1-1-2001

Similarities and Differences in Eyewitness Testimonies of Children Who Directly Versus Vicariously Experience Stress

Marc A. Lindberg Ph.D.
Marshall University, lindberg@marshall.edu

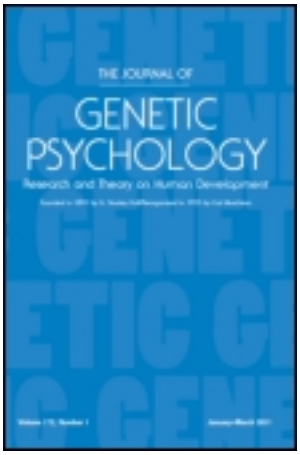
Susan Jones

Lisa McComas Collard

Stuart W. Thomas

Recommended Citation

Marc A. Lindberg, Susan Jones, Lisa McComas Collard & Stuart W. Thomas (2001): Similarities and Differences in Eyewitness Testimonies of Children Who Directly Versus Vicariously Experience Stress, The Journal of Genetic Psychology: Research and Theory on Human Development, 162:3, 314-333.



The Journal of Genetic Psychology: Research and Theory on Human Development

Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/vgnt20

Similarities and Differences in Eyewitness Testimonies of Children Who Directly Versus Vicariously Experience Stress

Marc A. Lindberg [a], Susan Jones [a], Lisa McComas Collard [a] & Stuart W. Thomas [a]

[a] Marshall University

Available online: 30 Mar 2010

To cite this article: Marc A. Lindberg, Susan Jones, Lisa McComas Collard & Stuart W. Thomas (2001): Similarities and Differences in Eyewitness Testimonies of Children Who Directly Versus Vicariously Experience Stress, The Journal of Genetic Psychology: Research and Theory on Human Development, 162:3, 314-333

To link to this article: http://dx.doi.org/10.1080/00221320109597486

The Journal of Genetic Psychology, 2001, *162*(3), 314–333

Similarities and Differences in Eyewitness Testimonies of Children Who Directly Versus Vicariously Experience Stress

MARC A. LINDBERG
SUSAN JONES
LISA McCOMAS COLLARD
STUART W. THOMAS
Marshall University

ABSTRACT. This study tested questions of ecological validity by comparing the eyewitness testimonies of children directly experiencing a painful inoculation experience with those of children in a yoked-control group who vicariously experienced the inoculation on videotape. The study involved 86 5-year-olds, divided between 2 groups: the experiential and yoked control. The experiential group was followed through a health department with a video camera as they received diphtheria, pertussis, tetanus (DPT), and oral polio inoculations. They were tested immediately, 20 min later, and 1 month later. Each child in the yoked-control group merely watched the videotape of his or her counterpart in the experiential group, made similar ratings of pain, and was given the same tests and suggestions. Stress and personal experience affected items congruent with the stressor to produce flashbulb-like memories, with slower rates of forgetting for some items, such as nurse identifications, and greater suggestibility for other items, such as estimates of needle size. These and the apparently conflicting results in the literature were said to make sense when personally experienced stress was viewed from S.-A. Christianson's (1992) interactive perspective rather than as a single ubiquitous variable.

Key words: children, direct experience of stress, eyewitness testimonies, vicarious experience of stress

CAN OUR STUDIES of eyewitness testimony be generalized to real cases of abuse? That question has been hotly debated in the laboratory, clinic, and courtroom, and it involves several related issues and methodologies. In explorations of children's eyewitness memories, one of two research strategies has generally been employed. Either researchers have tested children's recall of slides or videotapes, or they have tested children's memories for personally experienced and often painful medical procedures. Therefore, an important question is, "In what ways do the memories of children viewing videotapes parallel those who actually experience painful or stressful events?" The answer to this question is particularly

important when seeking to generalize results from the laboratory to cases of eyewitness testimony involving stressful or traumatic events, such as physical or sexual abuse (Ceci, 1991; Goodman, Quas, Batterman-Faunce, Riddlesberger, Riddlesberger, & Kuhn, 1996; Goodman, Rudy, Bottoms, & Aman, 1989; Yuille & Wells, 1991). Steward and Steward (1996, pp. 11, 27), for example, have proposed that "the results of bystander eyewitness research may underestimate the report of the child who has been physically and/or sexually abused," and that "distress may be a critical mediator that filters children's reporting of remembered events." Pynoos (1992) and Yuille and Tollstrup (1992), on the other hand, have reasoned that personal injury or penetration can cause attention to be more focused on internal rather than external stimuli, leading to relatively poorer memories for some of the events perceived. Finally, others have suggested that stress may not operate in such a global fashion and that its effects are primarily on attention and storage of events, increasing memories for items central to the stressor itself (Christianson, 1992; Easterbrook, 1959; Heuer & Reisberg, 1990; Walker, 1958). Our purpose in the present study was to develop a paradigm to begin explorations of similarities and differences between children who directly versus vicariously experienced the often stressful ordeal of preschool inoculations.

Several studies involving children have already documented that compared to bystanders, participants have improved memory and reduced suggestibility (e.g., Baker-Ward, Hess, & Flannagan, 1990; Jones, Swift, & Johnson, 1988; Rudy & Goodman, 1991; Tobey & Goodman, 1992). One potential limitation with these studies, however, was that the children who were in the direct-experience conditions did not experience high degrees of stress and pain that often accompany physical or sexual abuse. Thus, although these studies examined the effects of experience, they did not study that part of the ecological validity question assessing differential levels of perceived pain and stress because they only tested memories for events, such as getting a "tattoo" sticker on their arm, playing "Simon says," and so on.

What are the effects of stress on memory and suggestibility? The answers to these questions are less clear. Many studies have found either no effects or interfering effects of stress on memory (e.g., Bugental, Blue, Cortez, Fleck, & Rodriguez, 1992; Merritt, Ornstein, & Spicker, 1994; Ornstein, Gordon, & Larus, 1992; Peters, 1991). However, Goodman, Hirschman, Hepps, and Rudy (1991) found higher levels of recall associated with higher levels of stress. In a second

The authors thank Martin Amerikaner, Mary Chapman, and Peter Ornstein for their helpful comments and suggestions regarding earlier versions of this article. They also thank University Pediatrics and Joseph Werthammer, Gilbert Ratcliff, and Joseph Evans; Stone and Thomas Department Stores; and the Department of Health in Huntington for their assistance. Karla Dick, Deborah Childress, and Carolee McCallister helped with the data coding and analysis.

Address correspondence to Marc A. Lindberg, Department of Psychology, Marshall University, Huntington, WV 25755; lindberg@marshall.edu (e-mail).

study, Goodman et al. (1991) compared children who received inoculations to children who only received a decal rubbed onto their arm and leg. They found that the less stressed children who were not inoculated recalled more incorrect information than those who had been inoculated. However, because the to-be-remembered information was qualitatively different, strong conclusions were not possible. In another interesting study, Quas et al. (1999) found that high levels of stress were associated with poorer general free-recall performance but superior cued-recall performance and lowered suggestibility.

Why were the results from the different studies so different? One reason might be that Goodman (1991) tested specific features related to the stressor itself (questions about where they were touched and identifications of the nurse) as compared to the other investigators, who typically tested either overall recall, recall of all details added together (cf. Merritt, Ornstein, & Spicker, 1994), or details unrelated to the stressor (cf. Peters, 1991). In line with Christianson's (1992) theory, it could be reasoned that if an item tested was congruent with the stressor, then better retention would be the most likely result. If, on the other hand, the item was peripheral to the stressor, then either no differences or perhaps even poorer retention would result. According to Christianson (1992), the reason for this improved recall for features congruent with emotional events is that they receive preferential initial processing because of increased attention (Easterbrook, 1959). Christianson (1992) said that a second way in which information congruent with the stressor could lead to superior retention would be through increased elaborative processing that would create stronger traces. Both the increased attention and elaborative processing views would predict that information congruent with the stressor would show less steep forgetting curves.

This line of reasoning is similar to the "flashbulb hypothesis" (Brown & Kulik, 1977), which proposes that a highly emotionally arousing event can produce a vivid memory of the details of that event. Since the original research by Brown and Kulik (1977), there have been many other studies whose findings have supported the flashbulb hypothesis (e.g., Bohannon, 1988; Christianson, 1989; Winograd & Killenger, 1983), including several in the literature on children (e.g., Terr, 1983, 1996; Warren & Swartwood, 1992). However, all were studies of real-life events, and, as Christianson (1992) noted, one major problem was the lack of a comparable low-stress event. Without such a control event and tests of memories at different delays, it was impossible to determine whether flashbulb memories were truly distinct entities that had the same rates of forgetting as other memories. Furthermore, most flashbulb memory studies have typically studied discoveries of emotional events (e.g., the JFK assassination, the Challenger explosion), rather than personally experienced stressful events (Winograd & Neisser, 1992). Thus, most studies on different notions of flashbulb memories have not compared personally experienced emotional events with vicariously experienced events. In the present study, we sought to provide a paradigm to do this and offer initial descriptive data on such comparisons.

The present study was designed to explore factorial combinations of the following: (a) different levels of stress (measured by subjective pain of the inoculation), (b) experience (those who received the shot versus those in the yoked-control group who watched the experimental group child's procedures on videotape), (c) time of test (tests at 5 min, 20 min, and 1 month), and (d) type of memory trace tested (details congruent versus unrelated to the stressor). Thus, children in a natural or quasi experiment were compared to children in a yoked-control group, who merely viewed the same things vicariously on a videotape. It was hypothesized that when details congruent with the stress were tested (e.g., identifications of the nurses), those directly experiencing the stress would show better memories than those vicariously experiencing the painful inoculations (Christianson, 1992). However, in line with Lindberg, Kiefer, and Thomas's (2000) distinction between memories for details versus inferences and their notions of congruence, if memories for judgments and inferences were tested that were related to perceived pain (e.g., "size of needle"), then it was hypothesized that stress would act like a suggestion (or more properly an auto-suggestion), and those directly experiencing it would be more likely to overestimate its size (Bruner & Goodman, 1947). If items were tested that were extremely peripheral to the stressors, then differences in memory performance might favor the yoked controls.

Method

Participants

The average age of the children was 5.05 (SD = .42) years, and 42 boys and 44 girls participated. The children participated in either of two conditions, experiential or yoked control, with the restriction that an equal proportion of boys and girls were placed in each counterbalancing condition. The experiential group participants were children who were required to get their prekindergarten inoculations. In the first phase of testing, 22 participants for the experiential group were recruited from daycare centers and preschools with the help of a local pediatric practice. The children were from predominantly upper-middle-class homes. After receiving informed consent forms signed by the parents, the experimenters called and scheduled a time for the child's preschool DPT and oral polio inoculations at the health department. Because of difficulties in obtaining enough participants, the experimenters waited for 2 months (i.e., until 4 weeks before the beginning of school) and recruited at the health center as children and their parents came in for free inoculations. To provide incentives for participation in the second wave of testing, those who participated were offered a $10.00 gift certificate from Stone and Thomas Department Stores and a $1.00 lottery ticket. An additional 21 children for the experiential group were recruited in this fashion. These children were from a lower income area than the first sample and all were White. They were 4.08 months older than the first group of participants.

The children in the yoked-control group, who had not yet received the inoc-

ulations at the time of testing, were tested 1 year later. Waiting 1 year allowed for the yoked-control population to be selected without confounds, where yoked-control children could have known the experimental children, and also afforded the possibility of matching more closely age and daycare center. The children in the yoked-control group were selected from the same daycare centers and preschools as the children in the experiential group, or they were from the same neighborhood. Thus, each child in the yoked-control group was yoked to one of the 43 children in the experiential group, in terms of daycare center or preschool, neighborhood, age, race, or sex. There was only one Black child in the experiential group and therefore only one in the yoked-control group. At the end of the experiment, the yoked-control group received a small course on trips to the doctor and inoculations to reduce any fears that could have been produced from seeing children who appeared to be terrified and crying.

Procedures

When the children and their mothers and fathers arrived at the health department, one of the experimenters greeted them and tested the child's knowledge of colors as he or she waited in the lobby. All children had to demonstrate knowledge of color labels by correctly pointing to the color patches on the card, a color card that would later be used to test their retention of the colors of the things witnessed. The parents were instructed to hold their child on their laps in each of the rooms and remain silent while the child was being interviewed. The experimenter operating the camera began filming as soon as the child entered the room to have his or her temperature taken. The male experimenter operating the camera followed them into the room and continued filming from behind the chair where the child was seated. The camera followed the head movements of the child in an attempt to approximate the child's visual field. Views of the child's face were not included, then, and what was looked at by each child in the experiential group was placed in the center of the camera angle for the child's counterpart in the yoked-control group. After a typical wait of about 2 min in the temperature room, the temperature nurse placed a white thermometer under the arm of the child. After the thermometer was removed, the nurse told the child his or her temperature.

When the child was led to the next room for his or her injection, the experimenter operating the camera followed and again stood behind the chair where the child was again seated on the parent's lap. The door was closed and the nurse first gave the oral polio vaccine, which was approximately 1/2 oz of liquid, which tasted like fruit-flavored Koolaid. The nurse then administered the DPT injection with a 12-cm syringe and needle in the arm. Filming stopped when the child left the room.

After the injection, the child and the parent were taken to the interview room. A female experimenter greeted them, and the child was allowed to choose between several small candy bars and immediately eat the one selected to help

reduce stress and change the climate after just having been hurt by a stranger in the previous room. Children were told that if they answered the questions as well as they could, they would be given another candy bar when they went home. To establish rapport, the experimenter first asked the children to give their name and then to recall a favorite television program. The experimenter then asked them to try and remember everything that happened to them from the point where they entered the building until that moment. If the child did not say anything, the experimenter prompted with a question about what happened when they first came into the building, and then questions prompting what happened next were asked throughout their free recall. All responses were tape-recorded and later written down verbatim. The experimenter then introduced the child to the 6-point Likert scale of smiley faces. The faces ranged from one with a big smile to one with a big frown and five tears. The children were asked to point to a face that best described how much fun it would be to go through the shot experience again. After this question, half were given the leading question, "Where did they place the *yellow* thermometer when they took your temperature?" The other half were asked, "Where did they place the thermometer when they took your temperature?" (Here it should be noted that the thermometer was white.) The half that did not receive the leading question about the thermometer were given the question "Where did they place the *green* needle when they gave the shot?" The half that received the leading question about the thermometer were given the question "Where did they place the needle when they gave the shot?" (Here it should be noted that the shot device was blue.) The repeated-measure feature of the leading question manipulation was to increase sampling of questions by having each child participate in both the experiential and control groups. The next question asked them to show with their hands how big the syringe ("plunger thing and needle") was. They were instructed to indicate by spreading their hands, and the experimenter measured how far their hands were spread apart. After this 5-min testing period, they and their mothers or their fathers were taken to a large waiting room where the caretakers read Dr. Seuss books to the children for 20 min.

After the 20 min, the children and parents returned to the testing room, where the same experimenter gave a different test of memory that was designed to examine the manipulations of the first test and also test additional information. Questioning again began with a rapport-building process by asking the children for their favorite story and character in the story, and then giving them the same free-recall question, with the same type of prompting, that began the 5-min testing session. The children were then questioned about the colors of the lab coats of the temperature and shot nurses. The children were instructed to point to one of seven colors on the card they had seen in the waiting room, colors that were preselected to match the colors of the uniforms of the nurses and other colors in the inoculation environment. They were then shown a 6-point smiley-face card with six smiley faces ranging progressively from one with a straight mouth, to one with slightly turned-down corners of the mouth, to ones with bigger frowns, to one

with two tears, to one with a big frown with five tears going down from the eyes. They were instructed to point to how painful it was when the shot was given, when temperature was taken, and when the polio liquid was injected into the mouth. After these questions, they were asked what was on the table where the syringes were kept, and then they were asked to point to the color card to indicate the colors of the thermometer and the syringe. The experimenter then assessed memory for the identifications of the temperature and shot nurses. She showed the child a lineup of six pictures of the nurses that had given the shots and taken temperatures along with other nurses sampled from a local hospital selected to be somewhat similar in age to the shot nurse. The temperature nurse stood out from the lineup in that she had gray hair and was about 25 or more years older than the other nurses, who all appeared to be around 40. In this way, a fair lineup was presented for the shot nurse, and to test for generalizability of findings across different testing formats, the temperature nurse test was used to maximize performance and explore if such biased tests were differentially affected by stress and experience. After this lineup questioning, the children were asked to point to the colors of the temperature nurse's and shot nurse's shoes. The experimenter then showed the child a picture of the lab table from the injection room, where the syringes were kept, and asked, "What three things were not on the table when you were in the room?" This recognition testing was done several questions after the recall question to attempt again to maximize performance and explore if such tests showed different patterns of results. The eight things that were on the table were cotton swabs, needles, alcohol, a metal container containing tongue depressors, a container for discarded needle tips, and small bottles containing the inoculation fluid. The three things that did not belong were fairly obvious: a white coffee cup, a large jar of coffee creamer, and a large yellow tub of baby wipes. The experimenter prompted until three things were listed. The experimenter then said, "It would please me if you said that this didn't hurt at all. Big kids don't think that it hurts very much." They were asked to point again at a 6-point smiley-face card, going from a big smile to one with a big frown with five tears to indicate how much fun it would be to go through the procedures once again. It should again be noted that this was said after the aforementioned pain-of-needle rating was obtained.

One month later, the same camera man (the senior author) and a different female experimenter visited the homes of the children in the experiential group to administer the same questions asked at the 20-min test. The different experimenter was used because, typically, in forensic interviews, different people do the later questioning (Goodman et al., 1989). However, the questions themselves were the same as those in the 20-min interview.

Because of a change in procedures at the health department, the second group of 21 children in the experiential group went through a slightly different set of experiences than the first. With these different children, the temperatures were no longer taken, and the children went directly into the shot room. They also received

a tuberculin test in addition to the oral polio and the more painful DPT shots. The tuberculin test involved having the nurse prick the skin with a very small needle. Most children do not perceive this to be as painful as the DPT shot. Thus, these children, in contrast to those in the first experiential group, got a relatively painless shot with a very small needle before getting the more painful DPT inoculation. Another difference in this group was that they were tested 1 month later in their daycare center. Any effects analyzed or discussed due to these differences in procedures will be termed "procedures." Other than these differences, the children went through the same procedures and questions as did the first experiential group.

Each child in the yoked-control group was matched with one of the 43 children in the experiential group, based on daycare center, race, sex, and age. In a testing room in the daycare center, each yoked-control group child watched the videotape of the experiential-group child with whom he or she was matched. The video was clear and displayed the field of vision of the experiential-group child from the beginning to the end of the procedures. Because the experiential group was not told ahead of time that they would be tested for memory, the yoked-control group was merely told to watch this film very carefully. They received the memory tests in a different room. Apart from these differences in viewing conditions, the conditions, questions, time intervals, and so forth, of the Yoked controls were matched to those of their Experiential counterparts.

Coding

To explore possible differences in memory that could result from various levels of stress, it was necessary to develop measures for the perceived stress independent variable. Two approaches to measuring the construct of stress were used, one experimenter based and one child based. The experimenter-based measures were taken by experimenters who viewed the videotapes but were blind to the children's memory performance or subjective estimates of pain. They rated the taking of the thermometer, polio inoculation, and DPT shot according to the following scales: 1 = *willingly took,* 2 = *hesitantly took,* 3 = *mother pushed to take,* 4 = *fought mom and nurse.* The children's reactions to these procedures were rated on the following scales: 1 = *little reaction,* 2 = *whimpering,* 3 = *crying,* 4 = *crying and screaming.* The experimenters also used stopwatches to time the seconds of crying behavior that could easily be heard on the videotape and objectively recorded. The pain of needle construct, as determined by the objective measure of crying, was turned into a dichotomous independent variable by dividing the sample into two groups, those who cried and those who did not.

The child estimates of pain of needle were taken from the children's first rating of perceived pain of the needle according to a 6-point Likert scale that used a cartoon smiley face that went from a straight mouth to a frown with tears. These child-based subjective estimates of pain-of-needle were likewise split into two

groups, such that this measure could serve as a dichotomous independent variable. Those scoring 5 or 6 on the Likert scale, with 6 being the most painful, were defined as the *high-pain* children. Those rating the pain of the shots at 4 or below were defined as the *low-pain* children. These numbers were closest to a median split and this variable will be referred to as the subjective pain variable. The yoked-control group made similar judgments about their rating of how painful they perceived the procedures to be. Because these ratings were not necessarily tied to their counterpart in the experiential group, there was not a one-to-one matching between experiential and yoked-control group member, using this measure of perceived pain. However, because our purpose in the present study was to explore actual versus vicarious perceptions of pain using the same scale, this approach was most relevant for the kinds of comparisons outlined earlier. In the experiential group, there were 20 children in the low-pain group and 23 in the high-pain group. In the yoked-control group, there were 18 in the low-pain group and 25 in the high-pain group.

To explore how the two dependent variable measures of stress (cry vs. pain) were related, two different kinds of statistical tests were performed: chi-square tests and correlations. A 2 (cry vs. no-cry) × 2 (high-pain vs. low-pain) chi-square test showed that the two measures were significantly related, $\chi^2(1, N = 86) = 5.59$, $p = .02$, with 68% in the cry group rating the experience as high pain and 32% rating it as low pain.

The following were reasons why the subjective pain estimates were used and presented in the following presentations of results: (a) chi-square tests showed

TABLE 1
Correlations Between Measures of Pain and Reactions to the Procedures

	Rating							Crying
Measure	Took polio	React polio	Took shot	React shot	Pain ther	Pain polio	Pain shot	behavior to the shot(s)
React ther	.83**	1.00	.31	.47*	.76	.29	.60**	.68**
Took polio		.70**	.51**	.40**	.48*	.23	.40**	.76**
React polio			.32*	.41**	.76**	.22	.53**	.59**
Took shot				.70**	.01	.38**	.09	.66**
React shot					.19	.42**	.07	.70**
Pain ther						.10	.32*	.23
Pain shot							.10	.22*
Pain polio								.02

Note. The number of observations averaged 22 for the thermometer and shot measures, and 43 for the other measures. Took polio, took shots = ratings of the children's willingness to take oral polio vaccine and shot. React polio, react shot, react ther = ratings of children's reactions to vaccine, shot, and thermometer. Pain shot, pain ther, pain polio = children's ratings of pain.
*$p < .05$. **$p < .01$.

that the two measures were related; (b) the two measures correlated significantly with one another (see Table 1); (c) they revealed similar patterns of results, with the subjective pain measure being somewhat more robust than the cry statistic (cf. Merritt, Ornstein, & Spicker, 1994); (d) the subjective pain variable allowed factorial comparisons of experience and stress; and (e) the subjective estimates of pain might have more pragmatic utility in cases of abuse, where the child is often the only witness and it would be rather presumptuous for an adult to be the judge. The correlations between the various measures of pain and reactions to the procedures can be seen in Table 1.

Results

Preliminary analysis indicated no significant effects associated with the children's sex. Therefore, the data were collapsed across this factor. Age was also not used as an independent variable as it was discovered to be confounded with first versus second set of procedures. The average age in the second procedures (M = 5.27) was significantly higher than that in the first procedures (M = 4.93), $F(1, 81) = 8.06$, $p < .01$. Therefore, to analyze continuous variables, a regression analysis was first performed using age as the independent variable. Then in the analysis of variance, the variance in the dependent variable due to age was partialed out. In this way the predicted recall based on age was subtracted from the actual score leaving only the variance accounted for by the analyzed independent variable to be tested. For chi-square analyses of discrete variables, a data set was created that equated age across condition by randomly eliminating the oldest children from the discrepant conditions resulting in 73 children for these analyses.

Nurse Identifications

Although similar to children in the yoked-control group at the 20-min test, children in the experiential group were more accurate at the 1-month test in their identifications of the temperature nurse, $\chi^2(1, N = 38) = 7.94$, $p < .005$. These results may be seen in Figure 1. If there was differential forgetting between the experiential and yoked-control groups over the month interval, then one would expect that of those who got it right on the first test, more in the control group than in the experiential group would get it wrong on the second test. A Fisher's exact test found that the probability of correctly identifying the temperature nurse on the second test, given a correct on the first test, was higher ($p < .01$) for the experiential group than the yoked-control group. Of the 13 children in the experiential group who got it right on the 20-min test, 12 were also correct on the 1-month test. Of the 13 children in the yoked-control group who got it right on the 20-min test, only 6 were correct on the 1-month test.

The same pattern was found for identification of the nurse who gave the shots. Although similar on the 20-min test, children in the experiential group were

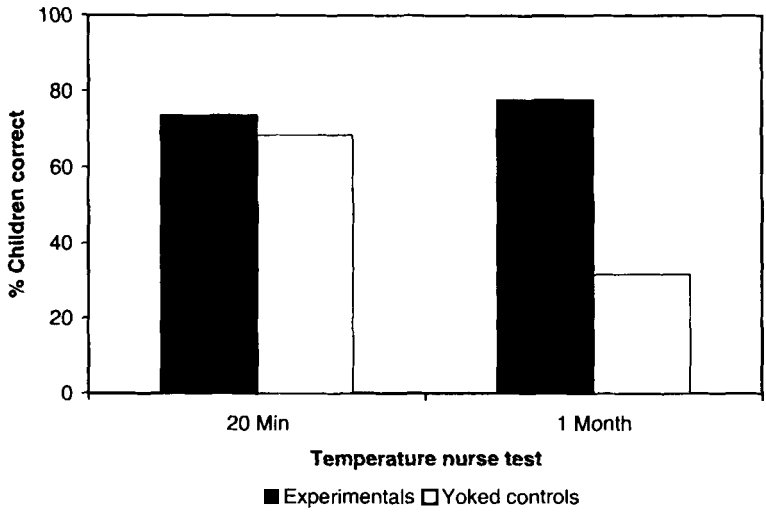


FIGURE 1. Percent of children correctly identifying the temperature nurse as a function of time of test and whether they directly experienced the shot (experiential group) or merely viewed it vicariously (yoked-control group).

more accurate in their identifications of the shot nurse at the 1-month test than were those in the yoked-control group, who merely watched the events on videotape, $\chi^2(1, N = 69) = 4.68, p = .03$. Those results can be seen in Figure 2. To more precisely test the possibility of differential rates of forgetting, a Fisher's exact test showed that the probability of identifying the shot nurse correctly on the second test, given a correct on the first test, was higher ($p < .05$) for the experiential group than the yoked-control group. Of the 12 children in the experiential who got it right on the first test, 7 got it correct on the 1-month test. Of the 8 members of the yoked-control group who got it right on the 20-min test, only one got it right on the 1-month test.

Clothing of Nurses

The children were asked to recall the colors of the shoes and coats worn by the temperature and shot nurses. There were no significant differences in coat color identification for either nurse. The temperature nurse's coat color was correctly identified out of the sample of seven colors on a card by 24% of the children at the 20-min test and by 32% at the one month test. The shot nurse's coat was correctly identified by 41% of the children at the 20-min test and by 16% at the 1-month test. The color of the temperature nurse's shoes was correctly identified by 11% at the 20-min test and by 27% at the 1-month test. A significant

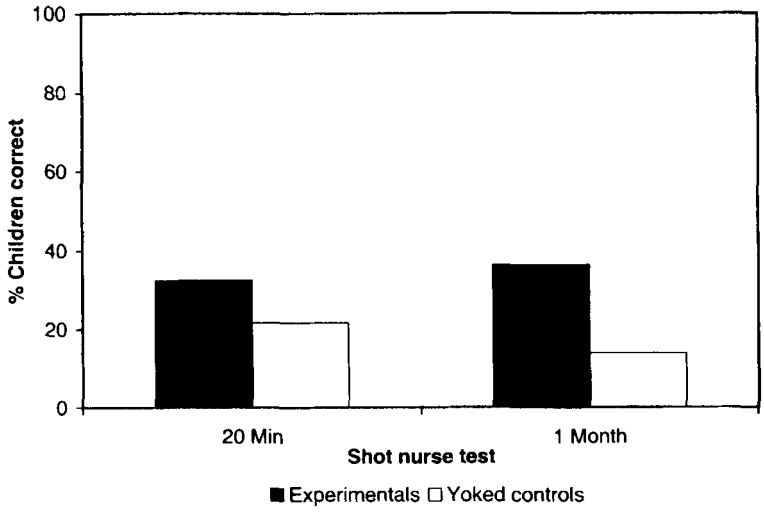


FIGURE 2. Percent of children correctly identifying the shot nurse as a function of time of test and whether they directly experienced the shot (experiential group) or merely viewed it vicariously (yoked-control group).

experiential versus yoked control difference for correct identifications of the temperature nurse's shoe color at the one month test was found, $\chi^2(1, N = 37) = 4.50$, $p = .03$. Of the 27% correct, 80% were in the control group and 20% were in the experiential group. In identifications of the shot nurse's shoe color, 12% were correct at the 20-min test and 16% were correct at 1 month. No other effects were significant.

Needle Size

The children were asked "How big were the plunger and needle thing that gave you your shot?" They were to estimate the size of the needle by using both hands. It was questionable whether the two procedures were equated for purposes of needle analyses. As noted, because of changes in procedure at the health department, children in the second group of procedures received a tuberculin test in addition to the DPT shot. Both the TB test and the DPT shot punctured the skin in two different places, and were different colors. Given that the children in the second experiment were given two shots, and the experimenters asked the same question about "the shot," it was possible that some children may have mistakenly reported the size, color, or placement of the TB test needle and plunger when asked for those details of the "needle." Thus, data from procedure 2 were not included in analyses regarding attributes of the needle.

To determine if the children differed in their estimates of needle size, a 2 (experience) × 2 (pain) × 3 (time of test) analysis of variance was performed. A significant interaction for experience by pain was found, $F(1, 34) = 7.67$, $p < .01$. Post hoc Tukey HSD tests revealed that children who directly experienced the procedures showed the perceived pain effect, with those rating the shot as painful estimating it to be $M = 20.00$ cm, $SD = 10.88$, and those who rated it as less painful, estimating it to be $M = 11.15$ cm, $SD = 9.20$. Such differences were not found for the controls, where those who perceived the shot to be more painful estimated it to be $M = 17.21$, $SD = 9.37$, compared to those who rated it to be less painful, $M = 22.03$, $SD = 13.07$. No other effects were significant.

Other Attributes of the Needle and Thermometer

One half of the children received leading questions about the color of the needle and the other half were given leading questions about the color of the thermometer. Because of the mentioned differences in procedures, only the first procedure group's data could be used in the other chi-square analyses of the attributes of the needle and thermometer. Because this left the analyses with too few potential children to examine interactions between perceived pain, sex, time of test, and so on, only simple two-way tests were performed. No significant effects were observed for the pain variable, leading questions, or time of test. It was found that 63% of the children correctly pointed out the color of the needle at 20-min and 68% did so at the 1-month test. Furthermore, 86% correctly reported its placement. With the thermometer, 63% correctly identified its placement (under the arm) at the 5-min test, 63% at the 20-min test, and 68% at the 1-month test. On tests of the color of the thermometer, it was found that the experiential group was more accurate than the yoked-control group on the first test of thermometer color, $\chi^2(1, N = 38) = 7.23$, $p < .01$, with 84% of the experiential group getting it correct on the first test and 42% of the yoked-control group getting it correct. This did not hold for the second test. No other differences, including effects of leading questions, sex, perceived pain, experiential versus yoked control, were significant for other attributes of the needle and thermometer.

Items on Table

Both recall and recognition were tested for items that were used during the inoculations and on a lab table in the room. The children were tested for recall by asking, "Remember that there was a table where the syringes and needles were kept? What other things were on the table?" Recognition testing for items on the table involved showing the children a picture of the lab table from the injection room and asking, "What three things were not on the table when you were in the room?" This was done several questions after the recall question. There were no significant effects of pain or experiential condition for either mea-

sure. Recall averaged 1.14 items, and recognition of the three things not on the table was 1.63.

Memories for the Gist

To test overall recall of the gist, all children were asked to say what happened from the time they entered the building until they left. They were prompted for what happened next throughout their recalls. The responses were scored and categorized by a group of raters in terms of the number of idea units. These were items referring to things in the room, things referring to procedures, things referring to persons encountered, and things referring to the pain. Interrater reliability was high (r = .91) in classifying these idea units. Total correct and incorrect idea units and words produced were analyzed by 2 (experience) × 2 (pain) × 2 (procedures) × 3 (time of test) analysis of variances. No significant main effects or interactions were found for total number of correct or incorrect words produced. Similarly, no significant effects were found for correct or incorrect idea units recalled. The children recalled an average of 3.81 correct idea units (SD = 2.05) and .25 (SD = .42) incorrect units. (These levels of recall were about 2 fewer than the same aged participants tested at 1 week, as reported by Ornstein, Shapiro, Clubb, Follmer, & Baker-Ward, 1997, who tested for a doctor visit that included many more possible things and procedures to recall.) There were no significant effects of any condition for differences in correct (M = 11.74, SD = 6.13) or incorrect (M = .76, SD = 1.28) total words produced. Basement effects precluded meaningful analyses of the type of word data.

To measure the children's general impression of what happened, we asked them, "Overall, how much fun would it be to go through all that again?" The children responded to a "smiley" face going from *a big smile* = 1, to *a big frown with tears* = 6. They were asked this question at the immediate test, the 20-min test, and the 1-month test. The only significant effect revealed was the main effect of pain, $F(1, 82) = 15.93$, $p < .01$. As would be expected, those who said it hurt a great deal rated the whole experience worse ("somewhat terrible" M = 3.62) than those who said it did not hurt as much ("fun" M = 2.18). Thus, their ratings of the pain of the shot were significantly related to their ratings of the whole experience, but the direct versus vicarious experience variable and the suggestions were not.

Discussion

This natural or quasi experiment that compared the experiential group, which directly experienced the painful DPT inoculations, with the yoked-control group, which merely viewed the experiences vicariously, more specifically tested Christianson's (1992) hypotheses that stress and direct experience do not operate in a ubiquitous fashion. The repeated-measure feature of this design, testing at 20 min

and at 1 month, was the first to specifically test the possibility of differential forgetting rates, a test deemed important by several investigators (Baker-Ward, Hess, & Flannagan, 1990; Brown & Kulik, 1977; Burke, Heuer, & Reisberg, 1992; Christianson & Hubinette, in press; Christianson & Loftus, 1987, 1991; Hosch & Bothwell, 1990; Howe, Courage, & Peterson, 1996; Jones, Swift, & Johnson, 1988; Tobey & Goodman, 1992; Wagennaar & Groeneweg, 1990; Yuille & Cutshall, 1986). For identifications of both the shot and temperature nurse, the Fisher exact tests found that the experiential group had superior memories and greater resistance to forgetting over the 1-month interval (Goodman et al., 1991). Thus, the findings were consistent and robust across different sample sizes and lineup characteristics, emphasizing the fact that these memory-by-experience interactions were consistent and not subject to small parametric variations in design. However, when items that were peripheral to the stressors were tested, for example, shoe color, then the yoked controls showed superior performance (Peters, 1991).

A related methodological point that had practical as well as theoretical implications involved the higher levels of performance on temperature-nurse identification as compared to shot-nurse identification. As was seen in the Method section, the temperature nurse stood out in the lineup because she was the only one to have gray hair and a red vest. She appeared to be at least 25 years older than the others, who all had brown or darker blonde hair. In the case of the shot nurse, a more valid test of recognition memory was performed and more than 75% of the children named the wrong person. However, in the biased test with the temperature nurse, only about 25% of the children named the wrong person. Was the better performance due to test conditions that tapped some thread of memory that was still there? What are the validities of partial versus whole memories? The present study could not say, and hypotheses on the validity of partial information (Fisher & McCauley, 1995) must be explored by further memory research.

Although the differential performance on nurse recall was robust, the nature of the present design did not allow for a precise test of the mechanisms that might be responsible for this finding. The data did, however, present some hypotheses for further studies to test. We observed that the degree of stress involved with each person to be identified, the shot nurse and the temperature nurse, both showed similar effects on recognition. Furthermore, the level of perceived pain did not enter into the results. We could, therefore, suppose that either the more extreme stress of the shot nurse carried over to bolster the memory of the temperature nurse, or that it takes personal experience with a relatively novel stressful event to help shield the trace from such rapid rates of forgetting. Alternatively, and in line with the favored hypothesis of Christianson (1992), we could suppose that the central features of the experience were rehearsed and more elaborately coded by those who experienced it directly. Thus, when testing at the 20-min interval, those directly experiencing the event may have processed their selection more fully. Which of the above hypotheses is correct must be worked out by further research. The present approach offers a good way to address these hypotheses.

To test these hypotheses, one should also use more analytic designs similar to those developed by Brainerd, Reyna, Howe, and Kingma (1990). At any rate, the present paradigm, by testing a vicarious versus direct personal experience with painful stimuli, and testing with different delays in the same design, found that videotape bystander studies may underestimate memory performance for items congruent with the stressor when assessed at longer delays.

If arousal and personal experience were associated with better recall, then how could they also have produced more inaccurate inferences of needle size? The finding of motivational states and congruency entering into overestimates in perception was found long ago by Bruner and Goodman (1947) during their work on estimating coin sizes. In the classic Bruner and Goodman study, poorer children estimated that a coin was bigger than wealthier children did. These results were also in line with the findings that flashbulb or emotion-eliciting events are not retained as photographically complete memories (Christianson, 1992; Heuer & Reisberg, 1990; McCloskey, Wible, & Cohen, 1988; Warren & Swartwood, 1992; Winograd & Killenger, 1983). The study by Bruck, Ceci, Francoeur, and Barr (1995) also found that children can be suggestible even about procedures central to their own bodies, such as touches and perceived pain, when tested at longer intervals after repeated suggestions. In the case of the needle, if children from the experiential conditions thought that the shot was a "big hurt," then they tended to estimate that the needle was a "big needle." If they did not, then those in the experiential group were fairly accurate in their estimations of needle size. The controls, on the other hand, were not as influenced by their vicarious estimations of pain.

It is important to discuss another cautionary methodological note that has theoretical implications for explanations of the needle-size data. In this study, we defined centrality and congruency in terms of congruency with the stressors: the needle and the assailants. It could have been that children who experienced the shot and high levels of distress chose not to focus on the needle. Thus, although it would have been in range for the camera to follow, some of these children may have chosen to either close their eyes or not follow it as closely with their eyes. In this way, they may have never encoded the size of the needle, or perhaps encoded it similarly to the controls and more poorly than their counterparts in the low-pain condition. Therefore, they may have based their judgments on guesses rather than on some representation of pain plus the actual needle size. Another example of this hypothesis as used to explain the coin estimate data of the Bruner and Goodman (1947) study would be that the poorer children may not have remembered the exact size of the coin due to their lack of experience through poverty. (Here it should be pointed out that it has been clearly demonstrated that blending is not a very viable hypothesis explaining suggestibility [Lindberg, Kiefer, & Thomas, 2000].) If one were able to use eye movements in similar designs, one might be able to better define centrality than was done in the present study or in other studies and better test the encoding interpretation. By using such micro-

genic designs, one might be able to more specifically operationally define centrality and therefore more specifically test the strength of the arguments advanced here. Thus, more analytic microgenic designs must be carried out to observe individual differences in attentional focusing, such that the psychological mechanisms responsible for congruency effects can better be described.

In summary, one of the main goals of this study was to develop a paradigm to explore children's eyewitness testimony by comparing those who participated in a stressful procedure with those who merely viewed the actions on a videotape. The children were very accurate on most aspects of the needle (placement and color), indicating that both groups focused on the needle, or "weapon" (cf. Cutler, Penrod, & Martens, 1987; Kramer, Buckhout, & Eugenio, 1990). Over 80% of the children correctly identified the placement of the needle, and over 65% correctly identified the color of the needle, the color of the thermometer, and the placement of the thermometer. Leading questions, personal experience, and perceived pain did not seem to be large factors in these analyses. This was consistent with the findings of Melton (1992) and Saywitz, Goodman, Nicholas, and Moan (1991), where retention was found to be quite good for the retention of central details. The seemingly conflicting results on stress and personal experience and those in the literature made good sense if the notion of congruency was combined with the stressors and Christianson's (1992) and Easterbrook's (1959) theories of attentional focusing. Notions of stress and pain operating ubiquitously as a central mediator (Steward & Steward, 1996) or in some simple Yerkes-Dodson (1908) fashion did not seem to apply. If the item was tested at longer delays and was congruent with the stressor, then better retention was found with assailant identification (cf. Goodman et al., 1991). In this sense, the laboratory studies using videotapes did not parallel situations in which the children experienced the stress themselves when testing occurred at long delays. However, if the item to be tested was peripheral to the stressor (items on the table, colors of the nurses' coats), or if a global measure of memory for gist was used, then few if any differences between yoked controls who merely viewed the procedures on videotape and those who experienced the pain and events themselves were found (cf. Baker-Ward, Gordon, Ornstein, Larus, & Clubb, 1993). If the item was very peripheral to the central stressor as with the color of the temperature nurse's shoes, then the yoked-control group showed better performance (cf. Peters, 1991). If inferences about needle size were made, then those directly experiencing more stress were more suggestible. Finally, the yoked controls vicariously experiencing the stress were not more prone to the suggestions and did not differ on memories for the gist. These results support those theorists who have suggested that stress should not be treated as some ubiquitous variable. In conclusion, questions of vicarious experiences being ecologically valid were found to be too general. Thus, one cannot discount laboratory studies as being ecologically invalid nor can one easily conclude that they directly mirror cases in which children have experienced abuse or physical pain themselves. Rather, focusing on

what different experiences can do to lead to variations in attention, personal relevance, rehearsal, and motivation seems to be the most fruitful approach.

REFERENCES

Baker-Ward, L., Gordon, B. N., Ornstein, P. A., Larus, D. M., & Clubb, P. A. (1993). Young children's long-term retention of a pediatric examination. *Child Development, 64*, 1519–1533.

Baker-Ward, L., Hess, T. M., & Flannagan, D. A. (1990). The effects of involvement on children's memory for events. *Cognitive Development, 5*, 55–70.

Bohannon, J. N. (1988). Flashbulb memories for the space shuttle disaster: A tale of two theories. *Cognition, 29*, 179–196.

Brainerd, C. J., Reyna, V. F., Howe, M. L., & Kingma, J. (1990). The development of forgetting and reminiscence. *Monographs of the Society for Research in Child Development, 55*, Nos. 3–4.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition, 5*, 73–99.

Bruck, M., Ceci, S. J., Francoeur, E., & Barr, R. (1995). I hardly cried when I got my shot! Influencing children's reports about a visit to their pediatrician. *Child Development, 66*, 193–208.

Bruner, J. S., & Goodman, C. C. (1947). Value and need as organizing factors in perception. *Journal of Abnormal and Social Psychology, 42*, 33–44.

Bugental, D., Blue, J., Cortez, V., Fleck, K., & Rodriguez, A. (1992). Influences of witnessed affect on information processing in children. *Child Development, 63*, 774–786.

Burke, A., Heuer, F., & Reisberg, D. (1992). Remembering emotional events. *Memory & Cognition, 20*, 277–290.

Ceci, S. J. (1991). Some overarching issues in the children's suggestibility debate. In J. Doris (Ed.), *The suggestibility of children's recollections: Implications for eyewitness testimony* (pp. 47–55). Washington, DC: American Psychological Association.

Christianson, S.-A. (1989). Flashbulb memories: Special, but not so special. *Memory and Cognition, 17*, 435–443.

Christianson, S.-A. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin, 112*, 284–309.

Christianson, S.-A., & Hubinette, B. (in press). Hands up! A study of witnesses' emotional reactions and memories associated with bank robberies. *Applied Cognitive Psychology*.

Christianson, S.-A., & Loftus, E. F. (1987). Memory for traumatic events. *Applied Cognitive Psychology, 1*, 225–239.

Christianson, S.-A., & Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Cognition & Emotion, 5*, 81–108.

Cutler, B., Penrod, S., & Martens, T. (1987). The reliability of eyewitness identification. *Law and Human Behavior, 11*, 233–258.

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66*, 183–201.

Fisher, R. P., & McCauley, M. R. (1995). Improving eyewitness testimony with the cognitive interview. In M. S. Zaragoza, J. R. Graham, G. C. N. Hall, R. Hirschman, & Y. S. Ben-Proath (Eds.), *Memory and testimony in the child witness. Applied Psychology: Individual, social, and community issues* (pp. 86–108). London: Sage Publications.

Goodman, G. S. (1991). Commentary: On stress and accuracy in research on children's testimony. In J. Doris (Ed.), *The suggestibility of children's recollections* (pp. 77–82). Washington, DC: American Psychological Association.

Goodman, G. S., Hirschman, J. E., Hepps, D., & Rudy, L. (1991). Children's memory for stressful events. *Merrill Palmer Quarterly, 37*, 109–158.

Goodman, G. S., Quas, J. A., Batterman-Faunce, J. M., Riddlesberger, Riddlesberger, M. M., & Kuhn, J. (1996). Predictors of accurate and inaccurate memories of traumatic events experienced in childhood. In K. Pezdek & W. P. Banks (Eds.), *The recovered memory/false memory debate* (pp. 3–28). San Diego, CA: Academic Press.

Goodman, G. S., Rudy, L., Bottoms, B. L., & Aman, C. (1989). Children's concerns and memory: Ecological issues in the study of children's eyewitness testimony. In R. Fivush & J. Hudson (Eds.), *What young children remember and why* (pp. 249–284). Cambridge, UK: Cambridge University Press.

Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & Cognition, 18,* 496–506.

Hosch, H. M., & Bothwell, R. K. (1990). Arousal, description, and identification accuracy of victims and bystanders. *Journal of Social Behavior and Personality, 5,* 481–488.

Howe, M. L., Courage, M. L., & Peterson, C. (1996). How can I remember when "I" wasn't there: Long-term retention of traumatic experiences and emergence of the cognitive self. In K. Pezdek & W. P. Banks (Eds.), *The recovered memory/false memory debate* (pp. 121–168). San Diego, CA: Academic Press.

Jones, D. C., Swift, D. J., & Johnson, M. A. (1988). Nondeliberate memory for a novel event among preschoolers. *Developmental Psychology, 24,* 641–645.

Kramer, T. H., Buckhout, R., & Eugenio, P. (1990). Weapon focus, arousal, and eyewitness memory: Attention must be paid. *Law and Human Behavior, 14,* 167–184.

Lindberg, M. A., Kiefer, J., & Thomas, S. W. (2000). Eyewitness testimony for physical abuse as a function of personal experience, development, and focus of study. *Journal of Applied Developmental Psychology, 21,* 555–591.

McCloskey, M., Wible, C.G., & Cohen, N. J. (1988). Is there a special flashbulb memory mechanism? *Journal of Experiential Psychology: General, 117,* 171–181.

Melton, G., (1992). Children as partners for justice: Next steps for developmentalists. *Monographs of the Society for Research in Child Development, 57*(5, Serial No. 229).

Merritt, K. A., Ornstein, P. A., & Spicker, B. (1994). Children's memory for a salient medical procedure: Implications for testimony. *Pediatrics, 94,* 17–23.

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

Ornstein, P. A., Shapiro, L. R., Clubb, P. A., Follmer, A., & Baker-Ward, L. (1997). The influence of prior knowledge on children's memory for salient medical experiences. In N. L. Stein, P. A. Ornstein, B. Tversky, & C. Brainerd (Eds.), *Memory for everyday and emotional events* (pp. 83–112). Mahwah, NJ: Erlbaum.

Peters, D. P. (1991). The influence of stress and arousal on the child witness. In J. L. Doris (Ed.), *The suggestibility of children's recollections* (pp. 60–76). Washington, DC: American Psychological Association.

Pynoos, R. (1992). *A model of post-traumatic stress disorder and memory in children.* Paper presented at the Developmental Psychobiology Research Group Conference, Estes Park, CO.

Quas, J. A., Goodman, G. S., Bidrose, S., Pipe, M., Craw, S., & Ablin, D. S. (1999). Emotion and memory: Children's long-term remembering, forgetting, and suggestibility. *Journal of Experimental Child Psychology, 72,* 235–270.

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., Goodman, G., Nicholas, G., & Moan, S. (1991). Children's memory for genital exam: Implications for child sexual abuse. *Journal of Consulting and Clinical Psychology, 59,* 682–691.

Steward, M. S., & Steward, D. S. (1996). Interviewing young children about body touch and handling. *Monographs of the Society for Research in Child Development, 61*(4–5).

Terr, L. (1983). Chowchilla revisited: The effects of psychic trauma four years after a schoolbus kidnapping. *American Journal of Psychiatry, 140*, 1543–1550.

Terr, L. (1996). True memories of childhood trauma: Flaws, absences, and returns. In K. Pezdek & W. P. Banks (Eds.), *The recovered memory/false memory debate* (pp. 69–80). San Diego, CA: Academic Press.

Tobey, A. E., & Goodman, G. S. (1992). Children's eyewitness memory: Effects of participation and forensic context. *Child Abuse and Neglect, 16*, 779–796.

Wagennaar, W. A., & Groeneweg, J. (1990). The memory of concentration camp survivors. *Applied Cognitive Psychology, 4*, 77–87.

Walker, E. L. (1958). Action decrement and its relation to learning. *Psychological Review, 65*, 129–142.

Warren, A. R., & Swartwood, J. N. (1992). Developmental issues in flashbulb memory research: Children recall the Challenger event. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: The problem of flashbulb memories* (pp. 95–120). New York: Cambridge University Press.

Winograd, E., & Killenger, W. A., Jr. (1983). Relating age at encoding in early childhood to adult recall: Development of flashbulb memories. *Journal of Experiential Psychology: General, 112*, 413–422.

Winograd E., & Neisser, U. (1992). *Affect and accuracy in recall: The problem of flashbulb memories* (pp. 95–120). New York: Cambridge University Press.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit formation. *Journal of Comparative Neurology and Psychology, 18*, 459–482.

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., & Tollstrup, 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. 202–216). Hillsdale, NJ: Erlbaum.

Yuille, J. C., & Wells, G. L. (1991). Concerns about the application of research findings: The issue of ecological validity. In J. Doris (Ed.), *The suggestibility of children's recollections: Implications for eyewitness testimony* (pp. 47–55). Washington, DC: American Psychological Association.

Received August 7, 2000