UC Davis
UC Davis Previously Published Works

Title

Discernment of Children's True and False Memory Reports: Police Officers and Laypersons.

Permalink

https://escholarship.org/uc/item/7tp952xp

Authors

Johnson, Hannah M
Block, Stephanie D
Shestowsky, Donna
et al.

Publication Date

2023-12-29

DOI

10.1177/08862605231220022

Peer reviewed

Discernment of Children's True and False Memory Reports: Police Officers and Laypersons

Hannah M. Johnson[1], Stephanie D. Block[1] (iD),
Donna Shestowsky[2], Joseph E. Gonzales[1],
Kristy L. Shockley[1,3], and Gail S. Goodman[2] (iD)

Abstract

Adults' ability to accurately evaluate children's statements can have far-reaching consequences within the legal system. This study examined the evaluations of police officers ("experts") and laypersons ("nonexperts") when presented with videotaped interviews of children aged 3 and 5 years who provided either true or false reports or denials. Participants were drawn from several counties in the eastern United States. Children's interview statements fell within four statement types: accurate reports, false reports, accurate denials, and false denials. Both groups of participants displayed overbelief in false denials. Several control variables predicted accuracy, including children's age and children's race. A significant interaction emerged: Experts (vs. nonexperts) had greater odds of being accurate when judging false reports (vs. false denials). These findings highlight the challenges adults face when distinguishing between various types of children's statements. The

[1]University of Massachusetts Lowell, USA
[2]University of California, Davis, USA
[3]The College of the Holy Cross, Worcester, MA, USA

Corresponding Authors:
Stephanie D. Block, Department of Psychology, University of Massachusetts, 860 Broadway Street, Lowell, MA 01854, USA.
Email: stephanie_block@uml.edu

Gail S. Goodman, Department of Psychology, University of California, One Shields Avenue, Davis, CA 95616, USA.
Email: ggoodman@ucdavis.edu

results have important implications for legal contexts, emphasizing that fact finders need to be mindful of the risks associated with both overaccepting false denials and accepting false reports.

In the legal system, children's statements often provide critical evidence in cases involving child maltreatment, custody, and domestic violence. In investigations of child sexual abuse (CSA), children's statements can be the primary evidence guiding the case toward criminal prosecution (Block et al., 2022). The way adults assess these statements can influence whether a case proceeds to the criminal courts, whether a perpetrator is held accountable, and whether children receive appropriate social services.

Police officers frequently play a pivotal role in determining the course of a case based on children's statements, especially in extrafamilial and criminal cases (Powell et al., 2010). Police officers serve as initial points of contact for children who witness or experience crimes and frequently play significant roles in determining whether allegations of alleged child abuse are prosecuted (e.g., Castelli & Goodman, 2014). Thus, police officers often act as "gatekeepers" of child abuse cases in the criminal justice system (Taylor, 2004). It is, therefore, important to examine how accurately they judge children's reports. Moreover, due to the significant role that laypersons can play within the legal system (e.g., serving as jurors) it is also important to understand how accurately they assess children's accounts. We addressed these issues by comparing how police officers ("experts") and laypersons ("nonexperts") discern children's true and false statements.

Past studies examined adults' evaluations of children's eyewitness testimony (e.g., Block et al., 2012; Brigham, 1998; Johnson et al., 2021). Several projects focused on adults' abilities to detect children's lies (e.g., Orcutt et al., 2001). Additional work has also investigated children's false statements that the children themselves may or may not believe to be true (e.g., Johnson et al., 2021). When children give a true report, it can be an "accurate report" or an "accurate denial." An accurate report occurs when children acknowledge an event that actually occurred; an accurate denial arises when children correctly deny an event that did not occur. Inaccuracies can be classified as either a "false report" or a "false denial" (Block et al., 2012; Otgaar et al., 2020). Children's false reports involve acknowledgment of events that did not occur, and false denials occur when children deny events that in fact did occur (Lyon et al., 2008).

False reports and false denials wreak havoc on the administration of justice. Incorrectly believing false reports can lead to false accusations of innocent defendants. When legal professionals mistakenly accept a child's false denial of abuse, it can result in the child not receiving the necessary protection and social services while allowing perpetrators to potentially harm others. Given the potentially grave consequences of inaccurate assessments by both police officers and laypersons, it is critically important to gain a deeper understanding of the factors influencing adults' ability to accurately evaluate children's statements.

Adults' Attempts to Detect Children's True Reports and False Reports

Adults are not much better than chance at accurately detecting adults' lies (Vrij & Baxter, 1999). Regarding adults' abilities to discern the veracity of children's statements, research findings are mixed (Block et al., 2012; Crossman & Lewis, 2006; Edelstein et al., 2006; Talwar & Crossman, 2012; Talwar et al., 2006). Although the video clips used in our study presented scenarios in which the children were not asked to lie, literature on detection of children's lying is relevant. Edelstein et al. (2006) examined adults' accuracy in discerning between true and false statements made by children and adults. The researchers found that adults were better at detecting lies made by children compared to lies made by adults. The variation in lie detection accuracy could potentially be attributed to the inclination to believe that adults are generally more truthful than children: Although participants were better at detecting lies by children compared to detecting lies by adults, participants were better at identifying truth statements in adults compared to those made by children. Overall, however, adults' accuracy of lie detection for both adults and children was no better than chance (Edelstein et al., 2006). These findings lead to the conclusion that adults are often inaccurate in their discernment of children's statements.

Other research has examined children's statements beyond those identified as "lies," finding that, in general, adults are better able to correctly judge accurate reports, accurate denials, and false reports compared to false denials. For false denials, adults are, on average, "confident" that the event did not occur, even when the event had in fact been experienced (according to parental report; Block et al., 2012; Domagalski et al., 2020; Johnson et al., 2021). These findings underscore the challenge adults face when evaluating young children's false denials compared to other types of statements. Given that children may deny child abuse despite having experienced it (Hartman et al., 2023), adults may sometimes believe children's false denials of abuse.

Police Officers' and Parents' Evaluations of Children's Reports

There may be more accurate discernment of children's reports for people who have special experience with children (Block et al., 2012; Crossman & Lewis, 2006). Regarding profession, generally an individual is deemed an expert when they possess specialized skills (Kahneman & Klein, 2009) (Lucassen & Schraagen, 2011). Thus, police officers may be viewed as experts to the extent they have specialized skill or training to accurately determine if a witness, including a child, is providing a true or false report. There is also the possibility that parenthood provides intuition or knowledge to help in deciphering children's true and false statements. The results of the literature on professionals and parents present a mixed picture on the issue of whether individuals with child experience better evaluate children's reports compared to those who have little to no experience with children (Brigham, 1998; Leach et al., 2004; Lindholm, 2008; Nysse-Carris et al., 2011). Further study is thus warranted. In our study, forensic training (on lie detection and/or forensic interviewing) and parental status were considered potentially important variables.

Professional Expertise. The relevant literature has categorized law enforcement officers (Leach et al., 2004; Lindholm, 2008; Nysse-Carris et al., 2011), forensic interviewers (Nysse-Carris et al., 2011), and other types of child welfare workers (Vrij et al., 2006; Goodman et al., 2002) as professionals. Such professionals may have received specialized instruction and training and/or interviewed hundreds of children. Many professionals in the child welfare and criminal justice systems are exposed to children's statements throughout their careers, and their decisions about whether the children are telling the truth or not can impact the trajectory of a case (e.g., Nysse-Carris et al., 2011). Yet, scant research has examined professionals' (e.g., police officers') abilities to detect when children are reporting true or false events.

A study by Nysse-Carris et al. (2011) is an example of this limited research. This work investigated differences between experts' (e.g., forensic interviewers' including a few police officers) and nonexperts' (undergraduate students') abilities to detect children's lies and found no significant relation between confidence and accuracy for either group and no significant main effects on expert status on detection accuracy. However, experts in contrast to college students more accurately detected lies in older children (5 and 6 year olds) than younger children (3 and 4 year olds). Similarly, when comparing social workers and undergraduate students, Goodman et al. (2002) found no significant differences in accuracy between the two groups in judgments of children's statements. However, undergraduates were more likely to believe male compared to female children. When examining law enforcement officers and undergraduate

students in their abilities to detect children's lies, police officers performed at below chance levels in their assessments of children's true and false statements (Leach et al., 2004).

Regarding children's lies, Gongola et al.'s (2017) meta-analysis of 45 studies on adults' ability to detect children's lies included 12 experiments with professionals as part of the sample. Professionals (vs. community members) were found to be only slightly better at evaluating children's reports (56% accuracy vs. 54% accuracy, respectively). Characteristics of children's statements vary greatly across studies, which might explain differences in findings regarding professionals' evaluations of children's reports. Some work has documented a positive correlation between confidence and accuracy; the more confident adults are about their evaluation, the more accurate they are in their judgments. But this relation has varied within and across studies (e.g., Leach et al., 2004).

One explanation for disparate findings within the literature might be explained by the concept of fractionation of expertise (Kahneman & Klein, 2009). This concept suggests that experts in many domains encounter such a diversity of situations that it is hard to become an "expert" in every type of situation. Kahneman and Klein (2009) argue that the two components needed to develop expertise are opportunity to practice the skills and the consistency of said skills. Although the professionals in these child discernment studies had more experience interacting with children than those in the relevant comparison groups, they did not necessarily have enough experience with children's statements to have consistent expertise.

Parental Status. Other research has delved into the issue of whether parental status influences judgment accuracy. Whereas several studies have failed to find parental status helpful in evaluating children's statements (e.g., Block et al., 2012), parents likely have heightened exposure to children compared to those without children. Brigham (1998) found that, when compared to nonparents, parents rated children as more honest and believed the children's denials when the incident involved theft. However, when children alleged CSA, nonparents believed their reports more often than parents. Additionally, Lewandowsky et al. (2012) observed that parents, compared to nonparents, viewed children as better source monitors (i.e., better able to identify whether a memory is real or from another source, such as a movie, book, or dream).

Perceptions of Children's Credibility and Believability

Existing research has attempted to elucidate the factors that affect adults' perceptions of the credibility and believability of children's reports, such as

children's age, the adults' attitudes toward cognitive development, and the valence of the event described by children. Among these factors, children's age has received a particularly high level of research attention. Studies have examined children as young as 3 years all the way to adolescence (Block et al., 2012; Brigham, 1998; Johnson et al., 2021; Nysse-Carris et al., 2011). Within this literature, children of a younger age are often judged to be less credible than older children (Laimon & Poole, 2008; Nysse-Carris et al., 2011; but see Bottoms & Goodman, 1994). Consequently, statements from younger (vs. older) children are more likely to be critically examined by adults.

The task of evaluating children's statements becomes more complex when the individual making the statement is not intentionally lying (Bruck & Ceci, 1999). Children's memories, especially for neutral or positive events, can be manipulated (Leichtman & Ceci, 1995). One reason younger children are often viewed as less credible relates to their potential susceptibility to suggestion (e.g., Ceci & Bruck, 1993). Younger children are also more prone to making source monitoring errors (Ghetti et al., 2002). Although under the right circumstances (e.g., free-recall interviewing techniques), young children may generate highly accurate reports (Brown et al., 2013; Chae et al., 2018), they also might make inaccurate or false statements for various reasons. These reasons include a limited grasp of the distinction between lying and honesty, cognitive developmental limitations, repeated suggestive questioning by adults, a reluctance to implicate someone else, feelings of embarrassment regarding the event, or being coached to intentionally lie (Lyon et al., 2008). Because preschool children are especially likely to have some of these difficulties, we examined perceptions of children aged 3 and 5 years.

The assessment of children's competency as witnesses is undoubtedly significant when it comes to both experts and nonexperts forming opinions about the believability and credibility of children. This aspect of children's competence is frequently a topic of consideration in legal proceedings, particularly when preschool-aged children are involved (e.g., Lyon & Saywitz, 1999; Saywitz et al., 1999). Children's cognitive ability is correlated with both age (e.g., Murnikov & Kask, 2021) and memory/suggestibility performance (e.g., Poole et al., 2014). Adults' attitudes toward cognitive ability could independently influence how individuals perceive children's capacity to respond to questions and resist suggestion. Therefore, in our current study, we gathered adults' assessments of the children's cognitive abilities.

Another factor associated with adults' ratings of the believability and credibility of children's reports relates to event valence (Block et al., 2012). Adults tend to be more accurate when they evaluate negative compared to positive events (Johnson et al., 2021). In summary, past research suggests

that children's age, attitudes about children's cognitive abilities, and event valence may all influence adults' decisions and thus were statistically controlled in this study.

The Present Study

Our research compared police officers' (experts') and matched laypersons' (nonexperts') evaluations of statements made by children aged 3 or 5 years. We expanded on previous work in several ways. First, the children had not been instructed to be dishonest. Prior work that examined professionals' abilities relied largely on deception detection in situations wherein children were coached or instructed to lie (Leach et al., 2004; Lindholm, 2008; Nysse-Carris et al., 2011). Second, in our study, professionals and matched laypersons evaluated multiple statement types (i.e., accurate report, false report, accurate denial, and false denial). Third, we also aimed to examine factors associated with the accuracy of adults' judgments of children's statements, including participant expertise (police officers vs. laypersons), participant parental status and experience with children (e.g., parent vs. nonparent), participant training in lie detection/forensic interviewing, and participants' attitudes about children's cognitive development. Previous research used the same stimuli as the current study but did not compare professionals' with nonprofessionals' judgments (Block et al., 2012; Johnson et al., 2021).

Although findings vary concerning whether experts (vs. nonexperts) are better at judging children's statements, some research suggests that experts are slightly more accurate (Gongola et al., 2017). Furthermore, there is a paucity of research on police officers regarding their abilities to evaluate various types of children's statements. Accordingly, we hypothesized that police officers would be more accurate than laypersons in discerning between true and false statements (i.e., true reports, false reports, false denial, and true denials) than laypersons. We predicted that police officers (perhaps because of their forensic training) would be more confident but that this greater confidence would not be associated with greater accuracy (Vrij et al., 2006). We also predicted, based on past research (Block et al., 2012), that both experts and nonexperts would be less accurate in their overall judgments of false denials.

Method

Participants

Eighty adults participated: 40 police officers (experts) and 40 laypersons from the community (nonexperts). The two groups were matched on several

key demographic variables. Their mean age was 41.98 years ($SD = 10.46$) for the expert group and 44.08 years ($SD = 12.18$) for the nonexpert group. The expert group was composed of 31 males out of 40; 32 out of 40 nonexperts were male. The race/ethnicity breakdown of the expert group was 80% Caucasian/White, 2.5% Asian, 2.5% Black, 5% Hispanic/Latino/a, 2.5% other, and 7.5% did not answer. For the nonexpert group, it was 72.5% Caucasian/White, 7.5% Asian, 2.5% Black, and 17.5% did not answer. Seventy percent of experts and 60% of nonexperts reported being parents. Fifty-five percent of experts reported having graduate degrees, and 62.5% of nonexperts reported having graduate degrees. None of the nonexperts had worked as police officers. Out of the 80 participants, 57.5% of experts indicated forensic training, whereas only 2.5% of nonexperts reported any type of forensic training.

Measures

Videotape Stimuli. The videotape stimuli were comprised of interview clips from a study by Schaaf et al. (2008), which examined true/false events reported by preschool children. In this earlier study, children were interviewed in a leading manner but were not instructed to lie. The research team ascertained whether the event occurred by obtaining reports from the children's parents. Each child was asked six questions, presuming that the event had been experienced, regardless of the child's answers. For true events, the child either correctly assented (accurate report) or incorrectly denied (false denial) experiencing the event. For false events, the child either incorrectly assented (false report) or correctly denied (accurate denial).

For the current study, we used the same video clip selection from the Block et al. (2012) study (24 video clips and 1 practice clip) but embedded them in an online survey platform (Qualtrics). To reduce the likelihood of fatigue, participants viewed only 12 out of the 24 videos as well as the practice video (13 video clips total). The 24 videos were split into 2 orders of 12 videos each. Participants viewed 1 set of the 12 videos, with sets counterbalanced across experts and nonexperts, as possible. To further reduce order effects, each set of 12 videos contained 2 versions of the same videos but in reverse order.

The 2 orders of 12 videos equally reflected child age, gender, event type (positive event [e.g., got a specific present for a birthday] or negative event [e.g., getting yelled at for falling out of the bathtub]), and report type. Each group of 12 videos contained 3 accurate reports (child correctly assented the event occurred), 3 false denials (child incorrectly denied the event happened), 3 false reports (child incorrectly assented the event occurred), and 3 accurate denials (child correctly denied the event happened) (Table 1).

Table 1. Video Clip Order and Characteristics.

Order 1	Gender	Age	Event Type	Race/Ethnicity	Report Type
Practice Clip	Male	3	Negative	Hispanic	Accurate Report
Clip 1	Female	3	Negative	White	False Denial
Clip 2	Female	5	Negative	White	False Report
Clip 3	Male	3	Negative	White	Accurate Report
Clip 4	Male	3	Negative	White	False Denial
Clip 5	Female	5	Positive	White	False Denial
Clip 6	Male	5	Negative	White	Accurate Denial
Clip 7	Female	5	Positive	White	Accurate Denial
Clip 8	Female	3	Negative	White	False Report
Clip 9	Female	5	Positive	Asian-American	Accurate Report
Clip 10	Male	5	Positive	White	False Report
Clip 11	Male	3	Negative	White	Accurate Report
Clip 12	Male	3	Positive	White	Accurate Denial

Order 2	Gender	Age	Event Type	Race/Ethnicity	Report Type
Practice Clip	Male	3	Negative	Hispanic	Accurate Report
Clip 13	Male	3	Positive	White	False Report
Clip 14	Male	3	Positive	White	False Report
Clip 15	Female	3	Positive	White	Accurate Denial
Clip 16	Female	5	Positive	Black	False Denial
Clip 17	Female	3	Negative	Asian- American	False Denial
Clip 18	Male	3	Negative	White	Accurate Denial
Clip 19	Male	5	Negative	White	Accurate Report
Clip 20	Female	3	Negative	White	Accurate Report
Clip 21	Male	5	Negative	White	False Denial
Clip 22	Female	5	Positive	White	Accurate Report
Clip 23	Male	5	Positive	White	False Report
Clip 24	Male	5	Negative	White	Accurate Denial

Children's Statement Evaluation Questionnaire. The Children's Statement Evaluation Questionnaire (Block et al., 2012) included a yes/no question asking whether the participant believed the event occurred and a Likert scale question assessing how confident the participant was in regard to that question (1=*extremely not confident* to 6=*extremely confident*). Seven additional questions per video clip were included (e.g., How accurate was the child? How believable was the child?), but as they did not produce significant or nonduplicative findings, analyses of them is not provided.

Attitudes Toward Child Witness Scale (ATCW). The questionnaire included 21 statements about child witnesses (Wrightsman et al., 2004). Participants

indicated the extent of their agreement to these statements on a seven-point Likert scale (1=*strongly disagree* to 7=s*trongly agree*). However, only two subscales were of interest in the present study: Competence and cognitive ability.

Demographic Questionnaire. This instrument included basic demographic questions (e.g., age, race/ethnicity, level of education, and profession) and included yes/no questions that assessed parental status and whether they had participated in lie detection/forensic interview training.

Procedure

The study was approved by the University's Institutional Review Board. All experts were recruited, with approval from the chief of police at each precinct, through a letter sent via email to precincts located in the northeastern United States; the experts volunteered to participate. We recruited a pool of laypersons by posting flyers in general community areas (e.g., libraries, with some being in the recruited experts' districts) and online using Craig's List. We selected laypersons who matched our police sample on the following demographic variables: age, level of education, sex/gender, and if they were a parent. The potential recruits sent the research team an email inquiring about participation and after screening, they were sent a private link to the survey, which was administered using Qualtrics.

 After we obtained their consent, participants received instructions emphasizing the importance of paying attention as they watched the videos. They then viewed the practice video and answered the Children's Statement Evaluation Questionnaire for that video. They subsequently viewed the remaining 12 video clips. Following each video clip, participants were prompted to answer the Children's Statement Evaluation Questionnaire for the video clip just viewed. After viewing all videos, they completed the ATCW scale and provided their demographic and background information. Participants were subsequently directed to a debriefing page. The study took between 45 and 60 min to complete. Gift cards were sent to participants after completion of the study.

Coding and Selection of the Variables for Analysis

Outcome Variable

 Judgment Accuracy. The main dependent variable of interest of the study was the participants' accuracy in judging the child's statement. Our first question asked participants if they believed the child experienced the event the

child was questioned about. Participants responded "yes" or "no." We later coded their responses to reflect whether their judgment was accurate or not. There were too few judgments for a d prime analysis.

Participant Control Variables

Age. Participant age was collected as a continuous variable.

Gender. Participant gender was coded as a dichotomous variable (i.e., male or female); we used male as the reference group for our analyses.

Participant Race/Ethnicity. This categorical variable captured how participants identified racially/ethnically, which was recoded for analysis as either White or person of color (POC). White served as our reference group.

Training. All participants were asked if they have any training regarding lie detection or forensic interviewing. Responses were coded as yes or no; we used "no training" as the reference group.

Confidence. Participants were asked to rate their confidence in their response about whether they believed the event occurred. This scale ranged from 1 (*extremely not confident*) to 6 (*extremely confident)*. The mean confidence score for each participant was calculated and subtracted from each of their event confidence scores. This calculation resulted in a within-subjects measure of participant confidence fluctuations from one event to another.

Parent Status. We statistically controlled for whether a participant was a parent or not a parent, with nonparents serving as our reference group.

Competence and Cognitive Ability Subscales. We included mean scores of these two subscales from the ATCW.

Child Characteristics from Video Stimuli Control Variables

Child Age. Children in the videos were either 3 or 5 years old, with 3 year olds serving as our reference group.

Child Race. The child's race was coded as White or POC; White was our reference group.

Event valence. Event type was a dichotomous variable that reflected whether the event the child was questioned about was a positive or negative event. Negative events served as our reference group.

Predictors of Interest

    Expert Status. Participants were either an expert (police officer) or a non-expert (layperson). Experts served as our reference group.

    Report Type. There were four report types present in the video stimuli: accurate reports, false reports, accurate denials, and false denials. False denials were used as the reference group.

Analysis Plan

All data processing and analyses were conducted using R ver. 3.6.3 (R Core Team, 2020). Generalized linear mixed-effects modeling (GLMM) was conducted using the *lme4* (Bates et al., 2015) and *lmerTest* (Kuznetsova et al., 2017) packages.

Identification of Judgment Accuracy Nesting Structure. We tested for variability in judgment accuracy due to the nesting of judgments within both participants and videos. Specifically, we fit three GLMM intercept-only models of participants' judgment accuracy where the intercept was nested within participants, videos, or both participants and videos. Nested models were compared using the likelihood ratio test (LRT) to determine which of these nesting structures were appropriate for the data (e.g., Raudenbush & Bryk, 2002).

Modeling Judgment Accuracy. We fit a series of GLMM models where judgment accuracy of children's statements was predicted by: (a) control variables and expert status (police vs. laypersons) (Model 1); (b) all Model 1 predictors plus report type (i.e., false denial, accurate denial, false report, and accurate report) (Model 2); and (c) all Model 2 predictors plus the interaction between expert status and report type (Model 3). These nested models were compared for significant improvement in model fit to the data using the LRT. Given a significant interaction between expert status and report type, we tested for all possible expertise-moderated effects of report type. For all models, we used false denial as the reference group for report type.

Results

Judgment Accuracy Nesting Structure

Model fit did not statistically differ between a video-only nesting and a video and participant nesting structure, $\chi^2(1) = 0$, $p = 1.000$. In contrast, the participant-only nesting structure fit statistically worse than a video and participant

Table 2. Selected Judgment Accuracy Model Coefficients.

Predictor	B	SE	p-Value	OR
(Intercept)	−1.26	1.19	.291	0.28
Video Set A	−0.50	0.29	.089	0.61
Participant Age	−0.01	0.01	.319	0.99
Participant POC	−0.26	0.26	.317	0.77
Participant Female	0.29	0.23	.197	1.34
Participant Nonexp	0.43	0.40	.280	1.54
Participant Parent	0.26	0.22	.231	1.3
Participant Training	−0.28	0.25	.269	0.76
Child Age 5*	0.89	0.31	.004	2.43
Child Female	0.66	0.35	.063	1.93
Child POC*	−1.28	0.63	.041	0.28
Pos. Event	0.09	0.32	.779	1.09
Confidence*	0.23	0.09	.009	1.26
Competence	−0.01	0.02	.788	1
Cog. Ability	−0.02	0.03	.516	0.98
ADenial*	3.62	0.62	<.001	37.41
AReport*	1.90	0.55	.001	6.65
FReport*	2.75	0.57	<.001	15.66
Nonexp:ADenial	0.13	0.59	.828	1.14
Nonexp:AReport	−0.19	0.49	.702	0.83
Nonexp:FReport*	−1.18	0.49	.016	0.31

Note. ADenial (accurate denial), AReport (accurate report), and FReport (false report) are compared with false denial, the reference report type. POC = Person of Color; Nonexp = Nonexpert; Pos. = Positive.
*$p < .050$.

nesting structure, $\chi^2(1)=274.55$, $p<.001$. Given the lack of improved fit when nesting in both participant and video compared to video-only, a video-only nesting structure was used for the remainder of the analyses.

Judgment Accuracy Predictors

Model comparisons indicated that Model 3 fit the data better than Model 2, $\Delta\chi^2(14)=31.91$, $p=.004$; Model 2 fit the data better than Model 1, $\Delta\chi^2(3)=27.39$, $p<.001$; and Model 1 fit the data better than the intercept-only model, $\Delta\chi^2(3)=27.39$, $p<.001$. Model 3 was interpreted as the best fitting model of judgment accuracy. Findings are shown in Table 2.
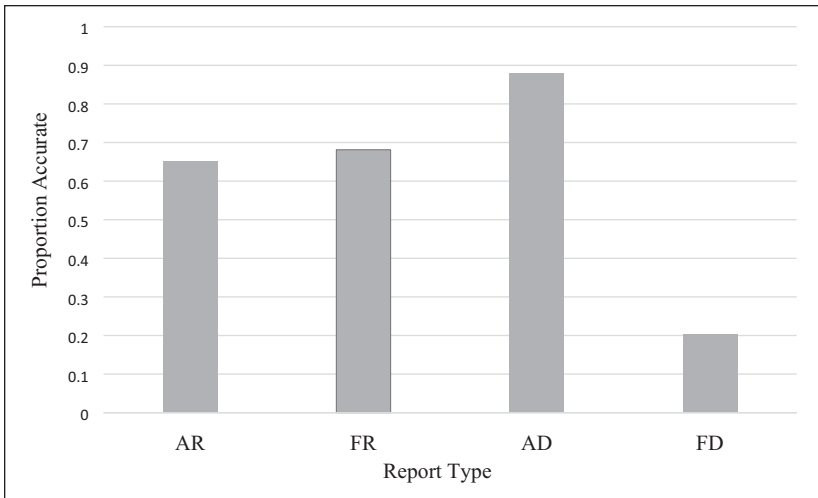
Figure 1. Participant accuracy across by report type.

Control Variables. One participant control variable, confidence in judgment, emerged as a statistically significant predictor of participants' judgment accuracy. Within-subject fluctuations in confidence of video judgments were related to greater judgment accuracy (Table 2); when relatively less confident in their judgments, participants were less likely to be correct, and when relatively more confident in their judgments they were more likely to be correct. No other participant control variable effects (i.e., age, race/ethnicity, and gender of the participant, parental status, training experience, and responses to ATCW scale) were detected as statistically significant (Table 2).

Two video control variables—child age and child race—were also detected as statistically significant predictors of participants' judgment accuracy. Participants were more accurate when evaluating videos of children who were 5 versus 3 years old and less accurate when evaluating videos of children who were POC versus White. No other video control variable effects emerged as statistically significant (Table 2).

Expert Versus Nonexpert Accuracy Judgments of Report Types. Significant main effects emerged for report type: Participants were statistically more accurate when judging accurate denials and accurate reports compared to false denials (Figure 1 and Table 2). Judging false reports did not significantly differ from the other report types.
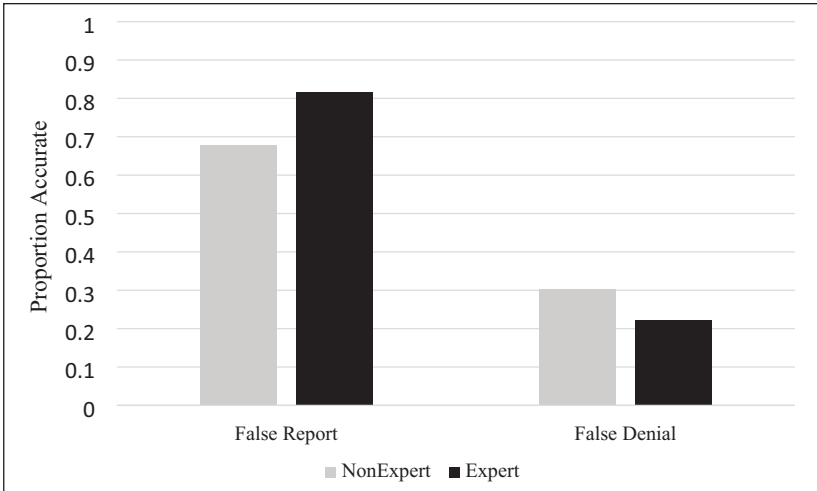
Figure 2. Expertise and judgment accuracy between false reports and false denials.

Of importance, although the main effect of expert status was not significant, a statistically significant interaction was detected such that the difference in judgment accuracy when comparing false denials and false reports was moderated by expert status. Specifically, the odds of making an accurate judgment when rating a false report compared to a false denial were 15.66 times greater for experts, but only 4.81 times greater for nonexperts (Table 2). Said differently, experts had a greater disparity in their judgment accuracy when comparing false report (81.6%) and false denial (22.1%) report types, whereas nonexperts had less disparity in judgment accuracy between false report (67.7%) and false denial (30.4%) report types (Figure 2). No other statistically significant interactions between report type and expert status were found.

Discussion

Adults, particularly those responsible for handling cases involving child witnesses or victims, often have the responsibility of assessing children's statements. The accuracy of this evaluation can influence the trajectory of the case. Our study built upon prior research on the discernment of children's true and false statements (e.g., Block et al., 2012) by examining the judgments of police officers and laypersons. Participants were over 80% correct

when evaluating accurate denials, almost 70% correct when evaluating false reports, over 60% correct when evaluating accurate reports, but only 20% correct when evaluating false denial statements. Our results regarding false denials are consistent with findings from previous research (Block et al., 2012; Johnson et al., 2021), which suggests that adults are particularly challenged when evaluating false denials, specifically when the children's statements are uncoached and the children were not instructed to lie. Wyman et al. (2018) also studied adults' abilities to detect false denials in coached 6- to 11-year-old children's statements and found that false denials were more accurately detected than false reports. However, it is important to note that in that study, the children were older and received coaching. In contrast, our study focused on preschool-aged children who were not coached. The results of our study contribute to this line of investigation by revealing that when assessing uncoached preschool-aged children who were interviewed about events, adults face challenges in distinguishing false denials from other types of reporting, regardless of whether they are police officers or laypersons.

Although the main effect of expertise was not significant, we did detect a significant interaction of report type and expertise. Specifically, whereas both nonexperts and police officers were more accurate in appraising false reports than false denials, there was a smaller discrepancy in accuracy between false reports and false denials for nonexperts (67.7% vs. 30.4%, respectively) compared to police officers (81.6% vs. 22.1%, respectively). Basically, laypersons demonstrated less variability in accuracy between false denials and false reports, both being closer to 50:50, whereas experts demonstrated greater variability in accuracy between false denials and false reports such that they were very accurate when judging false reports, but very inaccurate when judging false denials. Studies should replicate and extend the present findings to uncover the underlying reasons for these differences. Aligning with existing literature, factors such as prior forensic or lie detection training and participants' attitudes toward the competence and cognitive abilities of young child witnesses did not serve as reliable predictors of accurate judgments (Nysse-Carris et al., 2011).

We also examined the impact of children's characteristics in relation to participants' accuracy. We found that children's age and race were statistically significant predictors of accuracy. When children were younger (3 year olds), participants were less accurate in their judgments of the different report types. Research has identified that children under age 5 can be more susceptible to suggestibility and memory errors (e.g., Ceci & Bruck, 1993; Chae et al., 2018), which is one reason we focused on this age group. Children's statements can change in substance and delivery as they age due to cognitive development and other factors, which is why it is of value to examine

statements made by different age groups. We also found that when children presented in the videos were White, participants produced more accurate judgments across the report types. Given that our sample was composed mainly of White participants, the pattern we observed might reflect issues relating to cross-racial identification (Pezdek et al., 2003). Given that only three children who were individuals of color were included in our video clip stimuli, future research should further explore this possibility.

Our findings also revealed a positive relation between judgment confidence and judgment accuracy. Participants who exhibited higher levels of confidence were 1.26 times more likely to make accurate judgments compared to those who reported lower confidence levels. Given that prior research on the relation between confidence and accuracy had yielded mixed results (Gongola et al., 2017; Leach et al., 2004), future research should untangle the apparently complicated relation between these two variables.

Limitations and Future Directions

It is important to note that participants in our study were explicitly informed that the children involved were not directed to lie or to provide false information (although some did the latter). This situation differs from the typical courtroom setting wherein children are explicitly instructed to speak truthfully. Instead, the scenario in this study more accurately mirrors the way children typically convey information before they are summoned to testify in court.

Our research explored the statements of 3 and 5 year olds but lacked ecological validity in regard to child abuse claims. Future research should encompass a broader age range and incorporate questions related to child abuse to investigate the impact of expertise. Additionally, it should be noted that we aimed to mitigate potential fatigue and distractions by limiting the number of video clips participants viewed (12 out of 24). Although participation in this study may be no more tiring than legal situations (e.g., trials) that may involve evaluating children's statements, future research could profit from use of d prime and response bias analyses.

Recognizing the crucial role that adults' initial assessments of children's statements play, particularly within the legal context, future research should endeavor to explore various professional groups considered experts in domains related to children's reports (e.g., teachers, social workers, and child forensic interviewers) to expand our insights into how diverse professionals evaluate children's statements. Not all children disclose information to child interviewers, that is, to specially trained professionals who likely have considerable experience with children. Despite the potential lack of specialized training in child interviewing techniques, we chose to involve

law enforcement personnel in our study. We made this choice because police officers hold significant power and authority in shaping the course of cases involving extrafamilial and some intrafamilial matters based on their assessment of the truthfulness of children's statements. They also play pivotal roles within multidisciplinary teams within Children's Advocacy Centers.

With regard to diversity, future research should strive to gather a more extensive and diverse sample of professionals, encompassing various gender/gender preferences and racial backgrounds. This broader scope will yield a more comprehensive understanding of this crucial aspect of discernment, particularly with respect to women and individuals of color. Additionally, by including a larger and more representative sample, we can mitigate potential selection bias. If our findings are replicated, they would carry significant implications for the realms of criminal justice and child protection, as they suggest that participants encountered greater challenges when discerning statements from children of color.

In conclusion, the results of the present research contribute to the literature about the complexities of police officers' and laypersons' judgments of preschool-aged children's statements. Although much of the prior research focused on intentional false reports (e.g., Gongola et al., 2017), other studies revealed the challenges associated with identifying false denials (Block et al., 2012; Pipe & Goodman, 1991). The present study, which suggests that both police officers and laypersons exhibit reduced accuracy in identifying children's false denial statements, should motivate future action by researchers and practitioners alike. It highlights the need to direct more efforts toward developing interventions and training protocols aimed at enhancing the evaluation skills of both these groups when dealing with children's reports. In instances where adults fail to recognize crimes witnessed or experienced by children but denied by them, there is a significant risk that alleged perpetrators may avoid legal consequences and have increased opportunities to commit further offenses. Additionally, preschool-aged children who may require counseling or other support services might not receive the necessary assistance. These adverse outcomes could be alleviated through increased professional awareness of the risks associated with false denials and through policy changes aimed at enhancing professionals' capacity to evaluate the full array of potential report types.

Our work reinforces the idea that false denials present a unique challenge, not only to the general community but also to professionals tasked with assessing children's statements. Police officers, as first responders who are frequently called upon to quickly evaluate children's statements to make safety determinations, are among those who face this challenge. Future studies should explore the development of training protocols and

guidelines aimed at preventing hasty acceptance of false denials by fact finders and help the justice system avoid in avoiding pursuing false reports.

ORCID iDs

Stephanie D. Block (iD) https://orcid.org/0000-0002-6388-6721
Gail S. Goodman (iD) https://orcid.org/0000-0002-2691-640X

References

Bates, D., Machler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01

Block, S. D., Shestowsky, D., Segovia, D. A., Goodman, G. S., Schaaf, J. M., & Alexander, K. W. (2012). "That never happened": Adults' discernment of children's true and false memory reports. *Law and Human Behavior*, *36*(5), 365–374. https://doi.org/10.1037/h0093920

Bottoms, B. L., & Goodman, G. S. (1994). Perceptions of children's credibility in sexual assault cases. *Journal of Applied Social Psychology*, *24*(8), 702–732. https://doi.org/10.1111/j.1559-1816.1994.tb00608.x

Brigham, J. C. (1998). Adults' evaluations of characteristics of children's memory. *Journal of Applied Developmental Psychology*, *19*, 15–39. https://doi.org/10.1016/S0193-3973(99)80026-4

Brown, D. A., Lamb, M. E., Lewis, C., Pipe, M. E., Orbach, Y., & Wolfman, M. (2013). The NICHD investigative interview protocol: An analogue study. *Journal of Experimental Psychology: Applied*, *19*(4), 367–382.

Bruck, M., & Ceci, S. J. (1999). The suggestibility of children's memory. *Annual Review of Psychology*, *50*(1), 419–439. https://doi.org/10.1146/50.1.419

Castelli, P., & Goodman, G. S. (2014). Children's perceived emotional behavior at disclosure and prosecutors' evaluations. *Child Abuse & Neglect*, *38*(9), 1521–1532. https://doi.org/10.1016/j.chiabu.2014.02.010.

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, *113*(3), 403–439.

Chae, Y., Goodman, M., Goodman, G. S., Troxel, N., McWilliams, K., Thompson, R. A., Shaver, P. R., & Widaman, K. F. (2018). How children remember the Strange

Situation: The role of attachment. *Journal of Experimental Child Psychology*, *166*, 360–379. https://doi.org/10.1016/j.jecp.2017.09.001

Crossman, A. M., & Lewis, M. (2006). Adults' ability to detect children's lying. *Behavioral Sciences & the Law*, *24*(5), 703–715. https://doi.org/10.1002/bsl.731

Domagalski, K., Gongola, J., Lyon, T. D., Clark, S. E., & Quas, J. A. (2020). Detecting children's true and false denials of wrongdoing: Effects of question type and base rate knowledge. *Behavioral Sciences & the Law*, *38*(6), 612–629.

Edelstein, R. S., Luten, T. L., Ekman, P., & Goodman, G. S. (2006). Detecting lies in children and adults. *Law and Human Behavior*, *30*(1), 1–10. https://doi.org/10.1007/s10979-006-9031-2

Ghetti, S., Qin, J., & Goodman, G. S. (2002). False memories in children and adults: Age, distinctiveness, and subjective experience. *Developmental Psychology*, *38*(5), 705. https://doi.org/10.1037/0012-1649.38.5.705

Gongola, J., Scurich, N., & Quas, J. A. (2017). Detecting deception in children: A meta-analysis. *Law and Human Behavior*, *41*(1), 44–54. https://doi.org/10.1037/lhb0000211

Goodman, G. S., Batterman-Faunce, J. M., Schaaf, J. M., & Kenney, R. (2002). Nearly 4 years after an event: Children's eyewitness memory and adults' perceptions of children's accuracy. *Child Abuse & Neglect*, *26*(8), 849–884

Hartman, D., Wang, Y., Wu, Y., Goldfarb, D., Qin, J., Eisen, M. L., & Goodman, G. S. (2023). Child sexual abuse: A longitudinal study of disclosures and denials. *Child Maltreatment*, *28*(3), 462–475.

Johnson, J. L., Hobbs, S. D., Chae, Y., Goodman, G. S., Shestowsky, D., & Block, S. D. (2021). "I didn't do that!": Event valence and child age influence adults' discernment of preschoolers' true and false statements. *Journal of Interpersonal Violence*, *36*(1–2), 1–19. https://doi.org/10.1177/0886260517736276

Kahneman, D., & Klein, G. (2009). Conditions for intuitive expertise: A failure to disagree. *American Psychologist*, *64*(6), 515–526. https://doi.org/10.1037/a0016755

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). LmerTest Package: Tests in linear mixed effects models. *Journal of Statistical Software*, *82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13

Laimon, R. L., & Poole, D. A. (2008). Adults usually believe young children: The influence of eliciting questions and suggestibility presentations on perceptions of children's disclosures. *Law and Human Behavior*, *32*(6), 489-501. https://doi.org/10.1007/s10979-008-9127-y

Leach, A., 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*(6), 661–685. https://doi.org/10.1007/s10979-004-0793-0

Leichtman, M. D., & Ceci, S. J. (1995). The effects of stereotypes and suggestions on preschoolers' reports. *Developmental Psychology*, *31*, 568–578. https://doi.org/10.1037/0012-1649.31.4.568

Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing.

Psychological Science in the Public Interest, *13*(3), 106–131. https://doi.org/10.1177/1529100612451018

Lindholm, T. (2008). Who can judge the accuracy of eyewitness statements? A comparison of professionals and lay-persons. *Applied Cognitive Psychology*, *22*(9), 1301–1314. https://doi.org/10.1002/acp.1439

Lucassen, T., & Schraagen, J. M. (2011). Factual accuracy and trust in information: The role of expertise. *Journal of the American Society for Information Science and Technology*, *62*(7), 1232–1242.

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*(4), 914–929. https://doi.org/10.1111/j.1467-8624.2008.01167.x

Lyon, T. D., & Saywitz, K. J. (1999). Young maltreated children's competence to take the oath. *Applied Developmental Science*, *3*(1), 16–27. https://doi.org/10.1207/s1532480xads0301_3

Murnikov, V., & Kask, K. (2021). Recall accuracy in children: Age vs. conceptual thinking. *Frontiers in Psychology*, *12*, 686904. https://doi.org/10.3389/fpsyg.2021.686904

Nysse-Carris, K. L., Bottoms, B. L., & Salerno, J. M. (2011). Experts' and novices' abilities to detect children's high-stakes likes of omission. *Psychology, Public Policy, and Law*, *17*(1), 76–98. https://doi.org/10.1037/a0022136

Orcutt, H. K., Goodman, G. S., Tobey, A. E., Batterman-Faunce, J. M., & Thomas, S. (2001). Detecting deception in children's testimony: Factfinders' abilities to reach the truth in open court and closed-circuit trials. *Law and Human Behavior*, *25*, 339–372. https://doi.org/10.1023/a:1010603618330

Otgaar, H., Howe, M. L., Mangiulli, I., & Bücken, C. (2020). The impact of false denials on forgetting and false memory. *Cognition*, *202*, 104322. https://doi.org/10.1016/j.cognition.2020.104322

Pezdek, K., Blandon-Gitlin, I., & Moore, C. (2003). Children's face recognition memory: more evidence for the cross-race effect. *Journal of Applied Psychology*, *88*(4), 760–763. https://doi.org/10.1037/0021-9010.88.4.760

Pipe, M. E., & Goodman, G. S. (1991). Elements of secrecy: Implications for children's testimony. *Behavioral Sciences & the Law*, *9*(1), 33–41.

Poole, D. A., Dickinson, J. J., Brubacher, S. P., Liberty, A. E., & Kaake, A. M. (2014). Deficient cognitive control fuels children's exuberant false allegations. *Journal of Experimental Child Psychology*, *118*, 101–109. https://doi.org/10.1016/j.jecp.2013.08.013

Powell, M., Murfett, R., & Thomson, D. M. (2010). An analysis of police officers' decisions about whether to refer cases of child abuse for prosecution. *Psychology, Crime & Law*, *16*(8), 715–724. https://doi.org/10.1080/10683160903025828

R Core Team (2020). R. A language and environment for statistical computing (Version 3.6.3) [Software], https://www.R-project.org/

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (*Vol. 1*). Sage.

Saywitz, K. J., Snyder, L., & Nathanson, R. (1999). Facilitating the communicative competence of the child witness. *Applied Developmental Science*, *3*(1), 58–68. https://doi.org/10.1207/s1532480xads0301_7

Schaaf, J. M., Alexander, K. W., & Goodman, G. S. (2008). Children's false memory and true disclosure in the face of repeated questions. *Journal of Experimental Child Psychology*, *100*(3), 157–185. https://doi.org/10.1016/j.jecp.2007.09.002

Talwar, V., & Crossman, A. M. (2012). Children's lies and their detection: Implications for child witness testimony. *Developmental Review*, *32*(4), 337–359. https://doi.org/10.1016/j.dr.2012.06.004

Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2006). Adults' judgments of children's coached reports. *Law and Human Behavior*, *30*(5), 561–570. https://doi.org/10.1007/s10979-006-9038-8

Taylor, S. (2004). *Surviving the legal system: A handbook for adult & child sexual assault survivors & their supporters*. Coulomb Communications.

Vrij, A., Akehurst, L., Brown, L., & Mann, S. (2006). Detecting lies in young children, adolescents, and adults. *Applied Cognitive Psychology*, *20*(9), 1225–1237. https://doi.org/10.1002/acp.1278.

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*(1), 25–36. https://doi.org/10.1023/A:1008932402565

Wyman, J., Foster, I., Lavoie, J., Tong, D., & Talwar, V. (2018). Detecting children's false allegations and recantations of a crime. *Psychology, Crime & Law*, *24*(6), 652–671. https://doi.org/10.1080/1068316X.2017.1402018

Wrightsman, L. S., Batson, A. L., & Edkins, V. A. (Eds.). (2004). *Measures of legal attitudes*. Wadsworth.

Author Biographies

Hannah M. Johnson received her Ph.D. in Applied Psychology and Prevention Science. She is an Assistant Teaching Professor in the Psychology Department at the University of Massachusetts Lowell. Her research focuses on applied child maltreatment issues, specifically related to nonoffending caregivers of child victims of sexual abuse.

Stephanie D. Block received her Ph.D. in Developmental Psychology from the University of California, Davis and completed a postdoctoral fellowship at the University of North Carolina, Chapel Hill at the Center for Developmental Science and Injury Prevention Research Center. She is an Associate Professor of Psychology at the University of Massachusetts Lowell. Her research focuses on child maltreatment and encompasses her training in both developmental and applied science, as well as public health.

Donna Shestowsky was awarded a J.D. and Ph.D. in Psychology from Stanford University. She is a Professor of Law, Martin Luther King Jr. Research Scholar, and the Director of the Lawyering Skills Education Program at the University of California, Davis, School of Law. She is also Affiliated Faculty of the Psychology Department at the University of California, Davis. Her research centers on the intersection between

law and social psychology. Her legal and psychological commentary has appeared in national sources such as CNN, NPR, and the New York Times.

Joseph E. Gonzales earned his Ph.D. in Quantitative Psychology at the University of California, Davis. He is an Associate Professor in the Department of Psychology, an Associate at the Center for Women and Work, and a Faculty Researcher at the Center for Health Statistics at the University of Massachusetts Lowell. His research focuses on methods for studying multivariate dynamics and processes of change, latent construct modeling, and heterogeneity of process and measurement.

Kristy L. Shockley earned her Ph.D. in Applied Psychology and Prevention Science at the Univeristy of Massachusetts Lowell. She is currently a Visiting Professor in the Psychology Department at the College of the Holy Cross. She is a developmental psychologist with a focus on prevention research and community collaboration. Her overall research interest intersects developmental, community, and legal psychology as it relates to sexual abuse and criminal justice outcomes.

Gail S. Goodman received her Ph.D. from University of California, Los Angeles, and conducted postdoctoral studies at the University of Denver and the Université René Descartes in Paris, France. She is a Distinguished Professor of Psychology at the University of California, Davis. Her research concerns trauma and memory, memory development, child maltreatment, and children in the legal system. She has received many awards and grants for her research and writings and has published widely. Her research has been cited in U.S. Supreme Court decisions. She has consulted with numerous governments and agencies throughout the world on policies concerning child victims in the legal system.