



Published in final edited form as:

Law Hum Behav. 2004 December ; 28(6): 661–685. doi:10.1007/s10979-004-0793-0.

“Intuitive” Lie Detection of Children’s Deception by Law Enforcement Officials and University Students¹

Amy-May Leach^{2,6}, Victoria Talwar³, Kang Lee⁴, Nicholas Bala⁵, and R. C. L. Lindsay²

2Department of Psychology, Queen’s University, Kingston, Ontario, Canada.

3Department of Educational and Counselling Psychology, McGill University, Montreal, Quebec, Canada.

4Department of Psychology, University of California, San Diego, California.

5Faculty of Law, Queen’s University, Kingston, Ontario, Canada.

Abstract

Adults’ ability to detect children’s deception was examined. Police officers, customs officers, and university students attempted to differentiate between children who lied or told the truth about a transgression. When children were simply questioned about the event (Experiment 1), the adult groups could not distinguish between lie-tellers and truth-tellers. However, participants were more accurate when the children had participated in moral reasoning tasks (Experiment 2) or promised to tell the truth (Experiment 3) before being interviewed. Additional exposure to the children did not affect accuracy (Experiment 4). Customs officers were more certain about their judgments than other groups, but no more accurate. Overall, adults have a limited ability to identify children’s deception, regardless of their experience with lie detection.

Keywords

deception; lie detection

In North America, approximately 100,000 children serve as witnesses in the justice system each year (Bruck, Ceci, & Hembrooke, 1998). These cases are usually of a sensitive nature (e.g., involving allegations of child abuse) where there are no other witnesses or there is no evidence to corroborate the account. The responsibility of assessing the veracity of children’s testimony often falls upon frontline workers in the justice system. However, it is unknown whether law enforcement officials are actually able to detect children’s deception.

The majority of intuitive lie-detection research (as contrasted with polygraph studies) has focused on laypersons’ ability to identify adult liars. Examining these findings can offer a limited perspective on detecting children’s deception. Adult experimental scenarios often involve role-playing, with an individual being instructed to represent an emotion, opinion, or event either truthfully or untruthfully (e.g., DePaulo & DePaulo, 1989; Ekman & Friesen, 1974; Riggio, Tucker, & Throckmorton, 1988). Although some evidence suggests that adults are good lie detectors (e.g., DePaulo & Rosenthal, 1979), others imply that untrained observers cannot reliably detect deception (e.g., Ekman & Friesen, 1974). Given these mixed findings,

¹The authors acknowledge graduate and research grant support from the Natural Sciences and Engineering Research Council of Canada and the Social Sciences and Humanities Research Council of Canada.

⁶To whom correspondence should be addressed at Department of Psychology, Queen’s University, Kingston, Ontario, K7L 3N6; e-mail: 9aml1@qmlink.queensu.ca.

laypersons' ability to detect lies is, at best, doubtful (DePaulo, Zuckerman, & Rosenthal, 1980; Ambady & Rosenthal, 1992 for reviews).

It is possible that work-related experience with deception affects accuracy. Intuitively, law enforcement and justice officials, who often interact with dishonest individuals and may be specifically trained to detect lies, should perform well on lie-detection tasks. However, evidence to date shows that the majority of professional groups (e.g., police officers, customs officers, and FBI agents) cannot distinguish truth-tellers from lie-tellers any better than university students (DePaulo & Pfeifer, 1986; Ekman & O'Sullivan, 1991; Kraut & Poe, 1980). These findings suggest that experience and training do not necessarily improve lie-detection accuracy. Nevertheless, these factors do shape groups' perception of their own performance. A meta-analysis of lie-detection studies showed that confidence was unrelated to actual ability to detect deception (DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997). Despite their relatively poor performance, law enforcement officials are often more confident in their abilities than laypersons (e.g., DePaulo & Pfeifer, 1986; Frank & Ekman, 1997). Thus, "lie-detection experts" not only make incorrect decisions, but also are unable to recognize when they are doing so.

However, some groups consistently outperform others. High levels of performance have been observed in groups of Secret Service agents, CIA agents, sheriffs, and forensic clinical psychologists (Ekman & O'Sullivan, 1991; Ekman, O'Sullivan, & Frank, 1999). Recent research revealed that patients with damage to the left cerebral hemisphere were also good lie detectors (Etcoff, Ekman, Magee, & Frank, 2000). This superior performance may be explained by their reliance on nonverbal behavior.

Indeed, nonverbal cues have been shown to facilitate lie detection. DePaulo (1994) has reported that deceivers' behaviors (e.g., blinking) differ from those of truth-tellers. Micro-expressions (i.e., split-second glimpses of underlying behavior) may also be associated with deception because they are more difficult to control than verbal content or large facial movements (Ekman, O'Sullivan, Friesen, & Scherer, 1991). Successful lie detectors are more likely to cite nonverbal facial behaviors as the basis for their judgments (Ekman & O'Sullivan, 1991). Thus, detecting deception appears to require the observation of nonverbal responses.

The most important element in detecting deception may be the lie-teller's own ability to deceive. Individuals who act suspiciously are more likely to be labeled as deceptive (Bond et al., 1992). Successful deception often involves masking natural facial and demeanor cues associated with lying and displaying behaviors that are compatible with the falsehood. Adults' proficiency at controlling their external expressions may account for people's inability to detect their lies. However, the regulation of nonverbal behavior requires high levels of social and cognitive abilities that may develop, and be refined, over time.

Display rule theory posits that the management of nonverbal behavior must be learned through socialization. Children come to understand that internal affect does not always correspond to the external behavioral expression of emotion. This is accomplished with the use of display rules, or heuristics that determine the appropriateness of expressive behaviors in various contexts (Saarni, 1979). Young children are less likely to state display rules spontaneously and may have a weaker understanding of which facial expressions are most appropriate for given situations (Saarni & Von Salisch, 1993). If the understanding of display rules develops over time, so may the regulation of nonverbal responses. Thus, children may fail to regulate their nonverbal behavior when they tell lies, making their deceit easier to detect.

Suggestibility researchers have examined the detection of children's deception (Ceci & Bruck, 1993). This line of research involves providing children with false, or distorted, details and observing whether they incorporate these elements into their accounts of an event. Despite

extensive training and experience with children, clinicians and researchers cannot reliably assess children's accuracy following suggestion (Ceci & Huffman, 1997). Researchers (e.g., Goodman, Batterman-Faunce, Schaaf, & Kenney, 2002; Nysse & Bottoms, 2000) have shown that social workers, prosecutors, and untrained adults also have difficulty assessing the veracity of children's accounts. However, these findings may be due to the children's strong belief that the suggested events actually occurred. Thus, in a strict sense, these children were not lying because their accounts were not intentionally false.

Children's ability to knowingly and intentionally conceal the truth has rarely been studied. Feldman, Jenkins, and Popoola (1979) report that, although laypersons could not accurately classify teenagers and adults, they could detect 1st graders' lies. Feldman and White (1980) have also shown that older children are generally better at concealing deception than younger participants. Together, these studies appear to support the display rule theory position that children's ability to regulate their behavior and deceive improves with age. However, recent work suggests that young children are proficient at deception. When 3-year-olds were questioned about an actual transgression (i.e., peeking at a toy), few markers differentiated between the lie- and truth-tellers (Lewis, Stanger, & Sullivan, 1989). Moreover, undergraduates could not accurately detect these children's deception. Overall, these mixed findings do not offer a clear stance on adults' ability to detect children's deceit.

There are several limitations in previous research. Studies of deception often involved contrived scenarios in which individuals were instructed to behave "as if" they were lying (e.g., DePaulo, 1994; Frank & Ekman, 1997). The use of pretence may obscure the true behavioral practices related to deception in real-life situations. Another problem involves the use of lie-tellers who demonstrate noticeable cues to deception (Ekman & O'Sullivan, 1991; Ekman et al., 1999). Selecting the stimuli in this manner decreases generalizability because, in real life, lie-tellers may or may not display obvious nonverbal cues. Finally, previous experiments examining children's lie-telling in naturalistic settings (Lewis et al., 1989) only tested undergraduates' lie-detection abilities. Students' poor performance may have been due to their lack of experience with lie-detection rather than the children's ability to deceive.

The present research aimed to overcome the limitations of previous studies and to assess systematically adults' ability to detect children's deception. First, the deceptive scenario was naturalistic. Children were told not to commit a particular transgression (i.e., peeking at a toy). This temptation-resistance paradigm was adopted from Lewis et al. (1989). However, children were not required to provide specific responses about their actions (i.e., they were allowed to lie or tell the truth of their own volition). Second, a large number of child lie-tellers were included, regardless of their ability to deceive. Thus, no differentiations were made between successful and poor exemplars of deception to simulate real-life examples of lie-telling and ensure generalizability. Third, three different groups of adults were asked to detect children's deception. Students represented the performance of the average, untrained adult. Police officers were included because of their extensive lie-detection experience and training. The final group consisted of customs officers. These officials must conduct brief interviews with children, without background information, to determine whether they are being abducted from other countries. As a result, customs officers may rely on different cues than police officers (who often have longer periods of time in which to question individuals). Thus, the nonstudent groups offered unique life and work experiences that could affect their identification of children's lies.

The present research consisted of four experiments. Participants were shown videos of children who lied or told the truth about having committed a transgression. In Experiment 1, customs officers, police officers, and undergraduates were asked to differentiate between the child lie-tellers and truth-tellers (henceforth referred to as the "direct interview" condition). In Experiment 2, the children had made moral judgments about lying prior to answering questions

about their own transgression (the “moral discussion” condition). In Experiment 3, the children had been asked to promise to tell the truth before discussing the critical event (the “promise” condition). Again, in Experiments 2 and 3, groups were asked to discriminate between lie-tellers and truth-tellers. In Experiment 4, clips from all three types of interviews (direct, moral discussion, and promise) were rated by undergraduate students.

Several results were expected. First, it was hypothesized that law enforcement officials would be the most accurate due to their experience and training. Second, all participants were expected to perform best when the children had been alerted to the importance of truth-telling through moral discussions and promise-making. Third, a relationship between participants’ certainty and occupation was hypothesized, with law enforcement officials being more certain about their judgments due to their familiarity with the task. Finally, relationships between the amount of lie-detection experience and accuracy were expected.

EXPERIMENT 1

Method

Participants—In all, 25 police officers (5 women and 20 men, M age = 32.42 years, SD = 6.87) and 48 customs officers (20 women and 28 men, M age = 33.63 years, SD = 8.08) completed the study. Thirty-two undergraduate students (21 women and 10 men, M age = 23.22 years, SD = 6.17; 1 did not disclose his/her sex) participated in exchange for course credit.

Materials

Videotapes: The video segments had been recorded by a hidden video camera. In the temptation resistance paradigm, a female experimenter saw children (3- to 11-year-olds) individually. During this interaction, the experimenter was called out of the room. Prior to her departure, she asked the children not to peek at a toy while she was away. Upon her return, she asked the children three questions: (1) “While I was gone, did you turn your head to the side?” (2) “Did you move around in your chair?” and (3) “Did you peek to see who it [the toy] was?” There were three types of children. The first group of children peeked and lied about it (lie-tellers). The second group of children did not peek either of their own volition or because they were not given the opportunity to do so (truth-tellers). The third group of children peeked at the toy and admitted to the transgression (confessors).

A set of 80 video clips (M length = 17.50 s, SD = 6.66) was compiled of the exchange between the experimenter and children (regarding whether the child had peeked). Children’s upper bodies and faces were clearly visible. Several restrictions were used to assemble the videotapes. Two videotapes were made, with 40 video clips randomly assigned to each tape. Clips were randomly transferred onto videotape with the restriction that no more than three lie-tellers or truth-tellers appeared in a row. Each participant viewed only one tape.

Rating Scale: Participants were asked to rate whether each child was lying or telling the truth using a 7-point scale (1 = *definitely lying* to 7 = *definitely telling the truth*). This scale was selected to allow participants a full range of choices when they made their decision and afforded them the option of expressing their certainty about their judgments.

Procedure—Participants were tested individually or in small groups in a quiet room. A female experimenter randomly assigned participants to one of two videotape groups. The procedure was conducted in a single experimental session and took approximately 45 min to complete. Prior to the rating portion, participants were asked to provide personal information. Then, the experimenter explained the rating system, noting that about half of the children were

lying/telling the truth and that the clips were randomly presented. The videotape was shown with pauses in between the clips to allow participants the time to record their ratings.

Results

The purpose of this research was not to examine the development of successful deception. Age groups were not matched for size or behavior (i.e., lie-tellers and truth-tellers). The number of children included from certain age groups was low due to the opportunistic nature of the stimuli. Children had been allowed to respond of their own volition in the study from which the stimuli were produced, resulting in unequal ages of children who confessed, lied, and told the truth. This distorted sample may account for why preliminary analyses failed to reveal consistent age effects across experiments. Given that age was not a primary focus, and the issues related to unequal sample sizes, age effects will be considered separately later in this paper. Thus, data from children of all age groups were combined for all analyses (in both the present and subsequent experiments).

Preliminary analyses failed to reveal any significant effects involving participant sex. Thus, data from both male and female participants were combined for all subsequent analyses (this applies to all experiments). In addition, a reviewer suggested that participants' performance might have improved over the session. Comparisons between the first and second half of the trials failed to reveal significant practice effects for any of the experiments. Thus, only overall performance was examined in all analyses. Finally, participants' ratings of confessors were not analyzed. These children could be easily classified based solely on their verbal report (an admission) and their inclusion would have falsely inflated estimates of accuracy. Initially, these children's video clips were included as a control to assess participants' attention. Participants correctly identified most of these children as telling the truth ($M = .92$, $SD = .17$), suggesting that they were paying attention to the video clips. There were not enough confession trials, nor any false confessions, to justify a meaningful lie-detection analysis. As a result, only participants' classifications of 32 lie-tellers and 38 truth-tellers were used in the analyses.

Lie-Detection Accuracy—All responses were recoded to determine accuracy. Although the use of scales had been encouraged for signal detection analyses, analyses revealed no differences between the continuous and dichotomous data. For didactic purposes, it was easier to use dichotomous variables.

If the child was a lie-teller, individuals who rated the child as 1 (Definitely lying) to 3 (Likely lying) were given a score of "1." All others, who gave ratings of 5 (Likely telling the truth) to 7 (Definitely telling the truth), were assigned a score of "0." The reverse was true if the child was a truth-teller. Finally, all "not sure" (4) responses were eliminated from analyses because they could not be definitively classified as either correct or incorrect. This exclusion also occurred in all subsequent experiments. Scores ("1" or "0") for each judgment were summed across all children (or ratings) and averaged, yielding the overall accuracy score for each observer. Analyses were conducted on the mean scores (possible maximum score = 1.00; minimum score = 0). All effect sizes were calculated using meta-analytic r , as recommended by Rosenthal (1991).

A one-way ANOVA (with occupation as the independent variable) was performed on participants' overall accuracy score (see Fig. 1). There was a significant main effect of occupation, $F(2, 102) = 4.51$, $p < .05$, $r = .21$. Post hoc analyses revealed that both customs officers ($M = .49$, $SD = .08$, $r = .27$) and students ($M = .51$, $SD = .09$, $r = .36$) were significantly more accurate than police officers ($M = .44$, $SD = .08$), Tukey's HSD, $p < .05$. However, there were no differences between the overall accuracy of customs officers and students. One-sample t -tests were used to compare each group's mean accuracy to the level of chance (50%). The

overall accuracy of customs officers and students was at chance, whereas police officers performed significantly below chance, $t(24) = -3.53, p < .01$.

Signal Detection Theory Analysis—Accuracy reflects two separate aspects of the participants' decision-making process: (1) their actual ability to discriminate between truth- and lie-tellers (i.e., true-lie detection ability, usually referred to as d'), and (2) biases (i.e., the tendency to favor a particular response, such as categorizing children as lying, usually referred to as β). To extract this information, additional analyses were conducted using signal detection theory.

Discrimination: A one-way ANOVA, with occupation as the independent variable, was performed on the d' values, which is an index of discrimination ability (see Table 1). Police officers ($M = -0.29, SD = 0.46$) were better able to discriminate between the children than customs officers ($M = 0.00, SD = 0.45, r = .29$) and students ($M = 0.03, SD = 0.49, r = .33$), $F(2, 102) = 4.08, p < .05, r = .20$. Individual t -tests were performed to see whether the groups could actually differentiate between the children. In this analysis, the level of sensitivity was compared to 0 (no ability to differentiate between lie-tellers and truth-tellers). Unlike customs officers and students, police officers could reliably discriminate between the lie-tellers and truth-tellers, $t(24) = -3.13, p < .01$. However, the negative direction of the difference ($d' = -0.29$) suggests that officers did not label the groups correctly (i.e., they discriminated between the lie-tellers and truth-tellers, but tended to indicate that lie-tellers were telling the truth and vice versa).

Response Bias (β): Another aspect of detection is participants' response bias (β), or tendency to identify children as lie-tellers or truth-tellers, independent of their ability to discriminate between the two groups. A one-way ANOVA (with occupation as the independent variable) was conducted on the tendency to indicate that children were lie-tellers vs. truth-tellers (see Table 1). There were no significant effects, $F(2, 102) = 1.11, p = .33, r = .10$. Using t -tests, each β was compared to 1 (no bias). None of the groups differed significantly from 1; overall, their responses were not biased.

Certainty—The rating scale used also served as an indicator of the participants' certainty when making decisions. For example, rating a child as “definitely telling the truth” suggests greater certainty in the decision than indicating that the child was “likely telling the truth”. Thus, the rating scale was recoded to reflect the 4 levels of certainty (“definitely lying/telling the truth” = 3; “very likely lying/telling the truth” = 2; “likely lying/telling the truth” = 1; “not sure” = 0). The ratings were summed and averaged across all clips to yield a mean level of certainty.

A one-way ANOVA (with occupation as the independent variable) revealed a main effect of occupation, $F(2, 102) = 10.27, p < .001, r = .30$ (see Table 2). Post hoc analyses indicated that customs officers ($M = 1.91, SD = 0.35$) had higher levels of certainty than police officers ($M = 1.66, SD = 0.38, r = .32$) and students ($M = 1.56, SD = 0.35, r = .45$), Tukey's HSD, $p < .01$. Correlational analyses revealed that discrimination and certainty were unrelated for customs officers ($r(47) = -.04, p = .81$), police officers ($r(24) = .13, p = .55$), and students ($r(31) = -.28, p = .13$).

Other Variables—Participants' responses to questions related to their personal history (e.g., age, number of children) were recoded (“yes” = 1 and “no” = 0) and entered into a correlational analysis. Discrimination and years of lie-detection experience were not significantly correlated in Experiment 1 ($r(104) = -.05, p = .60$), Experiment 2 ($r(105) = -.04, p = .71$), or Experiment 3 ($r(125) = -.08, p = .37$). There were only a few, small, significant correlations between

discrimination and other variables. No correlations replicated across experiments and were likely due to chance. Thus, this information will not be reported in Experiments 2–4.

Discussion

Overall, participants could not accurately classify the truth-tellers and the lie-tellers. In fact, customs officers and students performed at chance, with police officers performing more poorly than the other two groups (i.e., below chance). However, police officers were the only group to accurately discriminate between children. None of the groups were overtly biased, indicating that their performance was not due to a tendency to favor a particular response. Despite their inability to identify children's statements correctly, customs officers were more certain about their decisions than the other groups.

Only police officers could recognize that there were two qualitatively different groups (lie-tellers and truth-tellers), but they mislabeled the lie-tellers as truth-tellers and truth-tellers as lie-tellers. This group could actually sense differences and did not simply respond that all children were lying, as indicated by the absence of bias. However, being able to differentiate between two groups does not necessarily translate into being able to recognize what is being discriminated. Perhaps police officers had identified markers for distinguishing between lie- and truth-tellers and based their decisions on an effective set of criteria, but misinterpreted the cues. This may account for police officers' poor accuracy, but above-chance discrimination ability.

Adults' inability to detect children's lies was not necessarily the reason for the low levels of accuracy in the present experiment. Rather, poor performance may have been due to the difficulty of the task. In particular, the video clips may have been too brief. Participants might not have been accustomed to making decisions within a short period of time. Even customs officers, who must identify deception quickly, are often given longer than 18 seconds (the average length of the clips in the present experiment) to render their judgments. In the shortest lie-detection tasks in the existing studies, participants were given more than 20 seconds to observe the potential deceivers (e.g., Keating & Heltman, 1994). In fact, several participants in the present experiment complained that the clips were too short to allow for correct categorizations. Thus, it is possible that the adults' true lie-detection abilities were obscured by the unreasonable demands of the task.

EXPERIMENT 2

Two issues were examined in Experiment 2. The length of the video clips was increased to determine whether this factor affected lie-detection. Also, the ability to detect children's lies was tested in a different context. In forensic situations, interviewers may engage in moral reasoning tasks before beginning a fact-finding interview for the purpose of establishing children's competence to testify (Haugaard, Reppuci, Laird, & Naful, 1991; Myers, 1996). Often, this will include a discussion of the concepts of lie- and truth-telling to determine children's level of understanding and their commitment to telling the truth (Lyon, 2000). Researchers have found that children's performance on these tasks is unrelated to their propensity to deceive (Bala, Lee, Lindsay, & Talwar, 2001). However, it is possible that moral discussions increase children's awareness of the importance of truth-telling and the negativity of lying and their emotional reactions to their own deceit. As a result, children may be more ill at ease when telling lies and their deception may become easier to detect.

Method

Participants—Twenty-eight police officers (4 women and 24 men, M age = 41.71 years, SD = 5.39) and 35 customs officers (18 women and 17 men, M age = 34.91 years, SD = 9.57)

participated. Forty undergraduates (28 women and 12 men, M age = 19.00 years, SD = 1.92) were awarded course credit for their participation.

Materials

Videotapes: Footage of children (different from Experiment 1) was compiled. In this study, 4- to 8-year-olds engaged in moral reasoning tasks. Specifically, a female experimenter read stories that involved lie-telling characters. For example, one story featured a girl who denied eating a forbidden candy while a teacher was out of the room. Following each story, the experimenter asked questions about the character's actions: (1) "Was what X [the character's name] said the truth or a lie?" (2) "Was it good or bad?" (3) "Was it a little good/bad or very good/bad?" (4) "Why?". The rest of the procedure was similar to that of Experiment 1 except for a few modifications. Due to the limited number of children who did not peek and, therefore, did not lie and the need to maintain a 50–50 lie-teller to truth-teller ratio, the number of clips seen by each participant was reduced. In all, 30 video clips (M length = 33.27 seconds, SD = 7.45) were produced from the recorded exchanges. A total of 20 clips appeared on each tape version (with the same truth-tellers and different lie-tellers on each tape).

To address the problems associated with the short time that observers had to familiarize themselves with the children prior to making their decisions, participants viewed an extra segment in which children were asked, and responded to, questions that were unrelated to the transgression (i.e., a neutral exchange). Each clip was divided into three parts: 1) a neutral exchange between the experimenter and the child; 2) black screen (to differentiate clearly between sections); 3) the experimenter asking the critical questions about the transgression and the children's responses. *It should be noted that participants did not view the moral discussion portion of the children's interviews.*

Procedure—The procedure was identical to that of Experiment 1 except for one modification: the female experimenter explained the presence of the additional neutral information. The experimenter emphasized that the children's responses to the neutral questions were completely unrelated to their propensity to peek and/or lie.

Results

Lie-Detection Accuracy—A one-way ANOVA revealed no significant differences between customs officers (M = .62, SD = .10), police officers (M = .64, SD = .13) and students (M = .63, SD = .13) in terms of accuracy, $F(2, 100) = .33$, $p = .72$, $r = .06$ (see Fig. 1). One-sample t -tests revealed that customs officers ($t(34) = 7.20$, $p < .001$), police officers ($t(27) = 5.85$, $p < .001$), and students ($t(39) = 6.71$, $p < .001$) all were more accurate than expected by chance.

Signal Detection Analysis

Discrimination: A one-way ANOVA did not reveal any significant group differences, $F(2, 100) = .36$, $p = .70$, $r = .06$ (see Table 1). When comparing d' to zero, customs officers (M = 0.66, SD = 0.56), $t(34) = 6.97$, $p < .001$, police officers (M = 0.80, SD = 0.73), $t(27) = 5.78$, $p < .001$ and students (M = 0.75, SD = 0.73), $t(39) = 6.46$, $p < .001$ reliably discriminated truth-tellers from lie-tellers.

Bias: A one-way ANOVA did not reveal any differences between groups in terms of bias, $F(2, 100) = .01$, $p = .99$, $r = .01$. t -tests comparing each β to 1 (no bias) revealed no significant bias in any of the groups.

Certainty—A one-way ANOVA revealed a significant group effect of certainty, $F(2, 100) = 12.32$, $p < .001$, $r = .33$ (see Table 2). Post hoc analyses (Tukey's HSD) revealed that police

officers ($M = 1.65, SD = 0.36$) and students ($M = 1.59, SD = 0.36$) were equally certain about their ratings. However, customs officers ($M = 2.05, SD = 0.54$) showed higher levels of certainty than both the police officers ($r = .39$) and students ($r = .46$) (Tukey's HSD, $p < .01$). The correlations between discrimination and certainty were not significant for customs officers ($r(34) = .20, p = .26$), police officers ($r(27) = -.18, p = .33$), and students ($r(39) = .10, p = .54$).

Discussion

All groups were significantly more accurate than expected by chance. Signal detection analyses revealed that all groups were sensitive to differences between lying and truth-telling children. In addition, there was no evidence of response biases in any of the groups. Finally, despite customs officers being very certain about their decisions relative to the students and police officers, their lie-detection performance was unrelated to certainty.

Findings in the present experiment differed from those reported in Experiment 1. Whereas participants in the first experiment performed either at or below chance levels, here the three groups were more accurate and discriminated between child lie-tellers and truth-tellers better than expected by chance. There are two possible explanations for these differences. First, since the clips in the present experiment included a neutral discussion, they were longer and may have afforded participants a greater amount of time to study the children's expressions. In addition, the groups may have compared children's behaviors during the neutral exchange and critical question sections. Familiarity effects were explored in Experiment 4. Second, increasing the salience of the moral implications of lying may have interfered with children's ability to regulate their expressive behavior, or increased their arousal, and made their deception easier to detect.

EXPERIMENT 3

It should be noted that discussing the moral implication of lying is only one method used to promote truth-telling in the legal system. In North American courtrooms, children are asked to promise to tell the truth prior to testifying. By emphasizing the importance of truth-telling in this way, children may also have difficulty successfully regulating nonverbal behaviors when lying. As a result, differences between lie-tellers and truth-tellers may be more evident, allowing observers to classify the children correctly. Experiment 3 examined this possibility.

Recent evidence suggests that children's verbal deception varies with the context of the interview. Talwar, Lee, Bala, and Lindsay (2002) revealed that the frequency of lying significantly decreased when children first promised to tell the truth (from approximately 80%, when they did not promise to tell the truth, to approximately 60% after promising to tell the truth). Given that having the children promise to tell the truth reduced their tendency to lie, it is possible that there are corresponding changes in children's nonverbal behavior if they decide to lie after promising to tell the truth. In this context, lie-tellers would not only have to contend with the negative implications associated with deception, but also the added stress of breaking a promise. This manipulation was expected to affect the children's ability to regulate their nonverbal responses, making their deception more apparent. In turn, law enforcement officials and students would be able to classify children at above-chance levels.

Method

Participants—The law enforcement group was comprised of 47 police officers (8 women and 39 men, M age = 33.74 years, $SD = 8.11$) and 40 customs officers (15 women and 25 men, M age = 35.38 years, $SD = 9.55$). The student sample was composed of 39 undergraduates (31 women and 8 men, M age = 19.10 years, $SD = 1.64$). Students received course credit for their involvement.

Materials

Videotapes: The videotapes featured different children from those in Experiments 1 and 2. Again, children were asked not to peek at a toy when an experimenter was out of the room. Upon the experimenter's return, she asked the children questions about the concept of a promise. Then, the experimenter asked the children to promise to tell the truth. Once this was done, she asked the children "While I was gone, did you peek to see who it [the toy] was?"

In all, a set of 39 video clips (M length = 32.47 seconds, $SD = 5.80$) was compiled. A total of 26 clips appeared on each tape (with the same truth-tellers and 13 different lie-tellers being used for each tape). As in Experiment 2, a set of three segments was produced for each child: 1) a neutral exchange; 2) black screen; 3) critical question about the transgression. The critical question was repeated on the video three times so that it would be comparable to the other two experiments in terms of length. *Footage of the children promising to tell the truth was not shown to the participants.* All other aspects of the videotapes were similar to Experiment 2.

Results

Lie-Detection Accuracy—A one-way ANOVA failed to reveal differences between customs officers' ($M = .57$, $SD = .10$), police officers' ($M = .59$, $SD = .08$), and students' ($M = .61$, $SD = .09$) accuracy, $F(2, 123) = .15$, $p = .24$, $r = .03$ (see Fig. 1). t -tests revealed that customs officers ($t(39) = 4.77$, $p < .001$), police officers ($t(46) = 7.82$, $p < .001$), and students ($t(38) = 7.13$, $p < .001$) were more accurate than expected by chance.

Signal Detection Analysis

Discrimination: An ANOVA did not reveal any differences between customs officers ($M = 0.37$, $SD = 0.53$), police officers ($M = 0.40$, $SD = 0.44$), and students ($M = 0.49$, $SD = 0.49$), $F(2, 123) = .69$, $p = .51$, $r = .07$ (see Table 1). However, customs officers, police officers, and students could reliably distinguish between lie-tellers and truth-tellers, $t(39) = 4.33$, $p < .001$, $t(46) = 6.23$, $p < .001$, and $t(38) = 6.28$, $p < .001$, respectively.

Bias: A one-way ANOVA revealed a significant effect of occupation, $F(2, 123) = 5.34$, $p < .01$, $r = .20$ (see Table 1). Post hoc analyses revealed that customs officers ($M = 1.05$, $SD = 0.16$) and police officers ($M = 0.95$, $SD = 0.12$) were differentially biased, $r = .32$. There was no difference between students ($M = 0.97$, $SD = 0.14$) and the other groups. Overall, police officers were significantly biased towards indicating that children were lying, $t(46) = -2.58$, $p < .05$. Customs officers and university students were not significantly biased in either direction, $t(38) = -1.36$, $p > .05$, and $t(39) = 1.88$, $p > .05$, respectively.

Certainty—An ANOVA revealed that customs officers ($M = 2.21$, $SD = 0.50$) were more certain about their decisions than police officers ($M = 1.85$, $SD = 0.46$, $r = .35$) and students ($M = 1.79$, $SD = 0.41$, $r = .43$), $F(2, 105) = 10.13$, $p < .001$, $r = .30$ (see Table 2). The difference between police officers and students was not significant. Finally, customs officers' ($r(39) = .26$, $p = .11$), police officers' ($r(46) = .17$, $p = .25$), and students' ($r(39) = .19$, $p = .26$) certainty was unrelated to sensitivity.

Discussion

All groups could accurately identify the two types of children at above chance levels. Signal detection analyses revealed that only police officers were slightly biased to categorize children as lie-tellers. Nevertheless, all groups could discriminate between lie-tellers and truth-tellers. Overall, asking children to promise to tell the truth appeared to facilitate the detection of children's lies. Although customs officers were the most certain about their decisions, certainty and discrimination ability were unrelated.

EXPERIMENT 4

Experiments 2 and 3 contained a possible confound. In both experiments, participants saw an experimenter interacting with the children in a neutral situation (prior to children's responses to the critical questions). Thus, it is unknown whether higher accuracy in Experiments 2 and 3 was due to a familiarity effect (i.e., previous exposure to the children in the neutral exchange section of the interview) or the type of interviews that the children received (i.e., the inclusion of moral reasoning tasks or promising to tell the truth). Experiment 4 was designed specifically to examine these possibilities.

All of the clips from the previous studies (Experiments 1–3) were included to compare participants' ability to identify deception across the three types of interviews. It was predicted that children who reasoned about the moral implications of lying (Experiment 2) or made a commitment to truth-telling (Experiment 3) would be easier to classify than those who had not (Experiment 1). Given the overall lack of group differences in previous experiments, only university students were used to test these hypotheses in this experiment.

METHOD

Participants—One hundred undergraduates (60 women and 33 men, M age = 19.54 years, $SD = 1.49$; 7 did not disclose their sex) participated in this experiment in exchange for course credit.

Materials

Videotapes: There were two experimental conditions. The “familiarization” condition contained all clips from the previous studies (M length = 24.60 seconds, $SD = 10.10$), plus short segments featuring the child's neutral exchange with the interviewer before the critical interview. These clips were divided into three segments: 1) a neutral interaction; 2) black screen; 3) critical questions about the transgression (“Did you peek?”). In the “nonfamiliarization” condition, clips contained only the critical questions (M length = 17.16 seconds, $SD = 6.48$).

Two videotape versions were constructed for this experiment. The first tape version was created by including the clips from the Version A tapes used in Experiments 1–3. The second tape was comprised of the clips used in Version B tapes from Experiments 1–3. In all, compiling the stimuli from the previous experiments produced a set of 81 video clips per tape (of which 35 clips were from Experiment 1, 20 clips from Experiment 2, and 26 clips from Experiment 3). The same truth-tellers appeared on both Tapes A and B (due to the limited number of children who did not transgress), but there were different lie-tellers on each tape. Clips were randomly transferred onto videotape with a restriction that no more than three children from any one type of interview (direct, moral discussion, or promise) appeared in a row. Each participant was assigned to one condition and saw only one tape.

Procedure—The procedure was similar to that used in Experiments 1–3, with one exception: participants were given a short break at the halfway mark in the rating task (clip number 41) to avoid fatigue. During this time, participants were asked to fill out a questionnaire requesting personal information. After the questionnaire was completed, the rating task continued.

Results

Lie-detection Accuracy—Students' accuracy was examined using a 2 (Familiarization) \times 3 (Interview type) mixed factors ANOVA (with the last variable as the repeated measure) (see Fig. 2). Only a main effect of interview type was significant, $F(2, 188) = 47.73, p < .001, r = .45$. As there are no known post hoc tests, nor effect size measures, designed for repeated

measures analyses, direct comparisons were made with paired *t*-tests. Participants were more accurate when rating moral discussion clips ($M = .66, SD = .11$) than direct interviews ($M = .54, SD = .09$), $t(95) = -8.86, p < .001$. In addition, participants were more accurate when rating promise clips ($M = .56, SD = .09$) than direct interviews, $t(95) = -2.14, p < .05$. Finally, participants were significantly more accurate when rating moral discussion clips than promise clips, $t(95) = 6.83, p < .001$. Thus, it may be said that participants were most accurate when viewing moral discussion interviews, followed by promise clips and, finally, direct interviews. Overall, participants were significantly more accurate than chance when rating children who received a direct interview ($t(95) = 3.85, p < .001$), children who had engaged in moral discussions ($t(95) = 14.59, p < .001$), and children who had promised to tell the truth ($t(95) = 5.92, p < .001$).

Signal Detection Analysis

Discrimination: A 2 (Sex) \times 2 (Familiarization) \times 3 (Interview type) mixed factors ANOVA (with the last variable as the within-subjects factor) was performed on the students' ability to differentiate between lie-tellers and truth-tellers (see Table 1). There was a significant main effect of interview type, $F(2, 178) = 32.40, p < .001, r = .39$ and an interaction between sex and interview type, $F(2, 178) = 3.19, p < .05, r = .13$. Further analyses revealed that male and female participants performed equally when rating children who had promised to tell the truth or engaged in the direct interview, but women ($M = 0.83, SD = 0.69$) were better able than men ($M = 0.53, SD = 0.58$) to discriminate between lie-tellers and truth-tellers when observing children who had engaged in moral discussions. There were no other significant main effects or interactions. However, participants could reliably distinguish between lie-tellers and controls who had engaged in moral discussions, $t(95) = 10.92, p < .001$ or promised to tell the truth, $t(95) = 3.33, p < .01$, but not children who had received a direct interview, $t(95) = 0.62, p = .54$.

Bias: A 2 (Familiarization) \times 3 (Interview type) mixed factors ANOVA (with the last variable as the repeated measure) was performed on participants' β (see Table 1). There were no significant main effects or interactions. Comparisons to 1 (no bias) revealed that students were significantly more likely to indicate that children were telling the truth when rating the moral discussion ($t(95) = 2.35, p < .05$) and promise ($t(95) = 2.86, p < .01$) clips. However, when they viewed direct interviews, there was no evidence of bias, $t(95) = 1.80, p = .08$.

Certainty—A 2 (Familiarization) \times 3 (Interview type) mixed factors ANOVA (with the last variable as the repeated measure) was performed on participants' certainty scores (see Table 2). Participants expressed significantly higher levels of certainty when rating moral discussion clips as opposed to those from the other types of interview, $F(2, 188) = 11.79, p < .001, r = .24$. Also, there was a significant interaction between the interview type and familiarization, $F(2, 188) = 3.37, p < .05, r = .13$. When they were not familiarized with the children, participants were more certain about their decisions when viewing clips of the moral discussion interviews than the other two types (and there were no differences in certainty scores between the latter types of clips). However, when they were familiarized with the children, students were more certain about ratings of the moral judgment and promise clips than the direct interview clips. There were no significant correlations between certainty and ability to differentiate between children in direct ($r(95) = .12, p = .24$), moral discussion ($r(95) = .11, p = .30$) and promise interviews ($r(95) = -.02, p = .88$).

Discussion

This experiment extended the findings of Experiments 1–3. When viewing all types of clips, participants could correctly classify the children above chance levels. Participants were most accurate when rating clips of children who engaged in moral reasoning tasks. Despite a bias

to report that children in moral discussion and promise clips were truth-tellers, participants were able to discriminate between the lie-tellers and the truth-tellers who had received these interviews. Familiarization with the children did not affect discrimination, indicating that findings in Experiments 2 and 3 were due to the type of interview rather than the length of exposure. Participants were more certain about their ratings of moral discussion and promise interviews when they had viewed extra footage of the children. However, certainty was unrelated to discrimination ability for all types of interviews.

One unexpected finding involved a sex difference in discrimination ability. Specifically, females outperformed males when they rated clips from the moral discussion, but not other, interviews. There is no ready explanation for this finding. Previous research on sex-related differences in accuracy is mixed (e.g., Forrest & Feldman, 2000; Zuckerman, DePaulo, & Rosenthal, 1981). In addition, no other sex differences were revealed in any of Experiments 1–3. This suggests that the present finding may be attributable to a Type I error.

Contrary to our predictions, familiarity did not improve overall accuracy or discrimination. It is possible that the clips were still too short to provide valuable information for lie detection, or that extra, neutral information might not promote successful lie detection. These possibilities must be addressed in specifically designed studies in the future. Nevertheless, Ambady and Rosenthal's (1992) meta-analysis has shown that the length of the interview does not affect accuracy. Even when the length and number of interviews are increased, researchers have failed to find improvements in performance (e.g., Granhag & Strömwall, 2001).

Consideration of Age Effects—Preliminary analyses failed to reveal consistent age effects across experiments. However, this may have been due to unequal samples within age groups. To address this issue, we conducted a logistic regression to determine whether the children's ages predicted accuracy. Age significantly predicted accuracy in Experiment 1 (Wald = 29.45, $p < .001$), Experiment 2 (Wald = 92.07, $p < .001$), and Experiment 3 (Wald = 11.77, $p < .01$). These findings were replicated in Experiment 4, Wald = 31.89, $p < .001$. Generally, participants were most accurate when rating the youngest children (see Table 3).

It is possible that deceptive abilities are refined with age. Yet, any firm conclusions are beyond the scope of this paper. In particular, there remain significant problems associated with the sample sizes. Analyses do not support the notion that easily detectable outliers are responsible for other condition effects (i.e., each experiment contained equal proportions of "blatant" lie-tellers and truth-tellers). However, due to the small age group sample sizes, the presence of even a single, poor lie-teller may have been responsible for artificially inflating age effects. In addition, it is not clear whether young children are less capable lie-tellers or observers have a greater ability to detect deception in this group. The present experiments were not adequate, nor designed, to examine the development of effective deception. At most, our findings suggest that this area of investigation should be the subject of future research.

Consideration of Arousal—We speculated that engaging in moral discussions or breaking a promise could provoke strong emotional reactions. Perhaps due to this duress, or arousal, children's lies were easier to detect. To test this hypothesis, a team of independent raters coded each child's arousal using a 9-point scale (1 = *not at all aroused*; 9 = *extremely aroused*). A 3 (Interview type) \times 2 (Lie vs. Truth) ANOVA failed to reveal any differences between truth-tellers and lie-tellers in terms of arousal, $F(1, 114) = .15, p = .70, r = .04$. However, there was a significant main effect of interview type, $F(2, 114) = 3.16, p < .05, r = .16$. Post hoc analyses revealed that children appeared more aroused when they had engaged in direct interviews ($M = 5.41, SD = .60$) than when they had promised to tell the truth ($M = 5.00, SD = .77$), Tukey's HSD, $p < .05$. There were no differences between children who had engaged in moral

discussions ($M = 5.31$, $SD = .77$) and any other type of interview. There were no significant interactions.

GENERAL DISCUSSION

The present set of experiments examined whether customs officers, police officers, and university students could detect children's deception. When children received a direct interview, the correct identification rates of the customs officers, police officers, and students were near chance levels (Experiment 1). However, when children had engaged in moral discussions (Experiment 2) or been asked to promise to tell the truth (Experiment 3), all groups could accurately identify the lie-tellers and truth-tellers above chance levels. Experiment 4 replicated the major findings of Experiments 1–3 and ruled out the possibility that the adults' accuracy in Experiments 2 and 3 was due to additional exposure to the children.

Signal detection theory analyses further confirmed that the three groups could discriminate between child lie-tellers and truth-tellers. Throughout most of the four experiments, adult participants did not display a response bias. When significant response biases were observed, they tended to be very small. Also, the biases observed in one experiment were not replicated in another experiment. With regard to discrimination ability (d'), discussing the moral implications of lying (Experiment 2) facilitated discrimination the most, followed by having children promise to tell the truth (Experiment 3). Increasing participants' exposure to the children did not improve discrimination (Experiment 4).

The significant condition effects are worth noting. Deception was consistently most detectable when, prior to lying, children had considered the moral implications of deceit. During story-telling sessions, the majority of children indicated that the characters were lying, which was "very bad". This moral discussion might have alerted children to the possibility that their own deceptive actions were equally negative. Whereas emphasizing the impropriety of lying may have affected children in the moral discussion interviews, emphasizing the importance of truth-telling (in the form of making a promise) may have had the same effect in the promise condition. We posited that each approach would increase children's arousal during deception, accounting for why their lies were easier to detect.

Yet, recent analyses failed to support our hypotheses about the effects of arousal. Lie-tellers were no more aroused than truth-tellers. More surprisingly, the only effect of interview type revealed that children who had promised to tell the truth were believed to be less aroused than children who had engaged in a direct interview. There are several explanations for our failure to find the expected arousal effects. First, it is possible that it is difficult to accurately assess arousal through observation alone. Second, the arousal related to deception may be as subtle as that resulting from cognitive dissonance. In that case, arousal did occur, but it was not readily detectable by physiological measures (i.e., it was only revealed through attributions) (e.g., Zanna & Cooper, 1974). Third, the interview manipulations may have increased arousal. However, children may simply have been actively suppressing it to a greater extent in the moral discussion and promise clips. In turn, this increase in cognitive effort may have led to the expression, or leakage, of the deception elsewhere. Finally, it is possible that our hypotheses are simply incorrect and that changes in arousal are not responsible for corresponding increases in lie detection. Of course, all of these alternatives are merely speculation and should be thoroughly tested in future studies.

It should be noted that, although participants' accuracy was above chance in some conditions, their overall ability to detect children's lies was poor. In the best displays of lie-detection accuracy (i.e., when rating moral discussion clips), adults classified less than 70% of the children correctly; over 30% of the children were labeled incorrectly. In fact, the average

accurate identification rate across conditions and groups was just slightly above the 50% chance level. These findings are similar to those reported in other studies of adults' deception (e.g., Ekman et al., 1999). This level of accuracy, although significantly different from chance, is not clinically or naturalistically meaningful. Most members of the justice system would surely be uncomfortable with such a low level of predictive success. Although there were not enough children who correctly (and incorrectly) admitted to having committed the transgression to perform full lie-detection analyses, consideration of the issue revealed that participants believed approximately 90% of confessors. Any firm conclusions about participants' willingness to believe truthful versus false confessions are beyond the scope of this paper, but the lack of perfect accuracy suggests that this issue should be explored further. Overall, the findings related to the study of adult deception may be generalizable to examinations of child lie-tellers.

As in previous studies with adult deceivers (e.g., Kraut & Poe, 1980; Porter, Woodworth, & Birt, 2000), there were few differences between the performance of our experienced and untrained groups. Overall, students performed as well as customs officers and, in one case, both of these groups were more accurate than police officers. Thus, experience and training did not appear to affect the successful identification of deception. However, it may be unrealistic to expect law enforcement officials to perform better than untrained adults. Officers rarely receive feedback about their accuracy (DePaulo & Pfeifer, 1986). Arrests and convictions are not necessarily indicative of guilt. Moreover, officials can never truly establish a base rate because they rarely know whether people they have not interviewed (or whom they have spoken to and let free) were lying. As a result, it is difficult for them to learn from encounters and adjust the criteria they use to detect deceit. Given this uncertainty, officers may acquire experience with lie detection without improving their skills. Furthermore, because lie detection is seen as an important skill in law enforcement, pressure to improve in the absence of feedback may encourage superstitious behavior (e.g., illusory correlations between behavior and perceived truthfulness). Such cues could distract officers from useful information that may permit improved performance. Thus, it is not surprising that officials do not perform better than laypersons. It is interesting to note that experience with children also does not improve adults' detection of children's lies. About half of the officers had children of their own. However, their lie-detection accuracy did not differ from those officers who did not have children. This result is consistent with Talwar and Lee's (2002) finding that parents could not detect children's lies.

Previous research has shown that law enforcement officials tend to be more confident than untrained observers even though confidence and accuracy are often unrelated (e.g., DePaulo & Pfeifer, 1986; DePaulo et al., 1997). The present findings are highly consistent with this conclusion. Daily work experience may be one reason for the group differences in certainty ratings. The customs officers were trained to detect deception with minimal information, which might have made them more comfortable with the present task. As a result, they were more certain about their decisions. On the other hand, the police officers were more accustomed to lengthy investigations and extensive information gathering. Similar to students, who had no training or experience, they may have been unfamiliar with the task demands (e.g., rapid decision-making with limited information) and more uncertain about their decisions.

To the best of our knowledge, the present study is the first in the literature that has systematically examined law enforcement officials' ability to detect children's lies. However, there are several limitations of the studies and further empirical research is urgently needed. The first issue is the generalizability of the results. In the present experiments, participants were asked to detect children's lies about peeking at a toy. There are certain benefits and disadvantages of this approach. Unlike previous studies, the present procedure allows children to lie out of their own volition and provides realistic samples of deception. One weakness is

the less-than-serious nature of the lie produced. Law enforcement officers are rarely asked to judge such minor transgressions. The majority of children who enter the justice system are interviewed about serious matters (e.g., child abuse). Thus, procedures (e.g., having children promise to tell the truth) must be examined in more legally relevant contexts in which children face as severe consequences as they would encounter in real life situations. Of course, ethical issues must be taken into account when designing such studies.

The second issue concerns the amount of information provided to the adult participants in the present experiments. In real life situations, law enforcement officials usually conduct their own interviews. Not only can they view reactions first-hand (rather than on videotape), but they can ask their own questions and follow-up on any inconsistencies. Direct encounters with children may enhance law enforcement officials' accuracy at detecting children lies. One reviewer suggested that the groups under investigation likely experienced different rates of lying and truth-telling in the real world (e.g., police officers might be exposed to more lie-tellers than university students). In turn, base rate expectancies provided in the instructions (e.g., that approximately half of the children presented would be lying), which were meant to reassure participants that they were not being tricked, might have artificially altered the groups' performance. Although there is no proof that the inclusion of base rates in the instructions actually impacted accuracy, other evidence suggests that it had little effect. First, providing base rates is typical of other studies of lie detection (e.g., Frank & Ekman, 1997). Second, due to the base rate fallacy, there was no real reason to expect that participants would actively use, or be sensitive to, the base rates. Perhaps if the purpose of the experiments were to determine the base rate expectancies of each group (and the resulting effect on accuracy), the instructions would have been flawed. Instead, the base rate information was needed to allow for a fair comparison, in terms of performance, across groups. There is no reason to think that base rate distortions are limited to groups (i.e., members within groups may hold dramatically divergent beliefs about the rate of lie-telling and truth-telling in the world). If participants were left to use their personal base rate expectancies, performance could have been distorted due to different expectations. This possibility remains to be examined in future studies. Finally, signal detection analysis is specifically designed to provide a measure of detection accuracy that is independent of response bias (base rate expectations). The d' values reported reflect the ability to discriminate between truthful and untruthful responses, independent of the individuals' beliefs about the overall likelihood of lying.

The third limitation is that it is not clear whether children revealed any markers of deception and whether adults consistently use these cues when making their decisions. Future studies are needed to examine children's nonverbal behavior during deception. For example, the child lie-tellers and truth-tellers in the present experiments could be compared in terms of their facial expressions and body movements. It is possible that participants' chance performance was partially influenced by outliers (i.e., children who were very easy and difficult to classify). Although no child was accurately identified (as a lie- or truth-teller) 100% of the time, further analyses could examine the variability, in terms of lie-telling ability, across children. Also, the adult participants could be asked about the cues they relied upon when making their decisions. Another set of studies arising from this issue could investigate whether lie-detection ability is stable. It may not be surprising that participants performed at chance if they were actually guessing. Measuring accuracy over time (both within and across sessions) could indicate whether performance is due to random fluctuations (e.g., luck) or enduring ability. Thus, analyses of the children's nonverbal behavior and adults' test-retest reliability at detecting children's deception would clarify whether adults' performance in the present set of experiments was due to their inability to detect deception at all or the ability of some children to lie effectively.

Despite the limitations of the present experiments, the findings have important implications for the legal system. Law enforcement officials' poor performance in the present experiments may debunk a common belief that children are unable to effectively deceive adults. Frontline workers should be made aware of these findings so that children's lie-telling skills are not underestimated. Our results indicate that interviewers must exercise caution when dealing with young witnesses, gathering concrete evidence rather than relying on instincts. Our findings speak to the danger of becoming overly confident about one's ability to detect children's lies simply because one has extensive experience with children or one's job calls for the determination of truth.

The condition effect is instructive for legal professionals who seek to construct interview procedures that more effectively elicit truthful testimony from children. When children testify in most North American courts, they must undergo a "competence examination." In this examination, children are asked to discuss the moral implications of lying and tend to be asked to take an oath or to promise to tell the truth. Talwar et al. (2002) have recommended that interviewers only ask children to promise to tell the truth. They proposed that "correctly" answering questions about lying and truth-telling should not be a precondition of children testifying. These suggestions were based on their findings that only having children promise to tell the truth decreased the incidence of lying; moral discussion did not change the rate of deception. However, the present research suggests that, although moral discussion does not increase the likelihood of truth-telling, the inclusion of moral reasoning tasks consistently facilitates the detection of children's lies. Given the relatively short amount of time needed for moral discussion, the possible benefits in terms of increased lie-detection accuracy seem to justify its continued inclusion in forensic interviews and court. Because having children promise to tell the truth appears to decrease lie production and increase lie detection, both practices (i.e., having children promise to tell the truth and moral discussions about truth and lie-telling) should be incorporated into child witness interview procedures.

REFERENCES

- Ambady N, Rosenthal R. Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychological Bulletin* 1992;111:236–274.
- Bala N, Lee K, Lindsay RCL, Talwar V. A legal and psychological critique of the present approach to the assessment of the competence of child witnesses. *Osgoode Hall Law Journal* 2001;38(3):409–452.
- Bond CF Jr, Omar A, Pitre U, Lashley BR, Skaggs LM, Kirk CT. Fishy-looking liars: Deception judgment from expectancy violation. *Journal of Personality and Social Psychology* 1992;63:969–977. [PubMed: 1460563]
- Bruck M, Ceci SJ, Hembrooke H. Reliability and credibility of young children's reports: From research to policy and practice. *American Psychologist* 1998;53:136–151. [PubMed: 9491744]
- Ceci SJ, Bruck M. Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin* 1993;113:403–439. [PubMed: 8316609]
- Ceci SJ, Huffman MLC. How suggestible are preschool children? Cognitive and social factors. *Journal of the Academy of Child and Adolescent Psychiatry* 1997;36:948–958.
- Depaulo BM. Spotting lies: Can humans learn to do better? *Psychological Science* 1994;3:83–86.
- Depaulo BM, Charlton K, Cooper H, Lindsay JJ, Muhlenbruk L. The accuracy-confidence correlation in the detection of deception. *Personality and Social Psychology Review* 1997;1:346–357. [PubMed: 15661668]
- Depaulo PJ, Depaulo BM. Can attempted deception by salespersons and customers be detected through nonverbal behavioral cues? *Journal of Applied Social Psychology* 1989;19:1552–1577.
- Depaulo BM, Pfeifer RL. On-the-job experience and skill at detecting deception. *Journal of Applied Social Psychology* 1986;16:249–267.
- Depaulo BM, Rosenthal R. Telling lies. *Journal of Personality and Social Psychology* 1979;37:1713–1722. [PubMed: 512835]

- Depaulo BM, Zuckerman M, Rosenthal R. Humans as lie detectors. *Journal of Communication* 1980;30:129–139. [PubMed: 7381035]
- Ekman P, Friesen WV. Detecting deception from the body or face. *Journal of Personality and Social Psychology* 1974;29:288–298.
- Ekman P, O'sullivan M, Frank MG. A few can catch liars. *Psychological Science* 1999;10:263–266.
- Ekman P, O'sullivan M, Friesen WV, Scherer KS. Face voice and body in detecting deceit. *Journal of Nonverbal Behavior* 1991;15:125–135.
- Ekman P, O'Sullivan M. Who can catch a liar? *American Psychologist* 1991;9:913–920. [PubMed: 1958011]
- Etcoff NL, Ekman P, Magee JJ, Frank MG. Lie detection and language comprehension. *Nature* 2000;405:139. [PubMed: 10821259]
- Feldman RS, Jerkins L, Popoola O. Detection of deception in adults and children via facial expressions. *Child Development* 1979;50:350–355.
- Feldman RS, White JB. Detecting deception in children. *Journal of Communication* 1980;30:121–128. [PubMed: 7381034]
- Forrest JA, Feldman RS. Detecting deception and judge's involvement: Lower task involvement leads to better lie detection. *Society for Personality and Social Psychology* 2000;26:118–125.
- Frank FG, Ekman P. The ability to detect deceit generalizes across different types of high-stakes lies. *Journal of Personality and Social Psychology* 1997;72:1429–1439. [PubMed: 9177024]
- Goodman GS, Batterman-Faunce JM, Schaaf JM, Kenney R. Nearly 4 years after an event: children's eyewitness memory and adults' perception of children's accuracy. *Child Abuse and Neglect* 2002;26:849–884. [PubMed: 12363335]
- Granhag PA, Strömwell LA. Deception detection based on repeated interrogations. *Legal and Criminological Psychology* 2001;6:85–101.
- Haugaard JJ, Reppuci ND, Laird J, Nauful T. Children's definitions of the truth and their competency as witnesses in legal proceedings. *Law and Human Behavior* 1991;17:645–659.
- Keating CF, Heltman KR. Dominance and deception in children and adults: Are leaders the best misleaders? *Society for Personality and Social Psychology* 1994;20:312–321.
- Kraut RE, Poe D. Behavioral roots of person perception: The deception judgments of customs inspectors and laypersons. *Journal of Personality and Social Psychology* 1980;39:784–798.
- Lewis M, Stranger C, Sullivan MW. Deception in 3-year-olds. *Developmental Psychology* 1989;25:439–443.
- Lyon TD. Child witnesses and the oath: Empirical evidence. *Southern California Review* 2000;73:1017–1074.
- Myers JEB. A decade of international legal reform regarding child abuse investigation and litigation: Steps toward a child witness code. *Pacific Law Journal* 1996;28:169–241.
- Nysse, K.; Bottoms, BL.; Lindsay, RCL. Experts' and novices' abilities to detect deception in children. Child witnesses: Competence and truth telling; Symposium presented at the American Psychology–Law Society Meeting; New Orleans, LA. 2000 Mar. (Chair)
- Porter S, Woodworth M, Birt AR. Truth, lies, and videotape: An investigation of the ability of federal parole officers to detect deception. *Law and Human Behavior* 2000;24:643–658. [PubMed: 11105477]
- Riggio R, Tucker J, Throckmorton B. Social skills and deception ability. *Personality and Social Psychology Bulletin* 1988;13:568–577.
- Rosenthal, R. *Meta-analytic procedures for social research*. London: Sage; 1991.
- Saarni C. Children's understanding of display rules for expressive behavior. *Developmental Psychology* 1979;4:424–429.
- Saarni, C.; Salisch, MV. The socialization of emotional dissemblance. In: Lewis, M.; Saarni, C., editors. *Lying and deception in everyday life*. New York, NY, US: The Guilford Press; 1993. p. 106-125.
- Talwar V, Lee K. Development of lying to conceal a transgression: Children's control of expressive behavior during verbal deception. *International Journal of Behavioral Development*. 2002

- Talwar V, Lee K, Bala N, Lindsay RCL. Children's conceptual knowledge of lying and its relation to their actual behaviors: Implications for court competence examinations. *Law and Human Behavior* 2002;26:395–415. [PubMed: 12182530]
- Zanna MP, Cooper J. Dissonance and the pill: An attribution approach to studying the arousal properties of dissonance. *Journal of Personality and Social Psychology* 1974;29:703–709. [PubMed: 4833431]
- Zuckerman, M.; Depaulo, BM.; Rosenthal, R. Verbal and nonverbal communication of deception. In: Berkowitz, L., editor. *Advances in Experimental Social Psychology*. Vol. Vol. 14. New York: Academic Press; 1981.

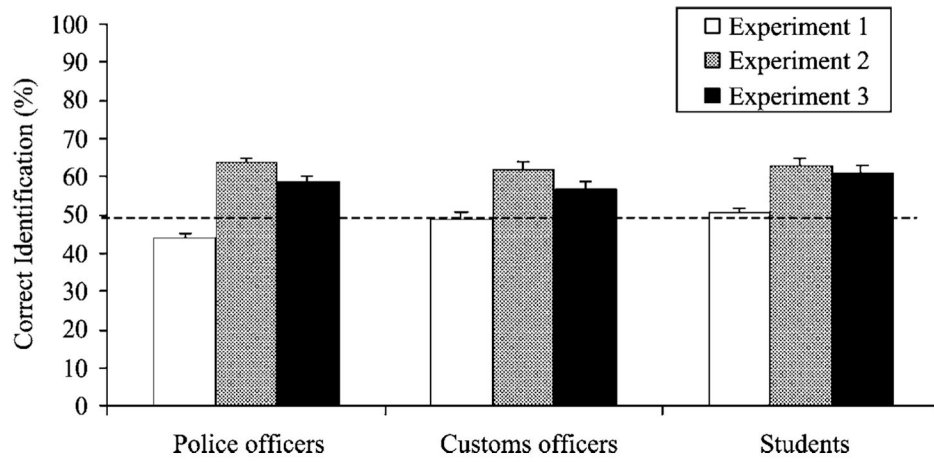


Fig. 1. Average percentage of correct identifications as a function of occupation and experiment.

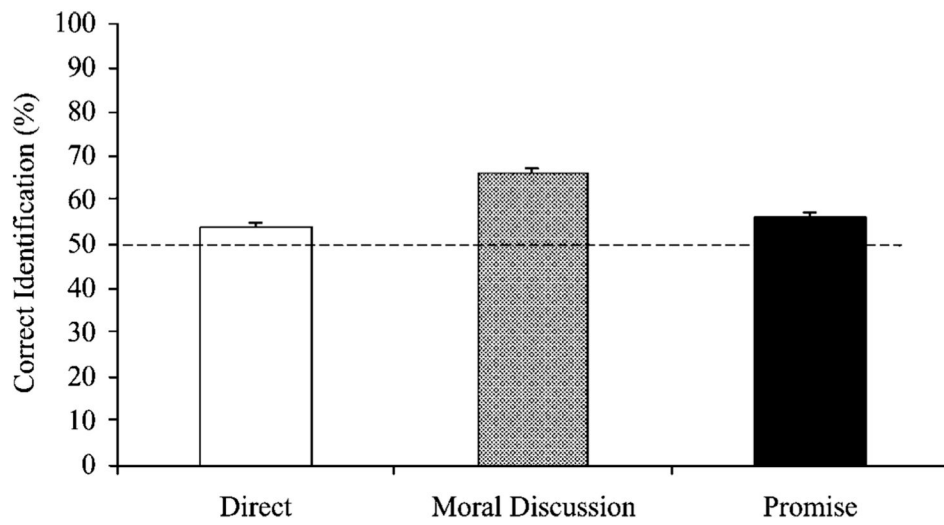


Fig. 2. Average percentage of correct identification as a function of the type of interview.

Table 1
Discrimination (d') and Bias (β) as a Function of Group and Interview

Group	d' (Lie-detection ability)		β (Bias)	
	M	SD	M	SD
Experiment 1 (Direct)				
Customs officers	0.00	0.45	1.02	0.19
Police officers	-0.29**	0.46	1.09	0.28
University students	0.03	0.49	1.01	0.16
Experiment 2 (Moral discussion)				
Customs officers	0.66***	0.56	1.06	0.33
Police officers	0.80***	0.73	1.07	0.42
University students	0.75***	0.73	1.07	0.40
Experiment 3 (Promise)				
Customs officers	0.37***	0.53	1.05	0.16
Police officers	0.40***	0.44	0.95*	0.12
University students	0.49***	0.49	0.97	0.14
Experiment 4				
Interview				
Direct	0.03	0.51	1.06	0.34
Moral Discussion	0.73***	0.66	1.16*	0.68
Promise	0.18**	0.52	1.10**	0.36
Familiar	0.28***	0.33	1.08	0.32
Unfamiliar	0.19***	0.33	1.04	0.31

Note. Statistical tests compared d' to 0 and β to 1.

*
 $p < .05$

**
 $p < .01$

 $p < .001$.

Table 2
Overall Certainty as a Function of Group and Interview Type

Group	Overall certainty	
	<i>M</i>	<i>SD</i>
Experiment 1 (Direct interview)		
Customs officers	1.91	0.35
Police officers	1.66	0.38
University students	1.56	0.35
Experiment 2 (Moral discussion)		
Customs officers	2.05	0.54
Police officers	1.65	0.36
University students	1.59	0.36
Experiment 3 (Promise)		
Customs officers	2.21	0.50
Police officers	1.85	0.46
University students	1.79	0.41
Experiment 4		
Interview		
Direct	1.38	0.29
Moral Discussion	1.49	0.34
Promise	1.43	0.32

Note. The rating scale reflected four levels of certainty (“extremely certain” = 3; “very certain” = 2; “certain” = 1; “uncertain” = 0).

Table 3
Accuracy and Frequency as a Function of Age Group

Age group	<i>N</i>	Overall accuracy	
		<i>M</i>	<i>SD</i>
Experiment 1 (Direct interview)			
3-year-olds	5	.75	.34
4-year-olds	9	.40	.29
5-year-olds	9	.47	.22
6-year-olds	15	.43	.23
7-year-olds	14	.51	.20
8-year-olds	5	.76	.34
9-year-olds	5	.32	.31
10-year-olds	5	.43	.29
11-year-olds	5	.39	.44
Experiment 2 (Moral discussion)			
4-year-olds	6	.85	.21
5-year-olds	8	.61	.23
6-year-olds	6	.64	.23
7-year-olds	5	.54	.26
8-year-olds	2	.34	.38
Experiment 3 (Promise)			
4-year-olds	10	.61	.20
5-year-olds	10	.70	.16
6-year-olds	9	.47	.21
7-year-olds	7	.47	.17
8-year-olds	2	.83	.34
Experiment 4			
3-year-olds	5	.79	.12
4-year-olds	23	.63	.28
5-year-olds	31	.56	.23
6-year-olds	24	.57	.27
7-year-olds	20	.52	.23
8-year-olds	7	.55	.22
9-year-olds	3	.41	.18
10-year-olds	3	.55	.29
11-year-olds	3	.54	.24

Note. Children for whom age information was not available were excluded from all analyses.