Routledge
Taylor & Francis Group

Check for updates

Factors influencing the perceived credibility of children alleging physical abuse

Meaghan C. Danby (iD), Stefanie J. Sharman (iD) and Bianca Klettke (iD)

School of Psychology, Deakin University, Burwood, VIC, Australia

Adults' assessments of the credibility of children's reports are affected by factors including the frequency of abuse, reporting delays and the child's age. The present study examined whether similar factors affect the perceived credibility of children reporting physical abuse, which is more common than sexual abuse. Two hundred and eight mock jurors read a simulated transcript of a child reporting physical abuse to police and made credibility ratings. Within each transcript, abuse frequency (once, repeated), reporting timing (recent, delayed), police question type (open, closed) and child age (6 or 10 years) were manipulated. The child was considered more credible when the abuse was only experienced once and reported shortly after it occurred, and when prompted with open questions. The child's age did not affect credibility judgments. Current findings support recommendations to prioritise open questions with children and provide evidence for extension of the benefits of open questions to children's credibility.

Keywords: child interviewing; child physical abuse; children's credibility; investigative interviewing; juror perceptions; police questioning.

Child abuse cases can be difficult to prosecute as corroborating evidence, such as medical evidence or testimony from other witnesses, is often absent (Hartley et al., 2013; Walsh et al., 2010). Given that the child's testimony is often the primary (or sole) form of evidence, the perceptions that police, prosecutors, judges and jurors hold regarding a child's credibility are vital to case outcomes. As such, it is important to understand the factors impacting adults' perceptions of children's credibility (see Bottoms et al., 2007, for a review).

To date, research examining adults' perceptions of children's credibility has focused primarily on children who reported sexual abuse. It has generally been shown that female mock jurors find child sexual abuse reports more credible than males, and that younger children are considered more honest about sexual abuse than older children (see Bottoms et al., 2007, and Hatton & Duff, 2016, for reviews). Few studies have considered the credibility of children reporting physical abuse, and have generally demonstrated a trend for children to be considered more positively when reporting physical than sexual abuse (Bornstein et al., 2007; Sheahan et al., 2021). For example, after reading a vignette depicting child abuse, male participants in Bornstein and colleagues' (2007) study found the described child more believable when the vignette involved physical abuse rather than severe sexual abuse.

This underrepresentation of physical abuse cases in the research regarding children's credibility is problematic because physical abuse represents a notable proportion of child abuse

cases (e.g. in Trocome et al., 2010, 20% of substantiated child maltreatment cases involved physical abuse while only 3% involved sexual abuse; a recent Australian Bureau of Statistics, 2019, survey found that 1.6 million children experienced physical abuse, and 1.4 million children experienced sexual abuse before the age of 15). Thus, the current study's aim is to explore factors that – based on the research examining adults' credibility judgments of children's sexual abuse reports – were predicted to affect adults' credibility judgments of children's physical abuse reports. Based on previous research, we explored the following factors: whether the abuse occurred once or was repeated and whether it was reported immediately or after a delay. We also investigated the type of questions used by police and the child's age. These factors are discussed in more detail below.

Abuse frequency and reporting delays

Children often experience physical abuse repeatedly before it is reported to an adult (Australian Bureau of Statistics, 2019; Trocome et al., 2010). One reason for the high incidence of repeated child physical abuse is children's reluctance to disclose their abuse to a trusted adult promptly (Foynes et al., 2009; Hershkowitz, 2006; Rush et al., 2014). Delayed reporting is particularly common in physical abuse cases (compared to cases involving sexual or other types of abuse; Foynes et al., 2009; Hershkowitz, 2006; Rush et al., 2014). For example, in an investigation of court-substantiated abuse cases, Rush and colleagues (2014) found that only 27% of child physical abuse victims had disclosed their abuse prior to any police investigation, compared to 67% in child sexual abuse allegations. Delays in reporting allow more opportunity for abuse to re-occur, and physical abuse may occur continually until authorities are made aware of it.

In terms of abuse frequency, few studies have compared the perceived credibility of children's reports of repeated versus single-instance events. In one study, Connolly and colleagues (2008) found that adults judged children's reports about an activity session that was repeated four times *less* credible than reports from children who experienced the activities only once. The authors suggested that this difference was possibly explained by a lower consistency and confidence in children's reports about the repeated event than about the single-instance event. In another study, Pozzulo and colleagues (2010) examined mock jurors' perceptions of simulated child sexual abuse case transcripts from complainants alleging repeated or single-instance abuse. They also considered the effect of delayed reporting, manipulating whether the complainant was still a child (short delay) or an adult (long delay). It is important to note that even in the short delay condition, the minimum delay was two years. They found no effect of either repetition or delay on mock jurors' judgments of credibility. The authors speculated that the lack of significant effects was potentially due to an absence of emotional expressions and detailed complainant descriptions in the study materials, which mock jurors may have expected.

Other studies have examined the effect of delay on credibility in studies of children's reports of single-instance sexual abuse; the results have shown a trend for mock jurors to perceive complainants as more believable when reports were made after shorter delays (e.g. under 1 year) than after longer delays (e.g. 20 years; Golding et al., 1995; Sugarman & Boney-McCoy, 1997). It is possible that mock jurors perceive children's memories to be more malleable over time, with long delays resulting in more memory decay and less accurate recall of an event. However, in contrast with adults' beliefs about children's memories, research has demonstrated that long-term memories for core details about personally salient and emotional events, such as abuse, tend to remain accurate over time, up to 20 years in some studies (Ceci & Bruck, 1993; Goldfarb et al., 2019; Goodman et al., 2017, 2019).

Taken together, the effects of event frequency and delay on adults' perceptions of children's credibility are mixed – and particularly for event frequency. However, there appears to be a trend for adults to perceive children's reports of sexual abuse made after shorter delays as more credible than reports made after longer delays.

Question type and age

While the frequency of abuse and timing of reporting cannot be controlled by the criminal justice system, the questions used to elicit children's accounts are dictated by the police, lawyers and judges that speak to them. Question type reliably affects the number of details in children's reports and their subsequent accuracy (see Larsson & Lamb, 2009, for a review). Closed questions request a particular detail to be reported and can be answered in very few words (e.g. 'Where did it hurt?', 'Did you tell anyone?'; Snow et al., 2009). Open questions encourage elaborate narrative responses without specifying the particular information to be reported (e.g. 'What happened next?', 'Tell me more about when John got angry'). Because open questions encourage more accurate (Dent & Stephenson, 1979) and detailed (Brown et al., 2013) responses than closed questions, open questions are commonly recommended for use by professionals working with children (Lamb et al., 2007; Powell et al., 2005). Despite these recommendations, police interviewers find using open questions difficult. Interviewers often rely too heavily on closed questions when speaking with children, especially with young children (e.g. under 7 years old; Sternberg et al., 2001; Thoresen et al., 2006; Westcott & Kynan, 2006).

Two studies have compared the effects of open and closed questions on children's perceived credibility, and have shown mixed results (Ruva & Bryant, 2004; Tubb et al., 1999). Tubb et al. (1999) asked mock jurors to read a hypothetical police interview conducted with a 9-year-old alleging a single instance of sexual abuse and manipulated the interviewer's questioning. No differences in credibility ratings were found between accounts elicited using open and closed questions. In the experiment conducted by Ruva and Bryant (2004) age was also manipulated; participants read mock court transcripts of a 6-, 10- or 22-year-old witness's account of one incident of a robbery-turned-murder elicited using either open or closed questions. A 6-year-old witness was rated as more credible when asked open than closed questions. No differences were found in the credibility of the 10- or 22-year-old witnesses across question types.

In studies that have not manipulated interviewer question types, children's age has often influenced their perceived credibility. While findings about children's age are somewhat mixed, there appears to be a tendency for younger children to be deemed more credible than older children when reporting on sexual abuse, presumably because their perceived sexual naïvety reduces suspicions that the report is fabricated (e.g. Davies & Rogers, 2009; see Bottoms et al., 2007). Conversely, older children are sometimes considered more credible than younger children when reporting on other events (e.g. play activities in Connolly et al., 2008), presumably due to expected age-related improvements in memory retrieval and accuracy (see Bottoms et al., 2007). In the only study to consider victim age in physical abuse cases, Sheahan and colleagues (2021) investigated mock jurors' assessments of a victim alleging either physical or sexual abuse from a soccer coach. When alleging that the abuse was sexual, the victim was considered more credible when presented as an adult than as a 12-year-old child, but when reporting that the abuse was physical, the victim was considered more credible as the child than as the adult. Results were surprising, and inconsistent with previous trends in the credibility of sexual abuse victims.

Current study

The aim of the current study was to examine factors predicted to affect mock jurors'

perceptions of children reporting physical abuse. Given the high occurrence of repeated incidents and delayed reporting in physical abuse cases, we manipulated these case factors to determine their impact on the rated credibility of children's police interviews. Since the questioning used during an interview is under the control of police, we also manipulated the questioning style used to elicit children's reports. Finally, given the relationship with question type and mixed effects to date, child age was also manipulated.

One of the main shortcomings in research addressing children's credibility is the fact that the perceived credibility of the child has been largely based on vignettes, which do not include actual testimony provided by the child (for a review see Voogt et al., 2019). To address this limitation, the current study used a simulated transcript of a child's investigative interview with a police officer. This method offered improved authenticity and highlighted the interviewer's question types. After reading the transcript, participants rated children's credibility. Another major limitation of past credibility research is the use of a single-item scale to measure perceived credibility (Voogt et al., 2019). Accordingly, the current study used a comprehensive 23-item scale broken down into five subcomponents: accuracy, believability, cognitive competency, reliability and truthfulness (see Voogt, Klettke, & Mohebbi, 2017), to determine whether there were any differences across the underlying factors.

We predicted that children would be judged more credible when reporting single-instance, rather than repeated, physical abuse (see Connolly et al., 2008; Deck & Paterson, 2020). We further predicted that children reporting physical abuse after a delay would be deemed less credible than those reporting abuse swiftly. Due to mixed previous findings about the effects of question type and age on credibility ratings, no directional hypotheses were made regarding these factors, but they were included for exploratory purposes.

Method

Participants

We recruited 245 adults living in the United States through Amazon Mechanical Turk (MTurk) to act as mock jurors. Samples drawn from MTurk have been shown to be representative of Americans (Buhrmester et al., 2011). Interview transcripts ranged from 797 to 1468 words in length ($M_{single} = 839$ words and $M_{repeated} = 1378$ words), and reading times were recorded for all participants. Proficient readers were expected to read up to 600 words per minute (Hill, 1981). Upon inspection of recorded reading times, 37 people were removed from the sample due to recorded reading times faster than 600 words per minute.

The final sample consisted of 208 adults ($n_{male} = 130$; $n_{female} = 77$; $n_{other} = 1$). The sample were aged 19–70 years ($M = 36.26$, $SD = 11.00$). Most participants had attained a high-school-level education (44.7%) or graduated from a 4-year college degree (42.3%), while fewer participants had completed a two-year technical school programme (10.0%) or postgraduate college degree (17.0%). Participants' level of experience with children varied: 18.8% had no regular interaction with children, 22.6% had less than one interaction with young children per month, 24.5% had up to one interaction per week, and 34.1% had multiple interactions per week. Participants' gender, education level and level of experience with children did not affect credibility ratings, ps $\geq$ .17. There was no relationship between reading time and any of the dependent variables in the final sample, rs $\leq$ .04, ps $\geq$ .55. Nor was there any relationship between transcript length and any of the dependent variables, rs $\leq$ .14, ps $\geq$ .05.

Materials

Interview transcripts

Mock jurors read a simulated interview transcript of a young girl reporting physical abuse to a police interviewer, which was based upon a real forensic interview with names and

specific details changed for anonymity. The research team systematically manipulated the interview transcript to create 16 versions using a 2 (abuse frequency: single instance, repeated) × 2 (report timing: recent, delayed) × 2 (question type: open, closed) × 2 (child age: 6 or 10 years) design. The manipulations made to the original interview transcript were informed by research on children's response patterns (e.g. Krahenbuhl & Blades, 2006; Lamb et al., 2003) and the researchers' own experiences interviewing over 500 children and reviewing over 200 forensic child interview transcripts themselves for other projects. All transcripts were then also reviewed for authenticity by two police officers with over 20 years' experience investigating child abuse cases collectively. Two of the transcripts can be viewed here: https://osf.io/z4gr5/?view_only=b170b896f43a4869b3fcf2028ff2532f

All transcripts began identically, with the police interviewer introducing herself, providing a few conversational rules for the interview (i.e. ground rules; see Brubacher et al., 2015) and inviting the child to discuss the topic of concern. In all conditions, the child responded that she wanted to talk about her babysitter, Maggie, and the interviewer prompted the child for a narrative of the last occasion on which she spent time with Maggie. In all conditions, the child reported that she sustained an injury to her arm because Maggie pushed her into an external glass door. All transcripts concluded in the same manner: the interviewer asked for any details the child had omitted from the account that would be required for evidence (e.g. particularisation details), including an indication of when each incident occurred.

Abuse frequency manipulation. In the single-instance condition, the child only reported the one occasion of physical abuse. In the repeated condition, the child was prompted to also provide information about a second time that the abuse occurred. The child reported that on an earlier occasion Maggie had pushed her onto her bedroom floor, twisting her arm, and locked her in her bedroom for the night. The child's report included present-tense statements and script language throughout (e.g. 'She grabs me'; 'Maggie always swears'; Fivush, 1984; Hudson & Shapiro, 1991) and difficulties in particularising occurrences (e.g. 'This time was just like the other one I told you'). This language was absent from the single-instance event condition.

Reporting time manipulation. In the recent reporting condition, the child reported that the abusive incident occurred the day prior to the interview (i.e. 'It happened last night'). When abuse was repeated, the second incident was reported to have happened within one week of the interview. In the delayed reporting condition, the child reported that the incident(s) had occurred two years earlier. The child made the timing of the abuse clear during her interview, by reporting the timing during her narrative account of the abusive incident(s) as well as at the conclusion of the interview when the interviewer clarified evidential details.

Question type manipulation. In the open question condition, the interviewer used exclusively open questions to obtain a narrative of the alleged abuse incident(s) (e.g. 'Tell me more about the part where she pushed you', 'What happened next?'). In the closed question condition, the interview only used closed questions to elicit information from the child (i.e. yes/no, e.g. 'Was anyone else home?', and specific/directive questions, e.g. 'Why did she push you?'; see Snow et al., 2009, for a description of open and closed question types). In both question conditions, the key details of the abuse were kept consistent; however, the elaboration and coherence provided were kept in-line with the type of question posed. Open questions elicited elaborate and coherent responses that provided a narrative of events, while closed questions elicited short answers (often one or two words; Brown et al., 2013; Lamb et al., 2007).

In both question conditions, the interviewer used some facilitators (e.g. 'Mmm hmm') intermittently to further the child's account. Facilitators comprised less than 10% of the interviewer's utterances. No prompts in either question type condition were leading or suggestive.

Child age manipulation. The age manipulation was made salient by the child stating her age at the commencement of the interview. The child's responses throughout the interview were also modified in accordance with her age; the 6-year-old child gave shorter responses with less sophisticated language than the 10-year-old child. The same key story details were present in both age conditions.

Credibility Scale

The Child Sexual Assault Victim Credibility Scale (hereafter referred to as 'the Credibility Scale') was used to collect mock jurors' perceptions of the simulated interview transcripts (see Voogt, Klettke, & Mohebbi, 2017). Based on a review of factors used to measure credibility, Voogt, Klettke, and Thomson (2017) proposed that five unique constructs underlie children's perceived credibility: accuracy, believability, competency, reliability and truthfulness. The Credibility Scale is a comprehensive measure of all five factors underlying the perceived credibility of child victims. It comprises 23 items across five subscales, in line with each construct of credibility. Each item is rated on a 6-point Likert scale, from $1 = strongly\ disagree$ to $6 = strongly\ agree$, with higher scores indicating higher credibility ascribed to the victim.

Previously, the Credibility Scale was validated based on transcripts involving child sexual abuse. The scale demonstrated good internal consistency for the five unique subscales of credibility in child sexual abuse cases with Cronbach's alpha coefficients ranging between .78 and .92 (see Voogt, Klettke, & Mohebbi, 2017). Typically, Cronbach's alpha values between .70 and .95 are deemed

acceptable (George & Mallery, 2003; Nunnally & Bernstein, 1994). While initially validated in the context of a sexual abuse case, the aim was to test in a physical abuse case in the current study, as sexual abuse is not mentioned specifically in the scale. We also verified the appropriateness of the Credibility Scale for use with our child physical abuse materials by inspecting the Cronbach's alpha coefficients for each subscale. Coefficients for the five domains of credibility ranged from .86 to .95, indicating suitable reliability for each subscale with our sample.

To ensure the individual items within a subscale were reliably measuring the appropriate construct, we also considered the effects of deletion of each item on the subscale alpha coefficient. Deletion of three items would have improved an alpha coefficient negligibly (by .01 on each of the accuracy, reliability and truthfulness subscales). Deletion of all other items either had no effect, or negatively impacted alpha coefficients. Based upon these findings we decided that the Credibility Scale was appropriate for use with our physical abuse interviews and retained all items.

Procedure

The research was approved by Deakin University's Human Research Ethics Board. The study was advertised on MTurk to all MTurk subscribers over 18 years old. Participants were invited to act as mock jurors in an anonymous online survey about adults' perceptions of child witness credibility. Mock jurors provided informed consent after reading the plain language statement and acknowledging that they were over 18 years. Participants were randomly allocated to a version of the interview transcript. After reading the transcript, the mock jurors rated their perceptions of the child using the Credibility Scale, which was presented one subscale at a time. Finally, mock jurors were asked demographic questions, including their age and gender. Participants were paid USD$2.25 for

Table 1. Untransformed descriptive statistics for mock jurors' ratings.

			Range	
	M	*SD*	Min	Max
Overall credibility	4.63	0.72	2.30	6.00
Accuracy subscale	4.48	0.79	2.00	6.00
Believability subscale	4.77	0.82	2.00	6.00
Competency subscale	4.54	0.88	1.33	6.00
Reliability subscale	4.46	0.87	1.75	6.00
Truthfulness subscale	4.80	0.69	2.29	6.00

participating in the study, which took on average 6.67 minutes for them to complete.

Analysis

A series of multiple regressions were used to explore the effects of abuse frequency (single-instance abuse = 0, repeated abuse = 1), reporting time (delayed reporting = 0, recent reporting = 1), child age (6 years old = 0, 10 years old = 1) and question type (open = 0, closed = 1) on mock jurors' credibility ratings. We first considered the effects of these factors on overall credibility ratings. Then, we considered effects on each subscale of the Credibility Scale separately. A post hoc power analysis indicated that with a sample of 208 participants and four predictors, there was sufficient power (.84) to detect small–medium effect sizes of R^2 = .06 in a regression.

Results

Preliminary analyses revealed significant negative skew of the data. Data were reflected, and a log transformation was applied. In Table 1 we present untransformed descriptive statistics summarising the samples' credibility ratings for ease of interpretation. All other presented statistics pertain to the transformed data.

Relationships between the variables

Correlations between the variables were examined (see Table 2). Mock jurors' ratings on each subscale of the Credibility Scale, as well as their overall credibility rating, were all highly positively associated with each other. These associations were expected because the subscales and the overall credibility rating were all derived from the same measure (i.e. the Credibility Scale). Reporting time had a small positive relationship with mock jurors' ratings of overall credibility, and with ratings on each subscale, indicating that delayed reports had lower ratings than recent reports. Question type had a small negative relationship with overall credibility ratings, as well as with each subscale except truthfulness, indicating that closed question interviews had lower ratings than open question interviews. Abuse frequency was negatively related to ratings on the reliability subscale, indicating that repeated abuse reports were rated as slightly less reliable than single-instance abuse reports.

Predictors of credibility ratings

Six multiple regressions were conducted to determine associations of abuse frequency, reporting time, question type and child age with mock jurors' ratings of overall credibility and ratings on the five Credibility Scale subscales. Bonferroni adjustments were applied to the regressions for the five subscales. There were no significant concerns with collinearity, with variance inflation factor (VIF) values ranging from 1.00 to 1.01. Unstandardised coefficient statistics are presented in Table 3 for all regression models.

Table 2.   Pearson correlations (two-tailed) between variables.

	2	3	4	5	6	7	8	9	10
1. Abuse frequency[a]	.03	−.02	−.07	−.12	−.10	−.09	−.09	−.16*	−.12
2. Reporting timing[b]		.01	.04	.25**	.24**	.23**	.29**	.24**	.18*
3. Child age[c]			−.03	−.08	−.08	−.08	−.05	−.06	−.09
4. Question type[d]				−.20**	−.22**	−.16*	−.23**	−.19**	−.14
5. Overall credibility rating					.92**	.93**	.88**	.94**	.91**
6. Accuracy subscale						.82**	.77**	.82**	.75**
7. Believability subscale							.80**	.83**	.82**
8. Competency subscale								.82**	.73**
9. Reliability subscale									.82**
10. Truthfulness subscale									

[a]Abuse frequency was coded as single-instance abuse = 0, repeated abuse = 1. [b]Reporting timing was coded as delayed reporting = 0, recent reporting = 1. [c]Child age was coded as 6 years old = 0, 10 years old = 1. [d]Question type was coded as open questioning = 0, closed questioning = 1.
*$p < .05$. **$p < .01$.

Table 3.   Unstandardised regression coefficients.

	Overall credibility		Subscales									
			Accuracy		Believability		Competency		Reliability		Truthfulness	
	B	SE	B	SE	B	SE	B	SE	B	SE	B	SE
Abuse frequency	−0.04	0.02	−0.04	0.02	−0.04	0.02	−0.04	0.02	−0.06	0.02	−0.04	0.02
Reporting timing	0.07	0.02	0.07	0.02	0.08	0.02	0.10	0.02	0.08	0.02	0.05	0.02
Child age	−0.03	0.02	−0.03	0.02	−0.03	0.02	−0.02	0.02	−0.02	0.02	−0.03	0.02
Question type	−0.06	0.02	−0.07	0.02	−0.06	0.02	−0.08	0.02	−0.07	0.02	−0.04	0.02

Overall credibility

The regression model for overall credibility was significant, $F(4, 203) = 8.26$, $p < .001$, and accounted for 14.0% (adjusted) of the variance in children's overall perceived credibility. Abuse frequency, reporting time and question type were significant predictors (frequency: $\beta = −.15$, $t = −2.35$, $p = .02$; reporting: $\beta = .27$, $t = 4.06$, $p < .001$; question type: $\beta = −.23$, $t = −3.52$, $p = .001$). Credibility ratings were higher when single-instance abuse was reported (rather than repeated abuse), when the report was made recently after the abuse (rather than after a delay) and when the interviewer elicited a narrative with open questioning (rather than closed). Child age was not a significant predictor, $p = 15$.

Accuracy subscale

The accuracy subscale regression was significant, $F(4, 203) = 8.04$, $p < .001$, and accounted for 12.0% (adjusted) of the variance in accuracy ratings. Abuse frequency, reporting time and question type were all significant predictors in the model (frequency: $\beta = −.14$, $t = −2.09$, $p = .04$; reporting: $\beta = .26$, $t = 3.86$, $p < .001$; question type: $\beta = −.25$, $t = −3.78$, $p < .001$). Accuracy ratings were higher when abuse was a single instance, the report was made recently after the abuse, and open questions were used to prompt the child. Child age was not significant, $p = .18$.

Believability subscale

The believability subscale regression model was significant, $F(4, 203) = 5.79$, $p < .001$,

and explained 8.5% (adjusted) of the variance in believability ratings. Reporting time predicted believability ratings, as did question type (reporting: $\beta = .24$, $t = 3.62$, $p < .001$; question type $\beta = -.18$, $t = -2.71$, $p = .007$). Believability ratings were higher when the report was made recently after the abuse, and open questions were used. Abuse frequency and child age were not significant, $ps \geq .09$.

Competency subscale

The model for mock jurors' competency ratings was significant, $F(4, 203) = 9.79$, $p < .001$, and accounted for 14.5% (adjusted) of the variance. Reporting time and question type significantly contributed to the model (reporting: $\beta = .31$, $t = 4.77$, $p < .001$; question type: $\beta = -.25$, $t = -3.90$, $p < .001$). Competency ratings were higher when the report was made recently after the abuse, and open questions were used to elicit a narrative from the child. Abuse frequency and child age were not significant, $ps \geq .06$.

Reliability subscale

The regression for the reliability subscale was significant, $F(4, 203) = 8.27$, $p < .001$, and accounted for 12.3% (adjusted) of the variance. Abuse frequency, reporting time and question type were significant predictors (frequency: $\beta = -.19$, $t = -2.89$, $p < .01$; reporting: $\beta = .26$, $t = 3.95$, $p < .001$; question type: $\beta = -.22$, $t = -3.40$, $p = .001$). Credibility ratings were higher for single-instance abuse, reports recently after the abuse and accounts elicited with open questions. Child age was again not a significant predictor, $p = .28$.

Truthfulness subscale

For the truthfulness subscale, the model was significant, $F(4, 203) = 4.42$, $p = .002$, and accounted for 6.2% (adjusted) of the variance. Abuse frequency, reporting time and question type were all significant predictors in the model (frequency: $\beta = -.14$, $t = -2.01$, $p =$ .04; reporting: $\beta = .19$, $t = 2.80$, $p = .006$; question type: $\beta = -.18$, $t = -2.32$, $p = .02$). Truthfulness ratings were higher when abuse was a single instance, the report was made recently after the abuse, and open questions were used to prompt the child. Child age was not significant, $p = .13$.

Discussion

The current study examined adults' evaluations of mock police interviews with children reporting physical abuse. As predicted, mock jurors found a child's report of physical abuse more credible when it described single instance abuse (rather than repeated abuse) and when the child reported abuse that had occurred recently (rather than two years earlier). We also found that mock jurors rated the child's account as more credible when open questions were used to prompt the child (rather than closed questions). We did not find an effect of child age on any dependent variable. Each finding is discussed in more detail below.

Abuse frequency and reporting delays

In the current study, reports of repeated physical abuse were considered less credible than reports of single-instance abuse. This same trend has been shown in a handful of previous studies investigating perceptions of children (Connolly et al., 2008) or adults (Deck & Paterson, 2020; Weinsheimer et al., 2017) reporting on innocuous activity session(s). The tendency for repeated event reports to be considered less credible than single-instance event reports may be due to the language used to report repeated events. After experiencing an event repeatedly, even young children develop general representations for the event in memory which alter how the event is reported (Fivush, 1984; Hudson & Shapiro, 1991). Statements in the present tense (e.g. 'She grabs me') and general temporal language (e.g. 'We always go to Sarah's house after school', 'We usually play games') are common in repeated

event reports, and differentiating separate incidents becomes difficult (e.g. 'This time was just like the other one I told you'; see Fivush, 1984; Hudson & Nelson, 1986; Roberts, 2002). This language was included in repeated abuse materials in the current study and may have conveyed that the child was not confident in her recall of the reported incidents, making mock jurors dubious over the accuracy of her memory. Indeed, lower perceived confidence has been shown to reduce credibility ratings of adults' repeated event reports previously (Weinsheimer et al., 2017). However, in the only study to measure children's language when reporting on repeated events, Connolly and colleagues (2008) did not find a relationship between credibility ratings and children's tense and pronoun use. Given children's difficulty in reporting some details about an incident from a repeated event, further exploration of the credibility of elements of children's language is needed in future.

An alternative explanation for our finding is that when assessing children's credibility, participants relied upon their own knowledge about event memory relative to the child's frequency of exposure to the event. Repeated experience with an event improves recall of details presented similarly each time (Powell & Thomson, 1996), and even children have demonstrated metamemorial knowledge of this process (Roberts & Powell, 2005). Knowledge regarding this benefit of repeated experience on memory may have affected adults' expectations of children's memory reports in the current study, where the child was expected to have a stronger memory of the abuse in the repeated than the single-instance condition. In the current materials, the first incident described in the repeated abuse condition transcripts was the same incident as that described in the single incident condition transcripts. Then, the second incident in the repeated abuse condition included the same number of key story details as did the first incident. The level of detail included per incident may have seemed enough for a child to

form a credible memory report of an incident experienced once, so the child was rated highly on the credibility scale. But participants may have expected more detail to be recalled per incident about an event experienced repeatedly, so the child was perceived as less credible in the repeated condition. Future work should consider measuring or manipulating the salience of metamemory knowledge such as the benefits (or the challenges) of repeated experience on memory to measure any impacts metamemorial knowledge has on the perceptions of repeated event reports.

In the current study, reports were deemed less credible when they were reported after a two-year delay rather than when abuse was reported immediately. This finding mirrors previous credibility assessments of children reporting sexual abuse, which show a general trend for longer reporting delays to be deemed increasingly less credible (e.g. Golding et al., 1995; Pozzulo et al., 2010; Sugarman & Boney-McCoy, 1997). Laypeople believe delayed reporting of abuse is uncommon for children (Shackel, 2008), likely reducing the perceived credibility of a delayed report. Mock jurors may also understand that memory accuracy and completeness can reduce over time, further negatively impacting the credibility of delayed reports. However, research on long-term memory has shown that salient events like abuse can remain accurate over time (Baugerud et al., 2014; Ceci & Bruck, 1993), and that false complaints from children are rare, even after a four-year delay in reporting the abuse (Goodman et al., 2002; Peterson, 2011). Educating jurors, police, lawyers and judges who evaluate children's testimony about the nature of memory regarding the commonality of delayed reports continues to remain important.

Question type and age

Mock jurors in the current study perceived the child's report to be more credible when open questions were used to prompt the child, rather than closed questions. Our mock transcripts

were designed to mimic the typical responses that children provide to open and closed questions. While the core event details were kept constant in each transcript, in the open question condition the child responded with greater verbosity, elaboration and narrative detail than the child in the closed question condition (see Larsson & Lamb, 2009). The relatively high level of detail about the abuse provided in the open question condition may have led the mock jurors to perceive that the child was convincingly recalling experienced abuse. Interviewer guidelines highly recommend the use of open questions with children (e.g. Lamb et al., 2007; Powell et al., 2005), and our results further support this recommendation.

Previous work using similar methodologies to those in the current study have found few differences in the perceived credibility of reports elicited with open versus closed questions (Ruva & Bryant, 2004; Tubb et al., 1999). Only Ruva and Bryant (2004) found that children were rated as more credible when answering open than closed questions, but this difference was only seen for the 6-year-old child in their study (no significant difference was seen for the 10-year-old or adult victims). Given that the interviewer's questioning style is a relatively easy factor to control in an interview, further exploration of the impact of question type on children's credibility is warranted.

We did not find an effect of child age on any dependent variable. It is possible that the age manipulation in the current experiment was not salient enough, or that manipulating the child's age as 6 years old versus 10 years old was not a large enough gap to cause any difference in credibility ratings. In sexual abuse cases, children under 12 are generally considered more credible than adults (Bottoms et al., 2007; Davies & Rogers, 2009; Nightingale, 1993). In the only prior examination of victim age in physical abuse cases, Sheahan and colleagues (2021) found that a 12-year-old victim was considered more credible than an adult victim. Physical and sexual abuse cases share many qualities; for example, both are typically perpetrated by a familiar person, both often lack physical evidence, both are likely to have delayed disclosures to an adult (Rush et al., 2014). Perhaps it is these qualities of interpersonal crimes that influence credibility ratings of children under 12. That is, young children are perhaps considered to be rather honest about victimisation committed by familiar adults.

Limitations and conclusions

There are a couple of limitations to the present study. First, participants were exposed to stimuli created for research purposes, not actual accounts from children. Although the transcripts were based on a true forensic interview, the independent variables were each separately manipulated by the research team. Interactions between independent variables were not considered, which may limit external validity. Further, the accuracy of the child's statement could not be compared to adult's perceptions of the child. Second, participants' credibility assessments during the experiment may not necessarily be an authentic comparison to assessments made during a high-stakes courtroom case. However, the current findings remain important and relevant because even small influences on credibility perceptions can contribute to a juror's reasonable doubt over allegations.

When children first report abuse to police in an investigative interview, that interview is often used as the child's evidence-in-chief should the case progress to court [e.g. Australian Law Reform Commission, 2010; Ministry of Justice (United Kingdom), 2011]. We found that two factors outside of interviewers' control could affect children's perceived credibility: credibility ratings were lower for children reporting repeated (rather than single) abuse and reporting after a delay (rather than reporting promptly). We suggest that future work should consider adults' metamemorial knowledge to examine any relationships with perceived credibility.

Importantly, the current study also demonstrated that the questioning police use in the investigative interview can impact mock jurors' perceptions of the child. Recommendations for child interviewers emphasise open questions as a particularly useful interviewing tool for eliciting accurate accounts of both single-instance and repeated abuse [Brubacher et al., 2014; Ministry of Justice (United Kingdom), 2011]. Current findings support these recommendations and extend the expected benefits of open questions to potentially advantage the child's credibility.

Ethical standards

Declaration of conflicts of interest

Meaghan C. Danby has declared no conflicts of interest
Stefanie J. Sharman has declared no conflicts of interest
Bianca Klettke has declared no conflicts of interest

Ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the Deakin University research ethics committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

Informed consent was obtained from all individual participants included in the study.

Disclosure of benefit or interest statement

The authors are not aware of any possible conflicts of interest, financial benefits, sources of financial support, corporate involvement, patent holdings, etc.

ORCID

Meaghan C. Danby http://orcid.org/0000-0001-7060-7121

Stefanie J. Sharman http://orcid.org/0000-0002-0635-047X
Bianca Klettke http://orcid.org/0000-0003-4602-2435

References

Australian Bureau of Statistics. (2019). *Characteristics and outcomes of childhood abuse*, (4906.0 - Personal Safety, Australia, 2016). https://www.abs.gov.au/ausstats/abs@.nsf/Lookup/by%20Subject/4906.0~2016~Main%20Features~Characteristics%20and%20Outcomes%20of%20Childhood%20Abuse%20(Feature%20Article)%20~30

Australian Law Reform Commission. (2010). *Family violence — a national legal response (ALRC Final Report No. 114)*. Commonwealth of Australia.

Baugerud, G. A., Magnussen, S., & Melinder, A. (2014). High accuracy but low consistency in children's long-term recall of a real-life stressful event. *Journal of Experimental Child Psychology*, *126*, 357–368. https://doi.org/10.1016/j.jecp.2014.05.009

Bornstein, B. H., Kaplan, D. L., & Perry, A. R. (2007). Child abuse in the eyes of the beholder: Lay perceptions of child sexual and physical abuse. *Child Abuse & Neglect*, *31*(4), 375–391. https://doi.org/10.1016/j.chiabu.2006.09.007

Bottoms, B. L., Golding, J. M., Stevenson, M. C., Wiley, T. R. A., & Yozwiak, J. A. (2007). A review of factors affecting jurors' decisions in child sexual abuse cases. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *The handbook of eyewitness psychology, Vol. 1. Memory for events* (pp. 509–543). Lawrence Erlbaum Associates Publisher.

Brown, D. A., Lamb, M. E., Lewis, C., Pipe, M., Orbach, Y., & Wolfman, M. (2013). The NICHD investigative interview protocol: An analogue study. *Journal of Experimental Psychology: Applied*, *19*, 367–382. https://doi.org/10.1037/a0035143

Brubacher, S. P., Poole, D. A., & Dickinson, J. J. (2015). The use of ground rules in investigative interviews with children: A synthesis and call for research. *Developmental Review*, *36*, 15–33. https://doi.org/10.1016/j.dr.2015.01.001

Brubacher, S. P., Powell, M. B., & Roberts, K. P. (2014). Recommendations for interviewing children about repeated experiences.

Psychology, Public Policy, and Law, *20*(3), 325–335. https://doi.org/10.1037/law0000011

Buhrmester, M., Kwang, T., & Gosling, S. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data. *Perspectives on Psychological Science*, *6*(1), 3–5. https://doi.org/10.1037/e527772014-223

Ceci, S. J., & Bruck, M. (1993). Suggestibility of the child witness: A historical review and synthesis. *Psychological Bulletin*, *113*(3), 403–439. https://doi.org/10.1037/0033-2909.113.3.403

Connolly, D. A., Price, H. L., Lavoie, J. A. A., & Gordon, H. M. (2008). Perceptions and predictors of children's credibility of a unique event and an instance of a repeated event. *Law and Human Behavior*, *32*(1), 92–112. https://doi.org/10.1007/s10979-006-9083-3

Davies, M., & Rogers, P. (2009). Perceptions of blame and credibility toward victims of childhood sexual abuse: Differences across victim age, victim-perpetrator relationship, and respondent gender in a depicted case. *Journal of Child Sexual Abuse*, *18*(1), 78–92. https://doi.org/10.1080/10538710802584668

Deck, S. L., & Paterson, H. M. (2020). Liars are perceived as more credible than truth-tellers who recall a repeated event. *Applied Cognitive Psychology*, *34*(3), 643–653. https://doi.org/10.1002/acp.3665

Dent, H. R., & Stephenson, G. M. (1979). An experimental study on the effectiveness of different techniques of questioning child witnesses. *British Journal of Social and Clinical Psychology*, *18*(1), 41–51. https://doi.org/10.1111/j.2044-8260.1979.tb00302.x

Fivush, R. (1984). Learning about school: The development of kindergartners' school scripts. *Child Development*, *55*(5), 1697–1709. https://doi.org/10.2307/1129917

Foynes, M. M., Freyd, J. J., & DePrince, A. P. (2009). Child abuse: Betrayal and reporting. *Child Abuse & Neglect*, *33*(4), 209–217. https://doi.org/10.1016/j.chiabu.2008.11.001

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference 11.0 update* (4th ed.). Allyn & Bacon.

Goldfarb, D., Goodman, G. S., Larson, R. P., Eisen, M. L., & Qin, J. (2019). Long-term memory in adults exposed to childhood violence: Remembering genital contact nearly 20 years later. *Clinical Psychological Science*, *7*(2), 381–396. https://doi.org/10.1177/2167702618805742

Golding, J. M., Sego, S. A., Sanchez, R. P., & Hasemann, D. (1995). Believability of repressed memories vignettes. *Law and Human Behavior*, *19*(6), 569–592. https://doi.org/10.1037/t38113-000

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. https://doi.org/10.1016/s0145-2134(02)00354-x

Goodman, G. S., Goldfarb, D., Quas, J. A., & Lyon, A. (2017). Psychological counseling and accuracy of memory for child sexual abuse. *American Psychologist*, *72*(9), 920–931. https://doi.org/10.1037/amp0000282

Goodman, G. S., Quas, J. A., Goldfarb, D., Gonzalves, L., & Gonzalez, A. (2019). Trauma and long-term memory for childhood events: Impact matters. *Child Development Perspectives*, *13*(1), 3–9. https://doi.org/10.1111/cdep.12307

Hartley, D. J., Mullings, J. L., & Marquart, J. W. (2013). Factors impacting prosecution of child sexual abuse, physical abuse, and neglect cases processed through a children's advocacy center. *Journal of Child & Adolescent Trauma*, *6*(4), 260–273. https://doi.org/10.1080/19361521.2013.836586

Hatton, V. A., & Duff, S. (2016). A systematic literature review: Which factors influence English and Welsh perceptions towards victims of child sexual abuse? *Journal of Sexual Aggression*, *22*(3), 275–290. https://doi.org/10.1080/13552600.2016.1191683

Hershkowitz, I. (2006). Delayed disclosure of alleged child abuse victims in Israel. *American Journal of Orthopsychiatry*, *76*(4), 444–450. https://doi.org/10.1037/0002-9432.76.4.444

Hill, J. K. (1981). Effective reading in a foreign language. *ELT Journal*, *XXXV*(3), 270–281. https://doi.org/10.1093/elt/xxxv270

Hudson, J., & Nelson, K. (1986). Repeated encounters of a similar kind: Effects of familiarity on children's autobiographic memory. *Cognitive Development*, *1*(3), 253–271. https://doi.org/10.1016/s0885-2014(86)80004-1

Hudson, J. A., & Shapiro, L. R. (1991). From knowing to telling: The development of children's scripts, stories, and personal narratives. In A. McCabe & C. Peterson (Eds.), *Developing narrative structure* (pp. 89–136). APA.

Krahenbuhl, S., & Blades, M. (2006). The effect of interviewing techniques on young children's responses to questions. *Child: Care, Health*

and Development, *32*, 321–331. https://doi.org/10.1111/j.1365-2214.2006.00608.x

Lamb, M. E., Orbach, Y., Warren, A. R., Esplin P. W., & Hershkowitz, I. (2007). Enhancing performance: Factors affecting the informativeness of you witnesses. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *The handbook of eyewitness psychology: Volume 1: Memory for events* (pp. 429–452). Erlbarum.

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. https://doi.org/10.1037/0022-006x.71.5.926

Larsson, A. S., & Lamb, M. E. (2009). Making the most of information-gathering interviews with children. *Infant and Child Development*, *18*(1), 1–16. https://doi.org/10.1002/ICD.573

Ministry of Justice (United Kingdom). (2011). *Achieving best evidence in criminal proceedings: Guidance on interviewing victims and witnesses, and guidance on using special measures.* http://webarchive.nationalarchives.gov.uk/20130128112038/http://www.justice.gov.uk/downloads/victims-and-witnesses/vulnerable-witnesses/achieving-best-evidence-criminal-proceedings.pdf

Nightingale, N. N. (1993). Juror reactions to child victim witnesses: Factors affecting trial outcome. *Law and Human Behavior*, *17*(6), 679–694. https://doi.org/10.1007/bf01044689

Nunnally, J., & Bernstein, L. (1994). *Psychometric theory*. McGraw-Hill Higher, INC.

Peterson, C. (2011). Children's memory reports over time: Getting both better and worse. *Journal of Experimental Child Psychology*, *109*(3), 275–293. https://doi.org/10.1016/j.jecp.2011.01.009

Powell, M. B., Fisher, R. P., & Wright, R. (2005). Investigative interviewing. In N. Brewer & K. D. Williams (Eds.), *Psychology and law an empirical perspective* (pp. 11–42). The Guilford Press.

Powell, M. B., & Thomson, D. M. (1996). Children's memory of an occurrence of a repeated event: Effects of age, repetition, and retention interval across three question types. *Child Development*, *67*(5), 1988–2004. https://doi.org/10.2307/1131605

Pozzulo, J. D., Dempsey, J. L., & Crescini, C. (2010). Factors affecting juror decisions in historic child sexual abuse cases involving continuous memories. *Criminal Justice and Behavior*, *37*(9), 951–964. https://doi.org/10.1177/0093854810373587

Roberts, K. P. (2002). Children's ability to distinguish between memories from multiple sources: Implications for the quality and accuracy of eyewitness statements. *Developmental Review*, *22*(3), 403–435. https://doi.org/10.1016/s0273-2297(02)00005-9

Roberts, K. P., & Powell, M. B. (2005). Evidence of metacognitive awareness in young children who have experienced a repeated event. *Applied Cognitive Psychology*, *19*(8), 1019–1031. https://doi.org/10.1002/acp.1145

Rush, E. B., Lyon, T. D., Ahern, E. C., & Quas, J. A. (2014). Reporting suspicion bias and abuse reporting. *Child Maltreatment*, *19*(2), 113–118. https://doi.org/10.1177/1077559514538114

Ruva, C. L., & Bryant, J. B. (2004). The impact of age, speech style, and question form on perceptions of witness credibility and trial outcome. *Journal of Applied Social Psychology*, *34*(9), 1919–1944. https://doi.org/10.1111/j.1559-1816.2004.tb02593

Shackel, R. L. (2008). The beliefs commonly held by adults about children's behavioral responses to sexual victimization. *Child Abuse & Neglect*, *32*(4), 485–495. https://doi.org/10.1016/j.chiabu.2007.04.016

Sheahan, C. L., Pica, E., & Pozzulo, J. D. (2021). Abuse is abuse: The influence of type of abuse, victim age, and defendant age on juror decision making. *Journal of Interpersonal Violence, 36*, 938–356. https://doi.org/10.1177/0886260517731316

Snow, P. C., Powell, M. B., & Murfett, R. (2009). Getting the story from child witnesses: Exploring the application of a story grammar framework. *Psychology, Crime and Law*, *15*(6), 555–568. https://doi.org/10.1080/10683160802409347

Sternberg, K. J., Lamb, M. E., Davies, G. M., & Westcott, H. L. (2001). The Memorandum of Good Practice: Theory versus application. *Child Abuse & Neglect*, *25*(5), 669–681. https://doi.org/10.1016/S0145-2134(01)00232-0

Sugarman, D. B., & Boney-McCoy, S. (1997). Impact of expert testimony on the believability of repressed memories. *Violence and Victims*, *12*(2), 115–126.

Thoresen, C., L⊘Nnum, K., Melinder, A., Stridbeck, U., & Magnussen, S. (2006). Theory and practice in interviewing young children: A study of Norwegian police interviews 1985–2002. *Psychology, Crime &*

Law, *12*(6), 629–640. https://doi.org/10.1080/10683160500350546

Trocome, N., Fallon, B., MacLaurin, B., Sinha, V., Black, T., Fast, E., & Holroyd, J. (2010). Chapter 1: Introduction. In *Canadian incidence of reported child abuse and neglect – 2008: Major findings* (pp. 7–11). Public Health Agency of Canada.

Tubb, A., Wood, J. M., & Hosch, H. M. (1999). Effects of suggestive interviewing and indirect evidence on child credibility in a sexual abuse case. *Journal of Applied Social Psychology*, *29*(6), 1111–1127. https://doi.org/10.1111/j.1559-1816.1999.tb02031.x

Voogt, A., Klettke, B., & Crossman, A. (2019). Measurement of victim credibility in child sexual assault cases: A systematic review. *Trauma, Violence & Abuse*, *20*(1), 51–66. https://doi.org/10.1177/1524838016683460

Voogt, A., Klettke, B., & Mohebbi, M. (2017). The development and validation of the Child Sexual Assault Victim Credibility Scale: An instrument to measure laypersons' perceptions of victim credibility. *Journal of*

Interpersonal Violence, *36*, 1–21. https://doi.org/10.1177/0886260517737554

Voogt, A., Klettke, B., & Thomson, D. (2017). The development of a conceptual model of perceived victim credibility in child sexual assault cases. *Psychiatry, Psychology and Law*, *24*, 760–769. https://doi.org/10.1080/13218719.2017.1315764

Walsh, W. A., Jones, L. M., Cross, T. P., & Lippert, T. (2010). Prosecuting child sexual abuse: The importance of evidence type. *Crime & Delinquency*, *56*(3), 436–454. https://doi.org/10.1177/0011128708320484

Weinsheimer, C. C., Coburn, P. I., Chong, K., MacLean, C. L., & Connolly, D. A. (2017). Perceptions of credibility for a memory report of a single versus repeated event. *Applied Cognitive Psychology*, *31*(4), 414–423. https://doi.org/10.1002/acp.3340

Westcott, H. L., & Kynan, S. (2006). Interviewer practice in investigative interviews for suspected child sexual abuse. *Psychology, Crime and Law*, *12*(4), 367–382. https://doi.org/10.1080/10683160500036962