

# The accuracy and suggestibility of children's memory for neutral and criminal eyewitness events

Joyce E. Ochsner, Maria S. Zaragoza\* and Karen J. Mitchell

*Department of Psychology, Kent State University, USA*

**Purpose.** A critical issue for psychologists interfacing with the legal system is knowing the extent to which laboratory studies of children's testimony generalize to eyewitness situations involving criminal events. The present study was designed to assess whether young children's eyewitness performance following a criminal event would differ from their performance following a neutral event, an issue that has received little empirical attention to date.

**Methods.** The accuracy and suggestibility of 30 first grade children who witnessed a staged event that culminated in a purse theft were compared to that of 30 first graders who witnessed the same event, but without the theft.

**Results.** Children who viewed the theft were found to be more accurate on various measures of recall and recognition than children who viewed the neutral event. On the other hand, children who viewed the theft were not always less suggestible than children who witnessed the neutral event.

**Conclusion.** Taken together, the results suggest that studies of neutral eyewitness situations are likely to underestimate children's memory performance in criminal eyewitness situations.

The capability of young children to provide accurate eyewitness testimony has been the subject of increasing and much warranted attention by the general public, as well as by social scientists and members of the legal profession. Struggles to prosecute crimes against children in which the child involved may be the only witness, such as with physical or sexual abuse, have helped bring to the fore issues pertaining to the accuracy and reliability of their eyewitness reports. In addition to testifying in abuse cases, children are increasingly being called upon to provide eyewitness testimony for other types of crimes (e.g. Ceci & Bruck, 1993a, b, 1995; Goodman, 1984; see also, Bottoms & Goodman, 1996 for a multinational collection of papers on this topic).

One critical issue for psychologists interfacing with the legal system is knowing the extent to which laboratory studies of children's testimony generalize to

\*Requests for reprints should be addressed to Maria S. Zaragoza, Department of Psychology, Kent State University, Kent, OH 44242-0001, USA.

eyewitness situations involving criminal events (Clifford, 1978, 1979; Davies & Flin, 1988; Tobey & Goodman, 1992). Although empirical studies of children's eyewitness memory/testimony have been rapidly accumulating over the past decade (see, for example, Bottoms & Goodman, 1996; Ceci, Ross & Toglia, 1989; Ceci, Toglia & Ross, 1987; Dent & Flin, 1992; Doris, 1991; Goodman & Bottoms, 1993; Zaragoza, Graham, Hall, Hirschman & Ben-Porath, 1995, for collections of recent papers on the subject; Ceci & Bruck, 1993b, for a review), we are unaware of any published empirical paper in which children's reports of criminal eyewitness events were directly compared with their reports of neutral eyewitness events. Hence, it is not yet known to what extent studies of neutral events accurately estimate children's performance in eyewitness situations involving criminal events.

There are several reasons, however, that one might expect children's eyewitness performance in criminal situations to differ from their performance in neutral situations. For example, criminal events about which children are asked to give an eyewitness report are undoubtedly going to be imbued with considerable importance and personal relevance (i.e. 'impact'; e.g. Yuille & Tollestrup, 1992). The perceived seriousness and importance of a criminal event could influence the child witness' memory performance at several points. First, if the child is aware of the event's significance during its occurrence, the child may pay more attention or process the event more elaborately (e.g. see Rudy & Goodman, 1991; Tobey & Goodman, 1992, for evidence consistent with this idea). This could occur both during the occurrence of the event and for some time afterwards, as the child retells and rehearses the circumstances of the event. Whether this intention to remember actually makes a difference in the amount recalled is unclear, however, because young children have only a limited repertoire of mnemonic strategies to draw upon for intentional memory tasks (Brown, Bransford, Ferrara & Campione, 1983; Naus & Ornstein, 1983). Second, it could affect the child's motivation to report the event accurately. In the very process of a child's being asked to recount his or her experience to an authority, the importance of the event will be heightened. Furthermore, the child is likely to know, or be made aware, that his or her testimony could have a significant impact on someone's life and perhaps the child's own life as well. If the child is more motivated to recall a criminal event, he or she may exert more effort in searching his or her memory. Whether this would actually result in more (and/or more accurate) details being reported is unknown. Albeit children are limited in their ability to make exhaustive memory searches (Chi, 1976; Naus & Halasz, 1979), recent evidence suggests that children's eyewitness reports can be improved (i.e. increased resistance to suggestion, increased accuracy) under some questioning circumstances (e.g. Memon, Holley, Wark, Bull & Kohnken, 1996; Saywitz & Moan-Hardie, 1994).

There is also some indication in the adult literature that studies of neutral events might underestimate the accuracy of participants' memory for criminal events (see Yuille & Tollestrup, 1992). For example, archival and field study data concerning memory for 'real' crime situations collected by Yuille and his colleagues suggests that events of high 'impact' (i.e. criminal events of consequence or direct personal relevance) tend to be well remembered, even over long delays (e.g. Yuille, 1986; Yuille & Cutshall, 1986; see Yuille & Tollestrup, 1992 for a review). Laboratory

research tends to support these findings, and to extend it by showing that memory for criminal events is often better than memory for a neutral event. For example, Hosch & Cooper (1982) found that victims and bystanders of a staged theft were significantly more accurate at identifying the suspect's photo than were control participants who witnessed a neutral event. It has also been demonstrated (Leippe, Wells & Ostrom, 1978) that the perceived seriousness of the event will affect memory performance, in that participants' photo identification of a suspect was more accurate the higher the perceived value of the stolen object.

Although no studies have directly compared children's memory performance in neutral and criminal situations, additional evidence can be gleaned from the results of studies that assessed children's accuracy and suggestibility for an eyewitness situation that involved a criminal event, namely, a staged theft (e.g. Brigham, VanVerst & Bothwell, 1986; King & Yuille, 1987, Expt 2). One interesting result obtained in both these studies was a lack of suggestibility to leading questions among children of various age groups. In contrast, studies of live neutral event situations (e.g. Goodman & Reed, 1986; King & Yuille, 1987, Expt 1), have found evidence that children are susceptible to the influence of leading questions. One possible explanation of this discrepancy is that the increased seriousness of the criminal event coupled with the importance of accurate recall may have made the children less prone to accept the misleading information. Alternatively, the novelty or importance of the criminal act itself might have led participants to pay more attention to details they would otherwise have overlooked (see Peterson, Moreno & Harbeck-Weber, 1993 for evidence consistent with this possibility). To the extent that participants are able to better resist suggestion when the event is well remembered (cf. Johnson & Foley, 1984; Loftus & Davies, 1983), factors that increase accuracy can be expected to reduce suggestibility. It is important to note, however, that because neither of these studies compared eyewitness performance in the criminal situation with performance in the neutral situation, it is difficult to know to what extent the nature of the eyewitness event was responsible for the children's resistance to suggestibility.

The present study was designed to assess whether young children's eyewitness performance following a criminal event would differ from their performance following a neutral event. To that end, we compared the accuracy and suggestibility of children who witnessed a commonplace live event with that of children who witnessed a staged purse theft. In the study, small groups of children observed a stranger enter their classroom. The stranger claimed he was looking for the principal. In the experimental or 'theft' condition the stranger stole the purse as he left the room, whereas in the control or 'no-theft' condition the stranger did not steal the purse. Participants in both groups had identical exposure to, and experience of, the stranger until the final seconds of the event. Thus, any differences in accuracy or suggestibility can be attributed to the cognitive processing that resulted from viewing the theft.

Following the eyewitness event, the children were asked to describe the stranger and his actions. Both types of questions were included because there is some evidence that children, like adults, remember actions better than appearance or clothing (e.g. Davies, Tarrant & Flin, 1989; Goodman, Aman & Hirschman, 1987),

thus leaving open the possibility that memory for these details might be differentially affected when comparing the theft and no-theft groups.

In addition to measuring accuracy in recall, we also assessed whether the children in the theft and no-theft groups would differ in their susceptibility to being misled. Suggestibility was measured using the paradigm developed by E. Loftus and her colleagues (e.g. Loftus, Miller & Burns, 1978). Misleading suggestions were embedded in objective questions about the event and participants were later tested on their memory for the details about which they had been misled.

## Method

### *Participants*

Participants were 60 first grade students from two parochial schools in two small midwestern US cities. All participants were Caucasian, aged between 6 years, 6 months and 7 years, 7 months, and lived in an area with predominantly blue-collar, lower middle-class families. Parents were first contacted by telephone so that they could be fully informed about the nature of the experiment. Detailed consent forms were then sent to interested parents.

Participants were randomly assigned to the theft and no-theft groups ( $N_s = 30$ ). Because of an over-representation of males in the classroom, there were 35 males (18 in the no-theft group and 17 in the theft group) and 25 females (12 in the no-theft group and 13 in the theft group).

### *Materials and procedure*

Participants were taken in groups of three to an empty classroom, where each participant was paired with a female experimenter and rapport was established. The entire procedure was audiotaped with three minicassette recorders placed in full view of the participants. Participants were told that the experimenters had developed a new kind of puzzle and wanted to know how well first graders could learn it. The 'puzzle' was a pegboard task patterned after the Animal House subtest of the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) (Wechsler, 1967). This task involves matching coloured pegs to animal pictures following a model that is provided, for example, putting a blue peg under each picture of a dog.

The experimental procedure involved six phases. Participants were (1) given a pegboard trial, (2) exposed to the eyewitness event, (3) asked to recall the event, (4) asked objective questions about the event, some of which presupposed misleading information, (5) given a second pegboard trial, and (6) given a final test of their memory for the event. Each phase is elaborated below.

*Phase 1: First pegboard trial.* A practice version of the pegboard task was administered to the children to familiarize them with the task. Following this, a different pegboard was administered which was later scored according to the standardized procedure of the WPPSI. Each participants' response time to complete the pegboard task was recorded. The task took approximately one minute to complete. Three versions of the pegboard task were used in the experiment and these were counterbalanced to control for order effects. The children were then asked to draw a picture for a few minutes, and were told that soon after they would do another pegboard. Two of the experimenters then left the room allegedly to procure supplies. The third experimenter went to an area of the room partitioned off so that she could not see the door, on the pretext of having to score the pegboards.

*Phase 2: Eyewitness event.* The live eyewitness event, described in detail below, involved a 40-year-old Caucasian male who entered the classroom allegedly looking for the principal. The event was the same for participants in the theft and no-theft groups with the exception that in the theft group the man took the experimenter's white purse from a chair near the door as he left. For both groups the entire event lasted 20 seconds.

The event contained six critical details about which the participants would later be tested, and across the experiment two versions of the critical detail were used. The critical details and the two versions of each were: (1) a magazine or newspaper that the confederate was carrying, (2) a pencil or pen that he dropped, (3) a jacket or sweater that he was wearing, (4) blue or brown pants that he was wearing, (5) he identified himself as either John Logan or John Foster, and (6) he sported a fake moustache or a beard (with no moustache). (The confederate had a real beard which he shaved for the moustache condition.) Two scenarios were enacted, differing only in the version of the critical item presented. Twelve participants in each of the theft and no-theft groups saw Scenario 1 and 18 participants in each group saw Scenario 2.

The witnessed event began when the male confederate entered the room and said in a friendly voice, 'Hi kids. My name is John Logan (or Foster) and I am looking for the principal'. Glancing around, he added, 'I see she's not in here. She was supposed to be down here'. He then dropped his pencil (or pen), set down his magazine (or newspaper) on a chair beside him, retrieved his pencil (or pen), and put it into his pocket. As he picked up the magazine (or newspaper) he commented, 'I'm sure clumsy today'. Fifteen seconds after he had entered the room, the experimenter signalled by one or two knocks whether the purse was to be taken. This kept the length of exposure to a constant 20 seconds across trials and, because the confederate was blind to the condition, it minimized unintentional bias. At this point the scenarios changed depending on whether it was the theft or no-theft condition.

In the theft condition, the confederate commented that he was leaving, made sure that he had the children's attention and then took the purse as he walked out the door. The experimenter immediately reappeared and, if the children did not point out the theft, said 'What did the man want?' The children invariably mentioned the theft, although their reactions varied from hesitation to nervous giggles to outright excitement. The experimenter initially disbelieved the children in as natural a way as possible, and allowed the children to convince her. At that point, she went quickly to the door, indicated that no one was around, and asked which way the man went. She then asked a fictitious person, out of the children's sight, to run to the principal's office to report it, and asked the other experimenters to help her ask the children what the man looked like 'because she did not see him'.

In the no-theft condition, once the confederate was signalled as to the condition, he commented that he was leaving, glanced around the room and left. The experimenter emerged and asked the children, 'What did the man want?' When they replied that he was looking for the principal, the experimenter went to the door and asked a fictitious person to help him find his way. The other experimenters returned and each explained to their child that, prior to redoing the pegboard, they wanted to know how well the children could remember the man who had been in the classroom, and so they were going to ask them a few questions.

*Phase 3: Recall.* The experimenters asked the children individually what happened, requesting that they talk quietly to minimize the possibility of others overhearing their responses. After participants finished their spontaneous reports, four questions were asked: (1) 'What did the man do?', (2) 'Did he do anything else?', (3) 'What did the man look like?', (4) 'Can you remember anything else?'. This encouraged their recall without cuing for specific details. The participants' responses were both recorded on audiotape and written down verbatim by the experimenter.

*Phase 4: Misleading post-event questioning.* Following recall, each child was asked 14 'yes/no' questions about the event, some of which contained misinformation. Across the experiment there were six misleading suggestions, although each participant was exposed to only three of these and the remaining three served as never-presented control items on the test.

Eight of the 14 questions were fillers and were the same for all participants. For each participant, three of the remaining questions were misleading in that they included a false presupposition about one of the critical items. The other three served as control questions by providing only neutral information about the critical item. For example, a misleading question (assuming that the confederate had identified himself as John Logan) was 'When John Foster came in, was he looking for the principal?'. Across participants, an equal number of children received misleading and neutral information about each of the critical items. (See Appendix A for the list of questions.)

*Phase 5: Second pegboard trial.* Another version of the pegboard was administered and the response time recorded. In addition to serving as a filler task, it was thought that comparing participants' performance on the first and second pegboard task might reveal greater distractibility among children who had witnessed the theft. Distractibility would be shown by slower performance among children in the theft condition when compared to their performance on the first pegboard trial.

*Phase 6: Final test questions.* A final set of six questions, designed to assess the children's suggestibility on the critical items, was administered. Participants were instructed to answer the questions on the basis of what they remembered seeing in the original event. Each test question was a two-alternative forced choice between the critical item originally seen and the item that had been presented as misleading information to participants who had been misled about it. For example, the participant was asked, 'Was the man's name John Foster or John Logan?' The same test questions were used for all participants, thus whether a particular question was a misled or control question depended on whether or not the participant had received misleading information about the item queried. Across participants the order in which the alternatives were presented was counterbalanced.

Finally, the children were debriefed. During debriefing, the theft group was told for the first time that this was a staged event. They were praised for any appropriate action they took and told what they should have done had it been a real crime situation. The no-theft group was thanked for their cooperation. All the children were asked not to tell their classmates until after school that day, but were told that their teachers and their parents knew about the experiment. All participants from a given classroom were tested in a single morning or afternoon time block to minimize chances of the participants' learning of the deception from their peers. The teachers monitored the classroom discussions during the testing and were not aware of any discussion about the experiment. None of the children reported having learned of it ahead of time. In an attempt to make participation in the experiment an educational experience, the following day a police officer accompanied the experimenter to meet with the entire class to further debrief the children and to discuss the appropriate actions to take after witnessing a real crime.

## Results

Preliminary analyses indicated that there were no differences in the performance of participants who viewed Scenario 1 and Scenario 2 on any of the dependent measures ( $ps > .05$ ), indicating that the particular version of the critical items viewed did not affect performance. Therefore, the data were collapsed across scenarios for all subsequent analyses.

Preliminary analyses also showed that exposure to the theft significantly impaired the children's performance on the pegboard task<sup>1</sup>. Although participants in the no-theft group were an average of 3.5 s faster on the post-event pegboard trial than on the pre-event trial ( $M_s = 78.5$  s and 75.0 s for the pre- and post-event trials, respectively), participants in the theft group were an average of 4.5 s slower on the post-event trial than on the pre-event trial ( $M_s = 74.8$  s and 79.3 s for the pre- and post-event trials, respectively). Analysis of variance (ANOVA) verified that the interaction between groups (theft and no-theft) and trials (pre- and post-event) was significant ( $F(1,56) = 4.73, p < .05$ ). *Post hoc* tests (Fisher's LSD) revealed that the interaction was due primarily to the fact that there was a significant difference between the pre- and post-event trials in the theft group ( $p < .05$ ), but not in the no-theft group ( $p > .05$ ). A related finding was that a significantly larger number of participants in the theft group (18 participants) than in the no-theft group

<sup>1</sup> Because only two errors were made, one in each group, accuracy was not taken into account in the final score as specified in the WPPSI scoring procedures. Thus, the final scores represent speed only.

**Table 1.** Mean number of correct and incorrect person and action descriptors recalled by participants in the theft and no-theft groups

	Correct		Incorrect	
	No-theft	Theft	No-theft	Theft
Person descriptors				
Person	0.94	1.20	0.30	0.27
Attire	0.60	1.00	0.50	0.40
Belongings	0.63	0.50	0.37	0.13
Total	2.17	2.70	1.17	0.80
Action descriptors	2.40	3.00	0.00	0.00
Total	4.57	5.70	1.17	0.80

(seven participants) took longer on the post-event trials than on the pre-event trials ( $\chi^2 = 6.86$ ,  $p < .01$ ). Thus, it appears that observing a criminal event slowed the children's performance on a subsequent task that was unrelated to the eyewitness event. Although the precise cause of this disruptive effect cannot be determined, it suggests that the theft children were disproportionately distracted which is certainly consistent with the idea that the children believed the theft was real.

### Accuracy

Separate ANOVA's were performed on two measures of accuracy: number of correct items recalled and number of incorrect items recalled. In each analysis, group (theft or no-theft) and sex (male or female) were included as between-subjects factors, and type of descriptor (person or action) was included as a within-subjects factor.

In general, the results showed that participants in the theft group remembered more information and were more accurate in their recall than participants in the no-theft group. Table 1 presents the mean number of correct and incorrect statements recalled by the theft and no-theft groups, divided according to whether they were descriptions of the confederate (including statements about his appearance, his clothing and his belongings) or descriptions of his actions. The theft group produced more correct statements in recall ( $M = 5.70$ ) than the no-theft group ( $M = 4.57$ ) as evidenced by a significant main effect for group ( $F(1,58) = 4.61$ ,  $p < .05$ ). Furthermore, the theft group appeared to produce fewer incorrect statements about the confederate than did the no-theft group ( $M_s = .80$  and  $1.17$ , respectively), although this trend did not reach statistical significance ( $F(1,56) = 3.15$ ,  $p = .08$ ). Clearly, the higher level of recall among the theft participants was not accompanied by a higher level of inaccuracies. There were no sex differences.

In general, there was a tendency for participants in both groups to recall correctly a larger number of action descriptors ( $M = 2.70$ ) than person descriptors

**Table 2.** Mean proportion of correct responses to misled and control items in the theft and no-theft groups

	Misled	Control	Difference
Theft group	.64(.28)	.78(.21)	.14
No-theft group	.49(.27)	.73(.21)	.24

Note. Standard deviations are in parentheses.

( $M = 2.43$ ), although this trend was not significant ( $p > .05$ ). It is worth noting, however, that participants' statements about the confederate's actions were more accurate than their statements about his appearance and attire. Whereas none of the participants recalled any of the confederate's actions incorrectly, the mean number of incorrectly recalled details about the confederate's appearance or attire was approximately one per participant (see Table 1).

### *Suggestibility*

The final test of the critical items consisted of a forced choice between the original and misleading items, the procedure typically used in laboratory studies of susceptibility to misleading suggestions (cf. Loftus *et al.*, 1978). One unexpected, yet interesting, finding was that when theft participants could not remember the answer to a question they were much less willing to choose between the two alternatives than the no-theft participants were. A number of participants from the theft group firmly stated that they could not remember the answer when asked the forced-choice questions, and when encouraged to guess, refused to do so. This resulted in 10 unanswered questions in the theft group (four misled items and six neutral items) and only one in the no-theft group (one neutral item). These were scored .5 in the analysis rather than 0 to account for the probability of a correct response if the participant had guessed. However, it should be noted that the results were unchanged when the missing values were scored as zero. Also of interest is the finding that seven of the 30 theft participants spontaneously corrected the misleading information while it was being presented while only one of the no-theft participants did. A Fisher Exact probability test confirmed that this difference is significant ( $p = .05$ ). The above findings suggest that participants who had witnessed the theft took their answers to these questions more seriously and were much less conforming than participants who had witnessed the neutral event.

Although the theft participants were, in a very real sense, less suggestible than the no-theft participants in their approach to answering the test questions, examination of the theft and no-theft participants responses on the final test reveals that the issue of group differences in suggestibility is somewhat more complex. Table 2 presents, for the theft and no-theft groups, the mean proportion of correct responses to the misled and control items. The data were submitted to an ANOVA with group and sex as between-subjects factors and condition (misled or control) as



a within-subjects factor. Consistent with previous research on children's eyewitness memory using this testing procedure (e.g. Ceci, Ross & Toglia, 1987), participants in both groups performed more poorly on the misled items than on the control items ( $F(1,56) = 20.33, p < .001$ ). Thus, it appears that both groups were influenced by misleading suggestions. Nevertheless, participants in the theft group performed more accurately overall ( $M = .71$ ) than those in the no-theft group ( $M = .61$ ), as evidenced by a significant main effect of group ( $F(1,56) = 5.37, p < .05$ ). Although the no-theft group would appear to be more suggestible, as evidenced by the greater misled-control performance difference in the no-theft condition ( $M = .24$ ) as compared to the theft condition ( $M = .14$ ), the group by condition interaction was not statistically significant ( $F(1,56) = 1.62, p > .05$ ). There were no sex differences.

Finally, we were also interested in assessing the relationship between participants' accuracy and suggestibility. Across both groups, the number of accurate responses given on recall was significantly correlated with accuracy on the misled items ( $r(58) = .31, p < .05$ ). Although participants' accurate recall included more than just the critical items they were misled about, each of the six critical items was in fact mentioned spontaneously in recall (though the number of participants who accurately reported a critical item prior to being misled varied from two to 14 participants per item). Perhaps reporting the information aloud rendered the misleading information ineffectual when it was presented later (Loftus, 1977). In fact, in none of the 48 instances in which a critical item was spontaneously mentioned on recall and followed by misleading information did the children alter their final response to agree with the misinformation. More interesting, perhaps, is the finding that of the 48 instances in which critical items were spontaneously reported, 30 were reported by children in the theft group as compared to 18 reported by children in no-theft group. Hence, it would seem that the theft group's greater resistance to suggestion is related to their greater accuracy in free recall.

## Discussion

In general, this study has shown that laboratory tasks employing neutral eyewitness events are likely to underestimate children's performance in eyewitness situations involving criminal events. Children who witnessed a theft produced more accurate information than participants who witnessed a neutral event without producing a greater number of inaccuracies; in fact, there was a tendency for participants in the theft group to produce fewer inaccuracies in recall than participants in the no-theft group. Greater accuracy on the part of children who witnessed the theft was also observed on the final forced-choice recognition test. These differences in productivity and accuracy are particularly noteworthy in light of the fact that the neutral event used as a control comparison in this study was more involving and realistic than many of the pictures, slides and film clips that have been employed in some eyewitness studies to date. Much greater differences in eyewitness performance might well exist between staged thefts and events that are not witnessed live.

Having established that participants who witnessed the theft performed more accurately than participants who witnessed the commonplace event, one important

remaining question is why these improvements in performance occurred. In attempting to answer this question it is important to keep in mind that the purse theft occurred during the final seconds of the eyewitness event, as the confederate was leaving the classroom. Viewing the theft could not have led participants to pay closer attention to the portion of the eyewitness event that occurred before the theft—the majority of the event. Recall that the bulk of the details they were tested about occurred prior to the purse theft. Consequently, it seems likely that an explanation of the theft participants' superior performance must involve processes which occurred after the event had been recorded in memory. One possible explanation is that participants in the theft group rehearsed the event more extensively, and that this additional rehearsal led to superior recall. Another possibility is that participants who had witnessed the theft were more motivated to recall the event accurately and therefore exerted more effort in retrieving details from memory and evaluating the accuracy of the details they retrieved. This motivation to be accurate may have also rendered participants in the theft condition more critical of the misinformation, thus accounting for the observed resisting of the misleading suggestions initially. An important question for future research is assessing the extent to which each of these factors might play a role in improving the performance of children who have witnessed a criminal event.

Another difference between criminal and neutral eyewitness situations is the higher level of stress presumably associated with criminal eyewitness situations, and it is therefore possible that arousal contributed to the present findings. Unfortunately, however, the present results cannot address the role of arousal/stress in children's eyewitness performance because arousal was not measured in this study. Moreover, the relationship between arousal/stress and memory is quite complex, with evidence of both positive and negative effects (and even null effects) on the memory performance of both children and adults (see Christianson, 1992a; Deffenbacher, 1983; Goodman & Hahn, 1987, for reviews; and see Christianson, 1992b for a recent collection of papers on this topic). Indeed, the more recent consensus appears to be that the stress/memory relationship is heavily dependent on other mediating variables (e.g. prior experience with the stressor, type of information queried, age and other individual differences; e.g. Goodman, Quas, Batterman-Faunce, Riddlesberger & Kuhn, 1994; Peterson & Bell, 1996; Vandermaas, Hess & Baker-Ward, 1993). Exactly how heightened arousal/stress affects the accuracy of children's testimony is still an open question.

Although we have argued on the basis of the present findings that studies employing neutral eyewitness events might underestimate children's performance in criminal eyewitness situations, our experimental design does not stimulate all aspects of an actual court case. In a court case, the interview process is quite different. There is generally a greater delay prior to questioning; in fact, months may elapse. Typically, there are several different interviewers, each potentially contributing a different source of bias. Furthermore, the interviewers are likely to have greater authority and the questioning process may not be as benign as that utilized in this study, increasing the demand characteristics of the situation. It is also likely, in an actual case, that stress will be intensified—both at encoding and retrieval. A child's stress may increase dramatically if he or she were to witness a criminal act

alone, rather than in a group. But, most importantly, children are often the victim rather than a bystander, so the degree of trauma associated with the event may be substantial. Understandably, their cognitive processing may be altered by extreme stress (cf. Pynoos & Eth, 1984; Terr, 1981, 1983).

Nevertheless, the results of the present study suggest that it cannot be taken for granted that children's eyewitness performance will be the same across various types of eyewitness situations. Children's behaviour following a criminal event differed both qualitatively and quantitatively from performance following a neutral event. The improvements in performance evidenced by children who witnessed the theft are important not so much because of their magnitude (the performance differences were, in fact, relatively small in absolute terms) but because of the consistency with which they were observed. Also notable is the fact that viewing the theft did not lead to improved performance on every task; children in the theft group showed impaired performance on the pegboard task. Although further research is needed to pin down more precisely the factors that influence children's performance in criminal eyewitness situations, it is clear that novel and important aspects of children's eyewitness behaviour emerged when their performance was assessed under the conditions of a criminal eyewitness event.

### Acknowledgements

The research reported here is based on a doctoral dissertation conducted by the first author in partial fulfilment of the requirements for the Doctor of Philosophy degree, Department of Psychology, Kent State University. Preparation of this article was supported in part by NIMH grant MH 43581 to Maria S. Zaragoza.

We are grateful to the principals and staff of St Gabriels School, Mentor, OH, and St Anselm's School, Chesterland, OH, for their assistance and cooperation throughout the project. We thank Betty May and Karen Davis for help with testing participants and we especially thank Charles Weikert for serving as the confederate in the experiment.

### References

- Bottoms, B. & Goodman, G. (Eds) (1996). *International perspectives on child witnesses: Psychological research and law*. Thousand Oaks, CA: Sage.
- Brigham, J. C., VanVerst, M. & Bothwell, R. K. (1986). Accuracy of children's eyewitness identification in a field setting. *Basic and Applied Social Psychology*, 7, 295–306.
- Brown, A. L., Bransford, J. D., Ferrara, R. A. & Campione, J. C. (1983). Learning, remembering, and understanding. In J. H. Flavell & E. M. Markman (Eds), *Handbook of child psychology*, vol. 3, pp. 77–166. New York: Wiley.
- Ceci, S. J. & Bruck, M. (1993a). Child witness: Translating research into policy. *Social Policy Report: Society for Research in Child Development*, vol. VII, Number 3. Ann Arbor, MI: SRCDC.
- Ceci, S. J. & Bruck, M. (1993b). The suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, 113, 403–439.
- Ceci, S. J. & Bruck, M. (1995). *Jeopardy in the courtroom: A scientific analysis of children's testimony*. Washington, DC: American Psychological Association.
- Ceci, S. J., Ross, D. F. & Toglia, M. P. (1987). Age differences in suggestibility: Narrowing the uncertainties. In S. J. Ceci, M. P. Toglia & D. F. Ross (Eds), *Children's eyewitness memory*, pp. 79–91. New York: Springer-Verlag.
- Ceci, S. J., Ross, D. F. & Toglia, M. P. (Eds) (1989). *Perspectives on children's memory*. New York: Springer-Verlag.

- Ceci, S. J., Toglia, M. P. & Ross, D. F. (Eds) (1987). *Children's eyewitness memory*. New York: Springer-Verlag.
- Chi, M. T. (1976). Short-term memory limitations in children: Capacity or processing deficits? *Memory and Cognition*, **4**, 559–572.
- Christianson, S.-A. (1992a). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin*, **112**, 284–309.
- Christianson, S.-A. (1992b). *The handbook of emotion and memory: Research and theory*. Hillsdale, NJ: Erlbaum.
- Clifford, B. R. (1978). A critique of eyewitness research. In M. Gruneberg, P. Morris & R. Sykes (Eds), *Practical aspects of memory*, pp. 199–209. London: Academic Press.
- Clifford, B. R. (1979). Eyewitness testimony: The bridging of a credibility gap. In D. P. Farrington, K. Hawkins & S. M. Lloyd-Bostock (Eds), *Psychology, law, and legal processes*, pp. 167–183. Atlantic Highlands, NJ: Humanities Press.
- Davies, G. & Flin, R. (1988). The accuracy and suggestibility of child witnesses. In G. Davies & Drinkwater (Eds), *Do the courts abuse children?: Issues in criminology and legal psychology*, no. 13. Leicester: British Psychological Society.
- Davies, G., Tarrant, A. & Flin, R. (1989). Close encounters of the witness kind: Children's memory for a simulated health inspection. *British Journal of Psychology*, **80**, 415–429.
- Deffenbacher, K. A. (1983). The influence of arousal on reliability of testimony. In S. M. A. Lloyd-Bostock & B. R. Clifford (Eds), *Evaluating witness evidence: Recent psychological research and new perspectives*, pp. 235–251. Chichester: Wiley.
- Dent, H. & Flin, R. (Eds) (1992). *Children as witnesses*. Chichester: Wiley.
- Doris, J. (Ed.) (1991). *The suggestibility of children's recollections*. Washington, DC: American Psychological Association.
- Goodman, G. S. (1984). Children's testimony in historical perspective. *Journal of Social Issues*, **40**, 9–31.
- Goodman, G. S., Aman, C. & Hirschman, J. (1987). Child sexual and physical abuse: Children's testimony. In S. J. Ceci, M. P. Toglia & D. F. Ross (Eds), *Children's eyewitness memory*, pp. 1–23. New York: Springer-Verlag.
- Goodman, G. S. & Bottoms, B. L. (Eds) (1993). *Child victims, child witnesses: Understanding and improving testimony*. New York: Guilford.
- Goodman, G. S. & Hahn, A. (1987). Evaluating eyewitness testimony. In I. Weiner & A. Hess (Eds), *Handbook of forensic psychology*, pp. 258–292. New York: Wiley.
- Goodman, G. S., Quas, J. A., Batterman-Faunce, J. M., Riddlesberger, M. M. & Kuhn, J. (1994). Predictors of accurate and inaccurate memories of traumatic events experienced in childhood. *Consciousness and Cognition*, **3**, 269–294.
- Goodman, G. S. & Reed, R. S. (1986). Age differences in eyewitness testimony. *Law and Human Behavior*, **10**, 317–332.
- Hosch, H. M. & Cooper, D. (1982). Victimization as a determinant of eyewitness accuracy. *Journal of Applied Psychology*, **67**, 649–652.
- Johnson, M. K. & Foley, M. A. (1984). Differentiating fact from fantasy: The reliability of children's memory. *Journal of Social Issues*, **40**, 33–50.
- King, M. A. & Yuille, J. C. (1987). Suggestibility and the child witness. In S. J. Ceci, M. P. Toglia & D. F. Ross (Eds), *Children's eyewitness memory*, pp. 24–35. New York: Springer-Verlag.
- Leippe, M. R., Wells, G. L. & Ostrom, T. M. (1978). Crime seriousness as a determinant of accuracy in eyewitness identification. *Journal of Applied Psychology*, **63**, 345–351.
- Loftus, E. F. (1977). Shifting human color memory. *Memory and Cognition*, **5**, 696–699.
- Loftus, E. F. & Davies, G. (1983). Distortions in the memory of children. *Journal of Social Issues*, **40**, 51–67.
- Loftus, E. F., Miller, D. G. & Burns, H. J. (1978). Semantic integration of verbal information into a visual memory. *Journal of Experimental Psychology: Human Learning and Memory*, **4**, 19–31.
- Memon, A., Holley, A., Wark, L., Bull, R. & Kohnken, G. (1996). Reducing suggestibility in child witness interviews. *Applied Cognitive Psychology*, **10**, 503–518.
- Naus, M. J. & Halasz, F. (1979). Developmental perspectives on cognitive processing and semantic memory structure. In F. I. M. Craik & L. Cermak (Eds), *Levels of processing*, pp. 259–288. Hillsdale, NJ: Erlbaum.

- Naus, M. J. & Ornstein, P. A. (1983). Development of memory strategies: Analysis, questions, and issues. In M. T. H. Chi (Ed.), *Trends in memory development*, pp. 1–30. Basel, Switzerland: Karger.
- Peterson, C. & Bell, M. (1996). Children's memory for traumatic injury. *Child Development*, **67**, 3045–3070.
- Peterson, L., Moreno, A. & Harbeck-Weber, C. (1993). 'And then it started bleeding': Children's and mother's perceptions and recollections of daily injury events. *Journal of Clinical Child Psychology*, **22**, 345–354.
- Pynoos, R. S. & Eth, S. (1984). The child as witness to homicide. *Journal of Social Issues*, **40**, 87–108.
- Rudy, L. & Goodman, G. S. (1991). Effects of participation on children's reports: Implications for children's testimony. *Developmental Psychology*, **27**, 527–538.
- Saywitz, K. J. & Moan-Hardie, S. (1994). Reducing the potential for distortion of childhood memories. *Consciousness and Cognition*, **3**, 408–437.
- Terr, L. (1981). Psychic trauma in children: Observations following the Chowchilla school bus kidnapping. *American Journal of Psychiatry*, **138**, 14–19.
- Terr, L. (1983). Chowchilla revisited: The effects of psychic trauma four years after a school bus kidnapping. *American Journal of Psychiatry*, **140**, 1543–1550.
- Tobey, A. E. & Goodman, G. S. (1992). Children's eyewitness memory: Effects of participation and forensic context. *Child Abuse and Neglect*, **16**, 779–796.
- Vandermass, M. O., Hess, T. M. & Baker-Ward, L. (1993). Does anxiety affect children's reports of memory for a stressful event? *Applied Cognitive Psychology*, **7**, 109–128.
- Wechsler, D. (1967). *Manual for the Wechsler Preschool and Primary Scale of Intelligence*. New York: Psychological Corporation.
- Yuille, J. C. (1986). Meaningful research in the police context. In J. C. Yuille (Ed.), *Police selection and training*, pp. 225–243. Dordrecht, The Netherlands: Martinus Nijhoff.
- Yuille, J. C. & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *Journal of Applied Psychology*, **71**, 291–301.
- Yuille, J. C. & Tollestrup, P. A. (1992). A model of diverse effects of emotion on eyewitness memory. In S.-A. Christianson (Ed.), *The handbook of emotion and memory: Research and theory*, pp. 201–215. Hillsdale, NJ: Erlbaum.
- Zaragoza, M. S., Graham, J. R., Hall, G. C. N., Hirschman, R. & Ben-Porath, Y. S. (Eds) (1995). *Memory and testimony in the child witness*. Thousand Oaks, CA: Sage.

Received 28 May 1997; revised version received 19 August 1997

## Appendix A Objective questions

### Misleading questions

- (1) Did the man wearing blue (brown) pants have light hair?
- (2) Who was John Logan (Foster) looking for? The principal?
- (3) Did he put the pencil (pen) he had in his pocket?
- (4) Did the man with the moustache (beard) wear glasses?
- (5) Did the man take off his jacket (sweater)?
- (6) Did the man put his magazine (newspaper) down on that chair?

### Neutral questions

- (1) Did the man wearing nice pants have light hair?
- (2) Who was the man looking for? The principal?
- (3) Did he put anything in his pocket?
- (4) Did the man wear glasses?
- (5) Did the man take off any of his clothes?
- (6) Did the man put something down on that chair?

*Filler questions*

- (1) Was he an older man?
- (2) Was he a big, fat man?
- (3) Did the man wear a white shirt?
- (4) Was the man's hair straight, not curly?
- (5) Did he wear a tie?
- (6) Did the man knock before he came in?
- (7) Did the man touch anyone?
- (8) When he left, did he say where he was going?

*Final test questions*

- (1) Was the man wearing blue pants or brown pants?
- (2) Did the man have a beard or a moustache?
- (3) Did the man say his name was John Foster or John Logan?
- (4) Was the man wearing a jacket or a sweater?
- (5) Did he have a pencil or a pen?
- (6) Was he carrying a newspaper or a magazine?