Title

Detecting children's true and false denials of wrongdoing: Effects of question type and base rate knowledge

Permalink

Journal

ISSN

Authors

Domagalski, Kirsten
Gongola, Jennifer
Lyon, Thomas D
et al.

Publication Date

DOI

Peer reviewed

Detecting children's true and false denials of wrongdoing: Effects of question type and base rate knowledge

Kirsten Domagalski[1], **Jennifer Gongola**[2], **Thomas D. Lyon**[2], **Steven E. Clark**[3], **Jodi A. Quas**[1]

[1]University of California, Irvine, CA, USA

[2]University of Southern California, Los Angeles, CA, USA

[3]University of California, Riverside, CA, USA

Abstract

One common and unfortunately overlooked obstacle to the detection of sexual abuse is non-disclosure by children. Non-disclosure in forensic interviews may be expressed via concealment in response to recall questions or via active denials in response to recognition (e.g., yes/no) questions. In two studies, we evaluated whether adults' ability to discern true and false denials of wrongdoing by children varied as a function of the types of interview question the children were asked. Results suggest that adults are not good at detecting deceptive denials of wrongdoing by children, even when the adults view children narrate their experiences in response to recall questions rather than provide one word answers to recognition questions. In Study 1, adults exhibited a consistent "truth bias," leading them toward believing children, regardless of whether the children's denials were true or false. In Study 2, adults were given base-rate information about the occurrence of true and false denials (50% of each). The information eliminated the adults' truth bias but did not improve their overall detection accuracy, which still hovered near chance. Adults did, however, perceive children's denials as slightly more credible when they emerged in response to recall rather than recognition questions, especially when children were honestly denying wrongdoing. Results suggest the need for caution when evaluating adults' judgments of children's veracity when the children fail to disclose abuse.

1 | INTRODUCTION

Studies of adult survivors of child sexual abuse and of child victims of sexual abuse consistently suggest widespread problems with under-reporting of abuse. Many child victims do not tell anyone about the abuse, especially right away, and some never tell until adulthood (Azzopardi, Eirich, Rash, MacDonald, & Madigan, 2019; Goodman et al., 2003; Scurich & John, 2019). Non-disclosure even occurs when victims are directly questioned by well trained forensic interviewers (Hershkowitz, Horowitz, & Lamb, 2005), highlighting the significance of the problem. Victims who fail to disclose or explicitly deny abuse when

questioned do so for many reasons, such as to protect a known perpetrator, embarrassment, fear of punishment, or other perceived negative outcomes (McElvaney, 2015; Sivagurunathan, Orchard, MacDermid, & Evans, 2019). Yet failing to disclose can lead to enormous consequences, including not only a lifetime of adverse effects on victims, but also increases in others' risk of harm (Chen et al., 2010; Coles, Lee, Taft, Mazza, & Loxton, 2015; Irish, Kobayashi, & Delahanty, 2010; Widom, Czaja, & Dutton, 2008). It is imperative, therefore, to continue to direct scientific effort towards testing methods of improving victim identification.

Adults are key to the identification process. They may suspect abuse and question a child about their suspicions, or they may hear someone else question a child, for instance, if they serve as factfinders in a criminal case. In both situations, adults must evaluate what the child said, or, in the case of non-disclosure, what the child did not say, to determine what, if anything, happened. The purpose of the present research was to examine how well adults can evaluate the accuracy of children's non-disclosures when they are questioned about alleged wrongdoing. We focused on situations in which children failed to disclose the wrongdoing either because it never happened (i.e., they are truthfully denying wrongdoing) or because they wish to conceal the event (i.e., they are falsely denying wrongdoing). Specifically, in two studies, we systematically compared adults' ability to detect non-disclosures by children asked either recall (open-ended) or recognition (closed-ended) questions about a play interaction that may have involved toy breakage.

Despite consistent findings from past research showing that adults are typically quite poor at detecting deception in children, including when children are omitting information from their reports (Gongola, Scurich, & Quas, 2017; Nysse-Carris, Bottoms, & Salerno, 2011), our study is significant and needed for several reasons. For one, research has yet to compare in a systematic manner whether adults' deception detection abilities differ depending on whether they see children answer recall questions versus recognition questions. This is a noteworthy gap in the literature given strong scientific and practical emphasis placed on the utility of recall questions in eliciting accurate accounts from children (Cronch, Viljoen, & Hansen, 2006; Lamb, Orbach, Hershkowitz, Horowitz, & Abbott, 2007; Powell, 2005). Second, and as we turn to next, there are reasons to suspect that recall and recognition questions might differentially affect children's true and false reporting tendencies, including when non-disclosures are involved, and, in doing so, affect how well adults can detect which nondisclosures are true and which are false.

Regarding the effects of recall versus recognition questions on children's reporting, research consistently suggests enhanced value of recall prompts. Recall prompts are those in which children are asked to provide narrative answers to open-ended questions asking what happened, such as "Tell me what happened when XX" and "Tell me more about XX" (La Rooy et al., 2015; Saywitz, Lyon, & Goodman, 2017). Children provide lengthier and more detailed answers to recall questions than to recognition (i.e., yes/no, single-word answer) questions, which only require children provide the necessary short answers, and to which children rarely elaborate. In addition, children's recall narratives are less prone to errors than their answers to recognition questions (Lamb & Fauchier, 2001; Lamb, Hershkowitz, Orbach, & Esplin, 2011; Saywitz, Camparo, & Romanoff, 2010).

However, recall questions are not without limitations, with one of the most common being a reduction in the content of what children report. That is, children often fail to disclose negative, distressing, or embarrassing information only in response to recall questions (Goodman & Quas, 1995; Saywitz, Goodman, Nicholas, & Moan, 1991). For instance, among children who experienced an invasive and distressing medical procedure that involved genital contact, very few openly disclosed the genital portion of the procedure in response to recall prompts (Goodman & Quas, 1995). Similar patterns are seen in studies of children's reports of transgressions and other negative experiences, including abuse, with sizeable numbers of children not telling when given general prompts asking them to describe or tell an interviewer what happened (Hershkowitz, Lamb, & Katz, 2014; Rush, Stolzenberg, Quas, & Lyon, 2017; Stolzenberg, McWilliams, & Lyon, 2017).

Because some children do not disclose or do not provide sufficient details in response to recall prompts, interviewers often turn to recognition questions ("Did he touch you over your clothes?"), also called closed-ended questions, to obtain needed information. Benefits of recognition questions have been noted. For example, in the medical procedure study by Goodman and Quas (1995) just mentioned, when recognition questions were asked about genital touch with the aid of medical props, approximately 80% of the children disclosed. Likewise, in a study of interviews with 4–7-year-olds about a recent broken toy incident, less than half of the children told when asked recall questions, whereas 80% did so when asked closed-ended questions that explicitly mentioned "something bad happening," (Quas, Stolzenberg, & Lyon, 2018).

Recognition questions, however, are also problematic. Most notably, they can increase errors in children's reports. Children tend to answer such questions even when they do not understand the questions, do not know the answer, or when no correct answer exists (e.g., the questions are nonsensical; Peterson, Dowden, & Tobin, 1999; Waterman & Blades, 2011; Waterman, Blades, & Spencer, 2004). In addition, recognition questions do not reduce non-disclosures to zero. Some children still deny negative experiences. In the study by Quas et al. (2018) of 4–7-year-olds questioned about toy breakage, although eventually 80% disclosed, nearly 20% remained silent. They did not tell the interviewer about breakage, despite her asking about playing with each toy in a pointed manner and asking a yes/no question about something bad happening. The interviewer in the study, like adults generally, had to evaluate the child's responses to determine whether the denial was believable.

Turning more specifically to question type and adults' detection abilities, in theory, such abilities may vary by whether children are denying wrongdoing in response to recall questions or in response to recognition questions. Recall questions, regardless of whether children are attempting to conceal information or not, place considerable cognitive demands on children. Recall prompts require children to conduct a memory search to remember multiple event details. Recognition questions, in contrast, give children a simpler option, agree or not an option that does not necessarily require memory searches. When children are attempting to generate a lie, they need to remember their experience sufficiently well to generate narrative details while concurrently monitoring their narrative so as not to disclose certain information. At times, children may also need to generate mentally alternative plausible details to substitute for the concealed information, all of which demands cognitive

resources that are otherwise used to control behavior and statements relevant to the lie (Kail, 1991; Vrij et al., 2008; Vrij, Fisher, & Blank, 2017). Thus, it is not surprising that providing false narratives has been associated with increases in verbal and non-verbal cues of deception (Talwar, Murphy, & Lee, 2007b; Vrij et al., 2008).

Answering recognition questions does not place the same level of cognitive demands on children that answering recall questions does. When denying information when answering recognition questions does not necessarily need children to have a strong memory for an event, and does not require children generate a narrative answer or provide plausible alternative details. Instead, children can simply answer "no." The reduced demands are consistent with the age at which this form of deception emerges—in the preschool years— well before the age at which children can generate plausible narratives as a way of lying (Newton, Reddy, & Bull, 2000; Talwar, Gordon, & Lee, 2007a; Talwar & Lee, 2002).

Children's early proficiency at this form of deception, specifically denying information in response to recognition questions, also means that leakage and hence deception detection is likely minimized. Indeed, adults have an extremely difficult time judging children's true and false denials of a range of events. According to a recent meta-analysis of adults' detection abilities (Gongola et al., 2017), across studies, adults' performance when discriminating between children's true and false statements was only slightly above chance (54%). Professionals who work with children were marginally better (56%). Most studies included in the meta-analysis had adults view scenarios involving children who answered either only recognition questions or recall and recognition questions combined. None compared whether adults' abilities varied between the two question types. Given the different demands placed on children during recall versus recognition question portions of interviews, it is important to ascertain whether adults' detection abilities vary depending on what types of question— and response—they hear.

Of course, adults' detection abilities may also be affected by their general beliefs about whether children are telling the truth in the first place. Prior work suggests that adults generally assume that children are telling the truth (e.g., Edelstein, Luten, Ekman, & Goodman, 2006; Gongola, Scurich, & Lyon, 2019; Talwar, Lee, Bala, & Lindsay, 2006; Westcott, Davies, Graham, & Clifford, 1991). Such an assumption makes sense, given that most children (and adults for that matter) are honest in conversations. Yet, when there is the possibility of deception, adults often underestimate individuals' tendency to lie, leading to a truth bias when evaluating responses (Bond & DePaulo, 2006; Gongola et al., 2017; Swann, 1984). Insight into whether this bias can be shifted, for instance by telling adults that children can lie or saying how often children lie, could be of considerable value. Kassin, Meissner, and Norwick (2005) demonstrated that it may be possible to do just this in a study of police officers' judgments about the veracity of confessions presented via videotapes. In Study 1, police officers were given no information about how often (i.e., the base rate at which) the true and false confessions occurred in the tapes. The officers demonstrated a truth bias (i.e., a tendency to believe the confession was true). In Study 2, however, officers were informed with the overall base rate of false confession in the videos. Here, the adults' truth bias was eliminated.

Although the participants of Kassin et al. were professionals judging possible confessions rather than laypersons judging children's statements, findings nonetheless indicate that it may be possible to adjust expectations surrounding truthfulness. It is important to establish whether this is in fact the case when children may be engaged in deception, and, of importance, whether these adjustments improve accuracy. We speculate that this may be especially likely when adults view children's responses to recall questions and are, as a result of base rate instructions, paying close attention to children's narratives.

2 | PRESENT RESEARCH

We assessed how the type of question asked (recall versus recognition) affected adults' abilities to detect whether children were telling the truth or lying when concealing information about toys breaking. In two studies, we had adults watch four videoclips of a child being questioned about a transgression. Each clip showed an interviewer asking a child either recall or recognition questions about what happened when the child had played with a confederate (Stolzenberg, et al., 2017). All four clips showed children answering either recall questions or recognition questions, and all children failed to disclose that toys broke. Two denials were true, and two were false. After viewing each video, adults rendered judgments about whether any toys broke and the child's credibility. In Study 1, adults were not given any information about how many of the videos involved situations in which toys actually broke. In Study 2, adults were given base rate information: In half of the videos the child was telling the truth, and in half the child was lying. Thus, both studies conformed to a 2 (type of question: recall versus recognition) × 2 (type of denial: true versus false) mixed factorial design, with type of question varying between participants and type of denial varying within participants.

3 | STUDY 1

Two hypotheses were advanced. First, we anticipated that accuracy distinguishing true and false denials would be higher among adults who saw children answer recall rather than recognition questions. Second, we expected a truth bias, such that adults' accuracy would be higher when judging true denials and lower when judging false denials.

3.1| Method

3.1.1 | Participants—Power analysis, using G*Power, with power $(1 - \beta)$ set at 0.80 and $\alpha = 0.05$, one tailed, indicated that a sample of at least 276 participants would be sufficient to detect a small to moderate main effect size (Cohen's $d = 0.30$) between recall and recognition groups in the main outcomes. This anticipated effect size was based on findings from prior similar studies of deception detection (Gongola et al., 2019). To account for potential attrition, 300 participants were recruited from CloudResearch (formerly TurkPrime), an online platform where surveys are posted for volunteers to complete for monetary compensation. Of these, seven participants exited the survey before answering any questions, and eight completed the study but were removed for failing a series of attention checks (Oppenheimer, Meyvis, & Davidenko, 2009). The final sample included 285 participants (61% female).

Participants' ages ranged from 18 to 82 (M = 38.98, SD = 12.12). A majority of participants reported their race to be White (81%), followed by African-American (8.07%), Asian (6.32%), American Indian or Alaska Native (2.81%), or other (1.75%). All participants were residents of the United States, over half indicated they were parents (59.64%), and 18% said they had jobs that involved working with children (e.g., teacher, social worker).

3.1.2 | Materials and procedure—Following informed consent, participants watched four videos (order randomized) of interviews with children, each of which showed a child answering questions about an interaction between the child and an adult confederate (for details of the original study of the interaction and subsequent interviews, see Stolzenberg et al., 2017). The interaction had taken place right before the interview and consisted of the child (age 4–7 years) and an adult confederate playing with toys. For half of the children, two toys broke while playing, and the adult confederate asked the child to keep the breakage a secret. For the other half of the children, no toys broke. Immediately after the interaction, the confederate left. An interviewer entered and asked the child about what happened, first with recall prompts and then with recognition questions probing about activities with each toy.

From the set of children who completed the original study, interviews from four children (two male, two female) were selected for inclusion as stimuli in the current research. None of these children disclose that a toy broke at any point in the interview. Two children's denials were true because no toys broke, and two children's denials were false because toys broke. It was important scientifically to hold the child constant across the two question type conditions. Thus, all children in the final stimulus sample denied toy breakage in *both* recall and recognition portions of the interview so that the same child could be included as stimuli for each of the two conditions. In this way, adult participants in each condition viewed different parts of the same child's video. By holding the child constant, we eliminated potential biases that could arise from viewing different children in the two question type conditions. However, this approach also meant that our videos were comprised of interviews with children who were particularly good deniers (truthfully or falsely), a point to which we return in the general discussion.

Each video was approximately two minutes in length and was edited to create versions in which the child was shown being asked only recall or only recognition questions. The number of questions asked in the recall condition varied slightly across children. All were first asked to "tell me everything that happened when the man came in while I was gone." Then follow-up prompts (e.g., "You mentioned the airplane, tell more about that.") asked about each toy the child mentioned (M = 7.75 follow-up questions). In the recognition condition, children were asked eight closed-ended questions about each toy that was played with duringthe interaction ("Did the airplane break?"). No additional portions of the interviews with children were shown.

In the current study, following consent, participants were randomly assigned to either the recall or recognition condition. They were given the following written instructions (important points emphasized via bold in the original text): "A child was alone with a stranger for several minutes. They played with six different toys. Some of the toys may or

may not have broken during play. Afterward, the child was interviewed about the play session. Your task is to evaluate the child's credibility and decide whether you think any of the toys broke."

Following the instructions, the interview videos were presented (order counterbalanced). After each video, participants were asked to indicate whether they thought any toys broke during the play session according to a six-point scale: 1, yes, I am certain a toy broke, to 6, no, I am certain that a toy did not break. Participants then rated how credible, reliable, believable, and likable the child was, each on a 1 (not at all) to 6 (extremely) scale. Finally, participants answered demographic questions (e.g., age, whether participants have children, whether occupation involved working with children) and attention check questions. At the end, participants were compensated and thanked for their participation.

3.2 | Results

3.2.1 | Preliminary analyses—
First, we tested whether gender, parental status, or having an occupation that involved working with children was related to participants' detection accuracy, discriminability, and perceptions of the child (see below for details of computation of outcome variables). Regarding gender, t-tests revealed that, although women and men did not differ in accuracy or discriminability, women ($M = 4.27$, SD = 0.89) viewed children more favorably overall than did men ($M = 3.94$, SD = 0.84), $t(283) = 3.15$, $p = 0.002$, Cohen's $d = 0.38$, regardless of the child's honesty. Parental status was related to overall accuracy and discriminability, but not perceptions of the child. Parents were more accurate (M = 0.56, SD = 0.20) than non-parents ($M = 0.48$, SD = 0.20), $t(283) = 3.27$, $p = 0.001$, Cohen's $d = 0.40$, and better able to discriminate between true and false deniers ($M = 0.12$, SD = 0.40) than non-parents were ($M = −0.04$, SD = 0.39), $t(283) = 2.56$, $p = 0.01$, Cohen's $d = 0.34$. Finally, occupation (working with children versus not) was unrelated to participants' accuracy, discriminability, or perceptions of the child. Of note, although participant gender and parental status predicted some outcomes, neither interacted with question type or denial type to affect responses. Thus, neither was included in the main analyses. We nonetheless return to these demographic differences in the general discussion.

3.2.2 | Proportion accuracy—
Four dichotomous scores, one for each child, were created based on participants' ratings on the six-point scale about whether they thought toys broke or not (i.e., certain toys broke to certain no toys broke). For the two interviews featuring *true denials* (i.e., no toys broke and children were honestly denying breakage), ratings indicating that no toys broke on the scale—that is, ratings of 4–6—were recoded as 1 for correct, while ratings of 1–3, indicating that toys broke, were recoded as 0 for incorrect. Conversely, for the two videos featuring *false denials,* ratings indicating toys broke (ratings of 1–3) were recoded as 1 for correct, and ratings indicating no toys broke when some actually did (ratings of 4–6) were recoded as 0 for incorrect. Scores were then averaged separately for the two true and false denial interviews to create accuracy scores for each participant.

The accuracy scores were entered as within-subject dependent measures into a 2 (question type: recall versus recognition) × 2 (denial type: true versus false) mixed model ANOVA.

Only the main effect of denial type was significant, $F(1, 283) = 27.02$, $p < 0.001$, $h_p^2 = 0.087$. Participants were more accurate identifying true denials ($M = 0.62$, SD = 0.37) than false denials ($M = 0.43$, SD = 0.37). The question type and question type × denial type interaction were non-significant $F(1, 283) < 1.10$, both $p > 0.29$. In fact, as seen in Table 1, participants' mean accuracy scores were similar between the recall and recognition conditions.

Follow-up single sample t-tests evaluated whether the true and false denial accuracy scores significantly differed from chance (0.50). Both did and in opposite directions. True denials were slightly but significantly above chance, $t(284) = 5.62$, $p < 0.001$, while false denials were slightly but significantly below chance, t(284) = −3.21, $p = 0.001$.

3.2.3 | Discriminability—To complement and extend our accuracy analyses, we tested more directly how well participants could discriminate between true and false denials. Discriminability, although related to accuracy, considers concurrently how well adults correctly identify false denials (i.e., hit rate) and how often adults incorrectly label true denials as being false (i.e., false alarm rate). Including the additional discriminability analyses, therefore, eliminates potentially lingering questions about whether our conclusions regarding adults' abilities are due to the specific types of scores we analyzed. Converging evidence across different indicators would suggest not.

Hit rates were calculated by taking the proportion of false denials correctly labeled as false, and false alarm rates were calculated by taking the proportion of true denials incorrectly labeled as false. The difference between the hit and false alarm rates (HR - FAR)[1] did not significantly differ between participants who viewed recall versus recognition questions, $t(283) = −1.05$, $p = 0.29$; see Table 1. However, collapsing across question type, the aforementioned truth bias again was evident. Participants rated a greater percentage of children's responses as truthful (59.74%) than actually were (50%), $t(284) = 5.27$, $p < 0.001$.

3.2.4 | Confidence—To ascertain whether participants' confidence in their judgments differed between question type conditions, we calculated confidence scores. Participants who were "certain" that a toy did or did not break were coded as 3, those who said that a toy "probably" did or did not break were coded as 2, and those who indicated a toy "maybe" did or did not break were coded as 1. Mean confidence composite scores were computed and compared between conditions. An independent samples t-test revealed that participants were no more confident in their judgements in the recall condition ($M = 1.72$, SD = 0.44) than in the recognition condition ($M = 1.80$, SD = 0.46), $t(283) = 1.63$, $p = 0.10$. Thus, viewing children's narrative recall responses did not lead participants to be more (or less) confident in their judgments, regardless of their judgements' actual accuracy, compared to viewing children's responses to recognition questions. When we recalculated average confidence ratings separately for all judgements in which adults said children were telling the truth and for all judgements indicating children were lying, adults were more confident when they

---

[1]We calculated discriminability using HR - FAR rather than d-prime because the number of stimuli in present study undermines the advantage for d-prime over HR - FAR. Having four truth-lie judgments per participant resulted in a higher number of hit and false alarm rates equal to 0 or 1, which are undefined in d-prime analyses and thus would need to be adjusted.

reported that children were telling the truth ($M = 1.91$) than when they reported that children were lying ($M = 1.48$), $t(165) = 6.57$, $p < 0.001$.

3.2.5 | Perceptions of credibility—In a final analysis, we evaluated whether participants' views of the child varied as a function of the types of question the child was asked or whether the child was honestly or falsely denying breakage. Immediately after each video, participants rated how reliable, credible, believable, and likeable the child was on scale from 1 (not at all) to 6 (extremely). Cronbach's alpha indicated high reliability across ratings for each child ($\alpha > 0.89$). The items were averaged to create credibility scores for each child, from which separate mean scores were computed for the two denial types. The credibility scores were entered into a 2 (question type) × 2 (denial type) ANOVA. The interaction approached significance, F(1, 283) = 3.62, $p = 0.058$, $h_p{}^2 = 0.013$. When children were honestly denying breakage, participants in the recall condition viewed children slightly more favorably than did participants in the recognition condition (Table 1). When children were falsely denying breakage, however, participants' perceptions were comparable regardless of what types of question they saw children answer.

3.3 | Discussion

In Study 1, we tested whether adults were better at discerning true and false denials of a transgression when they saw children answering recall rather than recognition questions. Our assumption was that leakage would be more evident when children narrated while concealing information about toys breaking than when children were only providing yes or no responses to recognition questions. Such was not the case. Participants' accuracy was unaffected by the type of question asked. In fact, across question type and regardless of whether children were answering truthfully or deceptively, participants' accuracy hovered near chance, a pattern consistent with prior work on deception detection with children (see Gongola et al., 2017, for a review).

A more consistent finding was participants' differential performance when viewing true versus false deniers. Participants were more accurate when evaluating true denials, often correctly believing the child, but less accurate when evaluating false denials, often incorrectly believing the child. We offer two potential explanations for these patterns, both of which may be operating concurrently. The first is a selection effect. As mentioned above, we may have selected children who were particularly good at concealing information by design. Children had to meet stringent criteria to be included in the study stimuli. This is especially true of children in the false denial condition, who must have denied toy breakage in response to both recall and recognition questions. In other studies, a sizable number of children who fail to disclose transgressions in response to recall questions do so when asked recognition questions, particularly about bad things happening (e.g., Lyon et al., 2014; Quas et al., 2018). Children who maintain denials across questions, therefore, are likely to be especially adept concealers and hence difficult to detect (Gongola, Quas, Clark, & Lyon, 2020; Gongola, Williams, & Lyon, under review). By holding the child constant between the question type conditions, we were able to conduct a rigorous test of the effects of question type. However, it will be important in future work to include different children, including those whose responses vary, to assess whether adults' detection abilities then vary.

A second possible reason why participants were more accurate when evaluating true denials and less accurate when evaluating false denials focuses not on the deceptive abilities in children per se, but instead on adults' evaluation tendencies. A truth bias may well have been affecting adults' judgments. Participants rated a greater percentage of children's responses as truthful than the actual base rate of truthful children in the study. Furthermore, adults were more confident when they indicated that children were telling the truth than when they indicated they were lying. This pattern makes sense given that people generally assume others are telling the truth rather than lying (Bond & DePaulo, 2006; Vrijet al., 2008), and in fact most interpersonal interactions involve honest communication (Street & Richardson, 2015). A similar truth bias was evident in the meta-analysis of adults' ability to detect deception in children by Gongola et al. (2017): Accuracy was higher when adults evaluated children's true statements than when adults evaluated children's false statements.

Whether such a bias would exist if adults were more skeptical or motivated to detect deceivers is an important follow-up question. That is, in Study 1 we informed participants that some toys may have broken but did not provide additional information. Thus, participants may have lacked strong expectations about the possibility of children lying, and as a result overlooked subtle deceptive cues evident in children's recall narratives. We hypothesize that, had we explicitly told participants that some children are lying or provided details on how many children lied, accuracy would have been higher, and the advantage of recall would have emerged. Kassin et al. (2005) demonstrated that base rate information about true and false confessions was effective at eliminating law enforcement's tendency to take adults' confessions at face value, another potential form of truth bias. The provision of base rate information about children's true versus false denials may similarly eliminate biases in adults' judgments and possibly improve accuracy. We tested this possibility in Study 2.

4 | STUDY 2

The design of Study 2 and the study materials were identical to those in Study 1, with the exception that participants were explicitly informed that 50% of the children were lying. We expected that, with this additional information, the truth bias would be eliminated. We further expected that, once this occurred, a question type effect would emerge such that participants would be better at discriminating true from false deniers in the recall condition than in the recognition condition.

4.1 | Method

4.1.1 | Participants—To ensure power to detect potential differences in the study's main outcomes we ran an updated power analysis based on observed effect sizes in Study 1. Power analysis, using G*Power, with power $(1 - \beta)$ set at 0.80 and $\alpha = 0.05$, one tailed, indicated that 398 participants were necessary to detect conservative main effects (Cohen's d = 0.25). To account for potential attrition, 450 participants were recruited from CloudResearch (formerly TurkPrime) in exchange for monetary compensation. Thirty-two respondents opened the study but exited before answering questions, and eight respondents completed the study but were removed for failing the attention checks. The final sample

included 410 participants (61.50% female). Participants' ages ranged from 18 to 80 (M = 39.37, SD = 12.96). A majority reported their race to be White (79.8%), followed by African-American (9.8%), Asian (5.1%), American Indian or Alaska Native (2.2%), or other (3.2%). All participants were residents of the United States, over half indicated they were parents (55.90%), and 19.3% said they had jobs in which they worked with children.

4.1.2 | Materials and procedure—Study 2 materials and procedures, including the four interview clips shown, were identical to those of Study 1, with the exception that base rate instructions were given to participants before the videos were shown. Specifically, at the start of the session, after consent but before the interviews were presented, the following written instructions were provided: "In this study, you will watch interviews with four different children. **Two of them lie at some point during the interview about a toy breaking, and two of them are always telling the truth.** For each video, your task is to evaluate the child's credibility and decide whether you think any of the toys broke" (bold included in instructions). In addition, after each video, prior to answering questions, participants received an additional reminder about base rates: "Do you think that any of the toys broke? **Keep in mind that 2 out of the 4 children in this study lie about toy breakage at some point during the interview.**" As in Study 1, immediately after each interview, participants indicated whether toys broke (1, certain toys did break, to 6, certain toys did not break) and rated the child's believability, reliability, credibility, and likeability.

4.2 | Results

The analytic approach largely followed that in Study 1. Accuracy, discriminability, and credibility scores were calculated in the same manner. For accuracy, we dichotomously coded whether participants were correct or not in their judgements on the six-point scale and averaged these separately for the true and false denials. Discriminability was calculated as hit rate minus false alarm rate, and credibility was computed by averaging participants' ratings of each child's credibility, reliability, believability, likeability (all $\alpha > 0.90$) and combining these based on denial type. Analyses tested for confounds, followed by analyses of accuracy, discriminability, confidence, and credibility. In addition, when possible, comparisons between Studies 1 and 2 are reported.

4.2.1 | Preliminary analyses—Consistent with Study 1, women (M = 4.01, SD = 0.76) viewed children as slightly more credible than did men (M = 3.83, SD = 0.82), $t(408)$ = −2.20, p = 0.03, Cohen's d = 0.23. Unlike Study 1, however, parental status was unrelated to participants' accuracy or discriminability scores. Also, occupation (working with children versus not) was unrelated to accuracy or discriminability, but was related to credibility: Participants who worked with children viewed them slightly less favorably (M = 3.75, SD = 0.83) than did participants who did not work with children (M = 3.99, SD = 0.77), $t(408)$ = −2.43, p = 0.02, Cohen's d = 0.30. Neither participant gender nor occupation interacted with the manipulated variables. We return to the demographic variables in the general discussion.

When we tested how well participants followed the base rate instructions by selecting two truth-tellers and two liars in their judgements, despite reminders, only 46.6% of the participants did as instructed. Of the remaining, 28.8% selected less than half of the children

as telling the truth and 24.7% selected more than half of the children as telling the truth. Comparisons of accuracy, discriminability, and child credibility between participants who did and did not follow instructions failed to reveal any differences. Thus, following instructions per se did not affect participants' judgments. Of primary interest in our study, however, was whether the provision of the base rate information improved participants' detection abilities.

4.2.2 | Proportion accuracy—When the accuracy scores were subjected to a 2 (question type) $\times$ 2 (denial type) mixed model ANOVA, only the main effect of denial type was significant, $F(1, 408) = 9.12$, $p = 0.003$. However, the direction of the effect was opposite to that uncovered in Study 1. Participants were significantly more accurate identifying false denials than true denials (see Table 1). Neither the main effect of question type nor the question type $\times$ denial type interaction was significant, both $F(1, 408) < 0.82$, both $p > 0.37$. Thus, in contrast to our hypothesis, even with base rate instructions about how often children were lying, participants' ability to discern when children were honestly or falsely denying breakage did not vary as a function of the types of question participants saw children answer.

Yet, giving participants base rate information was not without effects. Skepticism increased. Indeed, t-tests comparing accuracy scores with chance (0.50) revealed that accuracy for true denials was significantly below chance, $t(409) = -2.12$, $p = 0.035$, while accuracy for false denials was significantly above chance, $t(409) = 2.33$, $p = 0.02$. Finally, participants' overall accuracy did not differ between Study 1 and Study 2, $t(693) = 1.45$, $p = 0.15$ (Table 1).

4.2.3 | Discriminability—No significant differences emerged in discriminability scores (HR-FAR) in the recall versus recognition conditions, $t(408) = 0.55$, $p = 0.58$ (Table 1). Thus, although participants were more accurate in identifying false than true denials, they did not demonstrate a lie bias. Participants indicated that children were falsely denying breakage 51% of the time, a percentage comparable to the 50% rate at which children were actually lying, $t(409) = 1.11$, $p = 0.26$.

Discriminability scores in Study 2 did not significantly differ from those in Study 1, $t(693) = 1.50$, $p = 0.14$. However, participants in Study 2 had both higher hit rates (i.e., correct identification of false denials), $t(693) = 4.02$, $p < 0.001$, Cohen's $d = 0.33$, and higher false alarm rates (i.e., inaccurate labeling of true denials as false), $t(693) = 5.81$, $p < 0.001$, Cohen's $d = 0.47$, than in Study 1 (see Table 1).

4.2.4 | Confidence—As in Study 1, to ensure that confidence did not vary between conditions, participants' confidence in their judgements ("certain," 3; "probably," 2; "maybe," 1) were compared between conditions. No significant differences emerged (recall $M = 1.71$, SD = 0.41; recognition $M = 1.73$, SD = 0.46), $t(408) = 0.99$, $p = 24$). Thus, viewing children's narratives did not lead participants to report more (or less) confidence in their judgments, regardless of the judgements' actual accuracy. Likewise, Study 2 participants who received base rate instructions ($M = 1.73$, SD = 0.44) did not differ in confidence from Study 1 participants who did not receive such instructions ($M = 1.76$, SD = 0.45), $t(693) = 0.86$, $p = 0.39$.

Finally, we tested whether participants' confidence varied depending on whether they rated children as telling the truth or lying. We averaged confidence ratings separately for the children who participants said were telling the truth and the children who participants said were lying. Confidence did not differ when adults labeled children as truly denying ($M =$ 1.68) or falsely denying breakage ($M = 1.70$), $t(299) = -0.44$, $p = 0.66$.

4.2.5 | Perceptions of credibility—When participants' credibility scores were entered into the 2 (question type) × 2 (denial type) ANOVA, the question type $F(1, 408) = 5.23$, $p = 0.023$, $h_p{}^2 = 0.013$, and question type by denial type interaction, $F(1, 408) = 5.79$, $p = 0.017$, $hp2 = 0.014$, were significant. When children were honestly denying breakage, participants who saw children answer recall questions evaluated the children's credibility more favorably than did participants who saw children answer recognition questions. However, when children were falsely denying breakage, participants' evaluations did not vary depending on whether they saw children answer recall or recognition questions (Table 1). Finally, Study 2 participants viewed children as slightly less credible than did Study 1 participants $t(693) =$ 3.14, $p = 0.002$, Cohen's $d = 0.22$.

4.3 | Discussion

In Study 2, participants were explicitly informed that two children were telling the truth about toy breakage and two children were lying when denying toy breakage. The goals of providing this information were twofold: to reduce participants' truth bias tendencies and increase their deception detection accuracy. On the one hand, we were effective at eliminating the truth bias. On the other, however, our base rate instructions did not improve accuracy, perhaps because participants were now overly skeptical. Moreover, the type of question children answered, even with base rate information, did not alter participants' discriminability, which instead was poor in both the recall and recognition question conditions.

5 | GENERAL DISCUSSION

When adults suspect wrongdoing or harm, but a child fails to disclose or overtly denies any such experience, the adults are in a difficult position: They must evaluate the child's concealment as being either true or deceptive and render at times crucial decisions (e.g., whether to report abuse, whether to convict a defendant) accordingly. Our results are consistent with prior work (Bond & DePaulo, 2006; Gongola et al., 2017) in showing that adults' ability to differentiate true and false denials is quite limited. Of importance, though, our results extend this prior work by showing that this limited ability does not necessarily improve when adults are able to see a child provide narrative answers to recall questions rather than just simple yes/no or single-word answers to recognition questions. Despite the fact that recall questions are often touted as best practice when it comes to interviewing children and despite the fact that children's responses to recall questions are often more accurate than their responses to recognition questions, adults' detection abilities may not be altered. Furthermore, although providing adults with base rate information about the occurrence of false denials can reduce biases to overly believe a child, the information does not necessarily enhance accuracy.

We designed Study 1 to investigate whether adults' accuracy at judging the veracity of children's denials varied as a function of whether the adults saw children concealing information in response to recall questions or explicitly denying information in response to recognition questions. We assumed that, because of the demands associated with generating false plausible narratives while remaining cognizant of the need to avoid revealing specific details (Saykaly, Crossman, & Talwar, 2017; Vrijet al., 2008), children would display cues perceptible to adults, who in turn would be able to detect false denials, or at least detect them more easily than had the adults only seen children provide brief (e.g., "no") responses to recognition questions. However, such was not the case. Neither adults' accuracy nor discriminability varied as a function of the types of question children were asked.

The nature of the deceptive task, combined with our selection criteria for the children's videos, may have contributed to the lack of question type effect on adults' judgements. First, although the lies themselves were about a salient incident—a broken toy—the incident was embedded in a relatively interactive play session that had only recently occurred. Thus, children had a number of other true activities that they could recount, allowing them to describe other true details without having to expend effort generating false plausible alternative details for a majority of the interview. The latter type of deception is more difficult and does lead to leakage, including that which is perceptible in formal analyses of children's statements (Tye, Amato, Honts, Devitt, & Peters, 1999). It may also be that, at young ages, narrative reporting in general is somewhat demanding cognitively for children, who, regardless of deception, still need to conduct a memory search, form a coherent report, and recount this report in narrative format. All of this requires effort, which may have led to fewer differences in cognitive load between the true and false deniers in the study than might have been the case had a more demanding or complex lie been involved, or perhaps had older children had been included (who can more easily provide detailed narrative recall reports: Wyman, Foster, Crossman, Colwell, & Talwar, 2019). Additional research with varying experiences or ages would be worthwhile to identify whether there are conditions under which adults' abilities differ depending on how children are asked—and how they respond—to questions about past experiences.

Second, the children selected for inclusion were quite adept at maintaining their lies, having done so throughout their entire interview. Thus, while adults might be able to detect concealment in children prone to leakage who may not be able to maintain their lies (see, e.g., Quas et al., 2018), adults may not be able to discern the particularly good concealers from children truthfully denying the transgression. Of course, we only asked a small number of questions, all of which were straightforward and referred to a play activity that had just taken place. It is possible that even the best concealers, if asked to provide more elaborate denials in response to recall, would have generated cues that enabled adults to detect deception at higher rates than if they had only seen children responding to recognition questions.

As a side note, some hints of effects of question type were evident in adults' ratings of children's credibility. Adults viewed children more favorably when the adults saw children answer recall rather than recognition questions, with this being slightly more pronounced when children were honestly rather than falsely denying breakage. Adults may feel as

though they understand children more or that children's stories can be believed when they hear children narrate, a possibility worth exploring in the future. In other work, researchers have found that adults evaluate children's claims of abuse more favorably when children express emotion rather than display neutral expressions. Although emotional expressions likely affect adults' evaluations in unique ways, it may also be the case that more information about children, either from narrative answers or other modalities, leads adults to feel more positively (Cooper, Quas, & Cleveland, 2014). That the question type effect on credibility only reached significance in one of the studies, though, suggests that any effect is likely small, and that when narratives involve concealing information, adults' overall judgments are unlikely to be affected in meaningful ways.

In Study 1, the adult participants were alerted to the fact that some toys may have been broken and that children may or may not be telling the truth. However, adults were not explicitly told how often children were lying about toy breakage. The adults, in response, rated children as being honest—in this case in their denials of breakage —more often than children actually were, and were more confident overall when rating children as truthful than deceitful (i.e., adults were not as confident in their judgements that children were lying), thus suggesting a truth bias.

In Study 2, we attempted to mitigate this bias by telling the adults how often children were telling the truth or lying in the videos, that is, we gave adults base rate information about children's deception, imitating a successful manipulation employed by Kassin et al. (2005). Because truth biases can reduce discriminability, we anticipated that reducing the bias would enhance adults' deception detection discriminability and possibly accuracy, especially when they saw children's recall responses. Adults did, in some ways, become more accurate with the base rate instructions in that they no longer incorrectly assumed too often that children falsely denying breakage were being honest. However, adults also no longer thought that the children who were truthfully denying breakage were being honest. Finally, confidence ratings no longer differed based on adults' judgments of veracity but instead they were equally confident when they thought children were telling the truth and lying. Thus, general skepticism about children's statements emerged.

In line with this, comparisons between studies showed that adults rated the children in Study 2 as less credible than in Study 1, even though the children in the videos were identical across studies. Talwar et al. (2006) used a different approach to reduce truth bias and found similar results. Mock jurors were shown videos of a child testifying about a true or fabricated past event. Jurors who saw the direct examination portion tended to believe the child, regardless of the child's actual honesty, whereas jurors who saw the direct and cross-examination portions did not show this tendency. However, even with the truth bias reduced, mock jurors' deception detection accuracy remained low, similar to findings from our study. Given consistent trends suggesting that adults' skepticism is manipulable, it would be of interest to assess the impact of increasing adults' skepticism more broadly, including in terms of affecting adults' judgments of cases in which children's statements are included as evidence.

Overall, the costs and benefits of giving base rate information on adults' biases are complex and must be weighed carefully. Truth biases increase the risk of failing to identify children who experienced wrongdoing but deny its occurrence, a significant problem in abuse cases for which non-disclosure continues to be a particularly devastating problem (Alaggia, Collin-Vezina, & Lateef, 2019; McElvaney, 2015). Yet, overly skeptical perceptions of children's honesty can also be problematic. Skepticism may simply lead adults to ask a greater number of questions to find out what happened, if anything, to a child. Although even relatively young children are capable of answering repeated questions about an alleged transgression accurately (Lyon, Malloy, Quas, & Talwar, 2008), and their errors tend to be due to external influence or pressure to confirm or change answers (Goodman & Quas, 2008), adults sometimes believe that repeated questions reduce children's credibility and accuracy (Ruva & Bryant, 2004), which ultimately could damage a case.

Moreover, it is not entirely clear what "base rate" information is accurate and hence what should be provided. Certainly, there is some evidence that a sizable percentage (e.g., between 20% and 50%) of children fail to disclose or deny true abuse (Azzopardi et al., 2019; Malloy, Lyon, & Quas, 2007). However, this percentage varies depending on other characteristics of children and the type of abuse, and the specific percentage reported could alter adults' judgments in meaningful ways. Moreover, it is unlikely that adults have specific underlying assumptions per se about what base rates are. Instead, without additional evidence to indicate otherwise, our data, and those of other studies (e.g., Gongola et al., 2017; Talwar et al., 2006), seem to indicate that adults have a general tendency to believe what children say. How strong this tendency is (and whether adults feel that they could quantify it) would be an interesting question to explore. Overall, though, given that experts at times testify about the commonness of denials, non-disclosures, and even recantations, the effects of adults' implicit assumptions about children's honesty are important to understand.

Finally, in our research, accuracy and discriminability hovered near chance, both with and without base rate information. Thus, alterations to truth biases or skepticism ultimately do not help adults better discriminate between true and false denials by children. Given the consistency with which findings show that adults have limited abilities when it comes to detection, more efforts need to focus on whether actual differences exist in children's reporting when they are telling the truth or lying, and on training adults to detect those differences.

Turning to the demographics of our samples, several small but significant associations emerged between participant characteristics and judgments. Women viewed children as more credible than men did, a pattern consistent with prior studies showing, across various types of studies of juror decision-making in child sexual abuse cases and adults' perceptions of children's statements, that women hold more pro-child and pro-victim views than do men (Gabora, Spanos, & Joab, 1993; Quas, Thompson, & Clarke-Stewart, 2005; see Bottoms, Golding, Stevenson, Wiley, & Yozwiak, 2007, for a review). However, these favorable views did not translate into gender differences in accuracy or discriminability, which were comparable between genders. Also, parents were slightly better at discerning true and false denials than were non-parents. Consistent with this trend, Talwar, Crossman, Williams, and Muir (2011) found that adults who interact more with children have a slight advantage when

judging their statements relative to adults who do not (see also Talwar, Renaud, & Conway, 2015). Parents certainly interact with children on a regular basis or more so than do non-parents. However, even in our study, parents' overall accuracy was still only at 56%, a value only slightly above chance, and parents' increased accuracy was only evident in Study 1. Thus, at best, there may at times be a slight advantage for parents, though this is small and perhaps fleeting depending on other instructions. Finally, adults whose careers did not involve working with children held more favorable views of children than did adults who work with children. This difference emerged only in Study 2. Perhaps adults who work with children were more receptive to information about the possibility or commonness (50% of the time) of children lying, leading them to be slightly more critical in their perceptions. Adults' actual accuracy in evaluating children's honesty, though, was unaffected. Overall, the small magnitude of these effects, combined with their appearance and disappearance across studies, limit their practical utility when considering adults' deception detection abilities broadly.

As with all research, the generalizability of our results is limited by our study design, and future work on the topic of deception detection with children is clearly warranted. First, as mentioned, children in the videos answered limited numbers of recall and recognition questions about a brief interaction. It would be of interest to examine how adults evaluate children's true and false responses about more complex events, including their denials of entire events. Also of interest is how adults' evaluations of children's denials vary as a function of age, given that children's deceptive abilities in general vary with age (Talwar et al., 2007a), as do children's tendencies to conceal, deny, or take back claims of abuse (Goodman-Brown, Edelstein, Goodman, Jones, & Gordon, 2003; Malloy et al., 2007; Leach, Powell, Sharman, & Anglim, 2017). Finally, we included clips from four children. Other studies have included smaller (e.g., Talwar et al., 2006) or larger (e.g., Edelstein et al., 2006) numbers of stimuli and found similar results. While we do not know of any specific reasons why the number of stimuli might affect adults' judgements, it is an interesting question to explore.

Second, we focused exclusively on false denials in the current research. Non-disclosure is widely believed to be a much more pervasive problem than false claims of abuse, making our research particularly relevant to the more common type of situation that adults may encounter. Adults' detection abilities regarding children's false claims of wrongdoing may be affected by question type in ways not evident when children are falsely denying wrongdoing. Generating false narratives in response to recall questions about an untrue event, especially while maintaining a consistent and coherent story, is more difficult than answering recognition questions about the false event, leading to potential leakage that could be detectable to adults. Practically, though, such an investigation may be difficult with the type of interaction and transgression paradigm experienced by the children in this study. It is extremely difficult to induce children to make false claims about transgressions, even when pressured by adults in suggestive manners (Rush et al., 2017). Alternatively, children can in limited circumstances be coached to make false reports (Lyon et al, 2008); comparing adults' detection abilities when those reports are provided in response to recall versus recognition questions would be a useful addition to the current series of studies. Given the substantially larger problem of false denials than false allegations, however, there is a

compelling need to continue to direct attention toward understanding how best to detect undisclosed abuse to intervene and protect children. Our research represents an important contribution to this need.

In closing, adults such as police officers, judges, attorneys, social workers, and caretakers are often tasked with making crucial, potentially life altering judgments based on children's statements regarding sexual abuse. Given what we know about the high prevalence of non-disclosures, efforts should continually be made to improve adults' abilities to detect when children are reluctant to disclose information. While adults seem to have some correct notions regarding the types of interview condition that result in the most credible responses (e.g., non-suggestive questioning, open recall), deception detection efforts often fall short. Thus, it is important to continue to test interview protocols, individually or combined, that may help adults to better distinguish when children are concealing vital information, and thus ultimately prevent further victimization and a lifetime of adverse effects.

ACKNOWLEDGEMENT

REFERENCES

Alaggia R, Collin-Vézina D, & Lateef R (2019). Facilitators and barriers to child sexual abuse (CSA) disclosures: A research update (2000–2016). Trauma, Violence, & Abuse, 20, 260–283. 10.1177/1524838017697312

Azzopardi C, Erich R, Rash CL, MacDonald S, & Madigan S (2019). A meta-analysis of the prevalence of child sexual abuse disclosure in forensic settings. Child Abuse & Neglect, 93, 291–304. 10.1016/jxhiabu.2018.11.020 [PubMed: 30579645]

Bond CF Jr, & DePaulo BM (2006). Accuracy of deception judgments. Personality and Social Psychology Review, 10, 214–234. 10.1207/s15327957pspr1003_2 [PubMed: 16859438]

Bottoms BL, Golding JM, Stevenson MC, Wiley TR, & Yozwiak JA (2007). A review of factors affecting jurors' decisions in child sexual abuse cases. In Toglia MP, Read JD, Ross DF, & Lindsay RCL (Eds.), The handbook of eyewitness psychology, Vol. 1. Memory for events (pp. 509–543). New York, NY: Erlbaum.

Chen LP, Murad MH, Paras ML, Colbenson KM, Sattler AL, Goranson EN , … Zirakzadeh A (2010). Sexual abuse and lifetime diagnosis of psychiatric: Disorders systematic review and meta-analysis. Mayo Clinic Proceedings, 85, 618–629. 10.4065/mcp.2009.0583 [PubMed: 20458101]

Coles J, Lee A, Taft A, Mazza D, & Loxton D (2015). Childhood sexual abuse and its association with adult physical and mental health. Journal of Interpersonal Violence, 30,1929–1944. 10.1177/0886260514555270 [PubMed: 25411233]

Cooper A, Quas JA, & Cleveland KC (2014). The emotional child witness: Effects on juror decision-making. Behavioral Sciences & the Law, 32, 813–828. 10.1002/bsl.2153 [PubMed: 25537438]

Cronch LE, Viljoen JL, & Hansen DJ (2006). Forensic interviewing in child sexual abuse cases: Current techniques and future directions. Aggression and Violent Behavior, 11, 195–207. https://doi.org/10.1016Zj.avb.2005.07.009

Edelstein RS, Luten TL, Ekman P, & Goodman GS (2006). Detecting lies in children and adults. Law and Human Behavior, 30(1), 1–10. 10.1007/s10979-006-9031-2 [PubMed: 16729205]

Gabora NJ, Spanos NP, & Joab A (1993). The effects of complainant age and expert psychological testimony in a simulated child sexual abuse trial. Law and Human Behavior, 17, 103–119. 10.1007/bf01044540

Gongola J, Quas JA, Clark SE, & Lyon TD (2020). Adult perceptions of children interviewed with the putative confession instruction. Applied Cognitive Psychology.

Gongola J, Scurich N, & Lyon TD (2019). Effects of the putative confession instruction on perceptions of children's true and false statements. Applied Cognitive Psychology, 33, 655–661. 10.1002/acp.3483 [PubMed: 33574640]

Gongola J, Scurich N, & Quas JA (2017). Detecting deception in children: A meta-analysis. Law and Human Behavior, 41, 44–54. 10.1037/lhb0000211 [PubMed: 27685642]

Gongola J, Williams S, & Lyon TD Under-informative responding is associated with children's concealment of a transgression.

Goodman GS, Ghetti S, Quas JA, Edelstein RS, Alexander KW, Redlich AD, … & Jones DP (2003). A prospective study of memory for child sexual abuse: New findings relevant to the repressed-memory controversy. Psychological Science, 14, 113–118. 10.1111/1467-9280.01428 [PubMed: 12661671]

Goodman GS, & Quas JA (1995). De l'utilité des poupées fidèles sur la plan anatomique lors de l'interrogatoire d'enfants. Psychologie Française, 40, 295–301.

Goodman GS, & Quas JA (2008). Repeated interviews and children's memory: It's more than just how many. Current Directions in Psychological Science, 17, 386–390. 10.1111/j.1467-8721.2008.00611.x

Goodman-Brown TB, Edelstein RS, Goodman GS, Jones DP, & Gordon DS (2003). Why children tell: A model of children's disclosure of sexual abuse. Child abuse & neglect, 27, 525–540. [PubMed: 12718961]

Hershkowitz I, Horowitz D, & Lamb ME (2005). Trends in children's disclosure of abuse in Israel: A national study. Child Abuse & Neglect, 29, 1203–1214. 10.1016/jxhiabu.2005.04.008 [PubMed: 16260036]

Hershkowitz I, Lamb ME, & Katz C (2014). Allegation rates in forensic child abuse investigations: Comparing the revised and standard NICHD protocols. Psychology, Public Policy, and Law, 20, 336. 10.1037/a0037391

Irish L, Kobayashi I, & Delahanty DL (2010). Long-term physical health consequences of childhood sexual abuse: A meta-analytic review. Journal of Pediatric Psychology, 35, 450–461. 10.1093/jpepsy/jsp118 [PubMed: 20022919]

Kail R (1991). Development of processing speed in childhood and adolescence. Advances in Child Development and Behavior, 23, 151–185. 10.1016/s0065-2407(08)60025-7 [PubMed: 1767721]

Kassin SM, Meissner CA, & Norwick RJ (2005). "I'd know a false confession if I saw one": A comparative study of college students and police investigators. Law and Human Behavior, 29, 211–227. 10.1007/s10979-005-2416-9 [PubMed: 15912725]

La Rooy D, Brubacher SP, Aromaki-Stratos A, Cyr M, Hershkowitz I, Korkman J, … Lamb ME (2015). The NICHD protocol: A review of an internationally-used evidence based tool for training child forensic interviewers. Journal of Criminological Research, Policy and Practice, 1, 76–89. 10.1108/jcrpp-01-2015-0001

Lamb ME, & Fauchier A (2001). The effects of question type on self-contradictions by children in the course of forensic interviews. Applied Cognitive Psychology, 15, 483–491. 10.1002/acp.72

Lamb ME, Hershkowitz I, Orbach Y, & Esplin PW (2011). Tell me what happened: Structured investigative interviews of child victims and witnesses. Chichester, UK: John Wiley & Sons.

Lamb ME, Orbach Y, Hershkowitz I, Horowitz D, & Abbott CB (2007). Does the type of prompt affect the accuracy of information provided by alleged victims of abuse in forensic interviews? Applied Cognitive Psychology, 21, 1117–1130. 10.1002/acp.1318

Leach C, Powell MB, Sharman SJ, & Anglim J (2017). The relationship between children's age and disclosures of sexual abuse during forensic interviews. Child Maltreatment, 22, 79–88. 10.1177/1077559516675723 [PubMed: 27784813]

Lyon TD, Malloy LC, Quas JA, & Talwar VA (2008). Coaching, truth induction, and young maltreated children's false allegations and false denials. Child Development, 79, 914–929. 10.1111/j.1467-8624.2008.01167.x [PubMed: 18717898]

Lyon TD, Wandrey L, Ahern E, Licht R, Sim MP, & Quas JA (2014). Eliciting maltreated and nonmaltreated children's transgression disclosures: Narrative practice rapport building and a putative confession. Child Development, 85, 1756–1769, 10.1111/cdev.12223 [PubMed: 24467688]

Malloy LC, Lyon TD, & Quas JA (2007). Filial dependency and recantation of child sexual abuse allegations. Journal of the American Academy of Child & Adolescent Psychiatry, 46, 162–170. 10.1097/01.chi.0000246067.77953.f7 [PubMed: 17242619]

McElvaney R (2015). Disclosure of child sexual abuse: Delays, non-disclosure and partial disclosure. What the research tells us and implications for practice. Child Abuse Review, 24, 159–169. 10.1002/car.2280

Newton P, Reddy V, & Bull R (2000). Children's everyday deception and performance on false-belief tasks. British Journal of Developmental Psychology, 18, 297–317. 10.1348/026151000165706

Nysse-Carris KL, Bottoms BL, & Salerno JM (2011). Experts' and novices' abilities to detect children's high-stakes lies of omission. Psychology, Public Policy, and Law, 17, 76–98. 10.1037/a0022136

Oppenheimer DM, Meyvis T, & Davidenko N (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. Journal of Experimental Social Psychology, 45, 867–872. 10.1016/jjesp.2009.03.009

Peterson C, Dowden C, & Tobin J (1999). Interviewing preschoolers: Comparisons of yes/no and wh-questions. Law and Human Behavior, 23, 539–555. 10.1023/a:1022396112719 [PubMed: 10487148]

Powell MB (2005). Improving the reliability of child witness testimony in court: The importance of focusing on questioning techniques. Current Issues in Criminal Justice, 17, 137–143. 10.1080/10345329.2005.12036342

Quas JA, Stolzenberg SN, & Lyon TD (2018). The effects of promising to tell the truth, the putative confession, and recall and recognition questions on maltreated and non-maltreated children's disclosure of a minor transgression. Journal of Experimental Child Psychology, 166, 266–279. 10.1016/jjecp.2017.08.014 [PubMed: 28950167]

Quas JA, Thompson WC, & Clarke-Stewart KA (2005). Do jurors "know" what isn't so about child witnesses? Law and Human Behavior, 29, 425–456. 10.1007/s10979-005-5523-8 [PubMed: 16133948]

Rush EB, Stolzenberg SN, Quas JA, & Lyon TD (2017). The effects of the putative confession and parent suggestion on children's disclosure of a minor transgression. Legal and Criminological Psychology, 22, 60–73. 10.1002/bsl.2153 [PubMed: 28286409]

Ruva CL, & Bryant JB (2004). The impact of age, speech style, and question form on perceptions of witness credibility and trial outcome. Journal of Applied Social Psychology, 34, 1919–1944. 10.1111/jM559-1816.2004.tb02593.x

Saykaly C, Crossman A, & Talwar V (2017). High cognitive load during cross examination: Does it improve detection of children's truths and lies? Psychiatry, Psychology and Law, 24, 278–291. 10.1080/13218719.2016.1197816

Saywitz K, Camparo LB, & Romanoff A (2010). Interviewing children in custody cases: Implications of research and policy for practice. Behavioral Sciences & the Law, 28, 542–562. 10.1002/bsl.945 [PubMed: 20623723]

Saywitz KJ, Goodman GS, Nicholas E, & Moan SF (1991). Children's memories of a physical examination involving genital touch: Implications for reports of child sexual abuse. Journal of Consulting and Clinical Psychology, 59, 682–691. 10.1037/0022006x.59.5.682 [PubMed: 1955603]

Saywitz KJ, Lyon TD, & Goodman GS (2017). When interviewing children: A review and update. In The APSAC handbook on child maltreatment (4th ed., pp. 310–329). Los Angeles, CA: Sage.

Sivagurunathan M, Orchard T, MacDermid JC, & Evans M (2019). Barriers and facilitators affecting self-disclosure among male survivors of child sexual abuse: The service providers' perspective. Child Abuse & Neglect, 88, 455–465. 10.1016/jxhiabu.2018.08.015 [PubMed: 30219431]

Scurich N, & John RS (2019). The dark figure of sexual recidivism. Behavioral Sciences & the Law, 37, 158–175, 10.1002/bsl.2400 [PubMed: 30900348]

Stolzenberg SN, McWilliams K, & Lyon TD (2017). The effects of the hypothetical putative confession and negatively valenced yes/no questions on maltreated and nonmaltreated children's disclosure of a minor transgression. Child Maltreatment, 22, 167–173. 10.1177/1077559516673734 [PubMed: 27753611]

Street CN, & Richardson DC (2015). Lies, damn lies, and expectations: How base rates inform lie-truth judgments. Applied Cognitive Psychology, 29, 149–155. 10.1002/acp.3085

Swann WB (1984). Quest for accuracy in person perception: A matter of pragmatics. Psychological Review, 91, 457–477, 10.1037/0033-295x.9L4.457 [PubMed: 6505116]

Talwar V, Crossman A, Williams S, & Muir S (2011). Adult detection of children's selfish and polite lies: Experience matters. Journal of Applied Social Psychology, 41, 2837–2857. 10.1111/j.1559-1816.2011.00861.x

Talwar V, Gordon HM, & Lee K (2007a). Lying in the elementary school years: Verbal deception and its relation to second-order belief understanding. Developmental Psychology, 43, 804–810. 10.1037/0012-1649.43.3.804 [PubMed: 17484589]

Talwar V, & Lee K (2002). Emergence of white-lie telling in children between 3 and 7 years of age. Merrill-Palmer Quarterly, 48, 160–181. 10.1353/mpq.2002.0009

Talwar V, Lee K, Bala N, & Lindsay RCL (2006). Adults' judgments of children's coached reports. Law and Human Behavior, 30, 561–570. 10.1007/s10979-006-9038-8 [PubMed: 16941237]

Talwar V, Murphy SM, & Lee K (2007b). White lie-telling in children for politeness purposes. International Journal of Behavioral Development, 31, 1–11. 10.1177/0165025406073530. [PubMed: 18997880]

Talwar V, Renaud SJ, & Conway L (2015). Detecting children's lies: Are parents accurate judges of their own children's lies? Journal of Moral Education, 44, 81–96. 10.1080/03057240.2014.1002459

Tye MC, Amato SL, Honts CR, Devitt MK, & Peters D (1999). The willingness of children to lie and the assessment of credibility in an ecologically relevant laboratory setting. Applied Developmental Science, 3, 92–109. 10.1207/s1532480xads0302_4

Vrij A, Fisher RP, & Blank H (2017). A cognitive approach to lie detection: A meta analysis. Legal and Criminological Psychology, 22, 1–21. 10.1111/lcrp.12088

Vrij A, Mann SA, Fisher RP, Leal S, Milne R, & Bull R (2008). Increasing cognitive load to facilitate lie detection: The benefit of recalling an event in reverse order. Law and Human Behavior, 32, 253–265. 10.1007/s10979-007-9103-y [PubMed: 17694424]

Waterman AH, & Blades M (2011). Helping children correctly say "I don't know" to unanswerable questions. Journal of Experimental Psychology: Applied, 17, 396–405. 10.1037/a0026150. [PubMed: 22059954]

Waterman AH, Blades M, & Spencer C (2004). Indicating when you do not know the answer: The effect of question format and interviewer knowledge on children's 'don't know' responses. British Journal of Developmental Psychology, 22, 335–348. 10.1348/0261510041552710.

Westcott H, Davies G, Graham M, & Clifford B (1991). Adults' perceptions of children's videotaped truthful and deceptive statements. Children and Society, 5, 123–135. 10.1111/j.1099-0860.1991.tb00378.x

Widom CS, Czaja SJ, & Dutton MA (2008). Childhood victimization and lifetime revictimization. Child Abuse & Neglect, 32, 785–796. 10.1016/j.chiabu.2007.12.006 [PubMed: 18760474]

Wyman J, Foster I, Crossman A, Colwell K, & Talwar V (2019). The efficacy of free recall, cognitive load, and closed-ended questions when children are asked to falsely testify about a crime. Applied Cognitive Psychology, 33(4), 544–560.

TABLE 1

Mean (SD) accuracy, discriminability, and credibility perceptions as a function of denial type and question type in Study 1 and Study 2

| | Question type | | |
	Recall	Recognition	Overall
Study 1			
Accuracy	0.51 (0.20)	0.54 (0.21)	0.53 (0.20)
True denial	0.60 (0.38)	0.64 (0.37)	0.62 (0.37)
False denial	0.43 (0.36)	0.43 (0.37)	0.43 (0.37)
Discriminability	0.03 (0.39)	0.08 (0.42)	0.05 (0.40)
Perceptions	4.21 (0.86)	4.08 (0.90)	4.14 (0.88)
True denial	4.25 (0.92)	4.04 (0.97)	4.13 (0.95)
False denial	4.18 (0.95)	4.12 (0.96)	4.15 (0.96)
Study 2			
Accuracy	0.49 (0.25)	0.51 (0.24)	0.50 (0.24)
True denial	0.47 (0.34)	0.46 (0.36)	0.46 (0.35)
False denial	0.52 (0.35)	0.56 (0.35)	0.54 (0.35)
Discriminability	−0.03 (1.40)	0.05 (1.38)	0.01 (1.39)
Perceptions	4.03 (0.81)	3.86 (0.75)	3.94 (0.79)
True denial	4.05 (0.96)	3.78 (0.89)	3.90 (0.93)
False denial	4.02 (0.89)	3.94 (0.82)	3.98 (0.85)