Journal of Experimental Child Psychology **79,** 405–437 (2001)
doi:10.1006/jecp.2000.2612, available online at http://www.idealibrary.com on IDEAL®

The Credibility of Children's Testimony: Can Children Control the Accuracy of Their Memory Reports?

Asher Koriat and Morris Goldsmith

University of Haifa, Haifa, Israel

Wolfgang Schneider

University of Würzburg, Würzburg, Germany

and

Michal Nakash-Dura

University of Haifa, Haifa, Israel

In previous work with adults (A. Koriat & M. Goldsmith, 1994, 1996c), it was shown that people can enhance the accuracy of their testimony substantially when they (a) are effective in monitoring the correctness of their answers, (b) are free to control their reporting accordingly (i.e., to decide which pieces of information to volunteer and which to withhold), and (c) are given incentives for accurate reporting. A theoretical model was developed, which specifies the critical role of metacognitive monitoring and control processes in mediating free-report memory accuracy. The present study applies that model to examine the strategic regulation of memory accuracy by children. Three experiments indicate that both younger (ages 7 to 9) and older (ages 10 to 12) children can enhance the accuracy of their testimony by screening out wrong answers under free-report conditions but suggest a developmental trend in the level of memory accuracy that is thereby achieved. The implications of the results for the dependability of children's testimony in legal settings are discussed. © 2001 Academic Press

Key Words: memory; metamemory; monitoring; control processes; eyewitness testimony; recall; recognition; children.

An increasingly volatile issue in recent years concerns the accuracy or dependability of children's memory, particularly in the area of legal testimony, in which an ever-growing number of victims and eyewitnesses are children (e.g., Bruck &

This study is part of a project that was supported by a grant from the German-Israel Foundation for Scientific Research and Development (Grant No. I-412-019-4/95) and by the Max Wertheimer Minerva Center at the University of Haifa. We thank Shiri Pearlman-Avnion and Nira Shaked (Schwartz) for their help in setting up and running the experiments, and Hana Struminger for programming assistance.

Address correspondence and reprint requests to Asher Koriat, Institute of Information Processing and Decision Making, University of Haifa, Haifa, Israel. E-mail: Akoriat@psy.haifa.ac.il.

0022-0965/01 $35.00

Ceci, 1999; Flin, Boon, Knox, & Bull, 1992; McCauley & Fisher, 1995). Because of the greater involvement of child witnesses in legal settings, it is important to know whether their recollections of an event can be trusted. This is particularly so in cases, such as child molestation, in which the child's testimony provides the primary or sole source of evidence. To what extent, then, can children be counted on to give a complete and reliable account of past events?

Historically, there has been a bias in the judicial system against relying on the eyewitness accounts of children (see Goodman, 1984). In the United States, for example, some states have evidentiary corroboration rules mandating that the courtroom testimony of a child be confirmed by at least one other person (Ceci, Ross, & Toglia, 1987).

So far, however, research has failed to yield a consistent picture regarding the presumed inferiority of children's memory. In reviewing the literature, we found it useful to employ a long-standing distinction (Stern, 1904, reproduced in Neisser, 1982) between two principal criteria in terms of which memory can be evaluated: the amount of recalled information, and its accuracy (see also Koriat & Goldsmith, 1994, 1996a, 1996b, 1996c; Koriat, Goldsmith, & Pansky, 2000).

As far as the amount of remembered information is concerned, there seems to be a consensus that children are generally inferior to adults in the extent to which they can provide a complete account of past events. Thus, children were found to exhibit lower levels of recall and recognition performance in a variety of standard laboratory tasks, such as list learning and memory span (see, Schneider & Bjorklund, 1998). Also, in experiments using more complex, naturalistic stimulus materials (e.g., actual or filmed events), the amount of information recalled or recognized is generally found to increase with age (e.g., List, 1986; Marin, Holmes, Guth, & Kovac, 1979; Ornstein, Gordon, & Baker-Ward, 1992; Ornstein, Gordon, & Larus, 1992; Parker, Haverfield, & Baker-Thomas, 1986; Peterson, 1999; Poole & White, 1991, 1993), though in some cases even young children's memory is still remarkably good (e.g., Ornstein, Gordon, & Larus, 1992). The inferior performance of children is generally more pronounced when open-ended rather than specific questioning techniques are used (e.g., List, 1986; Marin et al., 1979).

Various theoretical accounts were proposed for why children might remember less than adults (see Schneider & Weinert, 1989), the most straightforward of which are that children's memory is inferior in terms of encoding, storage, and/or retrieval capabilities (e.g., Brainerd, 1985; Brainerd & Ornstein, 1991; Brown, 1979; Loftus & Davies, 1984). Other accounts, however, emphasize the development of knowledge structures and metacognitive skills (see Schneider, 1999; Schneider & Pressley, 1997): Increasing age brings with it increasing sophistication in the exercise of strategic memory skills, such as rehearsal, use of mental imagery, and semantic organization (Brown, 1979; Schneider & Pressley, 1989). Thus, growing competence in "knowing how to know" (Brown, 1975) may in part explain the growth of recall ability with age (Flavell, 1970).

Also, children may not have the relevant prior knowledge that would help them organize disparate elements into a cohesive whole, or to relate one set of events to another (Chi, 1978; Johnson & Foley, 1984; Lindberg, 1980; Ornstein, 1990; Ornstein & Naus, 1985; Siegler, 1983). Finally, it was proposed that children may lack the communication skills needed to spontaneously report the information that they remember, and that more structured questioning may be required in order to elicit the stored information (e.g., Fivush & Hammond, 1990; Gordon & Follmer, 1994; McCauley & Fisher, 1995; Ornstein, 1991; Poole & White, 1991).

The evidence that children typically recall less than adults could in itself cast doubt on the value of child testimony. However, the general concern about children's memory in forensic contexts relates not so much to the amount of information that children remember, but rather, to the accuracy or dependability of the information that children report. What is the evidence in this regard?

Here the findings are conflicting. On the one hand, several studies indicate that not only are children's memory reports less complete than adults', they are also less accurate. For instance, Poole and White (1993), who examined memory for a witnessed event after two years, found that children (aged 4 to 8 years) reported a higher proportion of false information in response to open-ended questions than did adults: Whereas only 7% of the information reported by the adults was incorrect, fully 20% of the information reported by the children was wrong. The children were also more likely to fabricate answers when questioned specifically about an unknowable detail, and many of the children attributed to one person actions that actually were performed by another—an error pattern with serious legal implications (none of the adults made this latter type of error). Similarly, Geiselman and Padilla (1988) found that children aged 7 to 12 were significantly less accurate than were adults in reporting details of a filmed liquor store holdup, exhibiting mean error rates of between 24% and 30%, depending on the type of interview (see also Dietze & Thomson, 1993).

Other studies, on the other hand, have failed to find evidence of the presumed inferior accuracy of children's memory, particularly when an open-ended free-recall questioning format is used. Thus, Cassel, Roebers, and Bjorklund (1996) found that in the free recall of details from a short video, commission errors were at floor levels for four age groups ranging from kindergarten to adults. Also, in comparing the effectiveness of the standard police interview and (revised) Cognitive Interview, McCauley and Fisher (1995) found second graders to exhibit high levels of accuracy (between 85% and 90% of the reported information was correct), comparable to other findings with adults. Similarly, high recall accuracy levels (between 87% and 97%) were observed by Marin et al. (1979) for four age groups ranging from kindergarten to college students (with kindergarten and first-grade children yielding the lowest error rates), and by Peterson (1999) for four age groups ranging from 3 to 13 years old (in an initial interview). Finally, Roebers and Schneider (in press), who used a cued recall format in testing children's (6, 8, and 10 years old) and adults' memory of details from a short video, found much lower accuracy rates overall (as low as 32%), and here the existence and magnitude of age differences depended in a complex manner on both the type of question used (misleading vs unbiased) and the nature of the queried information (see also Cassel et al., 1996).

In sum, with regard to age differences in the amount of information remembered, the results are fairly consistent in demonstrating the inferior performance of children. In contrast, with regard to memory accuracy, the results paint a much less consistent picture: “The literature on children’s eyewitness testimony is a story of contradictions, with developmental trends in accuracy, suggestibility and response to stress differing from study to study” (Poole & White, 1993, p. 844).

In searching for possible reasons for the different patterns of results, we considered the possible mediating role of metamemory processes. As just noted, metamnemonic skills were found to play an important role in mediating developmental differences in the amount of information remembered. Little is known, however, about their potential contribution to age differences in memory accuracy. In previous work with adults, Koriat and Goldsmith (1994, 1996c) showed that metamemory processes operating during memory reporting play a crucial role in determining the accuracy of the volunteered information (see also Barnes, Nelson, Dunlosky, Mazzoni, & Narens, 1999; Weingardt, Leonesio, & Loftus, 1994). Moreover, by taking such processes into account, they showed how apparently discrepant results in the adult eyewitness literature could be reconciled (see Koriat & Goldsmith, 1994). Perhaps these processes also underlie some of the age differences (or lack of them) observed in memory accuracy.

The experiments reported here are the first in a series of studies designed to uncover the contribution of metamemory processes operating at the reporting stage to age differences in memory accuracy. In what follows, we first present a general framework that was developed to understand how adults use metamemory processes to enhance the accuracy of the information that they report from memory. We then go on to present three experiments that examine the use of such processes by children.

A FRAMEWORK FOR THE STRATEGIC REGULATION OF MEMORY ACCURACY

Koriat and Goldsmith’s (1994, 1996c) work with adults is based on the assumption that in recounting past events, people do not simply report all that comes to mind but attempt to control their memory reporting in accordance with a variety of personal and situational goals, whether these involve aiding a criminal investigation, succeeding on an exam, or impressing one’s friends. Thus, people make strategic choices about which aspects of the event to relate and which to ignore, what perspective to adopt, what degree of generality or detail to use, and so forth (see Goldsmith & Koriat, 1999; Koriat & Goldsmith, 1996b). Such strategic control can have a substantial effect on the quality of adults’ memory reports, which raises the question: To what extent can children strategically control their memory reporting in accordance with specific goals?

Our focus in this article is on one particular means of strategic control, which we call *report option:* the option to choose which pieces of information to report and which to withhold. When a premium is placed on accurate reporting, for example in eyewitness testimony, adult rememberers attempt to enhance the accuracy of the information they report by selectively screening out answers that are likely to be incorrect. Consider, for example, a person who actually remembers the answers to 10 out of the 15 questions that are asked of her. If only the quantity of information matters, for instance, in the beginning stages of a police investigation, then she may provide answers not only to the 10 questions that she knows, but also to the remaining 5 questions that she is relatively unsure about (thus providing 10 correct and 5 wrong answers). However, if the accuracy of what she reports is at stake, for instance when testifying on the witness stand, then she will try hard to avoid providing answers that are likely to be wrong (responding instead "I don't remember"). Ideally, she will be able to volunteer only the 10 correct answers and withhold the 5 wrong answers, thereby achieving a completely accurate memory report.

This example brings out an essential difference between the two properties of memory mentioned earlier, quantity and accuracy. Memory quantity performance is *input bound*, reflecting the likelihood that an input question is answered correctly. Thus, in measuring memory quantity, we begin with the number of questions presented (input) and examine what percentage of them was answered correctly. This has been the predominant measure of memory performance in much traditional laboratory research. For instance, in the standard list-learning paradigms, memory is evaluated in terms of the proportion of input items that are correctly recalled or recognized. Memory accuracy performance, in contrast, is *output bound*, reflecting the likelihood that a *reported* answer is in fact correct. This measure was emphasized in more naturalistic research contexts, such as research on eyewitness testimony, in which the dependability of the information reported is of paramount concern (e.g., Cassel et al., 1996; Roebers & Schneider, 2000). Applying this contrast to the example just mentioned, whereas the quantity of information provided (10/15 = 67%) was the same in both cases, the accuracy of the information reported was much higher in the second case, in which the person exercised selective reporting (10/10 = 100%) than in the first case, in which she did not (10/15 = 67%). Note that 67% accuracy means that fully one-third of the information the person reports is wrong.

As is clear from the preceding example, the operational difference between the quantity and accuracy measures can emerge only under conditions of *free report,* that is, only when the rememberer has the option of deciding which items of information to report and which to withhold. This is true of most real-life memory situations. In contrast, under forced-report conditions, in which people are required to answer each and every question (as in standard forced-choice recognition tests), the input-bound (quantity) and output-bound (accuracy) memory measures are operationally equivalent.

The critical importance of report option for memory accuracy is well illustrated in connection with an empirical puzzle referred to as the “recall-recognition paradox” (Koriat & Goldsmith, 1994). The established wisdom in eyewitness research holds that testing procedures involving recognition or other directed forms of questioning can have “contaminating” effects on memory (e.g., through trace alteration or source confusion), and hence are less reliable in eliciting accurate reports than are free-narrative modes of testing (see, e.g., Brown, Deffenbacher, & Sturgill, 1977; Fisher, Geiselman, & Raymond, 1987; Gorenstein & Ellsworth, 1980; Hilgard & Loftus, 1979; Loftus, 1979, 1982; Loftus & Hoffman, 1989). This contrasts with the established superiority of recognition over recall testing in traditional laboratory experiments (e.g., Brown, 1976; Shepard, 1967; but see Tulving & Thomson, 1973). Because traditional laboratory experiments have focused primarily on the quantity of memory rather than on its accuracy, the paradox seems to imply an interaction between *test format* (recall vs recognition) and *memory property* (quantity vs accuracy), such that recognition testing yields better memory-quantity performance, whereas recall testing yields better memory-accuracy performance. This general pattern was indeed observed both with adults and with children (see Dent & Stephenson, 1979; Hilgard & Loftus, 1979; Lipton, 1977; Neisser, 1988).

This interpretation, however, is complicated by the possible contribution of *report option* (free vs forced reporting), because questioning procedures that differ in test format often differ in report option as well. For instance, in free-narrative and free-recall testing, people not only produce their own answers (test format), they also have the option to volunteer or withhold answers (free report). In contrast, in forced-choice recognition testing and some forms of directed questioning, witnesses are not only confined to choosing between the alternatives presented by the interrogator (test format), they are also required, explicitly or implicitly, to answer each and every question (forced report).

The experimental approach used by Koriat and Goldsmith (1994) to disclose the role of report option is also used in the present study with children, and so we will describe it in some detail. They orthogonally manipulated test format and report option, creating four types of memory testing methods: In addition to the standard methods of free recall and forced recognition, they also included two relatively uncommon procedures, forced recall, in which participants were required to provide an answer to all questions, and free recognition, in which they had the option of skipping over multiple-choice items. The same motivation for accurate reporting was maintained for all conditions (through monetary payoffs), and both quantity and accuracy scores were calculated. Thus, the design included all combinations of Test Format (2) × Report Option (2) × Memory Property (2) conditions.

The pattern of results that Koriat and Goldsmith obtained for adults can be summarized as follows: First, the “paradoxical” pattern was replicated, in which forced recognition yielded better quantity performance but poorer accuracy performance than free recall. Second, however, when test format and report option were disentangled, it became clear that whereas the superior recognition-quantity performance was indeed due to test format, the superior recall-accuracy performance was entirely due to the option of free report. Given the freedom to control their own memory reporting, participants were able to enhance the accuracy of their reports substantially compared to forced-report testing, and this was true for both the recall- and the recognition-testing procedures. In fact, contrary to the established wisdom in eyewitness research, memory accuracy was no better (and in some cases somewhat worse) under recall than under recognition testing when participants were allowed equal opportunity to screen their answers. Furthermore, free recognition turned out to be the more effective testing procedure, because it elicited more information than free-recall testing while yielding an equivalent accuracy rate. These results, then, call for greater caution in evaluating the effectiveness of different questioning techniques, because in most previous research with both adults and children, report option and test format have been confounded. When these two factors are unconfounded, report option rather than test format emerges as the critical factor in the assessment of memory accuracy.

The finding that report option is critical for memory accuracy raises the question of whether this is also true for children. Are children able to exploit the option of free report to enhance the accuracy of what they report? Can we trust an 8-year old child, for instance, to effectively censor what she reports, providing only those pieces of information that are likely to be correct? Will her performance be sensitive to specific incentives for accurate reporting? Will children also exhibit better overall performance (superior quantity and equal accuracy) for recognition as compared to recall testing when given the option of free report? Are there developmental changes in all of the aspects just mentioned?

These questions, which are the focus of the present study, have both theoretical and practical implications. On the theoretical side, a consideration of metacognitive processes can perhaps shed light on some potential sources of age differences in memory accuracy. Although, as mentioned earlier, metamnemonic skills have played a key role in understanding some deficits in children's memory-quantity performance, there has been much less attention to their potential contribution to memory accuracy. As far as the use of report option is concerned, it is unclear to what extent children are capable of strategically regulating their memory reporting in accordance with competing demands for maximizing both the quantity ("the whole truth") and the accuracy ("nothing but the truth") of their testimony. Such regulation presupposes a well-developed ability to differentiate between items of information that are likely to be right and those likely to be wrong (i.e., monitoring), and then to put that differentiation to use in deciding which items of information to volunteer and which to screen out (i.e., control).

Previous work on the monitoring and control processes of children suggests that children may be relatively deficient in either or both of these. With regard to memory monitoring, several studies have found developmental trends in the accuracy of prospective feeling-of-knowing judgments and judgments of learning (e.g., Lockl & Schneider, 2001; Schneider, Visé, Lockl, & Nelson, 2000; for reviews, see Schneider, 1985; Schneider & Bjorklund, 1998). Of more relevance to the present article, Pressley, Levin, Ghatala, and Ahmad (1987) found a developmental trend for grade-school children with regard to the accuracy of “retrospective” confidence, that is, the ability to monitor the correctness of the answers that one has produced. In addition, children may also be less able or less willing than adults to control their memory reporting on the basis of their subjective monitoring (see Dufresne & Kobasigawa, 1989). For example, several studies suggest that children are particularly reluctant to say “don’t know” in response to memory questions (e.g., Cassel et al., 1996; Dale, Loftus, & Rathbun, 1978; Geiselman & Padilla, 1988; Mulder & Vrij, 1996; Roebers & Schneider, 2000). Thus, one idea that was examined is that children are capable of screening out incorrect answers, but that they do not do so because they do not realize that “don’t know” is an acceptable response. In support of this contention, Mulder and Vrij (1996) found that explicitly instructing children aged 4 to 10 that “I don’t know” is an acceptable answer significantly reduced the number of incorrect responses to misleading questions (i.e., questions about events that did not in fact occur). Moston (1987) found that such instructions induced children aged 6 to 10 to make more “don’t know” responses, but in that study, this had no effect on the overall proportion of correct responses. On the other hand, Cassel et al. (1996) found that children (kindergartners, second graders, and fourth graders) exhibit a developmental trend and a greater tendency than adults to provide wrong answers to leading questions even when they are reminded that they have the option to say “don’t know” (see also, Roebers & Schneider, 2000).

Turning to the question of test format, although Koriat and Goldsmith’s (1994) results with adults indicate that recognition testing does not impair memory accuracy under free-report conditions, children are generally held to be more susceptible than adults to the effects of misinformation and leading questions (Bruck & Ceci, 1999; Cassel et al., 1996; Ceci & Bruck, 1993). Perhaps, then, test format might be a critical factor affecting the accuracy of children’s memory reports, even under free-report conditions.

Beyond their theoretical interest, these questions also have important practical implications: If young children can in fact strategically regulate their memory reporting to produce a more accurate record of past events, then they should be allowed—even encouraged—to decide for themselves which items of information to volunteer and which to withhold. The option of free report, combined with an explicit message that puts a premium on accurate reporting, will be expected to elicit more dependable testimony from a child than questioning methods that take away such control. By contrast, if children cannot be trusted to screen their own memory reports effectively, then perhaps forced-report questioning methods would be preferable (see Fisher, 1995).

EXPERIMENT 1: CHILDREN’S USE OF REPORT OPTION TO INCREASE ACCURACY

In Experiment 1, younger (second and third graders) and older (fifth and sixth graders) children were exposed to a computerized slide show and accompanying narration describing an event in the life of a family. Departing somewhat from the previous studies with adults (Koriat & Goldsmith, 1994), these more naturalistic stimulus materials were intended to better approximate the type of details that witnesses are questioned about, though, of course, the children were observers rather than participants in the "event," and levels of stress were relatively low. The two age groups were chosen in accordance with prior results indicating significant developmental improvements in monitoring and control processes during the grade-school years (e.g., Dufresne & Kobasigawa, 1989; Pressley et al., 1987; see Schneider, 1999, for a review).

Thirty questions about the story were then presented in either of two test formats (manipulated between participants), a recall or a multiple-choice recognition format. Report option was orthogonally manipulated as follows: Forced-report participants were required to answer all 30 questions, even if this required them to guess. Free-report participants, in contrast, were allowed to choose which questions to answer and which to skip. An explicit accuracy incentive was used to motivate selective reporting: Essentially, the children were rewarded for each correct answer but penalized by an equal amount for each incorrect answer. Memory performance in all conditions was scored for both quantity (input-bound percent correct) and accuracy (output-bound percent correct).

In addition, immediately following this first phase of the experiment, children in the forced-report condition were asked to take the same test again under free-report instructions, and conversely, children in the free-report condition were asked to take the test again under forced-report instructions. The design for the first phase of the experiment, then, conformed to a $2 \times 2 \times 2$ factorial: Age (younger, older) $\times$ Report Option (free, forced) $\times$ Test Format (recall, recognition), whereas the second phase was intended to provide additional information about the report-screening process.

This design allowed us to address the following questions: First, with regard to report option, can children regulate the accuracy of their memory reports by utilizing the option of free report? That is, can they enhance their memory-accuracy performance by screening out incorrect answers, thus exhibiting better performance under free- than under forced-report conditions? If so, how severe will be the resulting cost in memory-quantity performance? In previous work with adults, Koriat and Goldsmith (1994, 1996c) found that accuracy is generally enhanced only at the cost of a reduction in quantity performance, and that the extent of this quantity-accuracy trade-off depends, among other things, on monitoring effectiveness, that is, the ability to differentiate correct from incorrect answers. If children are deficient in this ability, then we might expect developmental differences in both the accuracy benefits and quantity costs of free memory reporting.

Second, with regard to test format, the experiment allowed us to examine whether children (unlike adults; see Koriat & Goldsmith, 1994, 1996c) are susceptible to memory contamination by recognition testing. If so, they may be less able to utilize the option of free report to screen out incorrect answers, and hence achieve lower levels of memory accuracy under recognition than under recall testing.

Method

Participants. One hundred sixty Israeli school children from an elementary school in northern Israel participated in the study. The children were mostly of middle-class and upper-middle-class socioeconomic background. The younger age group consisted of 40 second graders (mean age = 7.9) and 40 third graders (mean age = 8.9), with an overall mean age of 8.4. The older age group consisted of 40 fifth graders (mean age = 11.0), and 40 sixth graders (mean age = 11.9), with an overall mean age of 11.5. The children from both age groups were randomly assigned to the four Report Option (free vs forced) × Test Format (recall vs recognition) conditions (20 participants in each cell), with the constraint that there be an equal number of children from each grade and an approximately equal number of boys and girls (between 9 and 11 children of each sex) in each condition.

Materials and methods. A computerized slide show lasting about 5 min was developed. It consisted of 27 color photographs presented on a high-resolution (super VGA) screen controlled by an IBM-PC compatible computer. The pictures were presented at a rate of 8 s per picture, with about 1/2 s between pictures. Each picture was accompanied by a recorded narration spoken in Hebrew by a professional radio broadcaster, which was produced through speakers attached to the computer.

The show ("On the Way to the Picnic") depicts a staged incident in the life of a family preparing to go out on a picnic, in which the family cat climbed up an electricity pole and had to be enticed by various means to come down. The story begins with an introduction to each family member (the parents and three daughters), which is followed by a sequence of events culminating in the successful rescue of the cat.

A 30-item memory questionnaire concerning the slide show was developed in both a recall and a five-alternative recognition format. The questions related both to aspects that were central to the story and to peripheral details (e.g., Why did the cat…? What was the color of the mother's dress?). The questions in the two test formats were identical. However, whereas in the recall format each question was followed by a blank line for writing down the answer, in the recognition format it was followed by five alternatives, one of which was correct.

Procedure. The experiment was conducted in a quiet room at school during school hours. Either one or two children were exposed to the slide show at a time. The initial instructions, which were presented both visually and auditorily on the computer, told the children that they would hear a story accompanied by a series of pictures, and that their task is to try to remember as much as possible from the story, because when the story is over they will be asked to answer several questions about it. When the presentation was over, the children received a booklet that included the memory test and instructions appropriate for their individual condition. The experimenter read the instructions aloud while the children followed along in their booklet.

For the free-report conditions, the memory test was described as a game in which the child would win one bonus point for each correct answer but lose one point for each incorrect answer. However, the child need not answer all questions; for those questions skipped, the child would not gain any points but neither would the child lose any points. It was emphasized to the child that the child should deliberate whether to mark an answer or not, and that if the child was unsure about the answer then the child should consider skipping it because otherwise the child might lose a point. The child was told that when the study was over, the child would receive a prize whose value would depend on the number of points earned. Several queries were presented to ascertain that the child understood the rules of the "accuracy game" (e.g., "Suppose you answer 8 questions and 4 of these are correct, but 4 are wrong. How many points would you win?").

For the forced-report conditions, the instructions simply specified that for each question the child must choose/provide an answer, and that the child must do so even if the child had to guess. It was emphasized, however that even when the child does not know the answer to a question, the child should make an effort to choose/provide as accurate an answer as possible. At the end of the task, the experimenter ascertained that in fact no answer was left unanswered.

For the recognition conditions (both free and forced), each question was followed by five alternative answers, and the child indicated the child's choice by circling the correct answer. In the recall conditions, there was a space next to each question.

Immediately following the first test phase, the children were administered a second test phase. In this phase, children in the forced-report condition were asked to take the same test again, but under free-report instructions. Conversely, children in the free-report condition were asked to take the test again under forced-report instructions. The purpose of phase 2 was to help shed light on the nature of the free-report screening process.

The consent of the parents and of the school was obtained before beginning the study.

Results and Discussion

Memory-quantity scores were computed as the (input-bound) percentage of questions correctly answered by the participant, and accuracy scores were computed as the (output-bound) percentage of answers that were correct. As mentioned earlier, in the forced-report conditions, the two types of scores are operationally equivalent.

In scoring the recall answers, we employed two different criteria in deciding whether an answer was correct or not: Under a strict criterion, the answer was considered correct only if it matched or was virtually identical to the correct answer that appeared on the multiple-choice version of the test. Under a more liberal criterion, answers were accepted as correct if they were considered to retain the gist of the correct answer: For instance, "large blue bowl" was accepted in place of "blue wash basin," "on the car" was accepted in place of "on the jeep." and so forth. There were relatively few answers that were scored as correct under the liberal criterion but incorrect under the strict criterion: an average of 0.80

answers (out of 30 items) for the younger children and 0.65 answers for the older children. The results under both scoring rules were quite similar, and we will report the results in this and the later experiments using the strict scoring rule.

As noted earlier, the study was designed to include two age groups, "younger" (second and third grade) and "older" (fifth and sixth grade) children. Nevertheless, several preliminary analyses were conducted to compare the patterns of performance of the second and third graders, and of the fifth and sixth graders. These analyses did not yield any meaningful differences, hence the results will be reported in terms of the two intended age groups.

The mean quantity and accuracy scores for each Report Option × Test Format condition are presented separately for each age group in Table 1. Unless stated otherwise, all of the following statistical analyses are based on analyses of variance (ANOVAs) involving one or more of the three independent variables, Report Option, Test Format, and Age Group.

Memory quantity performance. We first examine the results from the perspective of the traditional quantity-oriented approach to memory. Previous studies have used either a free-recall or forced-choice recognition format and have generally found larger age differences for recall than for recognition testing (e.g., Cassel et al., 1996; List, 1986; Marin et al., 1979). In line with such findings, the older children in this experiment showed a marginal tendency toward better quantity performance than the younger children on the free-recall test, $F(1, 38) = 3.25$, $p < .08$, but there was no such tendency on the forced recognition test, $F < 1$. Examining the remaining two measures, a significant advantage for the older children was found both on the forced-recall test, $F(1, 38) = 25.67$, $p < .0001$, and on the free-recognition test, $F(1, 38) = 5.89$, $p < .05$. Overall, however, stronger age effects were found for recall than for recognition testing, $F(1, 152) = 4.02$, $p < .05$, for the interaction.

Memory accuracy performance. Turning now to the children's accuracy performance, several effects can be discerned. First, and most fundamental, is the

TABLE 1
Means and Standard Deviations of Quantity Scores (Input-Bound Percent Correct) and Accuracy Scores (Output-Bound Percent Correct) by Test Format and Report Option in Experiment 1

		Report option					
		Free report				Forced report	
		Quantity		Accuracy		Quantity or accuracy[a]	
Age group	Test format	*M*	*SD*	*M*	*SD*	*M*	*SD*
Younger	Recall	54.8	11.62	74.9	12.78	57.3	10.74
	Recognition	63.8	11.25	79.9	11.48	71.0	8.38
Older	Recall	62.3	14.51	80.8	11.34	73.0	8.71
	Recognition	71.8	9.52	88.1	9.32	72.3	11.24

[a]Quantity and accuracy are equivalent for forced report.

role of report option: Children in both age groups were able to enhance the accuracy of their memory reports substantially when given the option of free report. Overall, free-report accuracy (80.9) was higher than forced-report accuracy (68.4), $F(1, 152) = 55.92$, $p < .0001$, and this difference held for children in both age groups and for both recall and recognition testing. Thus, children as young as 7 years old are apparently able to utilize the option of free report to provide more dependable testimony, regardless of whether a recall or recognition test format is used.

The free-report advantage was not equivalent for all conditions, however, as indicated by a significant Age Group × Test Format × Report Option interaction, $F(1, 152) = 6.13$, $p < .0001$. For the recall test, the free-report advantage was greater for the younger children (18 percentage points) than for the older children (8 points), whereas for the recognition test, the advantage was greater for the older children (16 points) than for the younger children (9 points).

A second interesting finding is the developmental trend in the absolute level of accuracy attained under free-report conditions. A 2-way ANOVA on the free-report accuracy scores yielded significant effects for age group, $F(1, 76) = 7.72$, $p < .01$, and test format, $F(1, 76) = 5.88$, $p < .05$, with no interaction, $F < 1$. Indeed, for both test formats, older children exhibited higher free-report accuracy performance than younger children. Thus, not only could the younger children generally produce fewer correct answers than the older children (see earlier quantity results), the answers that they chose to provide under free-report conditions were also less likely to be correct: Overall, fully 23% of the younger children's answers were incorrect, compared to about 16% for the older children. It is important to note that this developmental trend in memory accuracy was exhibited under conditions in which a common explicit incentive for accurate reporting was provided.

Finally, consider the effect of test format. Consistent with earlier findings with adults (Koriat & Goldsmith, 1994, 1996c), there is no support for the common belief that recognition testing is inherently contaminating. On the contrary, when children in the recall and recognition conditions were given an equal opportunity to screen out incorrect answers (i.e., under free-report conditions), it was in fact recognition testing that yielded more accurate memory reports (84.0) than recall (77.9) (significant, as just reported). Thus, in comparing the accuracy of free-recall and forced-recognition testing (see earlier discussion of the recall-recognition "paradox"), the superior accuracy of free-recall (77.9) over forced recognition (71.7) is again shown to stem from the effect of report option rather than test format.

The accuracy–quantity trade-off. In previous work with adults, Koriat and Goldsmith (1994, 1996c) found that the accuracy gain from the use of free-report option generally comes at a cost in quantity performance, suggesting that the enhanced accuracy is achieved by a selective but imperfect elimination of candidate answers that are likely to be wrong. As can be seen in Table 1, this trade-off pattern holds true for children as well. In comparing free- and forced-report performance across the two age groups and the two test formats, a 13-percentage-point increase in accuracy performance (statistically significant, as reported earlier) was accompanied by a 5-point decrease in quantity performance, which was also significant, $F(1, 152) = 9.13$, $p < .005$. The quantity cost was equivalent for the two age groups, $F < 1$, but there was a crossover interaction between age group and test format, $F(1, 152) = 4.63$, $p < .05$: Higher quantity costs were suffered by the younger children than by the older children under recognition testing, but the reverse was true under recall testing (neither of these simple effects reached statistical significance, however). Thus, considered together with the similar interaction reported earlier regarding the accuracy improvement, older children achieved higher accuracy gains with a tendency toward lower quantity costs than younger children under recognition testing, but younger children achieved higher accuracy gains with a tendency toward lower quantity costs than older children under recall testing. This pattern may simply reflect the different levels of baseline (forced-report quantity) performance: Younger children may have more to gain (in terms of accuracy) and less to lose (in terms of quantity) by screening out answers under recall than under recognition testing.

The screening process underlying the effects of report option. According to our theoretical framework, both the accuracy increase and the quantity decrease that ensue from giving children the option of free report stem from the children's ability to screen out incorrect answers. Indeed, the very fact that quantity performance tends to decrease under free-report conditions is important in discounting other types of explanations, for instance, that increased motivation or retrieval effort (stemming from the chance to win points toward a prize) is the primary source of the improved accuracy performance (for further evidence, see Koriat & Goldsmith, 1994, 1996c). Additional insight can be gained into the source of the observed pattern of accuracy and quantity performance by examining the data from the second phase of the experiment. Recall that in phase 2 of the experiment, the free-report children were asked to provide answers to all questions, even to those that they had initially skipped in phase 1.[1] Table 2 presents the mean number of answers withheld on phase 1 of the experiment, and the mean accuracy of those answers on phase 2, for each age group × test format condition.

As can be seen, the children volunteered an average of 23.4 answers (out of 30) on the free-report phase, withholding 6.6 answers. Only four children failed to exercise the option of selective reporting, not withholding any answers at all (one older and three younger recognition participants). More answers were withheld under recall testing (7.5) than under recognition testing (5.7), $F(1, 76) = 5.50$, $p < .05$, but there was no age difference, $F < 1$, and no Age × Format interaction, $F < 1$.

[1]Conversely, children in the forced-report conditions were asked to answer the questions again under free-report instructions in phase 2. Considering the data from both phases in all conditions essentially allows for within-participant analyses of the effects of report option. However, because the results of these analyses yield the same basic pattern as the between-participant analyses, they are for the most part redundant and will not be reported. We report here only those phase-2 results that help clarify the screening process used by the children in the free-report condition in phase 1.

TABLE 2
Means and Standard Deviations of the Number of Withheld Answers (in Phase 1) and the Accuracy of Withheld Answers (Based on Phase 2 Responses) in the Free-Report Conditions of Experiment 1

		Number of withheld answers		Accuracy (% correct) of withheld answers	
Age group	Test format	*M*	*SD*	*M*	*SD*
Younger	Recall	8.1	3.07	18.0	16.73
	Recognition	6.8	3.52	28.7	22.11
Older	Recall	6.9	4.02	13.5	16.11
	Recognition	5.8	2.18	37.0	31.13

How does report option enhance the accuracy of children's memory reports? If report option allows children to selectively screen out answers that are likely to be wrong, then the answers they choose to withhold shall be less likely to be correct than those they choose to volunteer. Indeed, the answers withheld on the free-report phase were far less likely to be correct (24.3%; based on Table 2) than were the volunteered answers (81.9%; based on Table 1), $F(1, 72) = 425.14$, $p < .0001$ (mixed ANOVA). Moreover, this was true for the vast majority of the children: Examination of the individual participants' performance indicated that fully 18 of the 20 children in the younger age group in each test-format condition achieved higher accuracy in the initial free-report phase than in the subsequent forced-report phase, compared to 19 of 20 children in the older age group under recall testing and 20 of 20 in the older age group under recognition testing. Overall, withheld recognition answers were more likely to be correct than were withheld recall answers, $F(1, 72) = 11.24$, $p < .005$, but there was no age group effect, $F < 1$, and no Age Group × Test Format interaction, $F(1, 72) = 1.59$, *ns*. The test-format effect is probably due simply to the greater chance of guessing a correct answer on the recognition test.

Considering the absolute number of correct answers that were unduly withheld on the free-report phase, it is interesting to note that the mean was quite low for both age groups (1.91 and 1.68 answers, for the younger and older children, respectively). The two age groups did not differ in this respect ($F < 1$), nor was there an effect for test format, $F(1, 72) = 2.09$, *ns*, or for the interaction, $F < 1$. Thus, although one might worry that perhaps children would misuse the option of free reporting, there is no indication in these data that they screened out an unduly large number of correct answers.

As noted earlier, it has been suggested that young children are particularly unlikely to say "don't know" in response to memory questions (see Cassel et al., 1996). Can this possibly account for the somewhat inferior accuracy of the younger children in this experiment? As mentioned earlier, the younger children did not provide more answers than the older children on the free-report phase, that is, they were no less likely to use the "don't know" option. However, because the younger children apparently had a lower number of correct answers available in memory (based on their lower forced-report quantity performance in phase 2 of this experiment, 65.8% vs 72.0% for the younger and older children, respectively, $F[1, 76] = 6.44$, $p < .05$; no interaction with test format, $F < 1$), one might, in fact, have expected them to provide fewer answers than the older children. Thus, the equivalent volunteering rates for the younger and older children actually suggests that the younger children were somewhat more liberal in their control policy, tending to provide answers that the older children would have withheld. We will return to this point in discussing the results of the following experiments.

EXPERIMENT 2: THE ENHANCEMENT OF MEMORY ACCURACY AS A FUNCTION OF ACCURACY INCENTIVES

Experiment 1 indicated that children as young as 8 or 9 years old are capable of utilizing the option of free reporting to enhance the accuracy of their memory reports. Experiment 2 examines whether they are also sensitive to the level of accuracy incentive, increasing the accuracy of their report even further when a heavier premium is placed on memory accuracy. With adults, Koriat and Goldsmith (1994, Experiment 3) found that a high incentive condition, in which participants received a monetary bonus for each correct answer, but forfeited all winnings if even a single incorrect answer was volunteered, yielded better accuracy performance than a moderate incentive condition in which the reward for each correct answer equaled the penalty for each incorrect answer. In Experiment 2, we ask whether children too can increase their accuracy further when a stronger incentive is provided, and if so, whether this is equally true for the younger and older children.

A second, related question is whether age differences in memory accuracy will be preserved when there is a very strong and explicit accuracy incentive. As noted previously (Koriat & Goldsmith, 1994, Experiment 3), a strong demand for accuracy is perhaps more characteristic of courtroom testimony,[2] so it is important to determine whether the developmental trend in memory accuracy observed in Experiment 1 will remain under high incentive conditions.

Finally, given the importance attributed to the questioning format in eyewitness research (particularly with children), it should be interesting to examine whether the pattern of age and incentive effects differs for the two test formats, recall and recognition.

Experiment 2 was similar in procedure to the free-report conditions of Experiment 1 except that high incentive instructions were used. The results will be compared to those of the free-report conditions of Experiment 1, which will represent a "moderate incentive" condition in this comparison.

[2]One might question to what extent the pay-off schedule in this experiment captures the type of accuracy motivation operating in real-life courtroom situations. We can only say that from our informal observations of the children, they appear to take quite seriously the threat of losing all their winnings if they provide even a single wrong answer.

Method

Participants. Eighty Israeli children participated in this experiment. They were taken from the same school as the children who participated in Experiment 1, approximately 1 year later. The younger age group consisted of 40 second graders (mean age = 8.0) and 40 third graders (mean age = 9.0), with an overall mean age of 8.5. The older age group consisted of 40 fifth graders (mean age = 11.1), and 40 sixth graders (mean age = 11.9), with an overall mean age of 11.5. The children from both age groups were randomly assigned to the two Test Format (recall vs recognition) conditions, with the constraint that there be an equal number of children from each grade and an approximately equal number of boys and girls (between 8 and 12 children of each sex) in each condition.

Materials and methods. The materials and methods were identical to those in Experiment 1 with the following two exceptions: First, only a free-report condition was used, with both a recall and a recognition test format. (However, as in Experiment 1, a second, forced-report phase followed, in which the participants answered all the questions again, including those that they had initially skipped.) Second, high incentive instructions were used in which it was explained that participants would win one point for each correct answer but would lose all winnings if even a single volunteered answer is wrong. Again, several queries were presented to the participants to ascertain that the rules of the "accuracy game" were clear.

Results and Discussion

Table 3 presents the mean quantity and accuracy scores for each age group for the two test formats and for the moderate incentive (Experiment 1) and high incentive (Experiment 2) conditions.[3]

The effects of accuracy incentive. Were the children sensitive to the higher level of accuracy incentive provided in this experiment? Comparing the results for Experiment 1 (moderate incentive) and Experiment 2 (high incentive), a 3-way ANOVA, Incentive × Age Group × Test Format, on the accuracy scores indicated that the children did in fact attain a higher level of accuracy in the high incentive condition (87.9) than in the moderate incentive condition (80.9), $F(1, 152) = 15.27$, $p < .0001$. None of the interactions involving incentive were significant, and indeed, the incentive effect was significant both for the older children, $F(1, 76) = 10.54$, $p < .005$, and for the younger children, $F(1, 76) = 5.94$, $p < .05$, analyzed separately. Thus, even young children are apparently able to regulate the accuracy of their memory reports, providing more accurate reports when given a stronger incentive to do so. It should be noted, however, that relatively few of the children succeeded in achieving 100% accuracy in the high incentive condition:

[3]Because the children were not randomly assigned to the two incentive conditions we compared their forced-report quantity performance (phase 2) to determine whether there was any difference in their ability to remember the details of the story. These scores averaged 68.9% and 67.6% for the low incentive (Experiment 1) and high incentive (Experiment 2) participants, respectively, $F < 1$.

TABLE 3

Means and Standard Deviations of Quantity Scores (Input-Bound Percent Correct) and Accuracy Scores (Output-Bound Percent Correct) for the Free-Report Conditions of Experiment 2 (High Accuracy Incentive) and Experiment 1 (Moderate Accuracy Incentive)

		Incentive							
		Moderate				High			
		Quantity		Accuracy		Quantity		Accuracy	
Age group	Test format	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Younger	Recall	54.8	11.62	74.9	12.78	49.3	11.06	80.3	12.52
	Recognition	63.8	11.25	79.9	11.48	62.5	11.49	88.4	13.98
Older	Recall	62.3	14.51	80.8	11.34	58.5	13.22	91.9	6.61
	Recognition	71.8	9.52	88.1	9.32	62.0	12.40	91.1	10.89

7 of 20 children from each age group in the recognition format, and 4 of 20 older children and 2 of 20 younger children in the recall format.

Age and test-format effects under a high accuracy incentive. Given the high premium placed on accurate reporting in courtroom situations, it is of interest to examine whether the effects of age group and test format observed under the moderate accuracy incentive in Experiment 1 still hold under the stronger accuracy incentive of Experiment 2. On the whole, it seems that they do. Although a very strong and explicit premium was placed on accurate reporting in Experiment 2, the older children still attained a higher level of accuracy (91.5) than the younger children (84.4), $F(1, 76) = 7.93$, $p < .01$. Thus, the observed age difference in memory accuracy does not appear to depend on the level of accuracy incentive. Unlike Experiment 1, however, there was a tendency toward a more pronounced age difference on the recall than on the recognition test, $F(1, 76) = 3.11$, $p < .09$. Also, unlike Experiment 1, there was no test-format effect on accuracy, $F(1, 76) = 2.03$, *ns*, though there was a tendency toward more accurate recognition than recall for the younger children, $F(1, 38) = 3.71$, $p < .07$. In any case, there was again no indication of memory contamination stemming from the recognition format.

The quantity-accuracy trade-off. Koriat and Goldsmith (1994, 1996c) observed a quantity-accuracy trade-off not only in comparing forced-report to free-report performance, but also in comparing moderate incentive with high incentive conditions. If children exhibit a similar dynamic, then we should expect the increased accuracy performance under the high-accuracy incentive to be achieved at an additional cost in memory-quantity performance. Indeed, on the average, the 7-percentage-point increase in accuracy for the high incentive condition relative to the moderate incentive condition (reported earlier) was accompanied by a statistically significant 5-point decrease in memory-quantity performance, $F(1, 152) = 7.34$, $p < .01$. Comparing the magnitude of this trade-off to that observed in Experiment 1, we again find support for the principle that was previously predicted and demonstrated for adults (Koriat & Goldsmith, 1996c): When people are merely given the option of free report (under a moderate accuracy incentive), they can achieve substantial gains in accuracy with relatively small reductions in quantity performance, but when given stronger accuracy incentives, improving accuracy further becomes relatively costly. In addition, there were the expected main effects on quantity performance of test format (recognition superior to recall), $F(1, 152) = 21.59$, $p < .0001$, and age group (older children superior to younger children), $F(1, 152) = 10.19$, $p < .005$. There were no significant interactions.

The screening process underlying the effects of high accuracy incentive. As in Experiment 1, some insight can be gained into the mechanisms underlying the effects of accuracy incentive by examining the phase-2 data. First, a 3-way ANOVA on the number of answers withheld by the children on the initial free-report phase indicated that, as expected, children in the high incentive condition withheld more answers on the average (10.0) than did children in the moderate incentive condition of Experiment 1(6.6), $F(1, 152) = 33.12$, $p < .0001$. There were no interactions with age group or with test format (all F's < 1). Thus, also under the high accuracy incentive, the younger children did not use the "don't know" option any more often than did the older children ($F < 1$). As pointed out in discussing the results of Experiment 1, given their poorer memory in terms of the number of available correct answers (mean phase-2 forced-report quantity performance was 64.6% vs 70.7% for the younger and older age groups respectively, $F[1, 76] = 5.65$, $p < .05$; no interaction with test format, $F < 1$), this again implies a more liberal volunteering policy on the part of the younger children.

The accuracy of the withheld answers in the high incentive condition was again quite low (30.8%) compared to the accuracy of the reported answers in this condition (87.9%), $F(1, 76) = 604.43$, $p < .0001$ (mixed ANOVA). It was, however, somewhat higher than in the moderate incentive condition (24.3%), $F(1, 148) = 3.94$, $p < .05$. Thus, the children seem to respond to the higher accuracy incentive not only by withholding more answers, but also by withholding answers that were (somewhat) more likely to be correct. The net result is a larger number of correct answers withheld in the high incentive condition (3.2) than in the moderate incentive condition (1.8), $F(1, 148) = 17.98$, $p < .0001$, reflecting the quantity cost of improved accuracy performance. These differences did not interact with age group or test format.

In sum, the results of Experiment 2 indicate, first, that not only can children at the ages studied here enhance the accuracy of their memory reports when given the option of free report (Experiment 1), but under free-report conditions they can enhance their accuracy even further when given a stronger incentive for accurate reporting. This was true even for the younger age group and for both test formats. Thus, both groups of children demonstrate an ability to regulate their memory reporting in a sensible manner, increasing the accuracy of their reports in accordance with explicit payoffs for memory accuracy.

Second, the older children evidenced higher memory-accuracy performance than the younger children even under the very strong accuracy incentive used in Experiment 2. This is an important finding, because as noted earlier, the pay-off schedule for this incentive is perhaps more similar to the situation of courtroom testimony, in which a high premium is placed on report accuracy. Although the data do not allow us to pinpoint the source of the observed age difference, there is some evidence suggesting that part of it may derive from the use of a more liberal control policy by the younger children.

EXPERIMENT 3: MEMORY ACCURACY AFTER A ONE-YEAR DELAY

Experiment 3 consisted of a retest, a year later, of some of the children who had participated in Experiment 1. It may be recalled that in Experiment 1, memory tests were administered shortly after the children were exposed to the slide show. In many forensic situations, however, there may be very long delays between the witnessed event and initial questioning about it, and subsequent testimony. Here we examine whether the effects that we have described so far are also obtained when memory testing is carried out again, long after the event has taken place.

There are several specific questions that will be addressed in Experiment 3. First, are children still capable of using report option effectively to enhance their memory accuracy after one year? Second, is the regulation of memory accuracy by children sensitive to the level of accuracy incentive even after one year? How does the level of memory accuracy attained after one year compare to that achieved in immediate testing, and are there still age differences in memory accuracy after one year?

Method

Participants. Of the 80 children who participated in the free-report conditions of Experiment 1, 51 children can be contacted and tested a year later, with between 11 and 14 children in each of the four age group $\times$ test-format conditions.[4]

Procedure. There was no presentation of the slide show. Children were reminded that they had seen a slide show a year earlier, and each child was tested using the same test format (recall or recognition) that the child was given a year earlier. There were three testing phases: Phase 1 provided free-report instructions with a moderate accuracy incentive as in Experiment 1. In phase 2, the children took the same test again under forced-report instructions. In phase 3, they answered the questions again under free-report instructions, but now with a high accuracy incentive as in Experiment 2.

[4]In view of the relatively high attrition rate, we compared the performance in Experiment 1 of the 51 children who participated in Experiment 3, with the performance of the 29 (free-report) children who did not. There were no differences in their free-report accuracy and quantity scores, nor in their forced-report (phase 2) quantity scores (all F's < 1). We conclude, then, that this sample was representative.

Results

Table 4 presents the mean quantity and accuracy scores for the children participating in this experiment. The table includes data on the children's memory performance upon immediate testing (Experiment 1), both free-report performance under a moderate accuracy incentive and forced-report performance (phases 1 and 2, respectively), as well as when retested one year later (Experiment 3), both free-report performance under moderate and high accuracy incentives (phases 1 and 3, respectively), and forced report performance (phase 2).

The effects of retention interval. What were the effects of retention interval on children's performance? We restrict our examination to the moderate incentive condition, which can be compared between tests at the two retention intervals. With regard to quantity performance, as would be expected, a 3-way mixed ANOVA, Retention Interval × Test Format × Age Group, indicated that forced-report quantity scores decreased dramatically over the one-year period, from an overall average of 69.4 in immediate testing to 38.1 after one year, $F(1, 47) = 258.37$, $p < .0001$. This decrease was more pronounced for recall than for recognition testing, $F(1, 47) = 4.76$, $p < .05$, for the interaction. There were no interactions with age group (all F's < 1). The decrement over time was even more dramatic for free-report performance ($F[1, 47] = 8.59$, $p < .01$, for the Retention Interval × Report Option interaction): Free-report quantity decreased from 63.8 in immediate testing to 28.2 after one year, $F(1, 47) = 283.86$, $p < .0001$, and this reduction tended to be somewhat more pronounced for recall than for recognition testing, $F(1, 47) = 3.75$, $p < .06$, for the interaction.

TABLE 4

Means and Standard Deviations of Quantity Scores (Input-Bound Percent Correct) and Accuracy Scores (Output-Bound Percent Correct) for the Children Participating in Experiment 3, as a Function of Age Group, Test Format, Time of Test (Immediate, Experiment 1; 1 Year Later, Experiment 3), Report Option (Forced, Free), and Accuracy Incentive (Moderate, Experiments 1 and 3; High, Experiment 3 Only)

			Immediate			One year later				
			Forced (phase 2)	Free moderate (phase 1)		Forced (phase 2)	Free moderate (phase 1)		Free high (phase 3)	
Age	Test format		QA[a]	QTY	ACC	QA	QTY	ACC	QTY	ACC
Younger	Recall	*M*	57.0	50.9	71.1	24.8	17.3	38.3	10.3	48.7
		SD	11.59	10.86	14.21	8.48	10.31	18.71	6.58	27.83
	Recognition	*M*	72.8	68.2	80.2	45.4	37.7	48.7	21.8	64.7
		SD	10.79	10.24	14.03	11.67	12.94	14.29	11.91	18.18
Older	Recall	*M*	68.5	64.1	84.6	29.5	18.2	45.8	12.8	69.1
		SD	10.68	14.48	8.00	10.44	9.87	17.76	5.75	26.36
	Recognition	*M*	79.5	72.1	88.4	52.9	39.5	65.6	23.1	78.1
		SD	10.12	10.75	9.54	7.61	12.19	15.28	11.05	17.42

[a]QTY, quantity; ACC, accuracy; QA, quantity or accuracy (equivalent for forced report).

More interesting is the finding that accuracy too decreased dramatically over the same period, from an overall average of 81.1% in immediate testing to only 49.6% after one year. Thus, even when explicitly given the option of free report and encouraged to respond “don’t know,” only about one-half of what the children reported after one year (under a moderate accuracy incentive) was correct. A 3-way mixed ANOVA, Retention Interval × Test Format × Age Group, yielded significant main effects for retention interval, $F(1, 47) = 168.94$, $p < .0001$, age group, $F(1, 47) = 13.06$, $p < .001$, and test format, $F(1, 47) = 11.40$, $p < .002$, with no significant interactions. Thus, accuracy decreased equally for children in both age groups.

Nevertheless, the option of free report still allowed children to increase their accuracy, even after one year. Comparing the free-report and forced-report accuracy scores at the one-year test (within participant), the children were substantially more accurate under free report (49.6%) than under forced-report conditions (38.1%), $F(1, 47) = 34.56$, $p < .0001$. There was a marginal trend toward greater accuracy improvement for the older children (14.5 percentage points) than for the younger children (8.4 points), $F(1, 47) = 2.52$, $p < .12$. There were no interactions with test format (F’s < 1). Once again, this accuracy improvement came at the price of reduced quantity performance for free report (28.2) compared to forced report (38.1), $F(1, 47) = 94.99$, $p < .0001$. In this case, the quantity reduction was more pronounced for the older children (12.3 percentage points) than for the younger children (7.6 points), $F(1, 47) = 5.22$, $p < .05$.

It seems, then, that in utilizing the option of free report at the one-year test, the older children sacrificed somewhat more quantity performance to enhance their accuracy performance, and, in fact, the older children did achieve substantially higher free-report accuracy scores (56.1) than did the younger children (43.9) after one year. A 2-way ANOVA, Age Group × Test Format yielded significant main effects for age group, $F(1, 47) = 6.98$, $p < .05$, and for test format, $F(1, 47) = 10.63$, $p < .005$, but no interaction. Thus, in addition to the accuracy advantage for the older children, recognition testing was more accurate (57.4) than recall testing (42.4), reinforcing the earlier findings of no memory contamination stemming from the recognition format. At the same time, there was no difference between the age groups in their free-report memory quantity scores after one year (28.9 vs 27.5 for the older and younger children, respectively), and no interaction with test format (both F’s < 1). There was, however, an age advantage for the forced-report quantity performance (41.2 vs 35.1 for the older and younger children, respectively), $F(1, 47) = 4.94$, $p < .05$.

Again, some insight into the dynamics of the screening process can be gained by examining the answers that were withheld in the free-report phase. First, the children withheld an average of 12.5 answers after one year, compared to only 6.4 on immediate testing, $F(1, 47) = 39.31$, $p < .0001$. Older children tended to withhold slightly more answers (13.9) than did the younger children (11.1) at the one-year test, though this difference was only marginal, $F(1, 47) = 2.14$, $p < .15$. There was no interaction with test format ($F < 1$).

Second, the older children withheld significantly more correct answers (3.9) than did the younger children (1.9) at the one-year test, $F(1, 43) = 14.60$, $p <$

.0005. Again, there was no interaction with test format ($F < 1$). Hence, the accuracy of the withheld answers was higher for the older children (27.5%) than for the younger children (17.4%), $F(1, 43) = 9.40$, $p < .005$, with no interaction ($F < 1$). This pattern suggests, perhaps even more clearly than in the earlier two experiments, that part of the accuracy advantage of the older children stems from the use of a more conservative control policy: Despite having more correct candidate answers available in memory (based on forced-report performance), they tend to withhold more answers, thereby achieving higher free-report accuracy performance than their younger counterparts.

The effects of accuracy incentive after 1 year. So far we have focused on the results obtained under a moderate accuracy incentive. Are children still sensitive to differential accuracy incentives after a one-year retention interval? Memory performance was compared between phase 1 (free report, moderate incentive) and phase 3 (free report, high incentive) of the experiment. A three-way mixed ANOVA, Incentive × Age Group × Test Format on the accuracy scores indicated that, indeed, the children achieved substantially higher accuracy scores under the high incentive (65.1) than under the low incentive condition (49.6), $F(1, 47) = 41.28$, $p < .0001$. Furthermore, both the younger and the older children were equally sensitive to the accuracy incentives after one year ($F < 1$). Nevertheless, the older children continued to exhibit substantially higher accuracy scores (73.6) than the younger children (56.7) under the strong accuracy incentive, $F(1, 47) = 7.07$, $p < .05$.

How was this improved accuracy performance achieved? As expected, substantially more answers were withheld under the high accuracy incentive (21.8) than under the moderate incentive (12.5), $F(1, 47) = 163.01$, $p < .0001$. This difference was more pronounced for the recognition test (20.5 vs 9.2) than for the recall test (23.1 vs 15.9), $F(1, 47) = 7.80$, $p < .01$, but there was no effect or interaction involving age group. Not only were more answers withheld, however, but more correct answers were withheld under the high incentive (6.4) than under the moderate incentive (2.9) condition, $F(1, 43) = 133.40$, $p < .0001$. Thus, the underlying dynamic is again an accuracy-quantity trade-off: The children achieved higher accuracy performance under the strong accuracy incentive by sacrificing memory-quantity performance (mean quantity scores were 17.0 vs. 28.2, for the high and moderate incentives, respectively), $F(1, 47) = 111.64$, $p < .0001$.[5] There were no interactions involving age group (F's < 1).

In sum, although previous findings with adults indicate that accuracy (but not quantity) performance is relatively stable over time (see Brock, Fisher & Cutler, 1999; Ebbesen & Reinick, 1998; Goldsmith & Koriat, 1999), the results of this experiment indicate that memory accuracy in children may be much less stable,

[5]Note that although part of the superior accuracy in the high incentive (phase 3) compared to the moderate incentive (phase 1) condition could conceivably be the result of hypermnesia (the children having had two testing/retrieval opportunities subsequent to the moderate-incentive phase), the fact that the free-report quantity scores were lower in the high incentive phase discounts this as the primary source of the results.

reaching levels of only 50% after one year (see also Peterson, 1999; Poole & White, 1993; Salmon & Pipe, 2000). Such a low level of accuracy is actually quite surprising. It implies that not only is less information accessible to children over time (the typical finding in quantity-oriented research), but also that children are either less able or less willing to screen out incorrect answers after a long retention interval. Nonetheless, children in both age groups were still able to utilize the option of free report to enhance their memory accuracy after a one-year interval and regulated their reporting in accordance with the operative accuracy incentive.

GENERAL DISCUSSION

This study focused on the issue of the accuracy or dependability of children's testimony. Most previous studies that examined developmental changes in memory performance focused on memory quantity (see Schneider & Bjorklund, 1998). This is not surprising in view of the fact that in traditional item-based memory research, memory performance has been assessed primarily in terms of the input-bound measures of percent recall and percent recognition (see Koriat & Goldsmith, 1996a, 1996b). These measures ignore errors such as intrusions or distortions (see Roediger, 1996). In recent years, however, there has been a proliferation of accuracy-oriented memory research both with adults and children, focusing on various ways in which memory can go wrong beyond the simple forgetting of studied items (see Koriat et al., 2000). This recent focus seems to be more responsive to many real-life phenomena, such as false and confabulated memories, the effects of postevent information and leading questioning on the accuracy of eyewitness testimony, and so forth. In particular, with respect to children's testimony, there has been much concern about the extent to which what children report can be trusted (see Ceci & Bruck, 1998).

Even so, the distinction between accuracy-based and quantity-based memory assessment has not been articulated sufficiently (see Koriat & Goldsmith, 1996a, 1996b). In this study, we utilized Koriat and Goldsmith's (1994, 1996c) distinction between input-bound quantity and output-bound accuracy measures. In the context of courtroom testimony, for example, the former may be seen to reflect the extent to which the person can tell "the whole truth," whereas the latter uniquely reflects the ability to tell "nothing but the truth" (see also Cassel et al., 1996; Dietze & Thomson, 1993). This distinction is subtle and often missed.

Another subtle but important difference between input-bound and output-bound memory-quantity and memory-accuracy measures concerns the extent to which memory performance is under strategic control. Results obtained with adults indicate that participants cannot generally improve their memory-quantity performance when offered special incentives to retrieve additional items from their memory (e.g., Nilsson, 1987). In contrast, participants can enhance their memory accuracy substantially when given the option of free report and incentives for accurate reporting (Koriat & Goldsmith, 1994, 1996c). The enhancement of memory accuracy, however, depends critically on the ability to monitor the veracity of the information that comes to mind, and to regulate memory reporting in accordance with both the monitoring output and the operative incentives for accuracy.

Do children have the requisite abilities for the strategic regulation of memory accuracy under free-report conditions? If they do, what level of memory accuracy can they attain under various testing procedures and different levels of accuracy incentive? To examine these questions we applied an experimental paradigm that was used with adults (Koriat & Goldsmith, 1994). Because this paradigm does not include the collection of confidence judgments (cf. Koriat & Goldsmith, 1996c), it provides only indirect information regarding the operation of the children's monitoring and control processes. Nevertheless, as a first step toward investigating the development of these processes, the experiments yielded a number of interesting findings that will be summarized and discussed in turn.

The Strategic Regulation of Memory Accuracy

Can children regulate their memory reporting to produce a more accurate record of past events when given the option of free report? The answer is clearly yes. The results of Experiment 1 indicate that children in the age range studied could enhance the accuracy of the information that they reported from 68% for forced report to 81% for free report (under the moderate accuracy incentive). Thus, when children are given explicit incentives for accuracy, they can utilize effectively the option to withhold responding (i.e., say "don't know") as a means of increasing the accuracy of their testimony.

It is important to note that the effects of report option were obtained even in the test that was conducted a year later (Experiment 3). Although memory accuracy was overall quite low, it did increase from 38% in forced-report testing to 50% in free-report testing with a moderate accuracy incentive.

These results have some bearing on the concern that has been raised by several authors that children may be generally reluctant to respond spontaneously "I don't know" to memory questions, a reluctance that may result in false reports (e.g., Cassel et al., 1996; Dale et al., 1978; Geiselman & Padilla, 1988; Mulder & Vrij, 1996; Roebers & Schneider, 2000). Although there may be a developmental trend in this respect (see next), our results indicate that children can make good use of the "don't know" option, at least when explicitly encouraged to do so. Similarly, Mulder and Vrij (1996) found that when preschoolers were explicitly told that a "don't know" response is acceptable, they were able to screen out more incorrect responses to misleading questions (but see Moston, 1987). Perhaps the pay-off schedule used in our experiments is another effective way of communicating to children the importance of accurate reporting, and the legitimacy of a "don't know" response. It remains to be seen whether children can benefit as much from the option of free reporting under "spontaneous" conditions, when no explicit rewards for accuracy are offered.

If children can strategically regulate their memory accuracy, then not only should they be able to enhance their accuracy when given the option of free report, they should also be able to improve their accuracy further when given stronger incentives for accurate reporting. Thus, a key finding is that memory accuracy was higher for the high incentive condition of Experiment 2 (88%) than for the moderate incentive condition of Experiment 1 (81%), and that this difference was obtained for both age groups. The implication is that even young children are sensitive to different levels of accuracy incentive and can adjust their performance accordingly.

It is also remarkable that the effects of accuracy incentive were obtained after a year interval, with memory accuracy averaging 50% and 65%, respectively, for the moderate and high incentive conditions. Although these results call into question the extent to which a child's memory report can be trusted after a long delay even under a strong accuracy incentive, on the whole the effects of accuracy incentives on both immediate and delayed testing testify for the sensitivity of children's memory accuracy performance to specific situational demands during memory testing.

As will be discussed next, the option of free reporting can be used to enhance memory accuracy only if children are capable of monitoring the correctness of the answers that come to mind, and put that monitoring to use in screening their answers. Thus, the observed improvement in memory accuracy associated with the use of report option under different levels of accuracy incentive implies that children in the age range studied are capable of monitoring the correctness of the memory responses that come to mind and tend to control their memory reporting accordingly.

The Mechanisms Underlying the Strategic Regulation of Memory Accuracy

A clue to the mechanisms underlying the strategic regulation of memory accuracy is provided by the results indicating an accuracy-quantity trade-off: The option of free report in Experiment 1 yielded a 13-percentage-point increase in memory accuracy, but this was achieved at the cost of a 5-point decrease in memory quantity. Similarly, the high incentive condition of Experiment 2 yielded a 7-percentage-point increase in accuracy in comparison to the moderate incentive condition of Experiment 1. However, this increase too came at the cost of a 5-point decrease in memory-quantity performance.

This pattern of results accords well to what has been observed for adults. It suggests that the primary or sole vehicle for improving memory accuracy is the screening out of answers that are likely to be incorrect. Indeed, the within-individual analysis of the results of Experiment 1 revealed that the answers withheld by the children in the free-report phase had a much higher likelihood to be wrong (82%) than were the answers volunteered (24%). A similar screening process also appears to underlie the effects of accuracy incentive: More answers were withheld in the high incentive condition of Experiment 2 (10.0) than in the moderate incentive condition of Experiment 1 (6.6), and the accuracy of the withheld answers was higher for the high incentive (30%) than for the moderate incentive condition (24%). The net result was higher accuracy for the answers volunteered in the high incentive condition, but lower quantity performance, as more correct answers were withheld.

It is interesting to note that the pattern of results obtained here conforms to that found for adults in one more respect: Research with adults has suggested that whereas moderate gains in accuracy can be achieved with relatively small reductions in quantity performance, improving accuracy further becomes relatively costly (Koriat & Goldsmith, 1996c). The results with children yielded a similar pattern: The increase in accuracy due merely to the option of free report (under a moderate accuracy incentive) was achieved at a relatively modest cost in quantity performance, whereas the further increase in accuracy that resulted from the use of a higher incentive was accompanied by a relatively heavy quantity cost. This exact pattern was not found for the results obtained a year later (for reasons that are not clear), but these results too exhibited a quantity–accuracy trade-off. Memory accuracy increased by 11 percentage points when children were allowed free-report option, but this was achieved at a cost of 10 percentage points in quantity. Higher accuracy incentive resulted in an accuracy benefit of 15 percentage points, but at the cost of 11 percentage points in quantity.

Altogether, the results with children are consistent with the model proposed by Koriat and Goldsmith (1996c) regarding the metacognitive processes underlying the strategic regulation of memory accuracy. Thus, although no independent measures (e.g., confidence ratings) were collected that reflect children's monitoring and control processes, the results strongly suggest the following conclusions: First, children in the age range studied are able to monitor the accuracy of the information that comes to mind. Second, they control their memory responding by withholding answers that are more likely to be wrong. Finally, they tend to adjust their response criterion in accordance with the operative pay-off schedule, adopting a stricter criterion when a stronger incentive for accuracy is operative. Of course, the collection of confidence ratings could potentially provide more direct information regarding the mechanisms mediating the control of memory accuracy by children as well as possible age differences in these mechanisms. Overcoming the technical difficulties of doing so with grade-school and perhaps even younger children remains a challenge for future research.

The Effects of Test Format

On the whole, the pattern of results obtained with regard to the strategic control of memory accuracy was quite similar for recall and recognition. For both types of test format, memory-accuracy performance was better under free-report than under forced-report conditions and under high than under low accuracy incentive.

Over and above these effects, however, a comparison of the absolute levels of accuracy performance that can be achieved under recall and recognition testing is important for the issue of the credibility of memory testimony obtained with different questioning techniques. As discussed earlier, results with adults have challenged the established belief in eyewitness research that recognition testing is inherently less effective in eliciting accurate reports than recall testing (Koriat &

Goldsmith, 1994, 1996c).[6] However, because of the possibility that children are more suggestible than adults (Ceci & Bruck, 1993, 1999), it was important to examine whether such is indeed the case with children too. Our results clearly indicate that when children are allowed the option of free report, their performance is at least as accurate under recognition as under recall testing. In fact, recognition testing yielded significantly better memory accuracy than recall in Experiment 1 (84% and 78%, respectively) as well as in Experiment 3 a year later, (57% and 42%, respectively). It would appear, then, that free recognition is the more effective questioning procedure even with children, because it elicits more information than recall questioning, and, at the same time, yields an equivalent or even better accuracy rate. It would be instructive to examine whether this is also true with younger children still (see Bruck & Ceci, 1999).

Age Differences in Memory Accuracy

Let us now turn to the differences in memory performance observed between the younger and older children in our sample.

As far as the standard measures of memory are concerned, percent recall and percent recognition, our results replicated previous findings in documenting an increase with age in the quantity of information recalled and recognized. The results also replicated the general finding of stronger age differences in free recall than in forced recognition.

Importantly, however, in addition to the age differences in memory quantity, there were also age differences in the accuracy of the information reported under free-report conditions. In Experiment 1, the overall accuracy achieved was 77% for the younger children compared with 84% for the older children. This age effect was observed for both recall and recognition. A similar age effect was observed in the high incentive condition of Experiment 2: Although both groups of children enhanced their memory accuracy in response to the higher accuracy incentive, the older children still attained a higher level of accuracy (91%) than the younger children (84%). Interestingly, age differences were also observed in testing after one year, with the older children achieving 56% accuracy compared with 44% for the younger children. Here, too, the age differences were maintained even in the high incentive condition (74% vs 57% respectively).

Several previous studies have also documented a developmental increase in output-bound memory accuracy (e.g. Cassel et al., 1996; Roebers & Schneider, 2000). What is important about the present findings is that the age differences in memory accuracy were obtained under conditions that provided explicit and

[6]As in the present study, Koriat and Goldsmith's (1994, 1996c) previous comparisons of recall and recognition accuracy have generally involved cued recall rather than the type of open-ended free-narrative recall format commonly used in eyewitness research. Commission errors are relatively rare in the latter type of questioning. However, equivalent accuracy performance (and superior quantity performance for recognition) has also been found when comparing uncued free recall and free multiple-choice recognition in a standard word-list memory paradigm (Koriat & Goldsmith, 1994, Experiment 2).

equivalent incentives for accurate reporting. Indeed, even when the younger children were strongly motivated to be accurate, and explicitly encouraged to screen out incorrect answers, their memory reports could still be trusted less than those of the older children.

What, then, might account for the different levels of achieved accuracy? One clue comes from the observation that in all three experiments the younger children volunteered at least as many answers as did the older children under free-report conditions. This, despite the fact that they had fewer correct answers available in memory, as indexed by their forced-report performance. Thus, perhaps in line with previous suggestions, the younger children were more liberal in their control policy than were the older children, which, under Koriat and Goldsmith's (1996c) model, shall lead to lower accuracy performance. In addition, however, the age difference in accuracy may also be due in part to the younger children being less effective in monitoring the correctness of their answers, or relying less strongly on their monitoring in deciding which answers to volunteer and which to withhold (see Koriat & Goldsmith, 1996c).

Nevertheless, despite the observed age differences, it is important to note that the absolute levels of accuracy achieved by even the youngest group in immediate testing are relatively high. The implication is that on immediate testing, the vast majority of what a young child freely reports can be depended on to be correct. Although the present study examined grade-school children, a similar conclusion emerges from Bruck and Ceci's (1999) recent review, which emphasizes that in the absence of suggestive questioning techniques, even young preschoolers can provide highly accurate memory reports.

The picture changes radically, however, after a relatively long retention interval (Experiment 3), where the likelihood that a reported piece of information was correct was no better than 50-50. This finding is in line with previous studies showing that children's memory accuracy deteriorates more strongly over time than does that of adults (Flin et al., 1992; Poole & White, 1993). In interpreting this finding, one should perhaps also consider the changes that children of these ages undergo during the course of a year, changes that are likely to influence their perception and understanding of previously experienced events (see Liben, 1977). Erroneous memories that fit the child's current cognitive schemata might be particularly difficult for the child to identify and screen out. Of course, similar considerations may pertain to adults' memories (cf. Ross, 1989).

In conclusion, this study indicates that children can regulate their memory reporting to produce a more accurate record of past events when they are allowed to screen out wrong answers and when they are explicitly motivated to do so. Furthermore, they are also sensitive to specific levels of accuracy incentive, increasing the accuracy of their reports further when a higher premium is placed on memory accuracy. The accuracy benefits of report option and of a stronger accuracy incentive were obtained even after a year delay. These results have important practical implications in suggesting that children can be entrusted with the option of free report, and that they will utilize this option in a sensible manner to produce more dependable testimony than when they are deprived of control over their memory reporting (i.e., than when they are pressured into providing answers). The results also have important theoretical ramifications in highlighting the role of metacognitive processes underlying memory reporting in children as well as in adults. Overall, the results were quite similar to those obtained for adults and accord with the model proposed by Koriat and Goldsmith (1996c) in suggesting the operation of monitoring and control processes that mediate the strategic regulation of memory accuracy. The continued investigation of these processes could help elucidate some of the findings reported here, such as the developmental trend in memory accuracy and the dramatic decrease in children's report accuracy that occurs over time.

REFERENCES

Brainerd, C. J. (1985). Three-state models of memory development: A review of advances in statistical methodology. *Journal of Experimental Child Psychology,* **40,** 375–394.

Brainerd, C. J., & Ornstein, P. A. (1991). Children's memory for witnessed events: The developmental backdrop. In J. Doris (Ed.), *The suggestibility of children's recollections* (pp. 10–20). Washington, DC: American Psychological Association.

Brock, P., Fisher, R. P., & Culter, B. L. (1999). Examining the cognitive interview in a double-test paradigm. *Psychology, Crime and Law,* **5,** 29–45.

Brown, A. L. (1975). The development of memory: Knowing, knowing about knowing, and knowing how to know. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 10, pp. 103–152). New York: Academic Press.

Brown, A. L. (1979). Theories of memory and the problem of development: Activity, growth and knowledge. *Levels of Processing,* 225–258.

Brown, E. L., Deffenbacher, K. A., & Strugill, W. (1977). Memory for faces and the circumstances of encounter. *Journal of Applied Psychology,* **62,** 311–318.

Brown, J. (1976). *Recall and recognition*. London: Wiley.

Bruck, M., & Ceci, S. J. (1999). The suggestibility of children's memory. *Annual Review of Psychology*, **50,** 419–439.

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. (1998). The ontogeny and durability of true and false memories: A fuzzy trace account. *Journal of Experimental Child Psychology*, **71,** 165–169.

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.

Chi, M. T. H. (1978). Knowledge structures and memory development. In R. S. Siegler (Ed.), *Children's thinking: What develops?* (pp. 73–96). Hillsdale, NJ: Erlbaum.

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

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

Dietze, P. M., & Thomson, D. M. (1993). Mental reinstatement of context: A technique for interviewing child witnesses. *Applied Cognitive Psychology,* **7,** 97–108.

Dufresne, A., & Kobasigawa, A. (1989). Children's spontaneous allocation of study time: Differential and sufficient aspects. *Journal of Experimental Child Psychology*, **47,** 274–296.

Ebbesen, E. B., & Rienick, C. B. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *Journal of Applied Psychology,* **83,** 745–762.

Fisher, R. P., Geiselman, R. E., & Raymond, S. D. (1987). Critical analysis of police interview techniques. *Journal of Police Science and Administration,* **15,** 177–185.

Fisher, R. P. (1995). Interviewing victims and witnesses of crime. *Psychology, Public Policy, and Law,* **1,** 732–764.

Fivush, R., & Hamond, N. R. (1990). Autobiographical memory across the preschool years: Toward reconceptualizing childhood amnesia. In R. Fivush (Ed.), *Knowing and remembering in young children* (pp. 223–248). New York: Cambridge Univ. Press.

Flavell, J. H. (1970). Developmental studies of mediated memory. *Advances in Child Development and Behavior,* **5,** 182–213.

Flin, R., Boon, J., Knox, A., & Bull, R. (1992). The effect of a five-month delay on children's and adults' eyewitness memory. *British Journal of Psychology,* **83,** 323–336.

Geiselman, R. E., & Padilla, J. (1988). Cognitive interviewing with child witnesses. *Journal of Police Science and Administration,* **16,** 236–242.

Goldsmith, M., & Koriat, A. (1999). The strategic regulation of memory reporting: Mechanisms and performance consequences. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII—Cognitive regulation of performance: Interaction of theory and application* (pp. 373–400). Cambridge, MA: MIT Press.

Goodman, G. S. (1984). Children's testimony in historical perspective. *Journal of Social Issues,* **40,** 9–31.

Gordon, B. N., & Follmer, A. (1994). Developmental issues in judging the credibility of children's testimony. *Journal of Clinical Child Psychology,* **23,** 283–294.

Gorenstein, G. W., & Ellsworth, P. C. (1980). Effect of choosing an incorrect photograph on a later identification by an eyewitness. *Journal of Applied Psychology,* **65,** 616–622.

Hilgard, E. R., & Loftus, E. F. (1979). Effective interrogation of the eyewitness. *International Journal of Clinical and Experimental Hypnosis,* **27,** 342–357.

Johnson, M. K., & Foley, M. A. (1984). Differentiating fact from fantasy: The reliability of children's memory. *Journal of Social Issues,* **40,** 33–50.

Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General,* **123,** 297–315.

Koriat, A., & Goldsmith, M. (1996a). Memory as something that can be counted vs. memory as something that can be counted on. In D. J. Herrmann, C. McEvoy, C. Hertzog, P. Hertel, & M. K. Johnson (Eds.), *Basic and applied memory research: Practical applications* (Vol. 2, pp. 3–18). Hillsdale, NJ: Erlbaum.

Koriat, A., & Goldsmith, M. (1996b). Memory metaphors and the real-life/laboratory controversy: Correspondence versus storehouse conceptions of memory. *Behavioral and Brain Sciences,* **19,** 167–228.

Koriat, A., & Goldsmith, M. (1996c). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review,* **103**, 490–517.

Koriat, A., Goldsmith, M., & Pansky, A. (2000). Toward a psychology of memory accuracy. *Annual Review of Psychology,* **51,** 481–537.

Liben, L. (1977). Memory from a cognitive-developmental perspective: A theoretical and empirical review. In W. Overton & J. Gallagher (Eds.), *Knowledge and development, Vol. I: Advances in research and theory* (pp. 149–203). New York: Plenum.

Lindberg, M. A. (1980). Is knowledge base development a necessary and sufficient condition for memory development? *Journal of Experimental Child Psychology,* **30,** 401–410.

Lipton, J. P. (1977). On the psychology of eyewitness testimony. *Journal of Applied Psychology,* **62,** 90–95.

List, J. A. (1986). Age and schematic differences in the reliability of eyewitness testimony. *Developmental Psychology,* **22,** 50–57.

Lockl, K., & Schneider, W. (2001). *Developmental trends in children's feeling-of-knowing judgments.* Manuscript submitted for publication.

Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard Univ. Press.

Loftus, E. F. (1982). Memory and its distortions. In A. G. Kraut (Ed.), *The G. Stanley Hall lecture series* (Vol. 2, pp. 123–154). Washington, DC: American Psychological Association.

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

Loftus, E. F., & Hoffman, H. G. (1989). Misinformation and memory: The creation of new memories. *Journal of Experimental Psychology: General*, **118,** 100–104.

Marin, B. V., Holmes, D. L., Guth, M., & Kovac, P. (1979). The potential of children as eyewitnesses: A comparison of children and adults on eyewitness tasks. *Law and Human Behavior*, **3**, 295–306.

McCauley, M. R., & Fisher, R. P. (1995). Facilitating children's eyewitness recall with the revised cognitive interview. *Journal of Applied Psychology*, **80,** 510–516.

Moston, S. (1987). The suggestibility of children in interview studies. *First Language*, **7,** 67–78.

Mulder, M. R., & Vrij, A. (1996). Explaining conversation rules to children: An intervention study to facilitate children's accurate responses. *Child Abuse and Neglect*, **20,** 623–631.

Neisser, U. (1981). John Dean's memory: A case study. *Cognition*, **9,** 1–22.

Neisser, U. (1988). Time present and time past. In M. M. Gruneberg, E. P. Morris, & N. R. Sykes (Eds.), *Practical aspects of memory: Current research and issues* (Vol. 2, pp. 545–560). Chichester, UK: Wiley.

Nilsson, L. G. (1987). Motivated memory: Dissociation between performance data and subjective reports. *Psychological Research*, **49,** 183–188.

Ornstein, P. A. (1990). Knowledge and strategies: A discussion. In W. Schneider & F. E. Weinert (Eds.), *Interactions among aptitudes, strategies, and knowledge in cognitive performance* (pp. 147–156). New York: Springer-Verlag.

Ornstein, P. A. (1991). Commentary: Putting interviewing in context. In J. Doris (Ed.), *The suggestibility of children's recollection* (pp. 147–152). Washington, DC: American Psychological Association.

Ornstein, P. A., Gordon, B. N., & Baker-Ward, L. (1992). Children's memory for salient events: Implications for testimony. In M. Howe, C. Brainerd, & V. Reyna (Eds.), *Developmental long-term retention* (pp. 31–50). Hillsdale, NJ: Erlbaum.

Ornstein, P. A., Gordon, B. N., & Larus, D. M. (1992). Children's memory for a personally experienced event: Implications for testimony. *Applied Cognitive Psychology*, **6,** 49–60.

Ornstein, P. A., & Naus, M. J. (1985). Effects of the knowledge base on children's memory strategies. In H. W. Reese (Ed.), *Advances in child development and behavior* (Vol. 19, pp. 113–148). Orlando, FL: Academic Press.

Parker, J. F., Haverfield, E., & Baker-Thomas, S. (1986). Eyewitness testimony of children. *Journal of Applied Social Psychology*, **16**(4), 287–302.

Peterson, C. (1999). Children's memory for medical emergencies: 2 years later. *Developmental Psychology*, **35,** 1493–1506.

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

Poole, D. A., & White, L. T. (1993). Two years later: Effect of question repetition and retention interval on the eyewitness testimony of children and adults. *Developmental Psychology*, **29**(5), 844–853.

Pressley, M., Levin, J. R., Ghatala, E. S., & Ahmad, M. (1987). Test monitoring in young grade school children. *Journal of Experimental Child Psychology*, **43,** 96–111.

Roebers, C. M., & Schneider, W. (2000). The impact of misleading questions on eyewitness memory in children and adults. *Applied Cognitive Psychology*, **14,** 509–526.

Roediger, H. L., III. (1996). Memory illusions. *Journal of Memory and Language*, **35,** 76–100.

Ross, M. (1989). Relation of implicit theories to the construction of personal histories. *Psychological Review*, **96,** 341–357.

Salmon, K., & Pipe, M. E. (2000). Recalling an event one year later: The impact of props, drawing and a prior interview. *Applied Cognitive Psychology*, **14,** 99–120.

Schneider, W. (1985). Developmental trends in the metamemory-memory behavior relationship: An integrative review. In D. Forrest- Pressley, G. MacKinnon, & T. G. Waller (Eds.), *Cognition, metacognition, and human performance* (Vol. 1, pp. 57–109). New York: Academic Press.

Schneider, W. (1999). The development of metamemory in children. In D. Gopher & A. Koriat (Eds.), *Attention and performance XVII. Cognitive regulation of performance: Interaction of theory and application* (pp. 487–514). Cambridge, MA: MIT Press.

Schneider, W., & Bjorklund, D. F. (1992). Expertise, aptitude, and strategic remembering. *Child Development*, **63,** 461–473.

Schneider, W. & Bjorklund, D. F. (1998). Memory. In W. Damon, D. Kuhn, & R. S. Siegler (Eds.), *Handbook of child psychology: Cognition, perception and language* (Vol. 2, pp. 467–521). New York: Wiley.

Schneider, W., & Pressley, M. (1997). *Memory development between 2 and 20* (2nd ed.). Hillsdale, NJ: Erlbaum.

Schneider, W., Visé, M., Lockl, K. & Nelson, T. O. (2000). *Developmental trends in children's memory monitoring: Evidence from a judgment-of-learning task*. *Cognitive Development*, **15,** 115–134.

Schneider, W., & Weinert, F. E. (1989). Universal trends and individual differences in memory development. In A. de Ribaupierre (Ed.), *Transition mechanisms in child development: The longitudinal perspective* (pp. 68–106). Cambridge, UK: Cambridge Univ. Press.

Shepard, R. N. (1967). Recognition memory for words, sentences, and pictures. *Journal of Verbal Learning and Verbal Behavior*, **6,** 156–163.

Siegler, R. S. (1983). Five generalizations about cognitive development. *American Psychologist*, **38,** 263–277.

Stern, W. (1904). Wirklichkeitsversuche, *Beitrage zur Psychologie der Aussage*, **2,** 1–31. Translation by U. Neisser. In U. Neisser (Ed.). (1982), *Memory observed: Remembering in natural context*.

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, **80,** 359–380.

Weingardt, K. R., Leonesio, R. J., & Loftus, E. F. (1994). Viewing eyewitness research from a metacognitive perspective. In J. Metcalfe (Ed.), *Metacognition: Knowing about knowing* (pp. 157–184). Cambridge, MA: MIT Press.

Received May 22, 2000; revised September 1, 2000; published online June 20, 2001