Anxious and Nonanxious Children's Face Identification

Ryan J. Fitzgerald[1], Heather L. Price[1], & Deborah A. Connolly[2]

[1]University of Regina

[2]Simon Fraser University

This is the peer reviewed version of the following article: Fitzgerald, R. J., Price, H. L., & Connolly, D. A. (2012). Anxious and nonanxious children's face identification. Applied Cognitive Psychology, 26, 585-593, which has been published in final form at http://onlinelibrary.wiley.com/doi/10.1002/acp.2833/full.

This research was generously supported by a Human Early Learning Partnership grant to the second and third authors, a Natural Sciences and Engineering Research Council PGS-D grant to the second author, and a Social Sciences and Humanities Research Council grant to the third author. The authors thank Heidi Gordon, Rachel Richards, Rebecca Roberts, Scott Currie, Caroline Greaves, Zina Lee, and Heath Mahoney for their assistance with data collection and coding. Finally, the authors also sincerely thank the instructors, parents, and children who participated in this research.

Correspondence concerning this article should be addressed to Ryan J. Fitzgerald, who is now at Department of Psychology, University of Portsmouth, King Henry Building, King Henry 1 Street, Portsmouth, United Kingdom, PO1 2DY. E-mail: ryan.fitzgerald@port.ac.uk

Abstract

The impact of stress on children's face identification is not well understood, partly because of the ethical and methodological challenges posed by this line of research. In the present research, such challenges were addressed by having 4-year-old and 5-year-old children (*n* = 80) participate in swimming lessons that were anxiety provoking for some, but not all, children. Information processing conditions were also manipulated by varying event frequency and retention interval. Children's identifications were examined after both a short (1.5–4 weeks) and long (1 year) delay. Anxiety was largely unrelated to the accuracy of children's swimming instructor identifications; however, after a long delay, anxiety had a negative effect on correct line-up rejections. In addition, the confidence–accuracy relation was influenced by the quality of the information processing conditions but only after a short delay. Implications for child witnesses are discussed.

Anxious and Nonanxious Children's Face Identification

On 5 June 2002, 14-year-old Elizabeth Smart was kidnapped from her Salt Lake City, Utah, bedroom in the middle of the night. The only witness was Elizabeth's 9-year-old sister, Mary Katherine, who pretended to lie asleep throughout the abduction. Although Mary Katherine initially reported that the kidnapper's voice sounded familiar, she did not get a good look at his face and was unable to identify the source of familiarity (Kone, 2003). The identity of the kidnapper remained elusive to Mary Katherine until 4 months after the incident, at which point she revealed to her parents, 'I think it might be Immanuel' (Smart, Smart, & Morton, 2003, p. 127). Her memory was triggered by a photograph of a muscular woman she saw in the Guinness Book of World Records. Five months after Mary Katherine's revelation, Elizabeth was found living in captivity with Brian Mitchell, a man who referred to himself as 'Immanuel' and had previously worked for the Smart family.

The photograph of a muscular woman may very well have provided Mary Katherine with a previously inaccessible retrieval cue, which in turn facilitated her recollection of the intruder's identity. Alternatively, the rules that govern memories formed under normal circumstances might not apply to memories of trauma (e.g. Christianson, 1992). Memories typically lose strength over time because of either decay or interference (Jonides et al., 2008), but the opposite was true for Mary Katherine. Mary Katherine's story demonstrates how memories that appear inaccessible shortly after a traumatic event may be retrievable at a later, less stressful time. Unfortunately, research on children's memory for people encountered under stress is rare, due in part to the logistical and ethical challenges associated with conducting such research. In the present research, we were able to study stress in children ethically by capitalizing on the naturally occurring stress that some children experience during their early encounters in a swimming pool. Children's memory of their swimming instructor was tested on two occasions: first, between 10 and 31 days after the event; then again, 1 year later. The children also provided a post-identification confidence assessment, enabling us to evaluate factors that influence the relation between confidence and accuracy.

Children's memory for stressful events

When studying the influence of stress on children's memory, researchers often design naturalistic studies to circumvent the ethical issues that arise when planning to induce stress in a laboratory. For example, children's memory of Hurricane Andrew has been examined at varying retention intervals (Bahrick, Parker, Fivush, & Levitt, 1998; Fivush, Sales, Goldberg, Parker, & Bahrick, 2004). Perhaps related to the notion that not all naturalistic designs involve a target face that can be later identified, most studies have tested children's recall of event details. The results of these recall studies have been mixed. Researchers have found positive effects of stress (e.g. Shrimpton, Oates, & Hayes, 1998), negative effects of stress (e.g. Brown et al., 1999), as well as no effects of stress (e.g. Ornstein, Gordon, & Larus, 1992). The limited number of studies that have investigated the influence of stress on children's identification accuracy have also produced mixed results.

In three studies, children's identification of their inoculators was assessed (Goodman, Hirschman, Hepps, & Rudy, 1991). In Study 1, children receiving inoculations were compared with a control group (who received a design on their arm). Although the control group was more accurate than the immunization group, Goodman et al. noted two limitations of their first study: (1) the difference was not statistically significant, probably because of the small sample size (n = 18) and (2) the two groups experienced qualitatively different events, which might have confounded the results. In Study 2, the sample consisted of a larger group of children (n = 48), all of whom received an immunization. Instead of including a control group, Goodman et al.

evaluated identification performance among different stress ratings for the same event. Once again, no significant differences concerning level of stress on identification performance were found; however, numerically, the low stress group was again more accurate than the high stress group. In Study 3, Goodman et al. matched a subset of participants from Study 2 to a group of control participants, who received a design on their arm, as with the control group in Study 1. In contrast with the first two experiments, the high stress group attained a correct identification rate that was numerically higher than the low stress group's correct identification rate; however, again, the difference between the two groups was not significant.

In the inoculation studies conducted by Goodman et al. (1991), the line-ups were always target present. Peters (1991), on the contrary, explored the effect of stress on children's identification from line-ups that did or did not include the target. In Experiment 1, children were given two identification tasks. The first task was to identify their dentist following a recent appointment (high stress condition). The second task was to identify the research assistant who administered the dentist identification task (low stress condition). Although stress had a negative effect on correct identifications from target-present line-ups, accuracy on target-absent line-ups was unaffected by stress. In Experiment 2, children had their pulse taken by a stranger. The interaction was videotaped, and the children's anxiety was rated based on behavioral observation methods. The results were similar to Experiment 1: anxiety had a negative effect on correct identifications but had no effect on target-absent line-ups. In Experiment 3, children who received an inoculation were contrasted with children who listened to a talk about inoculations (the same nurse served as the target for both groups). Again, children in the low stress condition identified their target more successfully than children in the high stress condition. However, like the first two experiments, no effect of stress was observed for target-absent line-ups. In

Experiment 4, children were left alone in a room with a box containing money. In the low anxiety condition, the children were told that a research assistant would be coming to collect the box. In the high anxiety condition, a research assistant visited the room and pretended that he was stealing the box. This manipulation had no effect on identification accuracy, regardless of whether the target was in the line-up or not. Thus, across experiments, stress had no effect on target absent and only sometimes affected target-present lineups.

In two subsequent experiments, Peters (1997) induced stress by setting off an unexpected fire alarm. In Experiment 1, while the alarm rang, a confederate entered the room and announced that there might be a fire to some of the children (high stress condition). Other children were exposed to a loud radio, at which point a confederate entered the room and asked to look out the window for a delivery truck (low stress condition). This time, stress had a negative effect on accuracy for both target-present and target-absent line-ups. In Experiment 2, Peters examined the interaction between retention interval and stress. This time, both groups of children were exposed to a fire alarm. Stress was manipulated by only providing advance warning of the alarm to some of the children. When children made identifications after a 15-minute delay, the results of Experiment 1 were replicated (i.e. a negative effect of stress); however, there was no difference in identification accuracy after a 6-month delay.

According to the most recent meta-analysis (Deffenbacher, Bornstein, Penrod, & McGorty, 2004), children’s identification accuracy when under low stress (55%) was better than when under high stress (42%). However, research on children’s ability to identify individuals encountered under stressful conditions is still in its early stages. In the meta-analysis, only nine studies included child samples, and nonsignificant results were found in four of those nine studies. Moreover, the literature in this area contains several inconsistent findings that could relate to the practice of comparing groups that experienced different events (e.g. Goodman et al., 1991, Study 1). A problem with such designs is that because of the qualitative differences in experiences between the two groups, certain event characteristics other than stress might have influenced the results (e.g. children who listen to an immunization talk might look at the target for a longer duration than children who receive an immunization). Although previous researchers have strived to make the treatment and the control groups' experiences as comparable as possible, the fact that they were not the same is inherently problematic.

In the present research, all children participated in the same event: private swimming lessons. Because of the natural variation in anxiety invoked during children's early experiences in the water, we were able to compare identification accuracy between anxious and nonanxious children who experienced the same event. Another novel aspect of the present research is that most children also made a second identification. This is a topic of considerable importance, given that it is not uncommon for an eyewitness to make identifications on more than one occasion in actual criminal investigations (Behrman & Davey, 2001). Moreover, at least three DNA exonerations have involved false identifications made by witnesses who participated in more than one identification test (Godfrey & Clark, 2010). Although repeated identifications have rarely been studied in children, there has been one report of an interaction between stress and delay on the accuracy of follow-up line-up identifications (Lindberg, Jones, Collard, & Thomas, 2001). In that study, identification accuracy was similar between the inoculation and control groups at immediate testing; however, the inoculation group outperformed the control group when they were retested 1 month later. In the present research, we tested children's identification accuracy on two occasions, separated by a 1-year delay. Thus, we were able to examine whether the benefits of stress observed previously would also be found after a much longer delay.

Confidence and Accuracy

Although empirical studies suggest a modest relation between eyewitness confidence and accuracy (Bothwell, Deffenbacher, & Brigham, 1987; Sporer, Penrod, Read, & Cutler, 1995), juries nonetheless consider confidence to be one of the most influential factors when determining eyewitness credibility (Penrod & Cutler, 1995). As such, understanding the factors that influence confidence–accuracy (CA) relations has obvious forensic implications. The optimality hypothesis (Deffenbacher, 1980) posits that the CA relation is largely determined by the conditions under which an event is encoded, stored, and retrieved (i.e. the information processing conditions). According to Deffenbacher, when information processing conditions are good, the CA relation should be strong. When information processing conditions are poor, the CA relation should be correspondingly weak. Bothwell et al. (1987) found support for the optimality hypothesis in a meta-analysis that showed the CA relation was strengthened when target exposure duration (i.e. an encoding condition) increased. However, the optimality hypothesis was both formulated and supported using research on adult samples. It might not apply to children, whose metacognitive abilities have not yet fully developed (Roebers, 2002).

Age differences in the CA relation for face identifications have been found on numerous occasions. Keast, Brewer, and Wells (2007) compared the CA relation in children (M_{age} = 12 years) and adults (M_{age} = 25 years) using an 11-point confidence scale (ranging from 0% to 100%) and concluded that the children’s confidence ratings were rarely diagnostic of identification accuracy. For example, only 30% of children who rated their confidence above 90% were accurate. Compared with adults, children were both less accurate and more overconfident. Adolescents have also been shown to have more advanced metacognitive abilities than children. Although Brewer and Day (2005) found comparable CA relations in children (M_{age} = 10 years) and adolescents (M_{age} = 16 years), the proportion of correct identifications within each level of a five-point scale (ranging from ‘really unsure’ to ‘really sure’) was substantially lower in children than in adolescents. For example, correct identifications were made by over 90% of adolescents who were really sure. In contrast, correct identifications were only made by 50% of children who were really sure. Brewer and Day concluded that in comparison with their adolescent counterparts, children were far more overconfident.

Children show a clear propensity for making overly confident identifications, and the factors that influence their confidence ratings are only beginning to be understood. When recall has been tested, question format has emerged as an important factor. Although children are able to differentiate confidence judgments between correct and incorrect answers effectively when responding to unbiased questions, they show no such differentiation when responding to misleading questions (Roebers, 2002; Roebers & Howie, 2003). The ability to differentiate in response to misleading questions appears to progress with age, as 10-year-old children performed better than 8-year-old children (Roebers, 2002; Roebers & Howie, 2003). Given the developmental trends observed in their study, Roebers and Howie (2003) speculated that children younger than 8 years old might be most vulnerable to misleading questions. Additional evidence of a developmental progression in metacognition has been found when children have been put under the influence of social pressure. Schwarz and Roebers (2006) had 8-year-old and 10-year-old children answer questions about a video after hearing the answers of an assertive confederate who ostensibly viewed the same video. They also had control groups answer the questions in the absence of a confederate. Although both age groups were able to differentiate their confidence between correct and incorrect answers accurately in the control condition, only the 10-year-olds were able to do so when under the confederate's influence. In contrast, the 8-year-olds appeared to have been swayed by the confederate's assertiveness, as they showed no difference in confidence as a function of accuracy.

Although researchers have identified some of the factors that influence children's ability to assign confidence ratings that correspond with accuracy, little is known about how information processing conditions might influence the CA relation in children and, in particular, very young children. Support for the optimality hypothesis has been found in adults, but it has never been tested in children. Therefore, in addition to investigating the effect of stress on identification, the present research examined whether or not the optimality hypothesis describes the CA relation in preschool-age children.

In the present study, processing conditions included two components: contact frequency and retention interval. Retention interval is one of the numerous information processing conditions hypothesized by Deffenbacher (1980) to have an influence on the CA relation. Although frequency was not specifically mentioned by Deffenbacher, it is highly related to exposure duration, an encoding condition with a demonstrated influence on the CA relation (Bothwell et al., 1987). Our manipulation differed from previous manipulations in that for children exposed to a longer duration, exposure was divided across several sessions. Specifically, some children participated in one lesson (25 minutes of exposure), and other children participated in four lessons (100 minutes of exposure) over the course of 2 weeks. Retention interval was manipulated by administering the identification task 4 weeks after the first lesson for all participants. Accordingly, children who experienced four lessons completed their first identification one and one-half weeks after their last exposure while children who had experienced a single swimming lesson completed the identification 4 weeks after their last exposure. Thus, compared with poor processing conditions, good processing conditions consisted of more exposure to the target and less time between the event and the identification. If the optimality hypothesis applies to children, then good processing conditions should strengthen the CA relation.

Method

Participants

Children aged 4 and 5 years old were recruited through advertisements in newspapers, schools, and daycare facilities. The initial sample was composed of 80 participants (M_{age} = 58.4 months, *SD* = 7.9; *n* = 45 boys). We repeatedly attempted to contact all children 1 year later. Fifty-eight children were located and agreed to participate in the follow-up study (M_{age} = 59.0 months, *SD* = 8.3 at the time of the original swimming lesson; *n* = 34 boys).

Physiological measures of anxiety were inappropriate because children engaged in physical activity during the lessons. Thus, we relied on subjective ratings to evaluate children's anxiety. Immediately following the first (or only) lesson, parents and instructors completed a scale that ranged from not at all anxious (1) to extremely anxious (9). Children indicated their anxiety via the Koala Fear Questionnaire (Muris, 2002; Muris et al., 2003). The questionnaire was administered by instructing children to indicate (retrospectively) how they felt before and during the lesson by pointing to a picture of one of three koala faces exhibiting varying levels of anxiety. The Koala Fear Questionnaire has been shown to have good internal consistency (α = .90), as well as convergent validity, in a sample of 4-year-olds to 6-year-olds (Muris, 2002). Within the same age group, it has also been shown to have good test–retest reliability [intraclass correlations (*ICC*) = .92] and correlate with an 'anxiety interview' (Muris et al., 2003).

Two independent raters coded video recordings of the lessons for six behaviors. Three behaviors were associated with comfort: engagement behaviors, laughter/smiling, and brave activities. The remaining three behaviors were associated with anxiety: physical avoidance, clinging, and resistance/ refusal to participate in activities. Interrater reliability (ICC_1) on 16% of the sample ranged from ‘good’ to ‘excellent’ (Cicchetti & Sparrow, 1981) on all six behaviors coded (.77 for engagement, .82 for laughter, .88 for brave activities, .71 for avoidance, .80 for clinging, and .84 for resistance). The raters also evaluated the child’s overall anxiety using the same scale that parents and instructors used.

Because of ‘excellent’ (.83; Cicchetti & Sparrow, 1981) interrater reliability (ICC_2) between the instructor and coder evaluations of anxiety and the hands-on nature of the instructor– child interaction, we relied exclusively on instructor evaluations to determine children’s placement into an anxiety condition. Children’s ratings of their own anxiety and parental ratings of children’s anxiety were excluded based on the limited range of children’s own assessments (80% of children selected ‘not at all anxious’ even if crying and refusing to get into the water) and a lack of attention to their child by many parents during the swimming lesson. Although we measured children’s anxiety on a continuous scale, behavioral coding indicated children belonged to one of three distinct groups: nonanxious, anxious, and mixed anxiety. Children who received a rating of 1 or 2 of 9 from their instructor were placed into the ‘nonanxious’ condition (n = 41). Children who received a 4 or higher were placed into the ‘anxious’ condition (n = 39). Children who received a ‘3’ (n = 14) were both anxious and nonanxious at times (e.g. crying before the lesson and smiling by the end). Although it would have been interesting to test hypotheses about the children with mixed anxiety, there were too few of these children to do so effectively. As such, data from the children with mixed anxiety are not reported.

To assess for any differences in personality between the two anxiety groups, the participant's parents completed the Strengths and Difficulties Questionnaire (Goodman, 1997) and the Children's Behavior Questionnaire—Very Short Form (Putnam & Rothbart, 2006). Both measures have demonstrated good internal consistency and validity (Putnam & Rothbart, 2006; Stone, Otten, Engels, Vermulst, & Janssens, 2010). Anxious and nonanxious children did not significantly differ on any of the dimensions within either of the personality measures. The only difference we observed between the two groups was their fear of the water.

Procedure and Materials

Children participated in either one (n = 42) or four (n = 38) swimming lessons, which all took place at the same facility. The one-on-one lessons were carried out in 25 minutes and involved a scripted set of activities. Parents sat in a nearby viewing area. The instructor was one of five women qualified to provide swimming lessons by the Red Cross. Children who received four lessons had the same instructor for each lesson and lessons took place twice per week for 2 weeks. Four weeks following the first or only lesson, an interviewer first questioned the children about lesson activities[1] and then administered the line-up identification tasks. Children had no advanced warning about the identification task.

All line-up members wore a bathing cap similar to the one always worn by the instructors during the lessons. Ten lineups were constructed: two for each instructor (one target absent and one target present). Although the line-ups were not pilot tested, nontarget line-up members were selected based on similarity to the target's appearance (gender, approximate age, and race were

[1] For complete results from the interview, please see Price and Connolly (2007).

all the same, hair was obscured by the bathing cap). Each array included six color photographs, approximately 400 × 400 in size, presented simultaneously. One limitation of using a simultaneous line-up is that when this procedure is used, children tend to be poor at rejecting target-absent line-ups (Pozzulo & Lindsay, 1998). Although alternative line-ups, such as the elimination procedure (Pozzulo & Balfour, 2006; Pozzulo, Dempsey, & Crescini, 2009; Pozzulo & Lindsay, 1999) and the wildcard technique (Zajac & Karageorge, 2009), have been shown to improve children's ability to reject target-absent lineups, we chose to employ a simultaneous line-up because it is the procedure most commonly used by law enforcement (Wells & Olson, 2003). In an attempt to minimize children's tendency to choose from target-absent line-ups, the interviewer instructed all children to 'look carefully at each of these pictures and tell me if one of these pictures is of your swimming teacher. You do not have to choose one; your teacher's picture may not be here.' Such instructions have been found to increase children's correct rejection rate (Pozzulo & Dempsey, 2006).

After the identification task, the interviewer asked the children to provide an indication of confidence. Because of the children's relative youth, we were concerned about their ability to understand how to rate their confidence on a Likert scale, as has been used in previous research with older samples (e.g. Brewer & Day, 2005). As an alternative, we asked the children to indicate their confidence by choosing from one of three pictures: thumbs down ('not sure'), thumbs sideways ('sort of sure'), or thumbs up ('really sure'). These responses were converted onto a scale of 1 to 3 and treated as a continuous variable.

Design

A 2 (processing conditions: good, poor) × 2 (line-up type: target absent, target present) × 2 (anxiety: high, low) × 2 (delay: weeks, years) quasi-experimental design was employed. To manipulate processing conditions, children were randomly assigned to participate in either one or four lessons. All children made identifications 4 weeks after the first exposure, so the time elapsed between the last lesson and the identification task was longer under poor processing conditions (4 weeks) than under good processing conditions (1.5 weeks). As a line-up type manipulation, children were randomly assigned to receive line-ups that did or did not include their instructor. Instructor evaluations, which correlated with behavioral coding, were used to divide the children into anxiety groups. One year after the first identification, a subsample of children made a follow-up identification (delay). Thus, processing conditions, line-up type, and anxiety were between-subjects variables, and delay was a within-subjects variable.

Results

First identification (delay of 1.5–4 weeks)

Most children made accurate identification choices. For target-present line-ups, the rate of correct identifications (65.8%) exceeded the rate of incorrect decisions[2] (31.6%), $z = 3.13$, $p = .001$. Of the 12 incorrect decisions, 10 were incorrect rejections and 2 were foil identifications. For target-absent line-ups, the correct rejection rate (54.8%) was higher than the foil identification rate (40.5%), but the difference was not significant, $z = 1.28$, $p = .20$. In addition, ‘not sure’ responses were made on one target-present lineup and on two target-absent line-ups.

A 2 (identification decision: correct versus incorrect) × 2 (line-up type: target present versus target absent) between-subjects analysis of variance (ANOVA) was performed with confidence as the dependent variable. This test revealed a significant interaction, $F(1,73) = 7.79$, $p = .007$, $\eta_p^2 = 0.13$, indicating that the presence or absence of the target influenced whether or not confidence was diagnostic of accuracy (Figure 1). On target-present line-ups, a positive CA

[2] Incorrect decisions on target-present lineups included both incorrect identification choices and incorrect lineup rejections.

relation was observed. That is, correct identifications were made more confidently than were incorrect decisions, $t(35) = 3.77$, $p = .001$, $r = .54$. In contrast, on target-absent line-ups, confidence had little relation to accuracy, $t(38) = 0.62$, $p = .54$, $r = -.10$.

Anxiety. Because of the relatively small sample size, Fisher's exact test was used for all comparisons between anxiety groups (Rubin, 2007). Anxiety was unrelated to decision accuracy (Table 1). This was true both for target-present line-ups, $p = 1.00$, phi = -0.02, and for target-absent line-ups, $p = .75$, phi = 0.08. Even when target-present and target-absent line-ups were collapsed to increase statistical power, the absence of a relation between anxiety and accuracy remained, $p = 1.00$, phi = 0.02.

Processing conditions. Fisher's exact test indicated processing conditions had no relation to correct identifications, $p = .72$, phi = 0.10. On target-absent line-ups, children who had poor processing conditions were more accurate than children who had good processing conditions; however, the difference was not significant, $p = .11$, phi = -0.30. To test the optimality hypothesis, we evaluated the influence of processing conditions on the CA relation (Table 2). For target-absent line-ups, the Pearson correlation coefficients between poor and good processing conditions did not significantly differ, $z = 1.02$, $p = .31$. Thus, the target-absent data provided no support for the optimality hypothesis. For target-present line-ups, however, the CA relation under good processing conditions was significantly stronger than the CA relation under the poor processing conditions, $z = 2.44$, $p = .01$.

Follow-up identification (delay of 1 year)

Identification accuracy was influenced by whether the target was present or absent. On target-present line-ups, correct identifications (24.1%) were less common than incorrect decisions (72.4%), $z = 4.13$, $p < .001$. The 21 incorrect decisions consisted of 17 incorrect rejections and 4 foil identifications. For target-absent line-ups, the correct rejection rate (62.1%) was higher than the foil identification rate (34.4%), $z = 2.16$, $p = .03$. In addition, two children made ‘not sure’ responses, one for each line-up type.

A 2 (identification decision: correct versus incorrect) × 2 (line-up type: target present versus target-absent) between-subjects ANOVA on identification confidence revealed no effect of identification decision, $F(1,52) = 0.00$, $p = .96$, $\eta_p^2 = 0.00$, no effect of line-up type, $F(1,52) = 1.85$, $p = .18$, $\eta_p^2 = 0.03$, and no interaction, $F(1,52) = 2.41$, $p = .13$, $\eta_p^2 = 0.04$. The CA relation was positive for target-present line-ups, but the difference in confidence between correct ($M = 2.43$, $SD = 0.53$) and incorrect ($M = 2.10$, $SD = 0.83$) decisions was not significant, $t(26) = 0.99$, $p = .33$, $r = .19$. The CA relation was negative for target-absent line-ups, but again, confidence on correct ($M = 2.38$, $SD = 0.70$) and incorrect ($M = 2.70$, $SD = 0.48$) decisions did not significantly differ, $t(26) = 1.25$, $p = .22$, $r = -.24$.

Anxiety. Line-up type influenced the relation between anxiety and identification accuracy (Table 1). On target-present line-ups, anxiety had no relation to identification accuracy, $p = 1.00$, phi = 0.04. By contrast, a significant association was observed for target-absent line-ups, $p = .046$, phi = 0.45. Specifically, anxious children were more apt to make foil identifications than correct rejections, whereas nonanxious children made more correct rejections than foil identifications. Thus, anxiety had a negative effect on the accuracy of target-absent line-up decisions.

Processing conditions. On target-present line-ups, processing conditions had no relation to identification accuracy, $p = 1.00$, phi = 0.00. On target-absent line-ups, more correct rejections were made under poor processing conditions than under good processing conditions, but the difference was not significant, $p = .13$, phi = -0.32. There also was no significant difference in

CA relations between the poor and good processing conditions, regardless of whether the target was present, $z = 0.31$, $p = .76$, or absent, $z = 0.37$, $p = .71$.

Effects of delay

The exact binomial calculation of McNemar’s test was used to test the significance of changes in accuracy rates between the first and follow-up identifications (Rosner, 2006). This test indicated accuracy shifts differed according to whether the target was present or absent from the line-up. On target-present line-ups, the correct identification rate declined from 62% at the first identification to 24% at the follow-up, $p = .001$. On target-absent line-ups, the correct rejection rates on the first (59%) and follow-up (62%) identifications were similar, $p = 1.00$.

Steiger’s (1980) test for correlated correlations was used to assess the difference between CA relations at the first and follow-up identifications. On target-present line-ups, the CA relation at the first identification ($r = .42$, $p = .03$) was stronger than the CA relation at the follow-up identification ($r = .19$, $p = .33$), but the difference was not significant, $z = 0.94$, $p = .35$. On target-absent line-ups, the CA relations at the first ($r = -.18$, $p = .37$) and follow-up ($r = -.24$, $p = .32$) identifications were similar, $z = 0.26$, $p = .79$.

Anxiety. In anxious children, the correct identification rate declined from 73% at the first identification to 27% at the follow-up, $p = .06$. A similar trend was found in nonanxious children, whose correct identification rate declined from 56% to 22% across the first and follow-up identifications, $p = .03$. For anxious children, the correct rejection rate declined from 60% at the first identification to 53% at the follow-up, $p = .25$. In contrast, for nonanxious children, the correct rejection rate improved from 57% to 86%, $p = .22$.

Discussion

For both target-present and target-absent line-ups, most children were accurate at the first identification (1.5–4 weeks of delay). At the follow-up identification (1-year delay), very few children made correct identifications. In contrast, children were actually quite adept at rejecting target-absent line-ups. Given that young children are typically poor at rejecting target-absent line-ups (e.g. Pozzulo & Lindsay, 1998), the high correct rejection rate was somewhat unexpected. However, the children were also prone to rejecting target-present line-ups. Thus, the delay appears to have led the children to adopt a conservative response bias, which of course was beneficial for target-absent line-ups and detrimental for target-present line-ups.

Anxiety

For the most part, anxiety was unrelated to the accuracy of children's identifications. When no significant differences between stress conditions were found in previous investigations, statistical power was identified as an issue (Goodman et al., 1991; Peters, 1991). Thus, it is important to consider the power of our study to detect differences, if they had existed. Given the size of our initial sample (n = 80) and an α rate of .05, we had a 77% chance of rejecting the null hypothesis if there had been a medium effect size (w = 0.30; Cohen, 1988). Thus, there was sufficient power to detect a difference if there had been a medium-sized effect of stress.

The only reliable association between accuracy and anxiety was at the follow-up identification, when anxiety had a negative effect on target-absent line-ups. Many studies in this domain have included only target-present line-ups, so there is a paucity of data available for comparison with this finding. When target presence has been manipulated, the results have been mixed. In his first round of studies, Peters (1991) found stress effects only when the target was present. In subsequent research, however, Peters found negative effects of stress on both line-up types after a short delay, but the effects were absent after a long delay. These findings contrast with our own on two accounts: (1) we found no effects of anxiety when the target was present and (2) the only reliable effect of stress we found was at the long delay; accuracy among anxious and nonanxious children was highly similar at the short delay.

There are some key differences between the present research and previous research that could explain the discrepant findings. For example, in previous investigations, the degree of interaction between the children and the target typically consisted of activities that were relatively passively experienced and short in duration. In contrast, children's participation in the present research was highly interactive and relatively long in duration. Even the children who had only one lesson actively engaged in over a dozen activities with the instructor over a 25-minute period. Thus, the present research might not generalize to line-up identifications that follow brief interactions, which might be the more prototypical eyewitness scenario; however, it may be more applicable to specific crimes that involve longer interactions (e.g. child abductions). Perhaps the most crucial departure from previous research is that all children in the present research experienced the same event. In many of the previous investigations, children who differed in stress also experienced slightly different events. The approach taken by previous researchers might correspond with some real-world scenarios, where different events induce different levels of stress. However, in terms of isolating the impact of stress, comparisons between different events is inherently problematic. When stress differences are confounded with event differences, it is difficult to ascertain whether the outcome was influenced by stress, event characteristics, or a combination of the two. In the present research, when anxious and nonanxious children experienced the same event, identification accuracy was generally unaffected by stress.

Processing Conditions

According to Deffenbacher's (1980) optimality hypothesis, the strength of the CA relation will depend upon the quality of the information processing conditions experienced by the witness. We found support for this hypothesis on target-present line-ups at the first identification, when the CA relation was stronger under good compared with poor processing conditions. However, processing conditions had no influence on the CA relation when the target was absent. One consideration is that the correct rejection rate was higher in children who had poor processing conditions relative to those who had good processing conditions. In other words, poor processing conditions had a positive impact on accuracy in the target-absent condition. At first glance, this finding may seem counterintuitive. Presumably, it should be good processing conditions that have a positive impact on accuracy. However, this might not always be the case for target-absent line-ups. If poor processing conditions lead to a poor memory trace, witnesses may be more likely to reject a line-up because of their feeling of a weaker memory. Thus, the optimality hypothesis might only apply to children when a correct identification is possible, as occurs in target-present line-ups only.

Processing conditions also had no influence on the CA relation at the follow-up identification; however, it could be argued that such a long delay made the processing conditions poor for all children. That is, the 1-year delay might have negated any benefit that having four lessons had on the strength of the CA relation. In support of this interpretation, the CA relation for target-present line-ups was weaker at the follow-up identification ($r = .19$) compared with first identification ($r = .42$); however, the difference between these correlations was not significant.

Effects of Delay

Children who correctly identified their instructor at the first identification were frequently incorrect at the follow-up identification, a trend that was uninfluenced by anxiety. In previous research on repeated testing of stressful memories (Lindberg et al., 2001), stress had a positive effect on follow-up identifications. Specifically, children in the inoculation group who correctly identified their nurse after 20 minutes almost always correctly identified the nurse again 1 month later. Of course, we employed a much longer delay between line-ups. If stress enhances retention, the effect might be limited to shorter retention durations.

To our knowledge, no previous investigation has examined the effect of such a long retention interval on children's face identification accuracy; however, children's recall memory has been tested after comparable delays. For example, Flin, Boon, Knox, and Bull (1992) tested 6-yearold and 9-year-old's recall of an argument about a staged accident 1 day after the event and then again after 5 months. Both groups showed a decline in memory performance over time, with a more pronounced effect of delay in 6-year-olds. A similar decline in recall performance was observed by Ornstein et al. (2006), who found 4-year-olds to 7-year-olds' memory of a pediatric examination was far worse after 6 months than it was immediately after the examination. The results have differed, however, when retention of stressful events has been tested after even longer delays. For example, Peterson (1999) tested 2-year-old to 13-year-old children's recall of medical emergencies and found remarkable retention after a 2-year delay (see Burgwyn-Bailes, Baker-Ward, Gordon, & Ornstein, 2001, for similar results after a 1-year delay). Although children's accuracy in the present research rarely improved over time, many children who were accurate after the short delay were also accurate after the long delay.

Limitations

There are several limitations of the present research that warrant consideration. For example, the processing condition manipulation included two components (frequency and retention interval) that were confounded with each other. Thus, it is unclear whether the observed effects were a consequence of both frequency and retention interval working together, or one of these processing conditions making an independent contribution. Had we chosen to manipulate only one processing condition, a more parsimonious interpretation of the results could have been made. However, when comparing identifications as a function of exposure frequency, one must have a consistent delay between either initial exposure and identification or final exposure and identification. This study was part of a larger study, and with all variables considered, it made more methodological sense to maintain the consistent delay between initial exposure and identification. Another limitation is that the line-ups were not pilot tested. Although the study data provided no indication of bias toward any particular line-up foil,[3] we had no *a priori* knowledge of whether or not the line-ups were biased. Finally, although our design provided a novel method of isolating the effects of anxiety on memory, it also precluded random assignment of participants into the anxiety conditions. There were no differences between the two groups on any of the personality traits that were measured, but the groups might have differed on some unmeasured characteristic. For example, we did not assess the children's trait anxiety. As such, the extent to which the observed anxiety was induced by the swimming lessons is unclear.

Conclusion

Mary Katherine Smart was unable to remember the identity of her sister's kidnapper until 5 months after the event took place, suggesting memory might be impaired immediately after

[3] Of the 16 foil identifications made from target-absent lineups, specific foils were chosen 0–3 times. On target-present lineups, with the exception of two foil identifications, the target was chosen on all positive identifications.

experiencing a stressful event and then resume its normal function after an extended period. Although our data are not consistent with this notion, the stress endured by Mary Katherine could have had a different impact than the stress experienced by the participants in empirical investigations. For example, in the present research, the swimming lessons were undoubtedly anxiety provoking for some children. However, the children's anxiety levels had probably returned to normal within hours of their participation. Mary Katherine, on the contrary, was constantly reminded of the terror she experienced by the notable absence of her sister. Another point to consider is that Mary Katherine probably spent those 5 months before she remembered the kidnapper's identity actively trying to do so. She knew that if she could remember his identity, it could help reunite her with her sister. In contrast, children in the present investigation had no incentive to motivate them to actively try to remember the identity of their swimming instructor. Future research could investigate how different durations of anxiety influence memory. Another possible research question could address the influence of incentives on long-term memory. If there is a sizable reward for remembering a stressful event, will memory be better at a long delay than at a short delay? Such research could shed more light on the curious case of Mary Katherine's spontaneous remembrance.

References

Bahrick, L. E., Parker, J. F., Fivush, R., & Levitt, M. (1998). The effects of stress on young children's memory for a natural disaster. *Journal of Experimental Psychology: Applied, 4,* 308-331.

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior, 25*, 475-491.

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72*, 691-695.

Brewer, N., & Day, K. (2005). The confidence-accuracy and decision-latency relationships in children's eyewitness identification. *Psychiatry, Psychology, and Law*, *12*, 119-128.

Brown, D. A., Salmon, K., Pipe, M. E., Rutter, M., Craw, S., & Taylor, B. (1999). Children's recall of medical experiences: The impact of stress. *Child Abuse & Neglect, 23*, 209-216.

Burgwyn-Bailes, E., Baker-Ward, L., Gordon, B. N., & Ornstein, P. A. (2001). Children's memory for emergency medical treatment after one year: The impact of individual difference variables on recall and suggestibility. *Applied Cognitive Psychology, 15*, S25-S48. DIO:10.1002/acp.833

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

Cicchetti, D. V., & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: Applications to assessment of adaptive behavior. *American Journal of Mental Deficiency, 86*, 127-137.

Cohen, J. W. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum Associates.

Deffenbacher, K. A. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior, 4*, 243-260.

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28*, 687-706. DOI: 10.1007/s10979-004-0565-x

Fivush, R., Sales, J. M., Goldberg, A., Parker, J., & Bahrick, L. E. (2004). Weathering the storm: Children's long term recall of Hurricane Andrew. *Memory, 12,* 105-114. DOI: 10.1080/09658210244000397

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

Godfrey, R. D., & Clark, S. E. (2010). Repeated eyewitness identification procedures: Memory, decision making, and probative value. *Law & Human Behavior, 34,* 241-258. DOI: 10.1007/s10979-009-9187-7

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

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry, 38*, 581-586. DOI: 10.1111/j.1469-7610.1997.tb01545.x

Jonides, J., Lewis, R. L., Nee, D. E., Lustig, C. A., Berman, M. G., & Moore, K. S. (2008). The mind and brain of short-term memory. *Annual Review of Psychology, 59*, 193-224. DOI: 10.1146/annurev.psych.59.103006.093615

Keast, A., Brewer, N., & Wells, G. L. (2007). Children's metacognitive judgments in an eyewitness identification task. *Journal of Experimental Child Psychology, 97,* 286-314.

DOI: 10.1016/j/jecp.2007.01.007

Kone, D. (2003, March 12). Elizabeth's road home. *48 Hours*. Retrieved from http://www.cbsnews.com

Lindberg, M. A., Jones, S., Collard, L. M., & Thomas, S. W. (2001). Similarities and differences in eyewitness testimonies of children who directly versus vicariously experience stress. *The Journal of Genetic Psychology, 162*, 314-333.

Muris, P. (2002). The Koala Fear Questionnaire: Its relationship with fear of storms and hurricanes in 4- to 14-year-old Antillean children. *Journal of Psychopathology and Behavioral Assessment, 24,*145-150.

Muris, P., Meesters, C., Mayer, B., Bogie, N., Luijten, M., Geebelen, E., . . . Smit, C. (2003). The Koala Fear Questionnaire: A standardized self-report scale for assessing fears and fearfulness in pre-school and primary school children. *Behaviour Research and Therapy, 41,* 597-617. DOI: 10.1016/S0005-7967(02)00098-0

Ornstein, P. A., Baker-Ward, L., Gordon, B. N., Pelphrey, K. A., Tyler, C. S., & Gramzow, E. (2006). The influence of prior knowledge and repeated questioning on children's long-term retention of the details of a pediatric examination. *Developmental Psychology*, *42,* 332-344. DOI: 10.1037/0012-1649.42.2.332

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

Penrod, S., & Cutler, B. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law, 1*, 817-845.

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

Peters, D. P. (1997). Stress, arousal, and children's eyewitness memory. In N. L. Stein, P. A. Ornstein, B. Tversky, & C. J. Brainerd (Eds.), *Memory for everyday and emotional events* (pp. 351-370). Mahwah, NJ: Erlbaum.

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

Pozzulo, J. D., & Balfour, J. (2006). Children's and adults' eyewitness identification accuracy when a culprit changes his appearance: Comparing simultaneous and elimination lineup procedures. *Legal and Criminological Psychology, 11*, 25-34. DOI: 10.1348/135532505X52626

Pozzulo, J. D., & Dempsey, J. (2006). Biased lineup instructions: Examining the effect of pressure on children's and adults' eyewitness identification accuracy. *Journal of Applied Social Psychology, 36*, 1381-1394.

Pozzulo, J. D., Dempsey, J., & Crescini, C. (2009). Preschoolers' person description and identification accuracy: A comparison of the simultaneous and elimination procedures. *Journal of Applied Developmental Psychology, 30*, 667-676. DOI: 10.1016/j.appdev.2009.01.004

Pozzulo, J. D., & Lindsay, R. C. L. (1998). Identification accuracy of children versus adults: A meta-analysis. *Law and Human Behavior*, *22*, 549-570.

Pozzulo, J. D., & Lindsay, R. C. L. (1999). Elimination lineups: An improved identification procedure for eyewitnesses. *Journal of Applied Psychology, 84*, 167-176.

Price, H. L., & Connolly, D. A. (2007). Anxious and nonanxious children's recall of a repeated or unique event. *Journal of Experimental Child Psychology, 98,* 94-112. DOI: 10.1016/j.jecp.2007.05.002

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment, 87*, 103-113. DOI: 10.1207/s15327752jpa8701_09

Roebers, C. M. (2002). Confidence judgments in children's and adults' event recall and suggestibility. *Developmental Psychology, 38,* 1052-1067. DOI: 10.1037//0012-1649.38.6.1052

Roebers, C. M., & Howie, P. (2003). Confidence judgments in event recall: Developmental progression in the impact of question format. *Journal of Experimental Child Psychology, 85,* 352-371. DOI:10.1016/S0022-0965(03)00076-6

Rosner, B. A. (2006). *Fundamentals of biostatistics* (6th ed.). Belmont, CA: Duxbury Press.

Rubin, A. (2007). *Statistics for evidence-based practice and evaluation* (1st ed.). Belmont, CA: Thompson Brooks/Cole.

Schwarz, S., & Roebers, C. M. (2006). Age differences in the effects of social influence on children's eyewitness performance and their metacognitive monitoring. *Journal of Experimental Child Psychology, 94,* 229-248. DOI:10.1016/j.jecp.2006.01.003

Shrimpton, S., Oates, K., & Hayes, S. (1998). Children's memory of events: Effects of stress, age, time delay and location of interview. *Applied Cognitive Psychology, 12,* 133-143.

Smart, E., Smart, L., & Morton, L. (2003). *Bringing Elizabeth home: A story of faith and hope*. New York: Doubleday.

Sporer, S. L., Penrod, S., Read, D., & Cutler, B. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence accuracy relation in eyewitness studies. *Psychological*

Bulletin, 118, 315-327.

Steiger, J.H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87,* 245-251.

Stone, L. L., Otten, R., Engels, R. C. M. E., Vermulst, A. A., & Janssens, J. M. A. M. (2010). Psychometric properties of the parent and teacher versions of the Strengths and Difficulties Questionnaire for 4- to 12-year-olds: A review. *Clinical Child and Family Psychology Review, 13,* 254-274. DOI: 10.1007/s10567-010-0071-2

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54*, 277-295. DOI: 10.1146/annurev.psych.54.101601.145028

Zajac, R., & Karageorge, A. (2009). The wildcard: A simple technique for improving children's target-absent lineup performance. *Applied Cognitive Psychology*, *23*, 358-368. DOI:10.1002/acp.1511

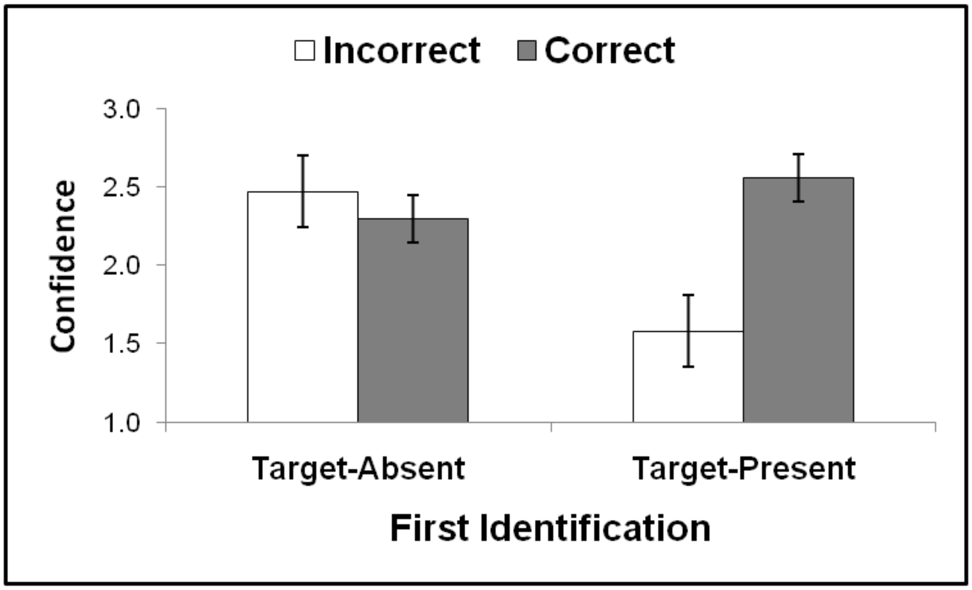

Figure 1. Mean confidence for correct versus incorrect identification decisions on target-absent and target-present line-ups at the first identification. Error bars correspond to +/- 1 standard error of the mean

Table 1

Identification Accuracy and Anxiety

		Target-Present		Target-Absent	
Identification	Accuracy	Nonanxious	Anxious	Nonanxious	Anxious
First	Correct	15	10	9	14
	Incorrect	7	5	8	9
	Not Sure	1	0	1	1
Follow-up	Correct	4	3	12	6
	Incorrect	14	8	2	9
	Not Sure	1	0	0	1

Values indicate frequency counts.

Table 2

Confidence-Accuracy Relations

		Processing Conditions			
		Poor		Good	
Identification	Target	*r*	*n*	*r*	*n*
First	Present	.30	19	.83*	18
	Absent	.04	21	-.30	19
Follow-up	Present	.25	16	.12	12
	Absent	-.20	17	-.35	11

*$p < .05$