Title
Detecting Deception in Children: A Meta-Analysis

Author
Gongola, Jennifer

Publication Date
2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA,
IRVINE

Detecting Deception in Children: A Meta-Analysis

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF ARTS

in Social Ecology

by

Jennifer Cynthia Gongola

Thesis Committee:
Assistant Professor Nicholas Scurich, Chair
Professor Jodi Quas
Professor JoAnn Prause

2016

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I wish to express my sincere gratitude to my committee chair, Assistant Professor Nicholas Scurich, for his thoughtful and expert guidance throughout this process. Thank you for being an inspiring, supportive, and overall exceptional mentor. I also thank my committee members, Professor Jodi Quas, for her invaluable input on child development, as well as Professor JoAnn Prause, for making the time and effort to educate me on the finer points of the meta-analytic approach. I offer my sincere appreciation for the learning opportunities provided by my committee.

ABSTRACT OF THE THESIS

Detecting Deception in Children: A Meta-Analysis

By

Jennifer Cynthia Gongola

Master of Arts in Social Ecology

University of California, Irvine, 2016

Assistant Professor Nicholas Scurich, Chair

Although research consistently reveals that children as young as three can use deception and will take steps to obscure truth, research concerning how well others detect children's deceptive efforts remains unclear. Yet, adults regularly assess whether children are telling the truth in a variety of contexts, including school, home, and legal settings, particularly in investigations of maltreatment. We conducted a meta-analysis to synthesize extant research concerning adults' ability to detect deceptive statements produced by children. We included 45 experiments involving 7,893 adult judges and 1,858 children. Overall, adults could accurately discriminate truths/lies at an average rate of 54.3%, which is significantly above chance levels. The average rate at which true statements were correctly classified as honest was higher (59.7%), while the rate at which false statements were classified as dishonest was at chance (49.4%). A small positive correlation emerged between judgment confidence and judgment accuracy. Professionals (e.g., social workers; police officers) slightly outperformed laypersons (e.g., college undergraduates). Finally, exploratory analyses revealed that the child's age did not affect the rate at which adults could discriminate truths/lies but it was trending in the expected developmental direction (i.e., adults were most accurate when judging younger children).

INTRODUCTION

Deception is an intentional act designed to impart a belief that the communicator knows is not true (Vrij, 2008). It is not categorically considered to be a negative act, and people lie for a range of prosocial reasons, including to minimize conflict in social interactions, to avoid hurt feelings, or to self-protect from negative evaluations by others (DePaulo, Kashy, Kirkendol, Wyer, & Epstein, 1996). Yet, in other situations, lying is done with the explicit intent to hide a transgression, hurt another, or avoid punishment or negative consequences. In the latter situations, determining truth can be essential, as the lie could have serious negative social, ethical, and legal implications. When that lie was reported—or believed to have been reported—by a child, the stakes associated with evaluating the lie can be incredibly high, as in situations in which abuse has been alleged and a legal case may ensue.

Although an impressive number of studies has examined whether people can in fact detect deception in children, results seem to vary widely, making it difficult to draw definitive conclusions regarding adults' deception detection abilities. One particularly effective way of integrating results of former studies that can enable clearer identification of important trends, as well as the magnitude of those trends, involves combining independent results via a meta-analysis. We took this approach in the present study in order to summarize extant results concerning adults' ability to detect children's deceptive statements. We further tested for potential moderators that could affect adult' detection abilities. We identified moderators based on literature concerning adults' detection abilities and children's development.

Adults tend to believe they are proficient at determining the truth, and a large number of studies has been conducted examining adults' perceptions of their ability to detect deception, what types of indicators adults use to make determinations about deception, and the actual accuracy of adults' judgments (for reviews see DePaulo et al., 2003; Vrij, Akehurst, & Knight, 2006; Vrij, Granhag, & Porter, 2010). Findings reveal first that adults are often confident in their deception detection abilities, and second that behaviors such as gaze aversion (avoiding eye contact), fidgeting, nervousness, incoherent responses, and facial expressions are all indicative of someone lying rather than telling the truth (Global Deception Research Team, 2006). And third, despite adults' confidence and perceptions of behavioral markers of deception, adults' actual accuracy when rendering judgments about deception is largely unimpressive, with accuracy rates hovering only slightly above chance (e.g., 55%), as reflected, for example, in several meta-analyses of adults' detection accuracy (e.g., Bond & DePaulo, 2006; Feeley, & Young, 1998; Kraut, 1980; Vrij, 2000). People are somewhat better at classifying honest statements as truthful (61%) than dishonest statements as lies (47%), a so-called "truth bias" (Bond & DePaulo, 2006; see also Vrij, 2000). However, more confident adults are not necessarily more accurate (DePaulo et al., 1997), and professionals (e.g., police officers or clinical psychologists) are often no more accurate than lay persons (DePaulo, Charlton, Cooper, Lindsay, & Muhlenbruck, 1997; Garrido, Masip, & Herrero, 2004).

Adults evaluate deception not only in other adults but, as mentioned, at times in children, whose lies carry potential significance for themselves, their family, and even community. A sizeable body of research has evaluated how well children can actually maintain lies. Evidence indicates, quite consistently, that even relatively young children can maintain at least some types of lies (Lyon, Malloy, Quas, & Talwar, 2008; Newton, Reddy, & Bull, 2000; Talwar & Lee, 2002), particularly those that involve denying an event rather than alleging a falsehood. At the same time, however, leakage, defined as verbal or nonverbal indicators of deception, is quite common at young ages (Feldman & White, 1980; Talwar & Lee, 2002). With age, children's ability to maintain a cogent lie, the range of types of lies (e.g., positive or prosocial and negative or transgression denials), and children's ability to control for potential leakage all increase (e.g.., Feldman & White, 1980; Feldman, Jenkins, & Popoola, 1979; Gadea, Aliño, Espert, & Salvador, 2015; Lewis, Stanger, & Sullivan, 1989; Talwar & Crossman, 2011; Talwar & Lee, 2002). In addition, children gradually become better able to elaborate on falsehoods with supporting statements and behaviors, thereby making their lie perhaps more plausible (Gordon, Talwar, & Lee, 2005; Talwar, Murphy, & Lee, 2007).

Given that children can in fact maintain a lie, it is not surprising that studies have investigated adults' ability to detect those lies. As mentioned, though, findings are mixed. Some studies suggest that adults are fairly adept at judging honesty versus deception in children, whereas other studies reveal near chance performance or response biases. In the current meta-analysis, we sought to synthesize the extant research and identify consistent trends across studies concerning adults' deception detection accuracy with children. We further evaluated whether several potentially key moderators affected adults' detection abilities, including the children's age, characteristics of the adult decision makers, such as their confidence and type of training, as well as certain study characteristics, for example, the type of lie and type of deception paradigm employed. Our main research questions were as follows: (a) *How well can adults detect children's deceptive statements overall and when children's statements are honest versus dishonest?* We were especially interested in whether and by how much adults' judgment accuracy exceeded chance levels. We were also interested in whether children's age would affect adults' detection abilities, a likely possibility given that, young children have difficulty masking leakage when attempting to engage in deception (Talwar & Lee, 2002), and adults may pick up on this leakage when rendering judgments. (b) *Is there a significant relation between judgment* *accuracy and confidence*? Although the adult deception detection literature tends to find no significant relations (DePaulo et al., 1997), the two may be related with children, largely as a result of adults being able to pick up on children's difficulty masking behavioral indicators of deception, which should increase the adults' accuracy and confidence concurrently. (c) *Do professionals (e.g., social workers, clinical psychologists, classroom teachers) and laypersons differ in how well they can detect deception in children?* In adults, group differences rarely emerge in deception detection accuracy. In children, some professionals have extensive training in cognitive development and children's competencies. These professionals may, therefore, be especially sensitive to leakage, or rather verbal and nonverbal cues indicative of deception (Crossman & Lewis, 2006; Talwar, Crossman, Williams, & Muir, 2011); although, it has been suggested that professionals might even display a bias towards believing children, given their advocacy role (Nysse-Carris, Bottoms, & Salerno, 2011).

A fourth question, which concerned whether the type of paradigm used to elicit deception affects adults' detection accuracy, was also pursued. However, analyses necessarily included smaller subsets of studies, making these mainly exploratory in nature. Deception procedures vary in the cognitive load they place on children (e.g., Case, 1992). This makes some deception tasks, including for example those requiring children to generate an alternative explanation rather than those that allow a simple confirmation or denial of behavior, potentially difficult and demanding cognitively, leaving fewer resources available for children to generate and then maintain a cogent lie. The increased cognitive load, in turn, is likely to affect younger more than older children, given the former's more limited working memory and cognitive capacity in general (Case, 1992; Chandler, Fritz, & Hala, 1989; Talwar & Lee, 2008), leading to potential inconsistencies in their lies (Quas, Davis, Goodman, & Myers, 2007) and leakage (Feldman, Jenkins, & Popoola, 1979;

Talwar & Lee, 2002). Accordingly, adults' deception detection accuracy may be highest when young children are asked to generate their own lies, especially those that also involved the creation of a false allegation.

We thus made several heuristic distinctions and grouped the studies into three types of paradigms: how the lie was generated, whether or not the lie was associated with a transgression, and what type of scenario was employed (see Table 1). First, children could be enticed to generate their own lie or be coached, either explicitly or implicitly, by an adult. The former includes authentic lies generated by the child in the moment, and the studies provide insight into how children behave when lying on their own accord. The later involves lying in response to a request from someone else, where adults either explicitly coach children regarding the content of their false statements (e.g., ask the child to misstate specific information about a recent event they experienced) or implicitly coach children by asking them to "trick" an interviewer but not providing explicit instructions on what to say. Second, the child could either lie about their own transgression, about another person's misdeed that they witnessed, or they could lie about an innocuous situation that did not involve any wrongdoing. The third paradigm distinction involves the type of scenario design used by the researchers. We will elaborate on each of the four types of scenarios in terms of the general sequence of events as well as how they fit within the first two types of paradigms.

The first and one of the most well-studied scenario is the classic temptation resistance paradigm (e.g., Lewis, Stranger, & Sullivan, 1989). Children are left alone in a room with an exciting toy placed behind them but are explicitly told not to turn around and not to look at the toy. Upon returning, the researcher asks children whether or not they turned around and looked at the toy, a question to which a majority of children lie (Crossman & Lewis, 2006; Talwar &

Lee, 2002; Talwar, Renaud, & Conway, 2015). Children's responses are typically recorded on video. More recent studies have added questions asking children to guess the name of the forbidden toy (e.g., Buzz Lightyear, Leach et al., 2009). Adult decision-makers then view several of these brief videotaped interviews in succession and indicate whether the child is telling the truth or not. This scenario always produces authentic lies that the child generates themselves in response to a minor transgression that they committed.

The second scenario we call the common events paradigm, in which children are asked to recall in detail a prior event (e.g., visiting a museum, magic show, bitten by a dog). These events are common in that they contain content about normally experienced activities rather than involving some type of transgression. Some children had experienced the event and hence are recalling a truthful experience. Others, however, have not but have been explicitly coached by an experimenter to lie and convince the interviewer that they had experienced that particular event (Talwar, Lee, Bala, & Lindsay, 2006). Also, children who did not experience an event may be implicitly coached to lie with a prompt asking the child to create a plausible, fictional event and convince the interviewer that it really happened to them. Adult participants view videotapes of the children's narratives and judge whether or not each child is telling the truth.

A third scenario we call the games paradigm. While the game itself varies, during the interaction, the adult may or may not have engage in specific behaviors with children, such as touching the child (e.g., on the stomach or nose). Either the confederate or another adult then explicitly coaches the children to tell the truth or to lie in a subsequent interview about that behavior. Typically, the studies that employ this scenario frame the behaviors as a transgression committed by the adult in the interview. In some of these studies, the coaching instructions included both false allegations when the behavior did not occur and false denials when the behavior did, thus allowing other adults to render judgments about deception for false negative and false positives within the same type of event.

A final scenario involves lab studies in which the child's parent or a researcher acts as a confederate and commits a wrongdoing, either intentionally (e.g., stealing a book) or accidentally (e.g., breaking a toy, Nysse-Carris, Bottoms, & Salerno, 2011; Tye, Amato, Honts, Devitt, & Peters, 1999). The adult then may explicitly coach the child to lie or tell the truth about what happened, sometimes offering an appealing toy as a reward for compliance, with some studies going further by the adult coaching children to blame an innocent third party. Other children are not given any coaching and, instead, will lie of their own volition. Videotapes of children denying the transgression to cover for the experimenter/parent, blaming someone else, or confessing what they witnessed are then shown to adult judges who render decisions about what they perceive to have happened.

In sum, the research on detecting children's deceptive statements appears mixed. In addition to assessing adults' average accuracy rates, we plan to examine any emerging trends from several theoretically important variables, such as the influence of adult's professional status, children's age, and study paradigm characteristics on accuracy.

METHOD

Study Selection

To be included in the analyses, studies must have had adult participants (herein "receivers") making judgments about the veracity of children's true and false statements (herein "senders") without assistance from detection aids (e.g., criteria-based content analysis [CBCA] or polygraphy). Studies were excluded if they were manipulating facial expressions only (e.g., no volume, Boerner, Chambers, Craig, Riddell, & Parker, 2013; Feldman, Jenkins, & Popoola, 1979; Morency & Krauss, 1982; Swerts, 2012; Swerts, van Doorenmalen, & Verhoofstad, 2013). These experiments were more cognitive in nature, often sacrificing ecological validity to examine more nuanced hypotheses regarding children's abilities to suppress affect, and thus the investigations are substantively different from the other child deception detection studies. Studies were not included if the children's false statements were not intentionally deceptive. Some studies, for example, induce false statements by misleading children with suggestive techniques (Block et al., 2012; Ceci, Huffman, Smith, & Loftus, 1994; Goodman, Batterman-Faunce, Schaaf, & Kenney, 2002; Leichtman & Ceci, 1995; Newcombe & Bransgrove, 2007) or the false statements may have occurred naturally through cognitive memory errors (e.g., encoding, storage and retrieval; Ball & O'Callaghan, 2001) and, while their statements were false, their intentions were not to purposely deceive the interviewer. Other inclusion criteria were as follows: We operationalized children as age 17 or under. Thus, we excluded studies in which receivers judged adult senders. The studies needed to provide a numeric metric for the rate at which receivers accurately (or inaccurately) detected sender's statements. Often times, this was expressed as an overall accuracy rate (i.e., the number of correct classifications divided by the total number of classifications made), though some studies decomposed the rate by the different types of classifications (e.g., false positive rate or false negative rate).

From each study, the following variables were coded (when possible): (a) number of (adult) receivers, (b) number of (child) senders, (c) receiver's professional status, (d) lie type (i.e., false positive or false negative), (e) study scenario, (f) method for generating false statements, and (g) transgression type (i.e., child's transgression, other's transgression, or no transgression). Accuracy rate (decomposed or not) and the corresponding standard deviations were noted, as well as any relevant moderator comparisons reported in the study, when provided.

A variety of methods was used to find published and unpublished studies of adult deception detection of child senders. We began with a computer-based search using PsycINFO, ProQuest, EBSCO, WorldCat, PsycLit and Google Scholar search engines for studies published prior to September 2015 with key words *accuracy, judgment, detect, child, deception, false statements, lie or truth*, with several variants and conjunctions of these terms. Once relevant studies were identified, their reference sections were examined for other relevant studies; the reference section of numerous non-empirical articles was also examined for potentially relevant studies. Additionally, an email request was sent to authors of published research requesting the provision of any unpublished studies, necessary statistical clarification, as well as recommendations for other researchers that could be contacted regarding possible unpublished research studies.

Table 1. *Descriptive Statistics of Included Studies*

Study	Child Senders			Adult Receivers			Moderators	
	Total	Males	Females	Total	Males	Females	Study Scenario	Generation of False Statements
†Talwar, Renaud & Conway (2015)	250	129	118	250	0	250	temptation resistance	self
⌂Cassidy (2015)	5	3	2	178	50	128	common events	explicit coaching
*Warren, Bakhtiar, Mulrooney, Raynor, Dodd, & Peterson (2015)	96	48	48	1074	239	733	common events	explicit coaching
Gadea, Aliño, Espert & Salvador (2015)	4	0	4	104	29	75	common events	implicit coaching

*†Saykaly et al. (2013) EXP1	36	13	23	48	13	35	common events	implicit coaching
*†Saykaly et al. (2013) EXP2	36	13	23	72	18	54	common events	implicit coaching
*†Warren, Dodd, Raynor, & Peterson (2012)	32	16	16	514	149	362	common events	-
*Shao & Ceci (2011)	24	11	13	129	16	113	games	self
*†⌂Nysse-Carris, Bottoms & Salerno (2011)	12	0	12	72	14	58	witness transgression	self
†⌂Talwar, Crossman, Williams, & Muir (2011)	16	8	8	150	34	116	temptation resistance & sour juice	explicit coaching
Landström & Granhag (2010)	108	65	43	240	65	175	common events	explicit coaching
*⌂Brunet (2009)	16	-	-	89	15	74	witness transgression	explicit coaching
Talwar, Crossman, Gulmi, Renaud & Williams (2009)	16	-	-	156	60	96	temptation resistance	self
Leach et al. (2009) EXP1	80	-	-	51	14	37	temptation resistance	self
†Leach et al. (2009) EXP3	48	-	-	197	33	164	common events	implicit coaching
Leach et al. (2009) EXP5	80	-	-	15	2	13	temptation resistance	self
Landström & Granhag (2008)	14	8	6	256	84	127	common events	explicit coaching
(Strömwall & Granhag, 2007)	44	20	24	88	35	53	common events	implicit coaching

*Strömwall, Granhag & Landström (2007)	30	11	19	60	20	40	common events	implicit coaching
Sumner-Armstrong & Newcombe (2007)	2	0	2	125	36	89	witness transgression	explicit coaching
Landström, Granhag & Hartwig (2007)	14	5	9	136	42	94	common events	explicit coaching
*Shao (2007)	9	1	8	87	44	43	witness transgression	-
Goodman et al. (2006)	3	3	6	370	170	200	games	explicit coaching
⌂Crossman & Lewis (2006)	58	29	29	64	19	19	temptation resistance	self
†Talwar, Lee, Bala & Lindsay (2006)	48	24	24	193	96	97	common events	explicit coaching
Edelstein, Luten, Ekman & Goodman (2006)	20	10	10	144	60	84	games	explicit coaching
†⌂Vrij, Akehurst, Brown & Mann (2006)	142	-	-	150	50	100	games	explicit coaching
Stromwall & Granhag (2005)	44	20	24	88	30	58	common events	explicit coaching
*†⌂Leach, Talwar, Lee, Bala & Lindsay (2004) EXP1	80	-	-	105	58	47	temptation resistance	self
*†⌂Leach, Talwar, Lee, Bala & Lindsay (2004) EXP2	30	-	-	103	53	50	temptation resistance	self

*†⌂Leach, Talwar, Lee, Bala, & Lindsay (2004) EXP3	39	-	-	126	72	54	temptation resistance	self
*†Leach, Talwar, Lee, Bala & Lindsay, (2004) EXP4	81	-	-	100	33	60	temptation resistance	self
⌂Bala, Ramakrishnan, Lindsay, & Lee (2004)	3	0	3	147	-	-	common events	explicit coaching
Talwar & Lee (2002)	101	40	61	156	37	119	temptation resistance	self
Talwar & Lee (2002)b	98	38	60	92	11	81	common events	self
*Orcutt, Goodman, Tobey, Batterman-Faunce & Thomas (2001)	70	33	37	987	497	490	games	-
*Tye, Amato, Honts, Devitt & Peters (1999) EXP1	28	-	-	115	37	78	witness transgression	implicit coaching
*Tye, Amato, Honts, Devitt & Peters (1999) EXP2	28	-	-	129	-	-	witness transgression	self
Jackson (1996)	200	-	-	200	-	-	common events	explicit coaching
⌂Chahal & Cassidy (1994)	4	2	2	60	6	54	common events	explicit coaching
†Vrij & Van Wijngaarden (1994)	82	-	-	82	-	-	sour juice	-

*†⌂Westcott, Davies & Clifford (1991)	32	16	16	32	10	22	common events	explicit coaching
Lewis, Stanger & Sullivan (1989)	33	15	18	60	-	-	temptation resistance	self
Lawrence et al. (in prep)	192	-	-	192	-	-	-	-
Johnson, Hobbs, Chae, Goodman, Shestowsky & Block (in prep)	107	-	-	24	-	-	-	self

Note. Not all child sender samples are independent, some studies share the same stimuli (child senders).

* Confidence-accuracy correlation

† Child age comparison

⌂ Professional accuracy

This process yielded 45 eligible experiments, of which 40 were published and 5 were unpublished (see Table 1). The earliest was dated 1989, and half were from 2007 or later. These studies included a total of 7,893 adult receivers and 1,858 child senders whose ages ranged from 3 to 15. Twelve experiments used some type of "professional receiver," which could be a classroom teacher, social worker, police officer, customs officer, clinician, researcher/psychologist, early education specialist, court judge, prosecutor, or other justice system professional. The common event paradigm was used in 19 studies; 12 used a temptation resistance paradigm, 5 used a game paradigm, and 6 had senders witness a transgression. The majority of studies examined accuracy at detecting false positives (28 experiments) rather than false negatives (13 experiments), while 4 used both types of lies.

We used Comprehensive Meta-Analysis (CMA) software to combine effect sizes from different data types and draw statistical conclusions. For each study, we report the unweighted mean accuracy rate and Cohen's *d*, because of its commonness in the social sciences and its ability to explain and convert between multiple types of effects (Cohen, 1992). When applicable, correlation coefficients (r) are reported as well. Random effects models were used to enable inferences to a broader population than the population from the studies that have been conducted (Schmidt, Oh, & Hayes, 2009; Hoyle, 1999). The Q statistic was used to evaluate differences between effect sizes with $\alpha = .05$.

RESULTS

Overall Accuracy

In 43 experiments, receivers ($n = 7,677$) classified messages as lies or truths and reported a mean percentage correct. Across these, the unweighted mean percentage correct was 54.34% ($SD = 8.2\%$), with a range of 32% to 68% and a median percentage of 55%. Comparisons of the observed accuracy rate to chance (45 experiments, $n = 7,893$) revealed levels in performance at detecting true and false statements greater than chance, Cohen's $d = 0.242$, 95% CI [0.119, 0.365]. Individual Cohen's *d* effect sizes ranged from -0.884 to 1.941 (see Figure 1). It should be noted that 22 of the 45 experiments on their own showed a statistically significant effect for accuracy above chance levels; of these, 6 of the studies examined the rate of inaccuracy and 16 studies examined the rate of accuracy. Given significant heterogeneity in effect sizes ($Q[44] = 616.45$, $p < .001$), we tested moderators that could potentially account for between-study differences.

Accuracy detecting false statements. Thirty-six studies provided percentage correct data in terms of classifying a false statement as dishonest. The unweighted mean percentage correct for lie classifications was 49.36% ($SD = 11.12\%$) with a range of 26% to 70% and median of 49.55%. Analyses directly comparing the accuracy of classifying false statements as dishonest to chance (41 effect sizes) was nonsignificant, Cohen's *d* = -0.094, 95% CI [-0.201, 0.014], indicating that, across all studies, receivers performed at chance when classifying false statements as dishonest. Individual Cohen's *d* effect sizes ranged from -1.432 to 0.547 (see top panel of Figure 1). This effect was statistically significant in 18 of the 41 individual effects. The unweighted mean percentage correct was calculated for false positives (*M* = 49.9%, *SD* = 11.85%) and false negatives (*M* = 47.17%, *SD* = 5.326%). Subgroup analyses were conducted on lie type (false positive or false negative) and the means did not significantly differ, $Q(1) = 0.198$, $p = .656$.

Accuracy detecting truthful statements. Thirty-six studies also provided percentages reflecting the adult receivers' ability to classify a true statement as an honest statement. The unweighted mean percentage correct was 59.65% (*SD* = 13.97%), with a range of 29% to 95% and median of 61%. Analyses directly comparing accuracy of classifying true statements as honest to chance (41 effect sizes) detected a statistically significant positive effect, Cohen's *d* = 0.338, 95% CI [0.157, 0.519] (see middle panel of Figure 1). A total of 26 of the 41 effect sizes on their own showed a significant difference from chance level at $p < .05$; Individual Cohen's *d* effect sizes ranged from -1.167 to 3.0, however the largest value was an outlier (Figure 1), specifically from the only study in which mothers judged their own children's statements. All of these were true positives, and just about every mother correctly believed their child's statement (Talwar, Renaud, & Conway, 2015). The unweighted mean percentage correct was calculated for true positives (*M* = 60.1%, *SD* = 13.54%) and true negatives (*M* = 53.43%, *SD* = 19.97%). Subgroup analyses conducted on truth type (true positive or true negative) revealed that these means did not significantly differ, $Q(1) = 2.362$, $p = .124$.

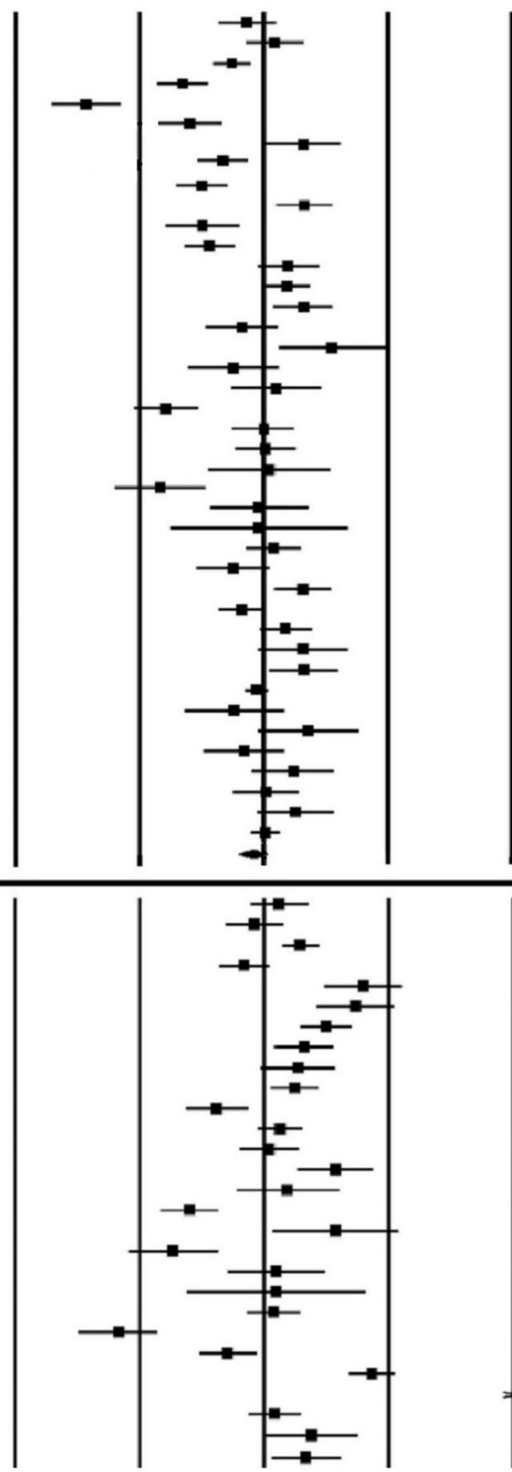
Figure 1. *Forest plot of Cohen's d individual and average effect sizes for false statements, true statement, and overall.*

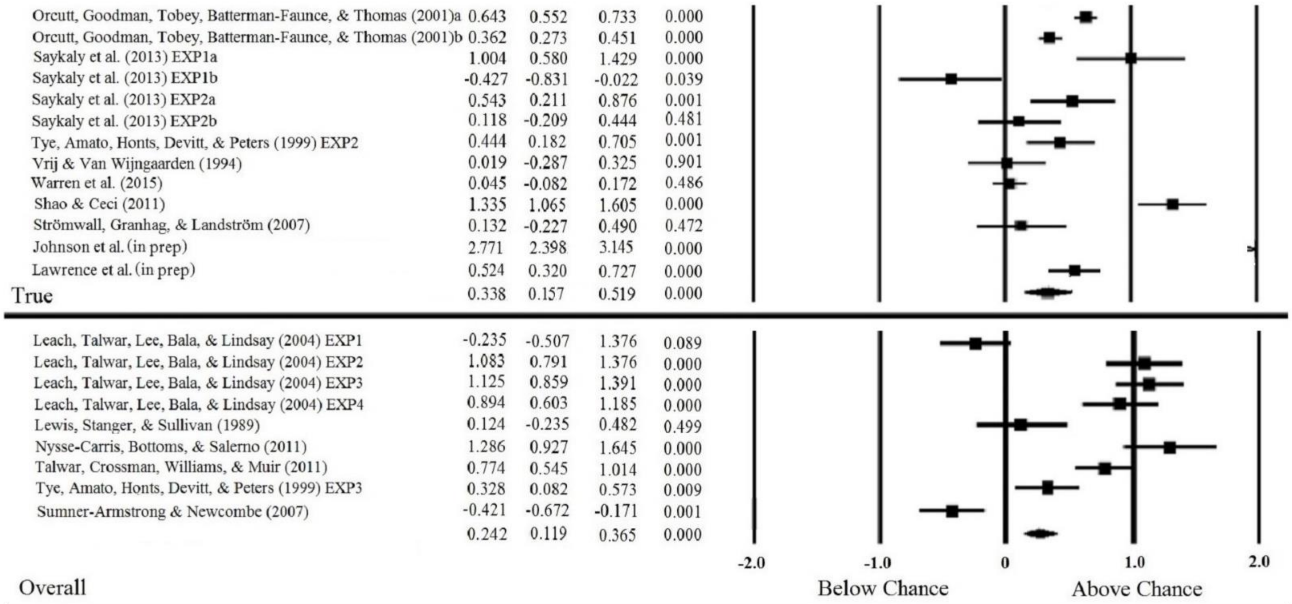
Accuracy

Study name	Std. dif in means	Lower limit	Upper limit	p-Value
Bala, Ramakrishnan, Lindsay, & Lee (2004)	-0.132	-0.361	0.097	0.259
Edelstein, Luten, Ekman, & Goodman (2006)	0.087	-0.144	0.318	0.461
Goodman et al. (2006)	-0.259	-0.404	-0.114	0.000
Leach et al. (2009) EXP3	-0.656	-0.859	-0.454	0.000
Shao & Ceci (2011)	-1.432	-1.705	-1.158	0.000
Johnson et al. (in prep)	-0.573	-0.846	-0.299	0.000
Shao (2007)	0.317	0.018	0.616	0.038
Lawrence et al. (in prep)	-0.333	-0.535	0.132	0.001
Talwar, Lee, Bala, & Lindsay (2006)	-0.500	-0.703	-0.298	0.000
Vrij, Akehurst, Brown, & Mann (2006)	0.324	0.097	0.552	0.005
Brunet (2009)	-0.495	-0.793	-0.196	0.001
Jackson (1996)	-0.440	-0.638	-0.241	0.000
Landström & Granhag (2008)	0.200	-0.045	0.446	0.110
Landström & Granhag (2010)	0.188	0.009	0.367	0.040
Landström, Granhag, & Hartwig (2007)	0.139	0.080	0.558	0.009
Strömwall & Granhag (2005)	-0.178	-0.474	0.118	0.239
Strömwall & Granhag (2007)	0.547	0.122	0.973	0.012
Strömwall, Granhag, & Landström (2007)a	-0.246	-0.605	0.114	0.180
Strömwall, Granhag, & Landström (2007)b	0.101	-0.257	0.459	0.582
Warren, Dodd, Raynor, & Peterson (2012)	-0.788	-1.045	-0.530	0.000
Warren, Dodd, Raynor, & Peterson (2012)a	0.000	-0.235	0.253	1.000
Warren, Dodd, Raynor, & Peterson (2012)b	0.011	-0.235	0.257	0.930
Westcott, Davies, & Clifford (1991)	0.044	-0.446	0.534	0.861
Crossman & Lewis (2006)	-0.834	-1.196	-0.473	0.000
Leach et al. (2009) EXP1	-0.038	-0.427	0.350	0.846
Leach et al. (2009) EXP5	-0.038	-0.754	0.677	0.916
Talwar & Lee (2002)	0.076	-0.146	0.298	0.500
Talwar & Lee (2002)b	-0.250	-0.540	0.040	0.091
Talwar, Crossman, Gulmi, Renaud, & Williams (2009)	0.135	0.091	0.538	0.006
Talwar, Renaud, & Conway (2015)	-0.184	-0.359	-0.008	0.040
Cassidy (2015)	0.175	-0.033	0.383	0.100
Chahal & Cassidy (1994)	0.314	-0.046	0.674	0.087
Gadea, Aliño, Espert, & Salvador (2015)	0.318	0.045	0.592	0.023
Orcutt, Goodman, Tobey, Batterman-Faunce, & Thomas (2001)	-0.060	-0.148	0.028	0.183
Saykaly et al. (2013) EXP1a	-0.237	-0.638	0.165	0.248
Saykaly et al. (2013) EXP1b	0.361	-0.042	0.765	0.079
Saykaly et al. (2013) EXP2a	-0.157	-0.484	0.170	0.347
Saykaly et al. (2013) EXP2b	0.237	-0.091	0.564	0.157
Tye, Amato, Honts, Devitt, & Peters (1999) EXP2	0.019	-0.239	0.278	0.883
Vrij & Van Wijngaarden (1994)	0.256	-0.051	0.563	0.103
Warren et al. (2015)	0.012	-0.102	0.125	0.840
False	-0.094	-0.201	0.014	0.087

Study name	Std. dif in means	Lower limit	Upper limit	p-Value
Bala, Ramakrishnan, Lindsay, & Lee (2004)	0.124	-0.104	0.353	0.287
Edelstein, Luten, Ekman, & Goodman (2006)	-0.077	-0.308	0.154	0.514
Goodman et al. (2006)	0.300	0.155	0.445	0.000
Leach et al. (2009) EXP3	-0.156	-0.354	0.042	0.122
Shao (2007)a	0.799	0.490	1.108	0.000
Shao (2007)b	0.732	0.425	1.039	0.000
Talwar, Lee, Bala, & Lindsay (2006)	0.496	0.293	0.698	0.000
Vrij, Akehurst, Brown, & Mann (2006)	0.324	0.097	0.552	0.005
Brunet (2009)	0.277	-0.018	0.537	0.066
Jackson (1996)	0.247	0.051	0.444	0.014
Landström & Granhag (2008)	-0.375	-0.622	-0.127	0.003
Landström & Granhag (2010)	0.132	-0.047	0.311	0.150
Landström, Granhag, & Hartwig (2007)	0.044	-0.194	0.281	0.720
Strömwall & Granhag (2005)	0.570	0.269	0.871	0.000
Strömwall & Granhag (2007)	0.193	-0.225	0.612	0.366
Warren, Dodd, Raynor, & Peterson (2012)	-0.596	-0.833	-0.358	0.000
Westcott, Davies, & Clifford (1991)	0.574	0.074	1.074	0.024
Crossman & Lewis (2006)	-0.727	-1.084	-0.369	0.000
Leach et al. (2009) EXP1	0.100	-0.288	0.488	0.614
Leach et al. (2009) EXP5	0.100	-0.616	0.816	0.784
Talwar & Lee (2002)	0.076	-0.146	0.298	0.500
Talwar & Lee (2002)b	-1.167	-1.479	-0.854	0.000
Talwar, Crossman, Gulmi, Renaud, & Williams (2009)	-0.284	-0.507	-0.061	0.013
Talwar, Renaud, & Conway (2015)	0.868	0.685	1.052	0.000
Talwar, Renaud, & Conway (2015)a	3.000	2.744	3.256	0.000
Cassidy (2015)	0.086	-0.122	0.293	0.420
Chahal & Cassidy (1994)	0.384	0.023	0.746	0.037
Gadea, Aliño, Espert, & Salvador (2015)	0.338	0.065	0.613	0.015

Std diff in means and 95% CI





Confidence-accuracy correlation. The correlation between receivers' judgment confidence and the accuracy of their judgment (17 experiments, see Table 1) showed a small significant effect, $r = 0.068$, 95% CI [0.001, 0.136]. Higher confidence was associated with a higher rate of accuracy (i.e., calling a false statement dishonest or calling a true statement honest). Individual correlations ranged from $r = -0.09$ to 0.287.

Professional and lay accuracy. There were 25 samples from 12 studies (Table 1) in which professionals ($n = 873$) classified statements as either honest or dishonest. Their unweighted mean percentage correct classifications was 56.07% ($SD = 7.1\%$), with a range of 44% to 70% and a median of 56%. Analyses comparing the professionals' accuracy rate to chance revealed a statistically significant effect, Cohen's $d = 0.491$, 95% CI [0.239, 0.743]. Fourteen of the 25 individual effect sizes on their own showed a significant difference from chance at $p < .05$. Individual Cohen's d effect sizes ranged from -0.998 to 1.721.

The accuracy rate of lay people, primarily undergraduates, was examined in 40 experiments (n = 6,380). Their unweighted mean percentage correct truth/lie classifications was 53.91% (SD = 8.89%), ranging 32% to 66.5%, median = 53.25%. Analyses directly comparing the accuracy rate to chance revealed a statistically significant effect, Cohen's d = 0.201, 95% CI [0.063, 0.340], with 20 of the 40 individual effects showed a significant difference on the own at $p < .05$. Individual Cohen's d effect sizes ranged from -1.200 to 1.941. Subgroup analyses, comparing professions and laypersons, revealed that the average effect sizes did significantly differ between groups, $Q(1) = 3.914$, $p = .048$, with professionals outperforming lay decision-makers.

Accuracy and children's age. A final set of analyses concerned whether adults' accuracy varied depending on the age of the sender. Sixteen studies (Table 1) conducted age comparisons between "younger" and "older" children. However, the ages that constituted "younger" versus "older" children varied across the studies: Some considered 3 to 5 year olds to be "young" while others consider ages 6 to 8 to be "young." In order to increase the consistency and the granularity of analysis, we grouped senders as follows: young (sender ages 3-5); middle (sender ages 6-9); and old (sender ages 10-15). This produced three types of age comparisons: young-middle, middle-old, and young-middle-old. The mean adult accuracy rate when detecting young senders was 60.05% (SD = 6.68%), the mean accuracy rate for middle senders was 56.73% (SD = 7.9%), and the mean accuracy rate for old senders was 52.45% (SD = 10.89%). Subgroup analyses between the three group comparisons revealed that the differences in the effect sizes were approaching significance $Q(2) = 5.594$, $p = .061$. Adults' accuracy rates did not differ between the young-middle age comparison (k = 8, p = .730), nor did they differ in the middle-old age comparison (k = 3, p = .939). Across the 5 studies that compared adult detection accuracy rates between all three young, middle and older groups of children, adults tended to be most accurate with the youngest age group relative to the oldest age group (Cohen's d = 0.238, 95% CI [0.112, 0.356]). Each group contains only a handful of experiments; therefore, this is an exploratory analysis of age comparisons.

Accuracy across paradigms. Subgroup analyses were conducted on the method for generating false statements (i.e., explicit coaching, implicit coaching, self-generated) and adults' average accuracy rates did not differ between these groups, $Q(2) = 2.864$, $p = .239$. Subgroup analyses were also conducted on transgression paradigm type (i.e., child's transgression, other's transgression, or no transgression) and scenario paradigm type (i.e., common events, temptation resistance, games and witnessing a transgression) and but the average effect sizes did not significantly differ as a function of these paradigm differences ($Q(2) = 0.027$, $p = .873$; $Q(3) = 0.937$, $p = .817$, respectively), and hence will not be discussed further.

DISCUSSION

The overarching purpose of the present meta-analysis was to synthesize findings from extant research on adult's ability to detect deception in children in order to generate clearer conclusions about how well adults can discern truthful and false statements provided by children. In all studies included in the meta-analysis, adults viewed children, usually on videotape, providing true or false answers to questions and attempted to classify those answers as honest or dishonest. Across studies and type of statement, the average mean percentage correct was significantly above chance, at 54.3%, and amounted to a small-medium effect size (d = 0.242). This average is comparable to that obtained in meta-analyses of studies examining adults' ability to detect deception in adults, with these percentages averaging between 54% and 57% (Bond & DePaulo, 2006; Aamodt & Custer, 2006; Kraut, 1980). Thus, in general, adults do not appear more proficient at detecting deception in children than adults, despite some reasons to suspect that children's more limited cognitive and deceptive abilities would render their true and false statements more easily discernable.

Subgroup analyses, though, revealed trends in terms of response biases that affected adults' deception detection accuracy. Consistent with the literature examining deception detection in adult senders (Bond & DePaulo, 2006), higher accuracy was detected when classifying children's true statements as such (59.7%) than when classifying false statements as such (49.4%). There were no significant differences in effect sizes for classifying true positives compared to true negatives, revealing a general truth bias, or tendency to believe proffered statements provided by children. One possible explanation for this involves a type of anchoring, in which most people believe that social interactions are honest and often fail to sufficiently adjust this inclination, thus resulting in a bias towards their initial positon (Vrij, Granhag, & Porter, 2010), which in this case would be that children are truthful. Further research is necessary to directly test this explanation for the truth bias, and whether it applies in the case of child senders. When judging deceptive statements by children, however, adults' performance did not differ from chance (49.4%), and this performance did not vary depending on whether the statement reflected a false positive or false negative. These findings could be interpreted as either that adults are not adept at detecting children's false statements or that children are somewhat effective in their use of deception (see Talwar, Crossman, Williams, & Muir, 2011). Overall, these two different levels of accuracy and tendency toward response biases have important implications for individuals who are charged with the difficult task of evaluating the veracity of children's statements, particularly those who do so in forensic settings. These professionals need to be informed of their potential biases, and that they ought not to place too much faith in their ability to detect deception in children (see also Vrij, 2002) but instead, when situations are warranted, to consider possibilities of both true and false statements, both in terms of true and false allegations and also true and false denials.

We were also interested in the links between judgment confidence and accuracy, given long-standing debates about whether confidence is informative as to the likely accuracy of judgments (DePaulo et al., 1997). Contrary to the adult deception detection literature, which has not uncovered any consistent associations, a modest relation emerged between how confident adults were in their evaluation of the veracity of children's reports and the accuracy of that evaluation. Despite this trend, however, the correlation, $r = 0.068$, was modest, and its practical significance is limited. We can only tentatively conclude that adults are more confident in their decisions when they have made a correct classification, but this finding should prompt further investigation into why this might be for children but not adults. Since accuracy rates tend not to differ between the two populations, there might be something unique about communicating with children that helps to inform both their confidence and their accuracy.

Two other important moderators we considered were characteristics of the receiver and the sender, namely the adult raters' training or expertise with children and the child sender's age. Although a lack of group variation has been uncovered in literature with adult senders (Vrij, 2004), professionals such as social workers, teachers, and even police officers, especially the first two, likely have more experience with children, and thus their average accuracy rates may be superior compared to lay persons. Indeed, the results revealed a statistically significant difference between professional (56%) and layperson (54%) overall accuracy rates. Practically, though, this advantage is slim, and professionals' lack of particular skill in detecting deception is especially concerning in light of evidence that laypersons often believe the professionals are highly skilled (e.g., Quas et al., 2005). Laypersons and professionals alike need to be made aware of their own (and each other's limitations) so that their evaluations are not given undue weight.

The ability to discriminate truths/lies as a function of the age of the child was trending in the expected direction. However, many of the constituent studies did not find a different rate of accuracy between "young" children and "old" children, and many of the studies that did find an effect on accuracy were inconsistent: some reported higher rates for young children (Westcott, Davies, & Clifford, 1991), and others reporter higher rates for old children (Vrij, Akehurst, Brown, & Mann, 2006). Adding to this confusion is no consistent operational definition of what ages correspond to "old" and "young" children. To make a consistent comparison across age groups, we utilized our own operationalization of young (age 3 to 5), middle (age 6 to 9) and old (age 10 to 15) children and re-analyzed the extant data. This regrouping yielded only a handful of studies per comparison group, hence the findings are only exploratory. This analysis revealed that adults were more accurate with the youngest children (60%) relative to middle (56.7%) and older (52.5%) aged children. Therefore, the children tended to adhere to the developmental expectations that they become increasingly skilled liars as they age. More research is clearly necessary to help reconcile the inconsistency of the findings on the relation between accuracy of deception detection and age of the child sender. There are likely to be a number of moderators (e.g., gender of child sender; type of lie; high or low stakes; etc.) that might interact with the age of sender; these interactive effects could influence both the ability of the child to deceive as well as the ability of an adult to detect that deception.

Limitations

Despite the unique contribution of the current meta-analysis in synthesizing the extant findings, this review must address the possible limitations common to similar meta-analyses.

Publication Bias. If only large or significant effects are published while non-significant findings are relegated to the "file drawer," it is possible that the effects reported here, as in other meta-analyses, could be inflated (Rosenthal, 1979). Efforts were made to reduce this possibility, though, by contacting published authors in an effort to locate unpublished findings, whether or not they are significant, and including all such findings in the analyses. To empirically test for the possibility of publication bias, we calculated Rosenthal's fail-safe N (Rosenthal, 1979), which gives the number of studies with a completely null effect that would need to be added in order to make the main effect no longer significant at $p < .05$. This number was estimated at 1,515 studies. As it is unlikely that this many unpublished studies are filed away, we can conclude with some confidence that bias is not present in these results. Since the primary focus of many of these articles was to examine how and why children lie and adult detection accuracy was added to supplement those conclusions, we found that nearly half of the reported effect sizes for adult's overall accuracy were not significant.

Internal validity. It is not uncommon to find a high level of heterogeneity in reviews of deception detection research (Bond & DePaulo, 2006; Sporer & Schwandt, 2007) because studies examining deception detection rarely use a uniform methodology and instead seek to examine a variety of possible factors that could potentially influence the results. Moderator analyses evaluated whether several key factors reliably affected the results. However, other factors, such as high and low stakes reports, length of time between the target event and interview, motivation, and presentation mode, were not included here due to insufficient numbers of studies testing such factors. The lack of inclusion could affect the reported results in at least two ways. First, not including them could increase error variance, thereby understating the reported effect sizes. Second, not including them could lead to important effects going undetected. A remedy to these concerns is to conduct additional research with more standardized methods. However, even with the heterogeneity of methodologies, the fact that some consistent trends emerged, and the fact that several of the findings reported here converge with those obtained with adult senders, lends support to the overall trends observed and the conclusions that were generated.

Suggestions for Future Research

Moderators and Methodology. Although accuracy did not vary significantly between paradigm types, future work disentangling other possible moderators would be informative. Of interest would be studies that address more specific paradigm comparisons, such as high stakes (witnessing a transgression) compared to low stakes (games); or salient events (e.g., injuries and dental visits) compared to common events. A few of these factors have been recently addressed. For example, Talwar, Crossman, Williams and Muir (2011) compared children's lies from different contexts, in this case pro-social (disappointing gift) compared to anti-social (temptation resistance) lies, and found increased detection accuracy for anti-social lies, particularly for female child senders. Furthermore, a study by Saykaly et al. (2013) compared stressful and non-stressful lies and found that participants were more accurate when they evaluated statements about non-stressful events.

Other factors involving the logistics of the child interviews could also have important effects on adult accuracy and should be considered in future research. Many studies only asked the children direct yes or no questions that resulted in interviews only seconds long. The interviews tended to be very brief (less than one minute) and did not include baseline information or responses from children. Comparing question types, interview length, narratives vs forced choice, and formal vs informal interview styles could shed light on factors that affect detection accuracy, particularly across age, given the dramatic effects that children's age has on their responses to different question types (see Lyon, 2014; Andrews et al., 2016). Whether these findings replicate or generalize to different contexts and different paradigms are important questions that need to be addressed by further research.

Despite some limitations, and despite the need for additional research, the present meta-analysis reveals first, that adults can detect deception in children, although their accuracy is only slightly above chance, and that accurate adults are slightly more confident in their decisions. Moreover, this review uncovered several significant trends across this literature. Adults were more accurate when detecting truthful statements, which suggests that adults are swayed by a type of truth bias in their judgments. Professionals who regularly work with children significantly, albeit only slightly, outperformed laypersons. However, laypersons and professionals alike appear to need more information about their own limitations in deception detection accuracy and about children's deceptive abilities in different contexts. Such information may help adults temper their own judgments and rely on other potentially corroborating information to decide whether or how truthful a child is when reporting personal information. A last notable trend is the attenuation of accuracy as the child ages. On average, adults tended to be more successful when detecting statements in younger, preschool-age children compared to older children, suggesting that children's capacity to maintain a lie is related to their developmental trajectory. Thus, while these findings in many ways echo the adult deception detection literature, there are also distinctive influences particular to detecting deception in children.

REFERENCES

Note: asterisks (*) indicate studies that were included in the meta-analysis.

Aamodt, M. G., & Custer, H. (2006). Who can best catch a liar? A meta-analysis of individual differences in detecting deception. *Forensic Examiner, 15*(1), 6–11.

Andrews, S. J., Ahern, E. C., Stolzenberg, S. N., & Lyon, T. D. (2016). The Productivity of Wh-Prompts when Children Testify. *Applied Cognitive Psychology*.

*Bala, N., Ramakrishnan, K., Lindsay, R., & Lee, K. (2004). Judicial assessment of the credibility of child witnesses. *Alta. L. Rev., 42*, 995–1017.

Ball, C. T., & O'Callaghan, J. (2001). Judging the accuracy of children's recall: A statement-level analysis. *Journal of Experimental Psychology: Applied, 7*(4), 331–345.

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.

Boerner, K. E., Chambers, C. T., Craig, K. D., Riddell, R. R. P., & Parker, J. A. (2013). Caregiver accuracy in detecting deception in facial expressions of pain in children. *Pain, 154*(4), 525-533.

Bond, C.F., & DePaulo, B.M. (2006). Accuracy of deception judgements. *Personality and Social Psychology Review, 10*(3), 214–234.

*Brunet, M.K. (2009). Why bullying victims are not believed: Differentiating between children’s true and fabricated reports of stressful and non-stressful events (Unpublished master’s thesis), University of Toronto.

Case, R. (1992). Neo-piagetian theories of child development. In R.J. Sternberg & C.A. Berg (Eds.) *Intellectual Development* (161-196). Cambridge University Press, New York, NY.

*Cassidy, T. (2015). Adult Judgments of Children's Veracity: The Role of Parenthood and Interpersonal Sensitivity. *Journal of Child and Adolescent Behavior, 3*(1), 1000178.

Ceci, S. J., & Huffman, M. C. (1997). How suggestible are preschool children? Cognitive and social factors. *Journal of the American Academy of Child & Adolescent Psychiatry, 36*(7), 948–958.

Ceci, S. J., Huffman, M. L. C., Smith, E., & Loftus, E. F. (1994). Repeatedly thinking about a non-event: Source misattributions among preschoolers. *Consciousness and Cognition, 3*(3), 388–407.

*Chahal, K., & Cassidy, T. (1994). Deception and its detection in children: A study of adult accuracy. *Psychology, Crime and Law, 1*(3), 237–245.

Chandler, M., Fritz, A. S., & Hala, S. (1989). Small-scale deceit: Deception as a marker of two-, three-, and four-year-olds' early theories of mind. *Child Development*, *60*(6), 1263-1277.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159.

*Crossman, A. M., & Lewis, M. (2006). Adults' ability to detect children's lying. *Behavioral Sciences & the Law*, *24*(5), 703–715.

DePaulo, B.M., Kashy, D.A., Kirkendol, S.E., Wyer, M.M., & Epstein, J.A. (1996). Lying in everyday life. *Journal of Personality and Social Psychology, 70*(5), 979–995.

DePaulo, B.M., Charlton, K., Cooper, H., Lindsay, J. L., & Muhlenbruck, L. (1997). The accuracy–confidence correlation in the detection of deception. *Personality and Social Psychology Review, 1*(4), 346–357.

DePaulo, B.M., Lindsay, J.L., Malone, B.E., Muhlenbruck, L., Charlton, K., & Cooper, H. (2003). Cues to deception. *Psychological Bulletin, 129*(1), 74–118.

DePaulo, B. M., Wetzel, C., Weylin Sternglanz, R., & Wilson, M. J. W. (2003). Verbal and nonverbal dynamics of privacy, secrecy, and deceit. *Journal of Social Issues, 59*(2), 391-410.

*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.

Feeley, T. H., & Young, M. J. (2000). Self-reported cues about deceptive and truthful communication: The effects of cognitive capacity and communicator veracity. *Communication Quarterly, 48*(2), 101-119.

Feldman, R. S., Jenkins, L., & Popoola, O. (1979). Detection of deception in adults and children via facial expressions. *Child Development, 50*(2), 350–355.

Feldman, R. S., & White, J. B. (1980). Detecting deception in children. *Journal of Communication, 30*(2), 121–128.

*Gadea, M., Aliño, M., Espert, R., & Salvador, A. (2015). Deceit and facial expression in children: the enabling role of the "poker face" child and the dependent personality of the detector. *Frontiers in Psychology, 6*, 1089.

Garrido, E., Masip, J., & Herrero, C. (2004). Police officers' credibility judgements: Accuracy and estimated ability. *International Journal of Psychology, 39*(4), 254–275.

Global Deception Research Team. (2006). A world of lies. *Journal of Cross-Cultural Psychology, 37*(1), 60-74.

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.

*Goodman, G. S., Myers, J. E., Qin, J., Quas, J. A., Castelli, P., Redlich, A. D., & Rogers, L. (2006). Hearsay versus children's testimony: Effects of truthful and deceptive statements on jurors' decisions. *Law and Human Behavior, 30*(3), 363–401.

Goodman, G. S., Tobey, A. E., Batterman-Faunce, J. M., Orcutt, H., Thomas, S., Shapiro, C., & Sachsenmaier, T. (1998). Face-to-face confrontation: Effects of closed-circuit technology on children's eyewitness testimony and jurors' decisions. *Law and Human Behavior, 22*(2), 165–203.

Gordon, H., Talwar, V., & Lee, K. (2005). Deception detection: Verbal and non-verbal cues in children between 7 and 11 years of age. In American Psychological Law Society Annual Conference.

Hoyle, R. H. (1999). *Statistical strategies for small sample research.* Thousand Oaks, CA: Sage Publications.

*Jackson, J. L. (1996). Truth or fantasy: The ability of barristers and laypersons to detect deception in children's testimony. Paper presented at the AP-LS Biennial Conference, Hilton Head Island, South Carolina.

*Johnson, J. L., Hobbs, S. D., Chae, Y., Goodman, G. S., Shestowsky, D. S., & Block, S. (in prep). Adults' evaluations of children's true and false reports and denials.

Kraut, R. (1980). Humans as lie detectors. *Journal of Communication, 30*(4), 209-218.

Lamb, M. E., Orbach, Y., Sternberg, K. J., Aldridge, J., Pearson, S., Stewart, H. L., ... & Bowler, L. (2009). Use of a structured investigative protocol enhances the quality of investigative interviews with alleged victims of child sexual abuse in Britain. *Applied Cognitive Psychology, 23*(4), 449-467.

Lamb, M. E., Sternberg, K. J., Orbach, Y., Esplin, P. W., Stewart, H., & Mitchell, S. (2003). Age differences in young children's responses to open-ended invitations in the course of forensic interviews. *Journal of Consulting and Clinical Psychology, 71*(5), 926-934.

*Landström, S., & Granhag, P. A. (2008). Children's truthful and deceptive testimonies: How camera perspective affects adult observers' perception and assessment. *Psychology, Crime & Law*, *14*(5), 381-396.

*Landström, S., & Granhag, P. A. (2010). In-court versus out-of-court testimonies: Children's experiences and adults' assessments. *Applied Cognitive Psychology, 24*(7), 941–955.

*Landström, S., Granhag, P. A., & Hartwig, M. (2007). Children's live and videotaped testimonies: How presentation mode affects observers' perception, assessment and memory. *Legal and Criminological Psychology, 12*(2), 333–348.

*Lawrence, H. J., Akehurst, L., Leach, A.-M., Cherryman, J., Vrij, A., Arathoon, M., & Vernham, Z. (in prep). Imposing cognitive load through gaze maintenance to discern truth and deception in child witnesses.

*Leach, A. M., Lindsay, R. C. L., Koehler, R., Beaudry, J. L., Bala, N. C., Lee, K., & Talwar, V. (2009). The reliability of lie detection performance. *Law and Human Behavior*, *33*(1), 96–109.

*Leach, A. M., Talwar, V., Lee, K., Bala, N., & Lindsay, R. C. L. (2004). "Intuitive" lie detection of children's deception by law enforcement officials and university students. *Law and Human Behavior*, *28*(6), 661–685.

Leichtman, M. D., & Ceci, S. J. (1995). The effects of stereotypes and suggestions on preschoolers' reports. *Developmental Psychology, 31*(4), 568–578.

*Lewis, M., Stanger, C., & Sullivan, M. W. (1989). Deception in 3-year-olds. *Developmental Psychology*, *25*(3), 439–443.

Lyon, T. D. (2014). Interviewing children. *Annual Review of Law and Social Science, 10,* 73-89.

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.

Morency, N. L., & Krauss, R. M. (1982). Children’s nonverbal encoding decoding of affect. In R. S. Feldman (Ed.), *Development of nonverbal behavior in children* (pp. 181–199). New York: Springer-Verlag.

Newcombe, P. A., & Bransgrove, J. (2007). Perceptions of witness credibility: Variations across age. *Journal of Applied Developmental Psychology*, *28*(4), 318–331.

Newton, P., Reddy, V., & Bull, R. (2000). Children's everyday deception and performance on false-belief tasks. *British Journal of Developmental Psychology, 18*(2), 297-317.

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

*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*(4), 339–372.

Quas, J. A., Davis, E. L., Goodman, G. S., & Myers, J. E. (2007). Repeated questions, deception, and children’s true and false reports of body touch. *Child Maltreatment*, *12*(1), 60-67.

Quas, J. A., Malloy, L. C., Melinder, A., Goodman, G. S., D'Mello, M., & Schaaf, J. (2007). Developmental differences in the effects of repeated interviews and interviewer bias on young children's event memory and false reports. *Developmental Psychology, 43*(4), 823-837.

Quas, J. A., Thompson, W. C., & Clarke-Stewart, K. A. (2005). Do jurors" know" what isn't so about child witnesses?. *Law and Human Behavior, 29*(4), 425–456.

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638–641.

*Saykaly, C., Talwar, V., Lindsay, R. C., Bala, N. C., Lee, K., Bertrand, M., & Nugent, S. M. (2013). Adults' ability to detect deception of stressful and non-stressful stories of children. *Psychology, Crime & Law*, *19*(10), 865–879.

Saywitz, KJ, Goodman, GS, & Lyon, TD (2002). Interviewing children in and out of court: Current research and practice implications. In J. Myers, L. Berliner, J. Briere, C. Hendrix, C. Jenny, & T. Reid (Eds.), *The APSAC Handbook on Child Maltreatment* (349-377). Thousand Oaks, CA: Sage.

*Shao, Y., & Ceci, S. J. (2011). Adult credibility assessments of misinformed, deceptive and truthful children. *Applied Cognitive Psychology, 25*(1), 135–145.

*Shao, Y. (2007). Credibility Assessment of Misinformed vs. Deceptive Children (Unpublished Master’s Thesis), Cornell University.

Schmidt, F. L., Oh, I. S., & Hayes, T. L. (2009). Fixed-versus random-effects models in meta-analysis: Model properties and an empirical comparison of differences in results. *British Journal of Mathematical and Statistical Psychology, 62*(1), 97-128.

Sporer, S. L., & Schwandt, B. (2007). Moderators of nonverbal indicators of deception: A meta-analytic synthesis. *Psychology, Public Policy, and Law, 13*(1), 1-34.

*Stromwall, L. A., & Granhag, P. A. (2005). Children's repeated lies and truths: Effects on adults' judgments and reality monitoring scores. *Psychiatry, Psychology and Law, 12*(2), 345–356.

*Strömwall, L. A., & Granhag, P. A. (2007). Detecting Deceit in Pairs of Children. *Journal of Applied Social Psychology, 37*(6), 1285-1304.

*Strömwall, L. A., Granhag, P. A., & Landström, S. (2007). Children's prepared and unprepared lies: can adults see through their strategies? *Applied Cognitive Psychology, 21*(4), 457–471.

*Sumner-Armstrong, C., & Newcombe, P. A. (2007). The education of jury members: Influences on the determinations of child witnesses. *Psychology, Crime & Law, 13*(3), 229-244.

Swerts, M. (2012). Let's lie together: co-presence effects on children's deceptive skills. In *Proceedings of the Workshop on Computational Approaches to Deception Detection* (55-62). Association for Computational Linguistics.

Swerts, M., van Doorenmalen, A., & Verhoofstad, L. (2013). Detecting cues to deception from children's facial expressions: On the effectiveness of two visual manipulation techniques. *Journal of Phonetics, 41*(5), 359-368.

Talwar, V., & Crossman, A. (2011). From little white lies to filthy liars: The evolution of honesty and deception in young children. *Advances in Child Development and Behaviour, 40*(140), 139-179.

*Talwar, V., Crossman, A. M., Gulmi, J., Renaud, S. J., & Williams, S. (2009). Pants on Fire? Detecting Children's Lies. *Applied Developmental Science, 13*(3), 119–129.

*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*(12), 2837–2857.

*Talwar, V., & Lee, K. (2002). Development of lying to conceal a transgression: Children's control of expressive behaviour during verbal deception. *International Journal of Behavioral Development*, *26*(5), 436–444.

*Talwar, V., & Lee, K. (2002). Emergence of white-lie telling in children between 3 and 7 years of age. *Merrill-Palmer Quarterly, 48*(2), 160-181.

Talwar, V., & Lee, K. (2008). Social and cognitive correlates of children's lying behavior. *Child Development*, *79*(4), 866-881.

*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.

Talwar, V., Murphy, S. M., & Lee, K. (2007). White lie-telling in children for politeness purposes. *International Journal of Behavioral Development*, *31*(1), 1-11.

*Talwar, V., Renaud, S. J., & Conway, L. (2015). Detecting children's lies: Are parents accurate judges of their own children's lies? *Journal of Moral Education*, *44*(1), 81–96.

*Tye, M. C., Amato, S. L., Honts, C. R., Devitt, M. K., & Peters, D. (1999). The willingness of children to lie and the assessment of credibility in an ecologically relevant laboratory setting. *Applied Developmental Science, 3*(2), 92–109.

Vrij, A. (2000). *Detecting lies and deceit: The psychology of lying and implications for professional practice*. Chichester, England: Wiley.

Vrij, A. (2002). Deception in children: A literature review and implications for children's testimony. In H. L. Westcott G. M. Davies (Eds.) *Children's Testimony: A Handbook of Psychological Research and Forensic Practice*, 175–194. Chichester, England: Wiley.

Vrij, A. (2004). Why professionals fail to catch liars and how they can improve. *Legal and Criminological Psychology, 9*(2), 159-181.

Vrij, A. (2008). Nonverbal dominance versus verbal accuracy in lie detection a plea to change police practice. *Criminal Justice and Behavior, 35*(10), 1323-1336.

*Vrij, A., Akehurst, L., Brown, L., & Mann, S. (2006). Detecting lies in young children, adolescents and adults. *Applied Cognitive Psychology*, *20*(9), 1225–1237.

Vrij, A., Akehurst, L., & Knight, S. (2006). Police officers', social workers', teachers' and the general public's beliefs about deception in children, adolescents and adults. *Legal and Criminological Psychology, 11*(2), 297–312.

Vrij, A., Granhag, P. A., & Porter, S. (2010). Pitfalls and opportunities in nonverbal and verbal lie detection. *Psychological Science in the Public Interest*, *11*(3), 89–121.

*Vrij, A. & Van Wijngaarden, J. J. (1994). Will the truth come out? Two studies about the detection of false statements expressed by children. *Expert Evidence: The International Digest of Human Behavior, Science and Law, 3*(2), 78–84.

*Warren, K. L., Bakhtiar, A., Mulrooney, B., Raynor, G., Dodd, E., & Peterson, C. (2015). Adults' Detection of Deception in Children: Effect of Coaching and Age for Children's True and Fabricated Reports of Injuries. *Behavioral Sciences & the Law, 33*(6), 784-800.

*Warren, K. L., Dodd, E., Raynor, G., & Peterson, C. (2012). Detecting Children's Lies: Comparing True Accounts about Highly Stressful Injuries with Unprepared, Prepared, and Coached Lies. *Behavioral Sciences & the Law, 30*(3), 329–341.

*Westcott, H. L., Davies, G. M., & Clifford, B. R. (1991). Adults' perceptions of children's videotaped truthful and deceptive statements. *Children & Society*, *5*(2), 123–135.