

Detecting Children's Lies: Comparing True Accounts About Highly Stressful Injuries with Unprepared, Prepared, and Coached Lies

Kelly L. Warren, Ph.D.[†], Elyse Dodd[†], Graham Raynor[†] and Carole Peterson, Ph.D.*

In this investigation, 514 university students judged whether children were telling the truth about highly emotional events. Eight children (half female, half 8–9 and the remainder 12–14 years old) had been injured seriously enough to require emergency room treatment and were interviewed a few days later. Each was yoked to three other children matched in age and gender who fabricated accounts under one of three conditions: lies that were unprepared, prepared (24 hours to prepare), and coached by parents. Participants were at chance when judging true accounts as well as unprepared and prepared lies. However, 74% of the coached lies were judged as true. Participants' confidence in their judgments, age, experience with children, and relevant coursework/training did not improve judgments. Copyright © 2011 John Wiley & Sons, Ltd.

INTRODUCTION

In the courtroom, children's credibility is often judged by jurors, lay persons who must determine for themselves whether children are lying or telling the truth. Likewise, police officers, lawyers, social workers, and judges routinely find themselves in the position of deciding on the truthfulness of children's accounts. Because of the practical importance of this decision-making, there has been a recent increase in research devoted to investigating the ability of adults to detect children's truthfulness (see review by Vrij, 2008).

Nature of the Event Being Recalled

An important issue is the sort of event being recalled. In the literature on children's memory, the salience, uniqueness, and emotionality of events are predictive of how well children recall them (see reviews by Bauer, 2006, 2007; Peterson, 2002; and Quas, Alexander, & Goodman, in press). For example, children as young as 3 years are surprisingly competent at recalling information about highly salient, unique, and emotionally arousing events. Children have long-term elaborate memories about stressful events such as personal injuries (Peterson & Whalen, 2001), natural disasters (Fivush et al., 2004), and painful medical procedures (Quas et al., in press). This body of research has had considerable impact on perceptions of children's memory competence in forensic

*Correspondence to: Carole Peterson, Ph.D., Psychology Department, Memorial University of Newfoundland, St. John's, Newfoundland, A1B 3X9 Canada. E-mail: carole@mun.ca

[†]Memorial University of Newfoundland

situations because criminal events are also likely to be highly salient and stressful. In contrast, relatively little research in lie detection has explored events with high salience, uniqueness, and emotion (but see Leach *et al.*, 2009).

If children's memory of highly salient and stressful experiences is much better than for more mundane ones, it is possible that there is a parallel advantage for salient, stressful events in terms of the detection of deception. That is, people may be better at determining the truthfulness of accounts of such stressful events, perhaps because they are likely to be richer in detail as well as more emotion laden. This possibility is the motivation for the current research, which takes advantage of a body of extant transcripts, specifically interviews of children injured seriously enough to need emergency room treatment (Peterson, 2010; Peterson & Whalen, 2001).

Detection of Child Deception

When children are first engaged in the forensic process, they are interviewed face-to-face. Likewise, courtroom appearances involve both seeing and hearing children. As a consequence, considerable research has explored how well people intuitively judge children's credibility when video-clips of children are played. Using these stimuli where children are both seen and heard, adults typically have accuracy rates little or no better than chance (Crossman & Lewis, 2006; Edelman *et al.*, 2006; Leach *et al.*, 2009; Leach *et al.*, 2004; Strömwall & Granhag, 2007; Strömwall, Granhag, & Landström, 2007; Talwar *et al.*, 2006; Vrij *et al.*, 2006a). Similar accuracy rates are shown by professionals such as police, teachers, and social workers who have experience working with children or in lie detection (Leach *et al.*, 2004; Vrij, Akehurst & Knight, 2006b).

Part of the explanation for this poor discrimination of children's honesty is that many people base their judgments in part on visual cues suggesting nervousness, agitation, or avoidance (Vrij *et al.*, 2006b). However, these cues are not reliable indicators of deception for either adults or children (Vrij, 2008; Vrij & Mann, 2001). According to Lindholm (2008), the ability to see interviewed children at times may be more problematic than helpful in judging truthfulness. He further stated that "judgments of accuracy based on transcripts rather than live testimony would increase the quality of legal decisions" (p. 1301).

Forensic interviews of children are almost always transcribed, and these transcripts often play important roles in forensic situations. Their strength is that individuals reading them can check back to earlier statements and make careful comparisons of what was said at different times in the interview. They are heavily used by police, prosecutors, defense lawyers, and expert witnesses, and in fact a large proportion of forensic effort surrounding child witnesses involves transcripts of child interviews conducted by police (B. Snook, personal communication, January 2010). Although jurors typically observe live testimony from children in court, they often review transcripts of testimony when deliberating. Thus, an understanding of the reliability of truth detection using transcripts has also been a focus of recent research.

This research on lie detection from transcripts has taken two directions. The most frequently investigated one relies on professionals who are trained to judge statement plausibility in transcripts by using either (or both) Criteria-Based Content Analysis (CBCA) or Reality Monitoring analysis (Vrij, 2008). The CBCA in particular is the most popular instrument for assessing the veracity of children's reports in forensic settings in many European jurisdictions (Vrij, 2005). However, it is less used in North America

and is even inadmissible in criminal court in many jurisdictions. For example, it is inadmissible in Canada on the basis of legal principle (it usurps the role of the trier of fact). Part of the reason for doubt in other jurisdictions is that some investigators have raised questions about its usefulness for young children (Blandon-Gitlin et al., 2005; Buck et al., 2002; Pezdek et al., 2004; Vrij, 2005; Vrij et al., 2004a).

The other direction is an exploration of intuitive lie detection from transcripts. The little extant research on this shows a success rate that is little better than chance (Tye et al., 1999), parallel to most findings of studies using video-clips. However, relatively little research has investigated intuitive lie detection from transcripts, despite their frequent use by many individuals in the forensic process. The current study addresses this gap.

Type of Interview

Much of the research on intuitive lie detection uses children's responses to yes/no questions or individual sentences. However, adults have been found to be more accurate in identifying truth-telling when narratives rather than these short responses are used (Leach et al., 2009; Vrij & Baxter, 1999). Theoretically, detailed descriptions of complex events are more cognitively demanding to generate and more difficult to monitor relative to short reports or those about less complex events. The cognitive load hypothesis (Vrij, 2000) states that deception is more effortful than honesty because it requires greater use of mental resources. This suggests that a child who is lying about a complex event may have more opportunities to make mistakes whereas a truthful child may have more opportunities to demonstrate mastery of the event. In the present study, children were interviewed about a complex event (injuries requiring medical treatment), and transcripts of their entire interview were used. Using entire interviews is also more similar to real-life situations than are judgments of sentences or short speech units.

Conditions of Account Fabrication

Although children as young as age 3 or 4 are able to spontaneously generate lies (Newton, Reddy, & Bull, 2000; Talwar & Lee, 2002), children in forensic situations typically provide reports nonspontaneously. Particularly relevant is both allowing children time in which to prepare their lies, and having someone coach children to lie. In terms of the effect of allowing children time to prepare lies, common sense might suggest that children who had time to think through what they want to say would seem more convincing than those who have not, at least if they are providing a deceptive account. Indeed, Strömwall et al. (2007) found that adults were poorer at judging the truthfulness of individual sentences from 11- to 13-year-old children's videotaped stories about emotional events if the children had 7 minutes to prepare (46% accuracy) versus no time to prepare (57% accuracy). However, the effects were modest at best.

The effect of coaching (helping children to prepare a story) on children's reports is even more important to understand, because children frequently experience some level of coaching (e.g., by parents) prior to testifying. However, few studies have investigated the effect of coaching on adults' ability to detect deception (Vrij et al., 2002, 2004b). A study of whether truthful versus fabricated accounts could be discriminated using the CBCA and/or Reality Monitoring instruments was done by Vrij et al. (2002, 2004b). These investigators asked 5–6-year-olds, 15-year-olds, and undergraduates to either lie or tell the truth about who rubbed information off a blackboard, and half were coached

in CBCA criteria. In both studies, CBCA scores did not differ between truth-tellers and liars for coached participants. Because scorers using the CBCA could not discriminate reports if participants were coached, one might assume that naïve adults would be unable to either, but this was not directly assessed. In contrast, Talwar *et al.* (2006) had children (4–7 years old) coached by parents and then asked them to provide truthful or fabricated accounts of relatively common but not universal events such as attending a wedding. Adults' ability to discriminate truth-tellers from liars using video-clips was at chance levels.

Overall, only one study asked lay persons to judge whether coached children were lying or not, and in that study video-clips were shown and there was no comparison with non-coached children. Moreover, to our knowledge only one study has explored the effect of giving children a chance to think about their fabrications, and in that study, children only had 7 minutes in which to prepare. In none of the studies were the target events the sort of highly salient, emotional, and unique events that memory researchers have found to be best recalled by children.

The Current Study

The current study attempts to address a number of the issues raised above. First and foremost, the target events were highly salient and emotional, namely injuries serious enough to require hospital emergency room treatment (mostly broken bones and lacerations requiring suturing). The children who were telling the truth were judged by their parents to have been extremely distressed at time of injury and were interviewed a few days later. Transcripts of their entire interview were used. Each injured child was yoked to three matched children, all of whom produced fabricated accounts of injuries similar to the truthful child's injury (e.g., if the truthful child broke his leg the yoked children were asked to fabricate accounts with a broken bone, with parental confirmation that they had never had a similar injury). A third of the children were asked to produce the fabricated account on the spot; a third were told that they would be interviewed the following day; and the remainder were coached by their parents for four days prior to their interview. Lay judges (university undergraduates) assessed truthfulness; the use of naïve judges parallels those courtroom scenarios in which jurors evaluate children's credibility.

For the most part, little extant research guides our hypotheses. Nevertheless, we expect adults to be relatively more successful in determining if injured children were telling the truth versus children who were in the prepared or unprepared fabrication conditions for two reasons. First, others have found adults to be more successful when children's narratives rather than short responses are used (Leach *et al.*, 2009; Vrij & Baxter, 1999), and the cognitive load hypothesis predicts that fabricating accounts of complex events should be more difficult than recounting true events (Vrij, 2000). We also expect adults to be more accurate when judging younger than older children, similar to Leach *et al.* (2004), although they used video-clips. We anticipate this especially for children in the prepared fabrication condition because we expect little advantage of a day's preparation for younger children whereas older children tend to be more successful at planning ahead (Friedman & Scholnick, 1997). In terms of coaching, prior research suggests that this helps children's lies to be more believable (Vrij *et al.*, 2002, 2004b), so we expect children who are coached to lie to be more likely to be judged as truth-tellers. However, we have no basis for predicting whether or not children who have been coached to lie are more convincing than children who have actually experienced highly stressful events.

METHOD

Participants

A total of 514 undergraduate university students were recruited to voluntarily participate in this study. Their mean age was 22.3 years with a range of 16 to 77. Of them, 362 participants were female and 149 were male. The participants were mostly Caucasian and came from mixed socioeconomic backgrounds. Furthermore, 24.7% of students claimed to have relevant academic courses or training experience working in forensic psychology, criminology, or other fields that were relevant to judging deception, and 8.2% identified themselves as parents. A draw for a \$100 prize was used as an incentive to participate in the study.

Procedure

Each questionnaire package handed out to participants included a copy of one of the 32 interviews, a consent form for the study, and a questionnaire. The questionnaire asked whether the participants thought the child was lying or telling the truth, and to rate their degree of confidence on a five-point scale from 1, "not at all confident," to 5, "very confident." Also collected was demographic information (age and gender) and information about what experience the participants may have had that was relevant to lie detection in children. Participants were asked to rate their experience with children on a five-point scale from 1 ("no experience") to 5 ("a lot of experience") as well as their academic courses and training background in areas relevant to lie detection such as criminology, social work, forensic psychology, police studies, or law on another five-point scale from 1 ("no relevant coursework/training experience") to 5 ("a lot of relevant experience"). Questionnaire materials were distributed in university classes and dormitories, and participants were assigned randomly to the different conditions. Students were asked to take them home and fill them out independently (without talking to others), and then return them to the next class or a collection box.

Materials

The stimulus materials were a collection of interviews with children who described receiving an injury. Interviews were tape recorded and transcribed verbatim, except for deletion of any identifying information. Half of the children were 8–9 years old ($n = 16$, $M = 8.6$ years, $SD = 0.5$), which constituted the 'younger age group'. The other half were 12–14 years old ($n = 16$, $M = 12.6$ years, $SD = 0.8$), the 'older age group.' The children were divided into four different conditions: true, unprepared lies, prepared lies and coached lies. Eight children provided true reports of injuries, four younger and four older, with half of each age group female and the remainder male. These reports had been randomly selected from interviews previously conducted with children who had been recruited in the emergency room because of physical injuries such as broken bones or lacerations requiring suturing. Only children who had been rated by their parents during parental interviews as highly distressed were included. Although they had been highly distressed at the time of injury, they had been interviewed a few days later (mean delay = 6 days) when they were no longer upset. During the interview,

children were first asked for free recall about the injury event and then the hospital event, and then they were asked a series of questions (mostly wh- in form) to elicit additional information about where the injury occurred, specifically how it happened and what caused it, who was there at the time and what they did, and so on. Details of this interview and the interviewing procedure have been published elsewhere (Peterson, 1999; Peterson & Bell, 1996; Peterson & Whalen, 2001).

There were three conditions of fabrication, with each child in each fabrication condition (randomly assigned) yoked to a child in the true condition. Thus, there were eight children in each fabrication condition (half younger and half female). These children were recruited through schools and personal contacts. They were asked to make up a story in which they sustained an injury similar to that of their yoked true counterpart. Parents confirmed that children in the fabrication conditions had never had a similar previous experience so that they could not use personal knowledge to provide details. They were all asked to “try and fool the interviewer”, which they all enjoyed doing as a challenge. (a) In the unprepared condition children were told what to lie about and asked to think about it briefly before the interview began. (b) Prepared children were told one day in advance what they should lie about, and asked to think about what they were going to say ahead of time. Their parents were asked to refrain from helping because we were interested in what their child would construct independently, and at the time of the interview parents confirmed that they had not helped. (c) Parents of children in the coached condition helped their child construct a coherent story every day for four days prior to the interview. They were instructed to go over their children’s story, to help them fill in details, and generally help them make the story believable. They did this twice a day for 10–20 minutes each day. They were given a brief list of suggestions that could be used to help their children construct their story, such as making sure they provided detailed information about where it happened, when it occurred, who was there, and specifically what happened. Children needed to talk about how they got hurt, the reactions of themselves and others, and what occurred up until the emergency room. This is the sort of information that children typically provide in interviews about actual injuries. Parents of children in the coached condition confirmed that they had indeed done the requested coaching. Children in all conditions were administered the same interview, almost all in their own homes. Although five interviewers (all senior-level undergraduate females) were used, all were trained and supervised by the same individual (Peterson), who has had many years of experience in training students to carry out child interviews. All aspects of the study were approved by Memorial University’s Interdisciplinary Committee for Ethics in Human Research.

RESULTS

Preliminary Analyses

Prior to conducting statistical analyses it was important to see if there were any obvious differences in the transcripts that could explain potential findings regarding a person’s ability to judge the veracity of a transcript, such as differing length or amount of detail. Each transcript was coded to determine its length in terms of the number of words as well as the number of people, objects, locations, activities, time references, emotions, and cognition details provided. A series of one-way ANOVAS was calculated on each

of the coded variables, to see if there were differences between transcripts in the four different conditions. All transcripts were similar in length (number of words: $M = 297.09$, $SD = 188.12$) and there was no difference in the number of overall details provided ($M = 63.53$, $SD = 31.05$). Moreover, there were no differences in the number of people ($M = 6.37$, $SD = 4.18$), objects ($M = 8.41$, $SD = 3.94$), locations ($M = 4.38$, $SD = 1.93$), activities ($M = 18.66$, $SD = 7.58$), time references ($M = 4.87$, $SD = 3.11$), emotions ($M = 1.13$, $SD = 1.07$), or cognition details ($M = 1.91$, $SD = 2.76$). Thus we assumed that differences across conditions were due to the condition and not due to systematic differences in the transcripts.

Judgments of True and False Reports

The proportion accuracies of adults' judgments about whether children's accounts were truthful are found in Table 1. Data are separately presented for males and females in each of the four experimental conditions. An overall assessment of the accuracy of veracity decisions revealed that participants were more likely to be incorrect (56.4%) than correct (43.6%) in their judgements, $\chi^2(N = 514, df = 1) = 8.48, p < .01$. Further analyses were conducted to determine what, if anything, could explain the differences in accuracy of a veracity decision.

After deciding on the veracity of the transcript presented, all participants indicated their confidence in their decision about the truthfulness of children's reports on a five-point scale from 1, not at all confident, to 5, highly confident, and the mean confidence rating was 3.62 ($SD = .71$). On parallel five-point scales participants' mean rating of how much experience they had with children was 3.30 ($SD = 1.12$), and their mean rating of how much coursework/training they had in relevant fields of study was 1.63 ($SD = 0.87$). These characteristics were not totally independent, of course. According to correlational analyses (Pearson's r), participants who were older were significantly more likely to have had more experience with children ($r = .28, p < .01$), were more likely to have had relevant coursework or training ($r = .11, p < .05$), and were more confident in their decisions ($r = .13, p < .01$). Those with more experience with children were also more confident in their decisions ($r = .20, p < .01$). In addition, women were likely to have had more experience with children than were men ($r = .27, p < .01$).

A step-wise regression with confidence, experience with children, relevant course experience, participant age, participant gender, and condition (true, unprepared lie,

Table 1. Proportion accuracies (and SDs) of adults' judgments about whether children's accounts were truthful

| Characteristics of children | True accounts | | Unprepared lies | | Prepared lies | | Coached lies | |
|-----------------------------|---------------|-----------|-----------------|-----------|---------------|-----------|--------------|-----------|
| | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> | <i>M</i> | <i>SD</i> |
| Older – All | .45 | .50 | .49 | .50 | .45 | .50 | .27 | .45 |
| Females | .57 | .50 | .56 | .51 | .50 | .51 | .19 | .40 |
| Males | .29 | .46 | .42 | .50 | .41 | .50 | .35 | .49 |
| Younger – All | .52 | .50 | .51 | .50 | .55 | .50 | .24 | .43 |
| Females | .53 | .51 | .42 | .50 | .53 | .51 | .31 | .47 |
| Males | .50 | .51 | .58 | .50 | .56 | .50 | .19 | .40 |
| All children | .48 | .50 | .50 | .50 | .50 | .50 | .26 | .44 |

prepared lie, and coached) as predictors was completed to determine their influence on the accuracy of a veracity decision. The model was significant $F(6, 502) = 3.46, p < .01, \eta^2 = .04$. However, the only variable included in the model was condition.

To assess the influence of condition we explored whether participants' judgments of the different types of report differed from chance (50% judged as true). To do this, one-sample *t*-tests were used to compare accuracy to the level of chance (50%) for each condition. As seen in Table 2, the only condition that differed from chance was the coached condition; only 25.6% of respondents correctly identified the coached fabricated stories as false. The accuracy for all other groups did not differ from chance.

In addition, follow-up goodness-of-fit chi-squares were also calculated on the data for each type of report separately, to see if the accuracy of judgments about the reports of older children and younger children, and of girls and boys, within each report condition differed from chance. For reports about true events and both unprepared and prepared fabrications, none of the follow-up χ^2 s were significant, although all were significant for children who were coached to lie. That is, participants were significantly worse than chance in their accuracy of assessing the truthfulness of girls', boys', older children's, and younger children's coached reports.

DISCUSSION

The premise motivating the current research is that laypersons would be more successful at detecting the truth of a report if the event was a highly salient and stressful one to the child. In this study the children providing truthful accounts had all suffered injuries that had been very painful and distressing. In fact, the parents of all the children (who witnessed their child's injury) rated their child as having been highly distressed. Likewise, the children themselves described their emotional reactions as lots of crying, screaming, or screeching. Moreover, the events had only occurred a few days previously so they were fresh in the children's minds. Although in this study we insured that the event was very recent to optimize child memory, this may not have been a necessary control because other research has shown that such highly emotional and salient events are recalled extremely well across delays measured in years (Peterson, 1999; Peterson & Whalen, 2001). The children were interviewed extensively with an interview that included both free recall and probe questions that inquired about who, where, what, when, and why. The interviewers also asked for details of the succession of events that occurred, including what the child had been doing prior to the injury, the injury itself, the responses of others,

Table 2. Summary of *t*-test discrimination accuracy for true, false unprepared, false prepared, and false coached reports

| Condition | Means | | <i>t</i> -test information | | | |
|------------------|----------|-----------|----------------------------|-----------|-------------------|------------------|
| | <i>M</i> | <i>SD</i> | <i>t</i> | <i>df</i> | Sig. (two tailed) | Cohen's <i>d</i> |
| True | .479 | .501 | -.502 | 141 | .616 | - |
| False unprepared | .504 | .502 | .088 | 126 | .930 | - |
| False prepared | .500 | .502 | .000 | 119 | 1.000 | - |
| False coached | .256 | .438 | -6.226 | 124 | .000** | .557 |

** $p < .001$.

and the events that occurred at the hospital. Thus, rather than isolated sentences being the stimuli being considered by the laypersons judging the children's reports, as has been used in some other research (e.g., Strömwall et al., 2007), a more complex report was available. Nevertheless, to our surprise, lay judges were unable to distinguish the reports as truthful rather than fabricated. This inability to judge the veracity of true accounts is similar to that found for children's reports of more benign events that are considerably less salient, such as playing with a house or erasing a blackboard (Lyon et al., 2008; Vrij et al., 2002, 2004b).

Perhaps the explanation for this unexpected finding is that injuries are simply not unique enough events. That is, although breaking an arm or requiring 12 sutures to treat a laceration are indeed relatively rare to any particular child, such injuries are relatively common in any population of children. Thus, even though particular children may not themselves experience such an injury, they see their friends and classmates sport casts and sutures, and such events are the sort of 'big news' events that are typically widely talked about (Peterson, 1999; Peterson & Whalen, 2001). Therefore, although the children who fabricated accounts of injuries had no direct experience with these sorts of events, they probably had prior indirect experience through familiar others. Nevertheless, it is still unexpected that those children who had been highly emotionally involved in these real-life events could not be discriminated from children who only pretended to be.

A second major finding is that the only accounts that laypersons were inclined to believe were the fabricated ones that had been coached by the children's parents. Three-quarters of the participants believed these coached reports to be truthful. This finding parallels other research that found coached reports to be relatively convincing, including to professionals using CBCA or Reality Monitoring criteria (Talwar et al., 2006; Vrij et al., 2002, 2004b). Vrij and his colleagues (2002, 2004b) informed children of specific criteria of the CBCA during their coaching sessions, which were conducted by trained research assistants. The investigators found that children who were knowledgeable about CBCA criteria were more convincing lie-tellers than children who had not had such specific coaching – i.e., coaching children to include information relevant to specific CBCA criteria was effective. In the present study coaching was done by parents, and although they were not given instructions relevant to CBCA scoring they were asked to make sure that the children's reports were detailed and contained information about who, what, when, where, and why.

We also found no difference in the ability of lay persons to discriminate the truthfulness of prepared and unprepared reports. More specifically, adult participants' judgments of both were not different from chance. This finding diverges from that reported by Strömwall et al. (2007). However, in that report, judges only decided about the truthfulness of children's individual statements, and they also had video clips of the children available to them. In the current study, judges had entire interviews available and thus could make their decisions on considerably more verbal information.

The fact that coached children were so successfully deceptive in comparison to children who generated a deceptive report on the spot (unprepared lie condition) or even a day later (prepared lie condition) may be interpreted as supportive of the cognitive load hypothesis (Vrij, 2000). This hypothesis states that deception is more challenging than truthfulness because lying requires greater effort to generate and maintain a coherent report. However, coached children had been rehearsed over the course of several days and thus had had plenty of opportunity to develop a coherent deceptive story. They also presumably had feedback from their parents when their stories contained inconsistencies.

Children who were not coached (even if they had had a chance to think about their story ahead of time) did not have such rehearsal and feedback, and thus were more likely to make mistakes that called into question the truthfulness of their accounts. In contrast, children who were telling the truth were just as likely to be judged as deceptive as the children in the unprepared and prepared lie conditions; this is not supportive of the cognitive load hypothesis. These children had no need to maintain a deceptive stance over the course of their interview since their reports (and similar reports by other children) have been found to be remarkably accurate in detail (Peterson, 1999; Peterson & Bell, 1996; Peterson & Warren, 2009; Peterson & Whalen, 2001).

The finding that parental coaching of children to lie is so effective has dismaying implications for the courts. Children often are coached prior to court appearances, by either or both parents and lawyers. Children may even be coached by parents prior to their first interview with police about target events. In the real world, however, children who are telling the truth are often coached too, not just those who are lying. Thus, it is important for future research to explore coached true reports of children as well as coached false reports. Such a comparison is more likely to reflect the reality of child testimony than a comparison of coached lies versus uncoached true accounts. Nevertheless, it is possible that children who are telling the truth are either uncoached or more minimally coached than children who deliberately intend to deceive, and so a study such as this one still is informative.

Overall, neither the age of the child nor of the adult lay-judges played a significant role; nor did the gender of the child or of the adult judge. We also had collected information on the experience of our adult lay-judges, experience both with children and with relevant coursework and training. This experience also had no significant effect on their success at discriminating liars from truth-tellers. The only effect such experience had was to make judges more confident in their decision, even though their confidence was not reflected in greater success at discriminating truthful accounts. However, it should be noted that the degree of experience of our adult lay-judges does not compare with that of police officers or other professionals who are trained in interviewing children. Moreover, only eight percent of our adult participants had the extensive experience with children that is acquired through being a parent. Thus, it is possible that more experienced judges could have been more successful at discriminating truthful from fabricated accounts.

In the present study, the naïve judges had available to them lengthy and complex interviews. These interviews were similar in structure to those recommended as 'best practice' for child interviews (American Professional Society on the Abuse of Children, 1990, 1997; Home Office, 2002). They included free recall first, followed by probed recall focusing on open-ended questions, which were only later followed up by more specific questions. Thus, the interviews were not dissimilar to those conducted in forensic situations. Furthermore, the target events were highly salient, emotional, and stressful – characteristics shared with forensic events although of course these events are not the same as those experienced by children suffering abuse or other situations likely to involve them with the courts. Although only eight children provided reports in each condition and it would be prudent for future research to include more child reports, no outliers were found when the reports were individually compared. In addition, the transcripts in the four conditions did not differ in terms of length or on a number of measures of elaborative detail. Although a few of the truthful child transcripts were collected several years previously, they did not differ from similar transcripts collected more recently (Peterson & Warren, 2009). Thus, it is unlikely that the null effects we found were due to mundane

causes. Finally, the sample size of naïve judges was large enough to provide appropriate statistical power.

It is possible that techniques such as the CBCA or Reality Monitoring procedures could have more successfully discriminated the various types of account, even though our lay judges could not. Such approaches are more disciplined in their approach to the transcripts than are the judgments of undergraduate students. Thus, a potentially fruitful direction for future research is to have professionals who are trained in using these techniques judge the transcripts. Another future direction is to expand the age of the children downward. It may be that younger children who fabricate reports are less successful than younger truth-tellers at convincing lay judges. Alternatively, it may be that coaching younger children makes them even more convincing relative to non-coached children. These questions also need to be addressed. A different direction is to present these findings to sitting judges, expert witnesses, and potential jurors in an effort to determine whether such research would affect their views about their own ability to detect truth from lies.

In summary, lay judges in the present study could discriminate truthful children from liars no more successfully when highly salient and emotional events were used than they could when children reported on the more mundane events that have often been used in prior research. Judges were no more successful than chance when determining the truthfulness of interviews by children who had suffered actual injuries, reports fabricated with a day to prepare, or those fabricated on the spot. Instead, it is the children who were coached to lie who were largely successful in convincing lay judges that they were telling the truth. Both of these findings have dismaying forensic implications. Juries are composed of lay individuals who must assess children's truthfulness, and the apparent success of coaching in convincing lay individuals that lies are true may result in unfortunate court decisions.

ACKNOWLEDGEMENTS

Preparation of this article was primarily supported by Grant 513-02 from the Natural Sciences and Engineering Research Council of Canada to C. Peterson. Additional funding came from the Memorial University Undergraduate Career Experience Program, the Student Work and Service Program, and the Summer Career Placement Program. We extend our thanks to the Janeway Hospital and its Emergency Room staff, and to all the recruiters, interviewers, transcribers, and data analyzers who participated. We also thank Penny Voutier for myriad helpfulness and Malcolm Grant for statistical consultation. Most of all we thank the parents and children who allowed us into their homes and cooperated so helpfully.

REFERENCES

- American Professional Society on the Abuse of Children. (1990). *Guidelines for psychosocial evaluation of suspected sexual abuse in young children*. Chicago, IL: Author.
- American Professional Society on the Abuse of Children. (1997). *Guidelines for psychosocial evaluation of suspected sexual abuse in young children (Revised)*. Chicago, IL: Author.
- Bauer, P. J. (2006). Event memory. In W. Damon & R. M. Lerner (Series Eds.) & D. Kuhn & R. S. Siegler (Vol. Eds.), *Handbook of child psychology: Vol. 2. Cognition, Perception, and Language* 6th ed. (pp. 373–425). New York: Wiley.

- Bauer, P. J. (2007). *Remembering the times of our lives: Memory in infancy and beyond*. Mahwah, NJ: Erlbaum.
- Blandon-Gitlin, I., Pezdek, K., Rogers, M., & Brodie, L. (2005). Detecting deception in children: An experimental study of the effect of event familiarity on CBCA ratings. *Law and Human Behavior*, 29, 187–197.
- Buck, J. A., Warren, A. R., Bettman, S. I., & Brigham, J. C. (2002). Age differences in criteria-based content analysis scores in typical child sexual abuse interviews. *Applied Developmental Psychology*, 23, 267–283.
- Crossman, A. M., & Lewis, M. (2006). Adults' ability to detect children's lying. *Behavioral Sciences & the Law*, 24, 703–715.
- Edelstein, R. S., Luten, T. A., Ekman, P., & Goodman, G. S. (2006). Detecting lies in children and adults. *Law and Human Behavior*, 30, 1–10.
- Fivush, R., Sales, J. M., Goldberg, A., Bahrack, L., & Parker, J. (2004). Weathering the storm: Children's long-term recall of Hurricane Andrew. *Memory*, 12, 104–118.
- Friedman, S. L., & Scholnick, E. K. (Eds.) (1997). *Why, how, and when do we plan: The developmental psychology of planning*. Hillsdale, NJ: Erlbaum.
- Home Office. (2002). *Achieving best evidence in criminal proceedings: Guidance for vulnerable or intimidated witnesses, including children*. London: Home Office Communication Directorate.
- Leach, A.-M., Lindsay, R. C. L., Koehler, R., Beaudry, J., Bala, N. C., Lee, K., & Talwar, V. (2009). The reliability of lie detection performance. *Law and Human Behavior*, 33, 96–109.
- Leach, A.-M., Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). "Intuitive" lie detection of children's deception by law enforcement officials and university students. *Law and Human Behavior*, 28, 661–685.
- Lindholm, T. (2008). Who can judge the accuracy of eyewitness statements? A comparison of professionals and lay-persons. *Applied Cognitive Psychology*, 22, 1301–1314.
- Lyon, T. D., Malloy, L. C., Quas, J. A., & Talwar, V. A. (2008). Coaching, truth induction, and young maltreated children's false allegations and false denials. *Child Development*, 79, 914–929.
- Newton, P., Reddy, V., & Bull, R. (2000). Children's everyday deception and performance on false-belief tasks. *British Journal of Developmental Psychology*, 18, 297–317.
- Peterson, C. (1999). Children's memory for medical emergencies: Two years later. *Developmental Psychology*, 35, 1493–1506.
- Peterson, C. (2002). Children's long-term memory for autobiographical events. *Developmental Review*, 22, 370–402.
- Peterson, C. (2010). "And I was very very crying": Children's self-descriptions of distress as predictors of recall. *Applied Cognitive Psychology*, 24, 909–924.
- Peterson, C., & Bell, M. (1996). Children's memory for traumatic injury. *Child Development*, 67, 3045–3070.
- Peterson, C., & Warren, L. K. (2009). Injuries, emergency rooms, and children's memory: Factors contributing to individual differences. In J. A. Quas & R. Fivush (Eds.), *Emotion and memory in development: Biological, cognitive, and social considerations* (pp. 60–85). New York: Oxford University Press.
- Peterson, C., & Whalen, N. (2001). Five years later: Children's memory for medical emergencies. *Applied Cognitive Psychology*, 15, 7–24.
- Pezdek, K., Morrow, A., Blandon-Gitlin, I., Goodman, G. S., Quas, J. A., Saywitz, K. J. et al. (2004). Detecting deception in children: Event familiarity affects Criterion-Based Content Analysis ratings. *Journal of Applied Psychology*, 89, 119–126.
- Quas, J. A., Alexander, K., & Goodman, G. S. (in press). Children's memory in social and emotional context: Recounting stressful life events. *Developmental Review*.
- Strömwall, L. A., & Granhag, P. A. (2007). Detecting deceit in pairs of children. *Journal of Applied Social Psychology*, 37, 1285–1304.
- Strömwall, L. A., Granhag, P. A., & Landström, S. (2007). Children's prepared and unprepared lies: Can adults see through their strategies? *Applied Cognitive Psychology*, 21, 457–471.
- Talwar, V., & Lee, K. (2002). Development of lying to conceal a transgression: Children's control of expressive behaviour during verbal deception. *International Journal of Behavioral Development*, 26, 436–444.
- Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2006). Adults' judgments of children's coached reports. *Law and Human Behavior*, 30, 561–570.
- Tye, M. C., Amato, S. L., Ronts, C. R., Devitt, M. K., & Peters, D. (1999). The willingness of children to lie and the assessment of credibility in an ecologically relevant laboratory setting. *Applied Developmental Science*, 3(2), 92–109.
- Vrij, A. (2000). *Detecting lies and deceit*. Chichester, UK: Wiley.
- Vrij, A. (2005). Criteria-based content analysis: A qualitative review of the first 37 studies. *Psychology, Public Policy, and Law*, 11, 3–41.
- Vrij, A. (2008). *Detecting lies and deceit: Pitfalls and opportunities*. Hoboken, NJ: Wiley.
- Vrij, A., Akehurst, L., Brown, L., & Mann, S. (2006a). Detecting lies in young children, adolescents and adults. *Applied Cognitive Psychology*, 20, 1225–1237.
- Vrij, A., Akehurst, L., & Knight, S. (2006b). Police officers' social workers', teachers' and the general public's beliefs about deception in children, adolescents, and adults. *Legal and Criminological Psychology*, 11, 297–312.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2002). Will the truth come out? The effect of deception, age, status, coaching, and social skills on CBCA scores. *Law and Human Behavior*, 26, 261–283.
- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004a). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal Behavior*, 24, 239–263.

- Vrij, A., Akehurst, L., Soukara, S., & Bull, R. (2004b). Let me inform you how to tell a convincing story: CBCA and Reality Monitoring scores as a function of age, coaching, and deception. *Canadian Journal of Behavioural Science*, *36*, 113–126.
- Vrij, A., & Baxter, M. (1999). Accuracy and confidence in detecting truths and lies in elaborations and denials: Truth bias, lie bias and individual differences. *Expert Evidence*, *7*, 25–36.
- Vrij, A., & Mann, S. (2001). Telling and detecting lies in a high-stake situation: The case of a convicted murderer. *Applied Cognitive Psychology*, *15*, 187–203.