
Child: Ya.	Child: Ya.
Mom: Who was wearing that?	Mom: Who was wearing that?
Child: mmm, a funny guy.	Child: mmm, a funny guy. <i>A live man with a firehat came through the window.</i>
Mom: You mean somebody alive? <i>A live man came in with the firehat on?</i>	Mom: He came in through the window?
Child: Ya. No he didn't come in with it, the firehat. He came in with no.	Child: Ya.
Mom: <i>He came in through the window?</i>	
Child: Yup.	

The third type of surface structure change involved alteration of lexical items that caused some confusion or amusement in the actual conversation. In the following example, the child is telling her mother about the magic trick.

Original Text	Altered Text
Child: <i>Agadazoop.</i>	Child: <i>Abracadabra.</i>
Like this, come disappear and it fliied in the pocket! And then it fliied back.	Like this, come disappear and it fliied in the pocket! And then it fliied back.
Mom: Wow! Wow! <i>Agadzazoop.</i> That was the magic word? (laughs).	Mom: Wow! you said the magic word <i>Abracadabra?</i>
Child: And then we had to, we saw it, then close it, then poof!	Child: And then we had to, we saw it then close it, then poof!

Because of the variability in the types of interviews for each of the 24 participants, it was

not possible to prespecify the number of spontaneous-prompted, speaker change, or lexical change items for each recognition test. However, as will be shown, the distribution of these different types of items were similar across the memory and control conditions, and the relative frequency of their occurrence did not correlate with mothers' memory performances. Each of the 20 passages was typed onto a separate page. The passages were arranged in terms of their original chronological order in the interview.

Follow-up memory test. A follow-up visit was arranged with the mother either in her home or in our laboratory. The follow-up visits were scheduled 3 to 4 days after the mother-child conversation. These follow-up interviews included only the mother. The interviewer for this follow-up session was not the same person who was present during the original mother-child conversation.

At the onset of the interview, all mothers were told that the purpose of the study was to examine parent-child conversations and mothers' memories of these conversations. Mothers were told that memories for ideas as well as for the exact wording of conversations would be tested. Mothers were asked if they had talked to their children about what had happened following the laboratory visit and, if so, what they had talked about. Mothers were tested for their memories of the original interview. First, they were asked to recall what happened in as much detail and in dialogue form. After the free recall test, the mother was asked what activities her child had told her about during the interview.

Next, the mother was given a recognition memory task. Each mother was told that we had transcribed her conversation with her child and that we had selected 20 different passages for her verification. The mother was asked to read each passage carefully to see if the wording was accurate and if she noticed any parts of the passage that had been omitted, added, or changed in any way. She was told that in some cases, we may have made mistakes when transcribing the videotape and that in other cases we deliberately made changes to parts of the passage. She was given a pen to mark any errors and when these were marked she was asked to indicate what was wrong. The mother was asked to think aloud as she checked each passage for its accuracy.²

Table 1
*Participant and Interviewing Style
Characteristics of Memory and Control Groups*

Characteristic	Memory group (<i>n</i> = 12)		Control group (<i>n</i> = 12)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Child's age (months)	51	4	50	6
Child's gender (% female)	50		42	
Memory test delay (hr)	94	25	89	21
Conversation measures				
Time (min)	24	9	23	7
Mother utterances	372	118	374	158
% on-topic utterances	74	12	80	13
Child utterances	269	87	255	125
% on-topic utterances	69	20	79	14
Questions (% on-topic utterances)	52	8	54	8
Feedback (% on-topic utterances)	24	7	25	5
Rehearsals (% on-topic utterances)	2	2	2	1

Results

Baseline Measures

The two groups (memory and control conditions) were matched in terms of age of child, gender of child, and delay between parent-child conversation and memory test (see Table 1). Additionally, equal numbers of mothers in both groups (42%) reported that they had talked to their child about the play session after leaving the interview. Usually, these were brief conversations that took place shortly after the interview.

Certain features of the mother-child conversations were coded in order to determine whether conversations in the memory group were structurally different from those in the nonmemory group. Each conversation was first parsed into turns (i.e., the speaker). The number of utterances in each turn was counted. An utterance generally contained a verb and was bounded by a pause. Sometimes, however, utterances could contain just one word (e.g., "Good"). Utterances were

² Mothers' confidence in their detections and corrections was measured on a 3-point scale. No significant correlations between confidence and accuracy were obtained, possibly because of the limited range of the scale. For this reason, the procedures and data are not presented in the text.

categorized as on topic or off topic. On-topic utterances were those that were related to the demands of the task (finding out what happened in the other room). Utterances that were unrelated to the task (e.g., disciplining the child, holding the child's attention, talking about events not related to the target activities) were coded as off topic.

We also counted the frequency of three different types of mothers' on-topic utterances: questions (e.g., "What did you do next?," "Did you remember the girl's name?"), feedback (confirmations or acknowledgments of the child's previous utterances; "Oh, you played with the toys," "I bet that was fun"), and rehearsals (during the interview, the mother repeated the activities that the child had told her about; "You told me that you used playdough"). High reliability of coding of the transcripts was easily obtained with this classification system (see Appendix A for details of reliability).

As shown in Table 1, there were no between-group differences (all $F_s < 1$) in terms of the number of maternal utterances, the number of child utterances, the percentage of utterances that were on topic, the length of the conversations (as measured by number of words spoken), or the duration of conversation (in terms of minutes). Mothers in both groups produced the same proportion of on-topic utterances that were questions, acknowledgments, and rehearsals.

Finally, the recognition tests of the two groups were well-equated in terms of all independent variables. Specifically, spontaneous-prompted changes accounted for 43% and 51% of the surface structure change passages for the memory and control groups, respectively; speaker changes accounted for 54% and 46% of the surface structure change passages for the memory and control groups, respectively; and lexical paraphrases accounted for 3% of the surface structure change passages for both the memory and the control groups. An analysis of variance (ANOVA) with 2 groups (memory and control) \times 3 item types (spontaneous-prompted, speaker change, and lexical paraphrase) as a repeated measure was carried out on the frequency of each type of recognition test item. The interaction between group and item type was not significant. Thus, the frequency of different surface structure change items was similar for memory and control groups. In addition, the number of words and speaker

turns for each passage were well-equated across groups as well as across the three different types of passages. Two separate 2×3 ANOVAs with repeated measures were carried out on the number of words in each passage and the number of speaker turns in each passage. The independent variables were 2 groups (memory vs. control) \times 3 passage types (verbatim vs. surface structure change vs. gist change). There were no main effects or interaction effects (all $F_s < 1.0$). On average, each recognition test passage contained eight speaker turns and 52 words.

Free Recall Data

Utterances in free recall. The number of utterances in the free recall was counted. Next, utterances were classified as dialogue utterances or as general utterances. *Dialogue utterances* were those that reflected the structure of the conversation by referencing the speaker(s) and, at times, the sequence of questions and answers. The following are examples of dialogue utterances: "I asked him to tell me what happened" or "I asked him many questions about what kinds of toys he played with." *General utterances* were those in which the mother only reported the content but not the context of the utterance. Thus, general utterances did not contain explicit information about who said what. The following are examples of general utterances: "I remember there was something about drawing a pig" or "The playdough was his favorite thing." The gist of each utterance was then coded as accurate or as inaccurate. This was accomplished by referring to the transcript of the original videotaped conversation to verify whether there was any utterance(s) that matched the gist of the mother's free recall.

Finally, three aspects of verbatim recall of the dialogue utterances were examined for accuracy: the speaker of the utterance, whether the utterance was spontaneous or elicited by questions, and lexical changes. Lexical changes involved changes in words that did not obviously change the gist and that might be considered to be paraphrases. For example, one mother recalled, "He wanted to go to Grandma's house," when in the original conversation, her son had said he wanted to go to his grandfather's house. Again, accuracy was determined by comparing the recall utterances with the utterances in the original

videotaped conversation. Interrater reliability for the coding of mothers' free-recall utterances was high (see Appendix A).

The data for the following analyses are shown in Table 2. Mothers in the memory and control groups recalled the same number of total utterances, $F < 1$.³ There were no between-group differences in the proportion of dialogue utterances recalled, although as can be seen from Table 2, mothers in both groups recalled more dialogue than general utterances, $F(1, 23) = 15.79$, $p < .01$. This latter result reflects the mothers' successful attempts to follow the instructions to report their memory of the conversation in dialogue form. Overall, 88% of all free recall utterances were correct in terms of gist. Correct memory for gist was similar for mothers in the memory and control groups.

The accuracy of surface structure (verbatim) recall for all utterances reported in dialogue form was examined next. For this analysis, the database included only that subset of utterances that were in dialogue form and that were correct in terms of their gist; hence, this represents a conservative test insofar as the incorrect gist utterances and general utterances were excluded. As was true in previous analyses, there were no

between-group differences: Overall, mothers' surface structure recall was correct for 74% of all dialogue gist-accurate utterances. In other words, for 26% of dialogue utterances in which the gist was correct, the mothers misreported the structure of the conversation. As can be seen from Table 2, most of the errors involved confusing spontaneous utterances with ones that had been prompted. There were few errors involving lexical items or reversals of speaker and listener.

Proportion of original conversation utterances recalled. At first glance, the data suggest that the mothers were able to remember a great deal of the conversation (approximately 27 utterances). However, given that there were an average of 591 utterances in the original conversations, only 5% of conversation was reported, and again, there was no difference between mothers in the memory and nonmemory group.

This figure of 5%, however, is probably too conservative a measure of the mothers' memories for several reasons. First, the baseline of total utterances in the original mother-child conversation contains a number of empty utterances (e.g., "OK," "That's good"). It would be unlikely that mothers would include these utterances in their free recall. Second, sometimes it took the mother several turns to figure out what her child was talking about. Although she might report in free recall, "I asked him several times what happened, and he finally told me that he played a game," this would only be counted as two utterances although it might represent 15 utterances (for example) in the actual conversation. Finally, as shown in Table 1, almost 25% of the utterances in the original conversation were off topic. These off-topic portions of the conversations were quite different in some respects from the on-topic portions in that the former reflected the mother's attempt to control the child's behavior whereas the latter reflected her attempt to question and obtain information. Off-topic utterances could occupy a large portion of the conversation and yet contain little information to recall

Table 2
Free Recall Data

Variable	Memory group ($n = 12$)		Control group ($n = 12$)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Total utterances	27.5	19.8	25.4	24.4
% Dialogue	70	30	70	59
% General	30	26	30	23
Gist correct				
Raw score	23	11	23	14
% raw score/total utterances	86	11	90	7
Surface structure correct				
Raw score	12.3	6.5	12.3	10.6
% raw score/correct gist utterances in dialogue form	71	14	76	14
Types of surface structure error				
% speaker	2	2	2	4
% spontaneous-prompted	21	26	16	16
% lexical	7	6	6	7

³ For this analysis and all others reported below, mothers who talked to their children about the play event after leaving the laboratory had the same levels of recall as mothers who did not talk to their children about the event. Because this variable of postinterview conversations did not interact with group, we report only the aggregated data.

(e.g., "I tried to get him to sit still" might occupy several minutes of the conversation in which the mother had one goal of trying to control her child's behavior), whereas the same proportion of on-topic utterances could generate a lot of information.

In order to deal with the above concerns, the following procedures and analyses were conducted to estimate mothers' free recall of the original conversations (see Appendix A for details of high interrater reliability).

Proportion of activities recalled. In the initial conversation, mothers elicited some information about most of the seven target activities ($M = 5.6$ activities for the memory group and 5.5 activities for the control group). At follow-up, when explicitly asked to list the activities that their children had talked about in the original conversation, mothers reported 66% of these. There were no between-group differences ($F < 1$).

This is an inflated measure of memory because the mothers often provided vague undetailed reports for which they were credited. Thus, a mother who reported, "There was something about playdough," was credited with one point even though the child may have provided a very elaborate explanation of the whole activity (e.g., there were molds of Sesame Street characters that when turned on a barber's chair would produce growing hair that could be cut with scissors). Therefore, the following, more detailed measures of the mothers' free recall of the content of the original conversations were devised.

Proportion of on-topic details recalled. On-topic utterances in the original conversation were coded for content details. A *content detail* was defined as a piece of information about the play activity that the mother learned about in the original interview. This exercise involved summarizing across turns what the child told the mother. For example, in the following series of 14 turns, the mother learned four content details about one activity—(a) stickers (b) were placed on a road, (c) the stickers were animals, and (d) they stuck to the road.

Mom: Did you play any games?

Child: Yeah. We put stickers on.

Mom: On what?

Child: On a road.

Mom: Stickers on a road? What kind of stickers?

Child: Like animal stickers.

Mom: Animal stickers?

Child: Uh-huh.

Mom: Yeah?

Child: Yeah.

Mom: Was it the kind of sticker thing that you could put them on and pull them off?

Child: No.

Mom: Once you stuck em they stuck? You couldn't take them off again?

Child: Yeah.

If a detail was repeated during the conversation, it was only counted once.

Next, we counted the number of on-topic details that mothers produced in free recall general or dialogue utterances. For example, the mother's recall utterance, "She sang along to a tape of Rubber Duckie," was assigned three details (sang, tape, Rubber Duckie). The number of on-topic details produced in free recall was divided by the number of on-topic details that were uncovered in the original interview. (As shown in Appendix A, reliability was high on both measures).

According to this measure of on-topic information (gist) recalled, mothers in the memory group and in the control group remembered 35% of all on-topic details revealed in the original conversation. This estimate of amount of information recalled is intermediary to the 5% estimate (obtained by dividing utterances in free recall by utterances in the original conversation) and the second estimate of 66% (obtained by dividing mother's recall of gross categories of a maximum of seven activities by child's discussion of these activities).

Proportion of off-topic episodes recalled. Another procedure for assessing the amount of the conversation recalled focused on off-topic utterances. Off-topic utterances in the original conversations were summarized into broad categories. For example, if a child kept repeating throughout the interview that she wanted to go home, and the mother tried a number of diversionary tactics to keep her attention, regardless of the number of utterances, this was categorized as one unit called *child wants to go home*. Other common units involved child wants to play with video and mother will not let him or child hits mother who disciplines child. Free-recall utterances that contained off-topic categories were counted (e.g., "She was quite restless at one point," "I remember telling her that she was going to get hurt if she kept climbing"). As shown in Table 3, mothers

Table 3
Percentage of Interview Recalled

Material recalled	Memory group (<i>n</i> = 12)		Control group (<i>n</i> = 12)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
On-topic details				
On-topic details in conversation	36.4	12.1	42.6	13.5
On-topic details in free recall	13.3	6.8	14.8	7.8
% on-topic details recalled	35	12	35	16
Off-topic units				
Off-topic units in conversation	6.5	2.2	5.2	3.2
Off-topic units in free recall	2.1	2.8	1.1	1.2
% off-topic units recalled	28	33	22	22
On-topic questions				
On-topic questions in conversation	40.8	10.2	43.2	15.9
On-topic questions in free recall	6.3	3.6	6.9	6.9
% on-topic questions recalled	16	9	17	17

tended to produce few free recall off-topic utterances. In terms of the proportion of the original material recalled, mothers in the memory group recalled 28% of off-topic events and mothers in the control group recalled 22% of off-topic events ($F < 1$).

To summarize, mothers recalled the content of more on-topic than off-topic details. They recalled approximately one third of all the details of the play session that they had learned about in the original conversation, and they recalled approximately one quarter of all off-topic events.

Proportion of questions recalled. The previous three measures (recall of activities, on-topic details, off-topic episodes) reflect mothers' memories of the content (gist) of the conversations. In order to provide some estimate of the mother's free recall of the structure of the conversation, we assessed her memory of the number of questions that she had asked her child.

We counted the number of on-topic free recall utterances that were dialogue utterances where the mother stated that she asked the child a question (e.g., "I asked her if she drew a pic-

ture"). Next, we counted the number of on-topic questions that the mother asked in the original interview. We eliminated all questions that were repeated or about the same topic from this measure (e.g., if the mother asked all the following questions, "What size was the ball? Was the ball big? Was the ball small?," this was counted as one instance of asking about the size of the ball). This procedure eliminated a large number of questions (e.g., even though mothers repeatedly asked their children, "What else happened?," this was only counted one time). The number of questions that the mothers recalled asking (at the memory test) was divided by the conservative estimate of the number of questions that the mothers asked in the original interview (see Appendix A for interrater reliability). As shown in Table 3, mothers in the memory group recalled 16% of all questions asked, and mothers in the control group recalled 17% of all questions ($F < 1$).

In view of the different ways we computed baseline data for these analyses, it is not possible to directly compare mothers' free-recall memories for gist with their memories for structure of the conversation (verbatim memory). Having said this, however, our analyses suggest that mothers recalled more of the content of on-topic details (35%) than of the structure of how these details were elicited (17%). In the Discussion, we return to the issue of evaluating how much the mothers were able to recall.

Free recall of surprising event. The final analysis of the free recall data was conducted to provide a more detailed account of mothers' memories of content and structure. As explained in the Method section, when children did not spontaneously reveal the surprising event of the man coming into the room for his firehat, the mother was first reminded to find out about the special event, and if this was not successful, the firehat was brought into the room as a cue. In our study, 18 of the 24 mothers required verbal prompts as well as the firehat prompt. These prompts led to parents finding out about the surprising event in 67% of the cases (12 of the 18 mothers were successful in finding out about the man with the firehat taking the crayon). When parents did find out, they themselves were surprised ("You mean a strange man came into the room?," "Someone took your crayon?"), or the firehat itself was the source of an object of

discussion and play. Nevertheless, only 50% of the prompted mothers who did find out about the surprising event reported it in free recall (6 out of 12), and for the 6 who did report it, only 2 reported how they found out ("I was reminded," "A firehat was brought into the room"). This analysis of the recall of one activity is a case study of how mothers' recall of certain events was not only poor, but in most cases, not contextualized; that is, the mothers' recall made it appear that the children spontaneously mentioned that a man came into the room to look for his firehat.

To summarize, overall mothers' free recall utterances of the original conversation were fairly accurate both in terms of gist and in terms of surface structure (verbatim aspects). As is the case in other studies, recall for gist was more accurate than for surface structure. Although their productions were fairly accurate, subsequent analyses revealed that the mothers omitted large portions of on-topic details, off-topic episodes, and the number of questions they asked during the original interview. The high rate of omissions raises the issue of whether mothers only produced utterances that they thought to be accurate and selectively omitted others about which they were less sure or whether, if prompted, they could accurately recall the details and the structure of the conversation. In order to address these issues, the next section assesses the accuracy of mothers' memories for gist and surface structure when they were explicitly asked about certain events.

Recognition Test Data

There were two primary measures, accurate detections and accurate corrections. For the altered passages, a response was scored as an accurate detection if the mother noted that there seemed to be "something wrong." For verbatim passages, an accurate detection was scored if the mother said that the passage "sounded correct." The second measure, accurate corrections, was more stringent. Here, not only did the mother have to indicate that something was wrong, but she also had to accurately indicate what was wrong. For the gist change passages, the mother did not have to provide the exact words of the original passage, but she did have to indicate how the meaning of the targeted section had been changed. For the prompted-spontaneous surface

structure change passages, the mother had to indicate that she had asked more questions to get the information than was evident in the test passage or, conversely, that she had originally not prompted the child even though the test passage indicated that she had. For surface structure change passages involving speaker changes, the mother had to reverse the speaker who had actually provided the original utterance (e.g., "My child didn't say that. I did."). Finally, for surface structure change passages involving lexical paraphrases, the mother had to pinpoint the item and indicate that the child had not used that word (e.g., "He didn't say abracadabra. He said agadazoop."). Detection and correction scores for the original verbatim items were identical. Sometimes, a mother would make more than one change in a passage. When this occurred, if she made a change to an altered section as well as to an unaltered section, she was still given credit.

Separate repeated measures ANOVAs were carried out on the percentage of correct detections and on the percentage of accurate corrections. For each analysis, the independent variables were group (memory vs. control) and the repeated measure of passage type (surface structure change vs. gist change vs. verbatim).

For the detection analysis, there was a significant main effect of question type, $F(2, 44) = 8.36$, $p < .001$. As shown in Table 4, and confirmed by planned comparisons, there were significantly more correct detections of gist-altered passages and of verbatim passages than of surface structure change passages. There was no significant main effect of memory condition and no significant interaction between group and passage type.⁴

Similar results were obtained when the depen-

⁴ This analysis was also carried out by counting the number of *yes* (i.e., this passage is identical to the original conversation) responses. In this analysis, *yes* responses are correct for verbatim passages only. There was no significant effect of memory condition, but there was a significant effect of passage type, $F(2, 44) = 14.98$, $p < .001$. Results of this analysis are consistent with results of previous studies of gist and verbatim memory. That is, participants were as likely to respond *yes* to verbatim passages (73%) as to surface structure change passages (62%). However, when changes were made to the gist, they were least likely to accept these passages as being in the original conversation (36%).

Table 4
Recognition Test Data

Variable	Memory group (<i>n</i> = 12)		Control group (<i>n</i> = 12)	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
% correct detections				
Surface structure items	37	20	40	20
Gist items	62	26	67	30
Verbatim items	70	36	75	19
% correct detections: surface structure items				
Speaker changes	35	29	42	36
Prompted-spontaneous	32	30	30	27
% accurate changes				
Surface structure items	23	14	27	20
Gist items	47	35	50	29
Verbatim items	70	36	75	19
% accurate changes: surface structure items				
Speaker changes	27	24	34	31
Prompted-spontaneous	20	23	25	23

dent variable of accurate corrections was examined. There was no main effect of memory condition, but there was a significant main effect of passage type, $F(2, 44) = 19.36, p < .001$. As shown in Table 4, and confirmed by planned comparisons, there were significantly more correct responses to verbatim passages (73%) than to gist-altered passages (48%). Surface structure change passages resulted in the most errors (25%).

Because the frequency of different types of surface structure change items (i.e., speaker changes, prompted-spontaneous changes, and lexical changes) varied across participants, further analyses were conducted to determine if the mothers' performances varied as a function of the density of different types of surface structure change items. First, the results of ANOVAs that compared the accuracy of mothers' detections and corrections of spontaneous-prompted items versus speaker change items (see Table 2)⁵ indicated that performance was similar for the two types of surface structure change items. Second, there were no significant correlations between the proportion of surface structure change items that involved spontaneous-prompted items and mothers' detection accuracy ($r = -.16$) and correction accuracy ($r = .04$). Thus, although there was

variability in the proportion of different types of surface structure change items across participants, this did not influence the level of mothers' performances.

Surprisingly, there were few significant correlations between mothers' performances on the various free recall measures and on the various recognition measures, with only two of significance. Both of these involved free recall of content details and recognition for gist. Specifically, the number of on-topic details produced in free recall was positively correlated with mothers' correct detections of gist-change passages ($r = .49, p < .01$) and with mothers' accurate corrections of gist-change passages ($r = .52, p < .01$). However, there were no other significant correlations. For example, gist errors in free recall did not correlate with gist detection on the recognition task ($r = -.06$) or with gist correction on the recognition task ($r = .03$). Correlations between surface structure errors in free recall did not correlate with responses to surface structure change items on the recognition test ($r = .11$) or with accurate changes to surface structure change items on the recognition test ($r = .01$). Thus, except for the relationship between free recall of content details with gist recognition, the free recall and recognition data appear to be tapping different components of memory.

Discussion

The purpose of this study was to examine specific elements of mothers' memories of interviews with their young children. The elements that we selected were ones that were deemed important in terms of reports that parents and other adults often give in forensic arenas about prior conversations with young children. Specifically, we examined whether mothers could remember the gist of conversations as well as the structure of the interactions that took place during a conversation with their young child 3 or 4 days earlier. The two major elements of structure that were examined were the strategies used to elicit statements from children and role assignment.

⁵ Lexical-change items were omitted from these analyses because there were only seven such passages across all 24 recognition tasks.

There were two primary measures to evaluate mothers' memories—free recall and recognition.

Mothers' free recall utterances were quite accurate in terms of gist (88% accuracy) and surface structure (74% accuracy). As was true for all analyses in this article, there were no effects of forewarning mothers to remember the conversation. The consistency of this null effect gives us confidence that warnings do not lead to better recall of parent-child conversations, at least if they are delivered in the manner of the present study.

Although the mothers' free recall utterances were quite accurate, their reports were incomplete in terms of reflecting both the original content (gist) and the structure of the conversation. First, in terms of content, depending on the measure selected, the mothers reported between 5% and 66% of the details from the original conversation. For reasons discussed above, measures of on-topic details (35%) and off-topic episodes (25%) reported from the original conversation seem to provide reasonable estimates of recall. Thus, on the basis of these latter estimates, the mothers clearly omitted major parts of the conversation, some of which seemed quite central to the play event. For example, only 50% of the mothers recalled their children's description of the special event involving the strange man who came into the room for his firehat and wanted the child's crayon.

The second aspect of the mothers' incomplete recall involves their reports of the surface structure of the conversation. Although mothers did not make many errors (26%), only 47% of all free recall utterances accurately mirrored the structure of the interview. The remaining utterances were either incorrect as to their gist or were not given in dialogue form. The latter case appears to reflect mothers' poor memory of the structure of the conversation. That is, although they were continuously prompted to present the content in dialogue form, 30% of all utterances did not specify the speaker or other contextual information (e.g., "And then there was something about playing with playdough."). This conclusion is supported by the fact that mothers recalled only 16% of the major questions they asked in the original conversation. Finally, as the case example demonstrates, mothers also omitted or forgot the source of their reports; they did not

mention that their children's reports of the special event were prompted by the research assistant bringing the firehat into the room.

To summarize one major conclusion of this study, mothers' free recall accuracy was quite good but incomplete both in terms of its content and more so in terms of its surface structure. To some degree, this pattern of data is similar to that found in the memory development literature. Namely, children's free recall reports are quite accurate. However, these reports are incomplete, omitting many important details (Ceci & Bruck, 1993). When children are asked more specific questions that require them to systematically search the contents of their memories, their accuracy decreases significantly (see Ceci & Bruck, 1995, for a review). As seen below, the recognition data from this study indicate that mothers behave similarly.

On the recognition task, which has been the traditional method for assessing memories of conversations, mothers were better able to retain information about meaning (or gist) than about surface structure. This finding is consistent with the earlier studies of adult verbal memory when sentences were presented in isolation rather than in the context of meaningful interactions (e.g., Sachs, 1967). In the present study, mothers were particularly poor at detecting alterations to the surface structure of the conversation. They misattributed words that their children had said to them (or vice versa), and they often could not tell whether the children's words were spontaneous or the result of much prompting and suggesting. The mothers were better at detecting passages where the meaning had been changed. Finally, when asked to correct passages that they thought had been changed, performance declined even further. That is, although they thought there was something "not right" with the detected passages, they accurately pinpointed the problem only for 25% of the surface change passages and for 48% of the gist change passages. As was found in a study of adults' memories of conversations with other adults (Kausler & Hakami, 1983), forewarning mothers to remember the conversation did not influence their accuracy on any of the recognition measures.

The mothers' relatively poor abilities to detect and to correct altered passages on the recognition tests has theoretical as well as applied implica-

tions. On the theoretical level, the results support prior findings that once the meaning of a sentence is extracted, memory for the exact words quickly fades from memory and all that remains is a gistified trace (Bransford & Franks, 1971). However, our conclusions are clearly discrepant with the findings of researchers who have examined memory for sentences within meaningful passages of discourse (e.g., Bates, Masling, & Kintsch, 1978; Kintsch & Bates, 1977; MacWhinney et al., 1982). In these previous studies, it was argued that verbatim memory for sentences is especially good when presented as part of meaningfully connected discourse and particularly when these are of high interactional value (e.g., insults or jokes). It could be argued that verbatim memory in our study was relatively poor (it was actually below chance for the detection measure) because we did not include passages of high interactional value. This is a difficult issue to evaluate in the context of the present study, because the sentence was not necessarily the unit of analysis, and the longer units that we used appeared to contain many instances of high interactional value. For example, for speaker-change and spontaneous-prompted surface structure items, the alterations involved several utterances each made by different speakers. However, the lexical change items most closely fit previous researchers' definitions of high interactional utterances—these were utterances that contained amusing words spoken by the child, such as “agada-zoop.” Unfortunately, there were not sufficient numbers of such items included across all recognition tests to adequately test the hypothesis of better memory for such items (there was a total of seven such items, only two of which were correctly detected by the mothers). Although we do not have measures of interactional value (which, according to MacWhinney et al., 1982, is difficult if not impossible to develop objective criteria for categorization), there was agreement among the test developers that the utterances selected for modification were not of low importance.⁶ Clearly, further research in this area is needed to more clearly define and assess interactional value in the context of adult-child interviews. It may turn out to be the case that in fact adults do have good memory for high interactional utterances within adult-child conversations. But, as shown in the present study, it may

be particularly difficult for them to pinpoint how a given utterance emerged or was located within the larger structure of an interview.

There is also another explanation for the seemingly different pattern of results. Perhaps they reflect differences in the paradigms used. That is, if we tested adults' memories of the structure of their conversations with other adults, perhaps they would do as poorly as the mothers in the present study. For example, they might make many source confusions regarding speaker and listener, and they might not remember whether utterances were prompted or the result of a series of questions. If this pattern of results were obtained, then our results would not reflect difficulties that are specific to remembering conversations with young children but would extend to conversations in general. Of course, it would be difficult if not impossible to conduct a study that directly compares adults' memories of conversations with children versus adults because of pronounced age differences in the structures and contents of these conversations. Nevertheless, until such data are available, the most conservative conclusion that can be drawn from the present study is that mothers have difficulty remembering the structure of conversations with their young children.

We now return to an evaluation of the quality of the mothers' memories of their conversations with their children. The evaluation is a function of the purpose and context of the recall. For example, if the mother's free recall was elicited by her husband's question, “What happened at the laboratory today with Janie?,” her statements as provided in the free recall portion of this study would be adequate for dinner table conversation. Her statements contained an accurate overview of nearly five out of seven of the activities as well as a smattering of the child's behavior during the interview. In situations such as these, there is

⁶ Although it is possible that the passages and the alterations were not salient enough, it is noteworthy that so many of the mothers were extremely disappointed in their performance on the test. When they were debriefed, many expressed concern about their lack of memory for their children's words despite the fact that they really had carefully listened to their children and that they had tried hard for this particular task.

little interest or purpose in providing the full set of details or even the structure of the conversation.

In other contexts, however, the mother's performance would not be evaluated as positively. Specifically, in forensic contexts, our results do raise concern about hearsay testimony that is provided by adults about prior conversations with young children. First, mothers' reports may not be very complete, and they may omit many details that may be important. Second, if probed about the contexts of certain utterances (for example, when a mother reports, "My child said that a man touched him."), our data indicate that the mother may not be able to accurately recall whether these were the child's own words or if her statement is a reconstruction of a conversation in which the child provided one-word answers to a series of direct and possibly leading questions from the mother. A genuine concern is that a mother might accurately report the gist of a child's disclosure but fail to recognize that this disclosure resulted from a host of repeated leading questions.

Failure to report the structure of the interview can result in an erroneous interpretation of the child's behavior by the fact finder. Because of concerns raised about the reliability of children's disclosures made under conditions of repeated prompting and suggestion (Ceci & Bruck, 1993), it may be important for jurors and judges to be given not only the gist of a child's disclosure but also the interrogative context in which it was elicited. As shown in the present study, it is this latter type of information that seems to be lost first. Although mothers at times may be able to accurately remember the structural features of the conversation, accuracy is highest if this information is provided spontaneously in free recall. If probed about these missing details, mothers may not be able to accurately provide the crucial information.

In conclusion, we discovered that memory for connected discourse between a preschool-aged child and a familiar caregiver may suffer from difficulties that inhere in the way adult-child conversations must be monitored and maintained. Adults thrust into the role of interviewers have difficulty keeping track of the source of utterances, the spontaneity of the utterances, and, at times, even their gist, only 3 or 4 days following the conversation. Such difficulties were not pre-

dictable from the literature dealing with college students' memories of lectures and television shows. As discussed above, the difficulties of the mothers in the present study may reflect the attentional demands of structuring an interrogative interview with very young children, or they may reflect general difficulties in remembering aspects of conversations, regardless of the age of the participants.

References

- Bates, E., Masling, M., & Kintsch, W. (1978). Recognition memory for aspects of dialogue. *Journal of Experimental Psychology: Human Learning and Memory*, 4, 187-197.
- Bransford, J. D., & Franks, J. (1971). The abstraction of linguistic ideas. *Cognitive Psychology*, 2, 231-380.
- Brown, A. S., Jones, E. M., & Davis, T. L. (1995). Age differences in conversational source monitoring. *Psychology and Aging*, 10, 111-122.
- Brown, A. S., & Murphy, D. R. (1989). Cryptomnesia: Delineating inadvertent plagiarism. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15, 432-444.
- Ceci, S. J., & Bruck, M. (1993). The suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, 113, 403-439.
- Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom: A scientific analysis of children's testimony*. Washington, DC: American Psychological Association.
- Fivush, R. (1993). Developmental perspectives on autobiographical recall. In G. S. Goodman & B. Bottoms (Eds.), *Child victims and Child witnesses: Understanding and improving testimony* (pp. 1-24). New York: Guilford.
- Garrod, S., & Trabasso, T. (1973). A dual memory information processing interpretation of sentence comprehension. *Journal of Verbal Learning and Verbal Behavior*, 12, 155-168.
- Gernsbacher, M. A. (1985). Surface information loss in comprehension. *Cognitive Psychology*, 17, 324-363.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3-28.
- Kausler, D., & Hakami, M. (1983). Memory for topics of conversation: Adult age differences and intentionality. *Experimental Aging Research*, 9, 153-157.

- Keenan, J., MacWhinney, B., & Mayhew, D. (1977). Pragmatics in memory: A study of natural conversation. *Journal of Verbal Learning and Verbal Behavior*, 16, 549-560.
- Kintsch, W., & Bates, E. (1977). Recognition memory for statements from a classroom lecture. *Journal of Experimental Psychology: Human Learning and Memory*, 3, 150-159.
- MacWhinney, B., Keenan, J., & Reinke, P. (1982). The role of arousal in memory for conversation. *Memory and Cognition*, 10, 308-317.
- Marsh, R., & Bower, G. (1993). Eliciting cryptomnesia: Unconscious plagiarism in a puzzle task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 673-688.
- McGough, L. (1994). *Fragile voices: The child witness in American courts*. New Haven, CT: Yale University Press.
- Poole, D., & White, L. (1991). Effects of question repetition on the eyewitness testimony of children and adults. *Developmental Psychology*, 27, 975-986.
- Sachs, J. (1967). Recognition memory for syntactic and semantic aspects of connected discourse. *Perception & Psychophysics*, 2, 437-444.
- Stafford, L., Burggraf, C., & Sharkey, W. (1987). Conversational memory: The effects of time, recall, mode, and memory expectancies on remembrances of natural conversations. *Human Communicative Research*, 14, 203-229.
- Stafford, L., & Daly, J. A. (1984). Conversational memory: The effects of recall mode and memory expectancies on remembrances of natural conversations. *Human Communicative Research*, 10, 379-402.

Appendix A

Coding Reliability

Interrater reliability was high for all the coding of this study.

Length and Structure of Initial Mother-Child Conversations

Ten sections, each containing 50 utterances, were randomly selected from 10 different interviews. There was perfect agreement between the raters on classification of utterances as rehearsals, feedback, and questions. As well, there was perfect agreement on classification of utterances as on topic or off topic.

Free Recall Utterances

Two raters coded the protocols of 8 participants for a total of 156 recall utterances. There was perfect agreement on the categorization of utterances as general or dialogue utterances. There was also perfect agreement on the number and type of dialogue errors. There was disagreement on gist errors in 4 of 156 instances.

Amount of Original Information Recalled in Free Recall

In order to establish reliability for the on-topic and off-topic details, two raters coded seven mother-child conversations and the same mothers' free recall. Reliability for number of on-topic content units that mothers learned about in the original interviews and the number of on-topic content units that mothers produced in free recall was very high ($r_s > .95$). There was almost perfect agreement on off-topic events and recall units ($r_s > .97$).

Six different interviews were selected to check reliability on the estimates of the number of on-topic questions mothers asked. The correlation between raters for number of questions asked in the original interview was .97.

Received September 23, 1997
Revision received July 27, 1998
Accepted July 29, 1998 ■