



Tracking labels for occurrences of alleged child abuse from police interviews to trials

Meaghan C. Danby^{1‡}, Becky Earhart^{2*,‡} , Sonja P. Brubacher²,
 Martine B. Powell², Jane Goodman-Delahunty³ and
 Nina J. Westera^{2†}

¹School of Psychology, Deakin University, Geelong, Australia

²Centre for Investigative Interviewing, Griffith Criminology Institute, Griffith University, Brisbane, Australia

³School of Psychology, Charles Sturt University, Manly, Australia

Purpose. Labelling (i.e., naming) individual occurrences of repeated abuse allegations with explicit and consistent terms may improve children's reporting of these offences. The aim of the present study was to track labels for occurrences of alleged child sexual abuse from the police interview to court proceedings.

Methods. We examined the labels used in the police interviews and trials of 23 child complainants (5–15 years old at interview). The initiator of each label (child, interviewer, lawyer, or judge), stage of the process in which the label was generated, and the type of information used to label specific occurrences of abuse were recorded. Any subsequent reuse or replacement of the labels was also recorded.

Results. Most labels were created by police interviewers. Few children generated labels. Most occurrences of abuse were labelled early in the legal process; 82% were first labelled either in the police interview or in the prosecution's opening statement. The labels were frequently replaced with alternate terms, with an average of three different labels for the same incident. After original labels were established for occurrences, they were just as likely to be replaced as they were to be reused. The most frequently observed label replacement was by defence lawyers during cross-examination.

Conclusions. Labels were used inconsistently throughout the police interview and trial. To give children the best chance of describing specific occurrences of abuse during legal proceedings, labels should be created from children's words wherever possible and used consistently thereafter by all justice professionals.

When children make allegations of abuse, they may be formally interviewed by police and eventually provide testimony in a trial. Notwithstanding the potential socio-motivational barriers associated with the testimonial process (Collin-Vézina, De La Sablonnière-Griffin, Palmer, & Milne, 2015; Parkinson, Shrimpton, Swanston, O'Toole, & Oates, 2002), children's developing cognitive skills can also introduce challenges into the elicitation of

*Correspondence should be addressed to Becky Earhart, Centre for Investigative Interviewing, Griffith Criminology Institute, Griffith University, Mt Gravatt Campus, 176 Messines Ridge Road, Mt Gravatt, QLD, 4122, Australia (email: r.earhart@griffith.edu.au).

†Deceased May 25, 2017.

‡Meaghan C. Danby and Becky Earhart contributed equally to the research and the paper.

their accounts. In child sexual abuse (CSA) cases, allegations often pertain to repeated abuse (e.g., Trocmé *et al.*, 2010). In several jurisdictions, children are required to provide some specific details (e.g., time, place, abusive acts, context) about individual occurrences in order for appropriate charges to be laid (i.e., particularization; Powell, Roberts, & Guadagno, 2007). Although laws in many places have been enacted to soften this requirement (i.e., by the creation of *persistent* or *continued* abuse charges), certain specific details are nonetheless required to prove two or more acts of abuse for these types of charges, which focus on an unlawful sexual relationship (e.g., Section 66EA of the NSW proposed *Criminal Legislation Amendment [Child Sexual Abuse] Act 2018 No 33* requires ‘two or more unlawful sexual acts with or towards a child over any period’ and s 229B of the *Criminal Code [Qld]* requires proof of three or more distinct occasions). Even where particularization is not mandatory, research suggests that the credibility of children’s testimony is enhanced when episodic details are provided (Burrows & Powell, 2013; Smith & Milne, 2011).

Cognitively, the task of providing details about individual occurrences of a repeated event is difficult for children, who (like adults) form strong memories for the generic representation (what usually happens) but become confused about the details that vary across occurrences (Connolly & Lindsay, 2001; Goodman-Delahunty, Nolan, & van Gijn-Grosvenor, 2017; Powell, Roberts, Ceci, & Hembrooke, 1999; Powell & Thomson, 1996; see Brubacher, Powell, & Roberts, 2014, for a review). Following repeated experiences, the most common error children make is to misattribute details from one occurrence to another (e.g., saying that digital penetration occurred the night the blue pyjama set was worn when it was actually the night the spotted nightshirt was worn). Young children have particular difficulty distinguishing between occurrences (Powell & Thomson, 1996; Powell *et al.*, 1999). According to the Source-Monitoring Framework (Johnson, Hashtroudi, & Lindsay, 1993), these errors occur due to inaccurate decision-making about the origin of a memory during retrieval. Memories for occurrences from a repeated event often share a high proportion of qualitative properties (e.g., abuse may always occur in the same location, with the same person, and make the victim feel the same way), causing particular difficulty in distinguishing between sources (occurrences) of recalled details (Lindsay, Allen, Chan, & Dahl, 2004; Roberts & Blades, 1999).

An alternative explanation to the Source-Monitoring Framework is fuzzy-trace theory (Brainerd & Reyna, 1990), which suggests the confusion of details across occurrences is due to faulty retrieval of the memory trace for the to-be-remembered occurrence. After repeated experience with an event, children construct a memory trace that contains the common elements or themes (e.g., abuse typically takes place when the child is in her pyjamas in bed; called a ‘gist trace’) in addition to specific episodic traces that contain the individual experiences at the event (e.g., the specific pyjamas worn each time; called ‘verbatim traces’). Memory errors can occur when a retrieved detail is gist-consistent but assigned to the wrong occurrence because the episodic trace has decayed, or the episodic trace for an alternative episode is incorrectly retrieved instead (Roberts, 2002).

Children’s difficulty in accurately particularizing occurrences of abuse may be more apparent when they are subjected to a challenging cross-examination (Zajac & Hayne, 2003). Therefore, strategies that minimize possible confusions when children give their accounts are of interest to memory researchers and legal practitioners alike. Clear labelling of individual occurrences with explicit, consistent, and (ideally) unique terms (e.g., ‘the first time’, ‘the time my sister was there’) is one technique expected to minimize confusions (Brubacher *et al.*, 2014).

Using labels to assist children's reports

Adults who question children may help them to more accurately recount an occurrence of abuse by providing them with appropriate support. Labelling occurrences is expected to support children's retrieval of the correct episodic memory trace by providing a retrieval cue, and minimize source-monitoring errors by clarifying which occurrence is being discussed. Previous laboratory work found that labels including contextual and temporal cues for an occurrence of a repeated classroom event (e.g., 'Tell me everything about *the last time* when you wore the different badge') assisted children to report more accurate details than labels using temporal cues alone (e.g., 'Tell me everything about *the last time*'; Pearse, Powell, & Thomson, 2003). The authors proposed that labels with contextual cues assisted children to mentally reinstate the context of the occurrence (e.g., picture it in their mind's eye) to facilitate the recall of additional details (Dietze & Thomson, 1993; Fisher & Geiselman, 1992).

For a label to act as an effective retrieval cue, it should align with how the to-be-recalled occurrence is encoded and stored in memory (Tulving & Thomson, 1973). Recommendations for forensic interviewers (Brubacher *et al.*, 2014; Powell & McMeeken, 1998) advise interviewers to adopt children's own labels for occurrences, or use children's words to create labels for future references to the occurrences, because children's words are expected to reveal the way their cognitive sources are organized and the specific episodic details they remember the best. Occurrences of repeated events may be organized in memory in numerous ways, such as along temporal dimensions (e.g., the first time, the last time), by location (e.g., the time in the kitchen, the time in the bedroom), or by salient abuse detail (e.g., the time he went under my clothes, the time he tried to kiss me). Verbalizing these categories as labels makes them explicit so as to minimize confusions about which occurrence is being recalled. Further, since the interviewer is naïve about the event, children are more likely than interviewers to select a unique label for an occurrence (Brubacher *et al.*, 2014).

Even when interviewers use children's labels, children may fail to report repeated events accurately. Nevertheless, labelling is an interviewing technique under the control of the questioner and reminds both parties of the occurrence under discussion (Orbach & Pipe, 2011). By using explicit and consistent labels based on the child's words, interviewers can reduce the potential for misunderstandings and increase the chance of children providing accurate testimony.

Despite these recommendations, forensic interviewers have been shown to ignore or replace children's labels in practice, a behaviour which was associated with children being less responsive to interviewer prompts (Brubacher, Malloy, Lamb, & Roberts, 2013). Brubacher and colleagues conducted the first examination of how labels for occurrences of sexual abuse are used in practice, describing the types of labels that arose in police interviews conducted in the United Kingdom with 5- to 13-year-old children alleging repeated abuse. The aim of the present study was to expand that research by tracking how labels are used during the police interview by the child and interviewer, as well as how they are later used by all parties in court (prosecutors, defence lawyers, and judges). The present study is the first to examine how labels are used across different stages of the criminal justice process.

Present study

Overall, the broad aim of the present study was to explore how often children generated their own labels for occurrences, and to what degree the labels remained consistent

throughout the investigation and legal proceedings, in order to identify qualitative patterns that could form the basis of recommendations for improvement. Based on the finding by Brubacher *et al.* (2013) that children and police interviewers created roughly equal numbers of labels during the forensic interview, we expected that children in the current sample would generate about half of the labels used in police interviews. We hypothesized that labels would frequently be replaced with other terms for the same occurrence at trial, as a high proportion of children's labels were replaced or ignored in previous research (Brubacher *et al.*, 2013). We conducted exploratory descriptive analyses to examine the type of information that was used to label events and tracked the longevity of labels throughout the criminal justice process.

Method

The study sample

Prosecutors' files and trial transcripts from a total of 154 cases involving allegations of sexual abuse that came to trial between 2011 and 2015 in Australia were received as part of a larger study for the Royal Commission into Institutional Responses to Child Sexual Abuse (Powell, Westera, Goodman-Delahunty, & Pichler, 2016). Ethics approval for the production of these files for research purposes was granted by the Human Research Ethics Committees of the collaborating universities. All files were de-identified. From this sample, only case files containing both transcripts of the police interview and full trial were considered. Within this group, cases in which the complainant was over 18 years of age at the time of reporting were excluded. Cases in which the complainant alleged a single abusive incident were also excluded. The final study sample consisted of transcripts of police interviews and trials of 23 complainants (19 girls and 4 boys) alleging multiple incidents of CSA. The complainants experienced between one and four police interviews ($M = 1.57$, $SD = 0.84$). Thus, the study sample included a total of 36 police interviews and 22 trials (two complainants testified in the same trial). The children were aged 5–15 years ($M = 12.13$, $SD = 2.87$) at the time of their police interviews and 6–17 years ($M = 13.39$, $SD = 3.12$) at the time of the trials. The mean delay in reporting the abuse was 206 days after the last occurrence of abuse ($SD = 325$ days; $n = 10$ [this information was unavailable for 13 complainants]). All interviews and trials took place in a single state in Australia.

All complainants alleged abuse by a single perpetrator. Of the 23 complainants, six children alleged CSA perpetrated by a parental figure (biological parent, stepparent, or mother's partner), eight alleged the offender was another member of their family, and for the remaining nine children, the alleged offender was a family friend or acquaintance. The most severe charge filed against each perpetrator was recorded as exposure ($n = 1$), non-penetrative indecency ($n = 13$), or penetration ($n = 9$).

Of the 23 children, 13 (57%) alleged fewer than five occurrences of abuse happened to them, and nine children (39%) alleged more than five happened to them; none of the children alleged exactly five occurrences. For one child, the frequency of the alleged abuse was unclear. Each complainant could particularize (i.e., provided specific episodic details about) two to five incidents of CSA ($n_{\text{two}} = 21$; $n_{\text{three}} = 19$; $n_{\text{four}} = 16$; $n_{\text{five}} = 5$; $M = 2.78$, $SD = 0.90$) during their police interviews, totalling 61 particularized occurrences across the 23 complainants. Occurrences of CSA that were particularized at trial but not previously reported during police interviews were excluded, since we could not track labels for these occurrences throughout all stages of the criminal justice process.

Most of the 61 particularized incidents had charges associated with them at trial ($n = 52$ incidents, 85%); only nine incidents (15%) that were particularized during police interviews were not subsequently charged. Of the 52 particularized occurrences with associated charges, the alleged perpetrator was acquitted of all charges in 29 occurrences (56%) and was convicted of at least one charge in 23 occurrences (44%). The use of labels for an occurrence of abuse was unrelated to whether charges were laid or the jury verdicts, $ps > .19$. Accordingly, legal outcomes are not considered further.

Coding of the transcripts

The coding scheme applied to the police interview and trial transcripts in the present study was adapted from that used by Brubacher *et al.* (2013). Coders recorded each label that was used to name a particularized occurrence of CSA. A label was a phrase that referred to a single occurrence which contained specific information about the time, location, abusive act, or other contextual details of the incident (e.g., 'the time Dad was out').

Coders noted any subsequent reuse of an identified label, recording each time an established label was reused by a new person, and/or used in a different stage of the legal process. Coders also recorded replacement labels, when a new label was used to refer to the same incident. For example, an incident labelled as 'the last time' during the police interview was referred to as 'the incident at his house' during cross-examination by the defence lawyer. Replacements were only coded when the new label contained different episodic information from the original label, rather than minor variations in the language used (e.g., changing 'the first time' to 'the first incident' was not considered a label replacement).

For both original and replacement labels, coders recorded who provided each label (i.e., who was the first person to use the episodic information in a labelling manner): the child, the police interviewer, the prosecutor, the defence lawyer, or the judge. The stage of the criminal justice process in which the label was generated was also noted: the police interview; the prosecutor's opening statement; the defence lawyer's opening statement; the complainant's evidence-in-chief; cross-examination of the complainant; re-examination; or the judge's closing statement. Evidence from other witnesses (i.e., other than the complainant) was omitted since they very infrequently used labels to refer to any specific occurrences of alleged abuse.

Labels were classified in one of five ways according to the type of information they included. *Temporal* labels included temporal information about the incident (e.g., 'the first time', 'the one on Monday'), *locational* labels referred to the location of the incident (e.g., 'the time at his house'), *abuse-related* labels referred to an abusive act performed during the incident (e.g., 'the time he kissed me'), *situational* labels referred to a contextual detail about the incident (e.g., 'the time dad's friend was over'), and *mixed* labels combined multiple label types (e.g., 'the last time he kissed me' combines temporal and abuse-related information).

Inter-rater reliability

Two coders randomly selected and coded seven police interview transcripts that were not part of the study sample for training purposes. For reliability, a further 10 police interview transcripts not included in the study were double-coded; three at the outset of coding and seven in the latter stages of coding to ensure that coders maintained reliability. The

percentage agreement for the number of labels present in each transcript was 83%. Kappa values for classifications of label type and who produced each label ranged from .86 to .89. Coders jointly reviewed every label that had been coded and resolved any disparities.

Results

First, we examined how frequently occurrences of alleged abuse were labelled. Of the 61 incidents discussed during police interviews, 59 were given explicit labels at some stage during the legal process (i.e., during the police interview or the trial). In total, 177 labels were generated for these 59 occurrences, 118 of which replaced an existing label for an occurrence. Each occurrence received between zero and eight labels (including the original label and any replacement labels; $M = 2.90$, $SD = 1.78$).

Who generated labels?

Of the 177 labels used in the sample, most were created by the prosecution and defence lawyers (see the first column in Table 1). Judges and children created labels least frequently. We separately explored who created the first label for an occurrence and who created replacement labels. The police interviewer was most frequently the first person to label an occurrence ($n = 27$, 46% of first labels), whereas defence lawyers most commonly created replacement labels ($n = 50$, 42% of replacement labels). Table 1 demonstrates the frequency of total, first, and replacement labels created by each party.

Next, we examined the replacement of labels more closely. Since judges only created labels during their closing statements (i.e., after the child complainants had completed their evidence at trial), and never created an original label for an incident, the 18 replacement labels created by judges were omitted from the analysis. Fifty-nine labels (one per occurrence) were never replaced because they were the final labels used for each occurrence in each case. For the remaining 100 labels, a 4 (label creator: child, interviewer, prosecution, defence lawyer) \times 4 (label replacer: child, interviewer, prosecution, defence lawyer) chi-square analysis was conducted. Fisher-Freeman-Halton exact values were used due to small cell sizes. This test generalizes Fisher's exact test to contingency tables greater than 2×2 (Conover, 1980). There was a significant association between the creator and the replacer of a label, $p = .01$, Cramer's $V = .40$. Children rarely replaced their own labels. Police interviewers and defence lawyers each replaced their own labels more frequently than expected values, and defence lawyers also replaced child-created labels more than expected values.

Table 1. Frequency of labels created by each party

Creator	Total labels		First labels for an occurrence		Replacement labels	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Child	23	12.99	11	18.64	12	10.17
Interviewer	36	20.34	27	45.76	9	7.63
Prosecutor	46	25.99	17	28.81	29	24.58
Defence lawyer	54	30.51	4	6.78	50	42.37
Judge	18	10.17	0	0.00	18	15.25
Total	177	100.00	59	100.00	118	100.00

Children's labels

In total, 13 children created 23 labels (six children created one label, four created two labels, and three children created three labels). There was no significant age difference between the 13 children who created labels ($M = 12.31$, $SD = 2.59$ at interview) and the 10 children who never created a label ($M = 11.90$, $SD = 3.32$ at interview), $t(21) = 0.33$, $p = .74$, Cohen's $d = 0.14$. Children most often created labels during their police interview (10 labels) and their cross-examination (nine labels). The remaining four labels were created during the complainant's evidence-in-chief. The majority of labels created by children were ultimately replaced ($n = 16$, 70% of child-created labels).

What type of information was used to label?

The majority of labels utilized information that was temporal, locational, or mixed different types of information (see first panel of Table 2). Temporal labels were the most frequent and were broken down as follows: 35% referred to the first occurrence, 14% referred to the last occurrence, 40% referred to another sequential occurrence (i.e., second, third), and 11% were other temporal references (e.g., 'the incident in 2010'; 'the time in April'). Abuse-related and situational labels occurred less frequently. We compared the type of labels first created for an occurrence to the type of labels created as replacements for an existing label. The original label created for an occurrence was most frequently a temporal label (see middle panel of Table 2), whereas replacement labels were more variable across different types (see final panel of Table 2). A 5 (type of label) \times 2 (original or replacement label) chi-square analysis showed no relationship between the type of label and replacement status, $\chi^2(4, N = 177) = 6.68$, $p = .15$, Cramer's $V = .19$. See Table 2 for the distribution of total, first, and replacement label types.

We next explored who created each type of label. Since judges created few labels ($n = 18$), these were omitted from the analysis in order to eliminate small cell sizes. Considering the remaining 159 labels, a 5 (type of label: temporal, locational, situational, abuse-related, mixed) \times 4 (creator: child, interviewer, prosecutor, defence lawyer) chi-square analysis found the type of label was significantly associated with who created it, $\chi^2(12, N = 159) = 32.98$, $p = .001$, Cramer's $V = .26$. Figure 1 shows the proportion of each type of label produced by each participant. Child complainants and prosecutors were more likely, and defence lawyers less likely, to create temporal labels, compared to expected values. Instead, defence lawyers were more likely to use mixed labels than expected. Prosecutors were less likely than expected to use abuse-related labels, whereas police interviewers used abuse-related labels more often than

Table 2. Frequency of labels of each type

Type	Total labels		First labels for an occurrence		Replacement labels	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Temporal	52	29.38	21	35.59	31	26.27
Locational	43	24.29	12	20.34	31	26.27
Situational	24	13.56	10	16.95	14	11.86
Abuse-related	18	10.17	8	13.56	10	8.47
Mixed	40	22.60	8	13.56	32	27.12
Total	177	100.00	59	100.00	118	100.00

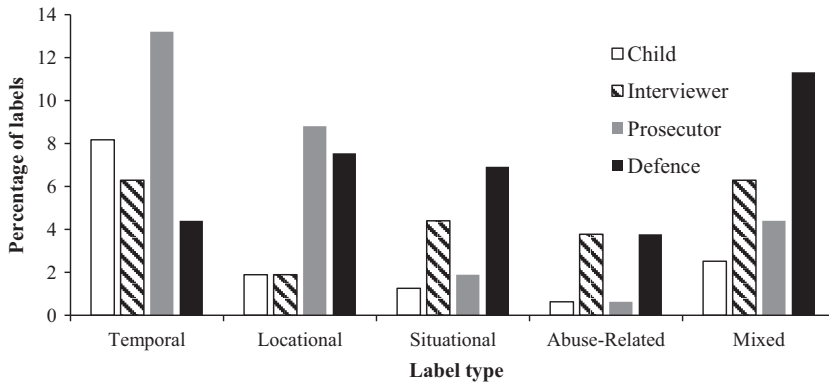


Figure 1. Types of labels produced by each party.

expected values. Police interviewers were also less likely than expected to use locational labels.

At which stage of the legal process were labels generated?

We considered the frequency of labels created during each stage of the criminal justice process. Of the 177 labels, most were created during either the complainant’s police interview ($n = 46, 26\%$) or the cross-examination of the child complainant ($n = 58, 33\%$), as shown in the first column of Table 3. To examine the creation of these labels more thoroughly, we next explored the first label appointed for each occurrence and subsequent replacement labels separately. Particularized occurrences were most frequently first appointed a label during the police interview, and all but 10 occurrences had been labelled by the conclusion of the prosecutor’s opening statement (see middle panel of Table 3). We compared the proportion of first labels that were nominated during the police interview to the proportion of first labels that were nominated during each stage of the trial using a series of z-tests. Significantly more first labels for an occurrence were generated during the police interview than in any phase of the trial, $z_s \geq 4.67, p_s < .001$. Conversely, replacement labels were most frequently generated during the cross-examination of the child complainant (see final panel of Table 3). The

Table 3. Frequency of labels created at each stage of the criminal justice process

Stage	Total labels		First labels for an occurrence		Replacement labels	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Police interview	46	25.99	37	62.71	9	7.63
Prosecution opening	27	15.25	12	20.34	15	12.71
Defence opening	5	2.82	0	0.00	5	4.24
Evidence-in-chief	15	8.47	4	6.78	11	9.32
Cross-examination	58	32.77	5	8.47	53	44.92
Re-examination	8	4.52	1	1.69	7	5.93
Judge’s closing	18	10.17	0	0.00	18	15.25
Total	177	100.00	59	100.00	118	100.00

proportion of replacement labels created during cross-examination was significantly higher than the proportion of replacement labels created during any other phase, $z_s \geq 4.97$, $p_s < .001$.

Tracking labels throughout the legal process

In this phase of analysis, we examined only the labels that were originally generated in the police interview, in order to follow them through the trial. Of the 46 labels generated during the police interview (including original labels and replacements within the interview[s]), 28 were never reused by another party or in another stage of the case. The remaining 18 labels created in police interviews were reused between one and six times ($M = 2.50$, $SD = 1.72$) throughout the ensuing trial.

Of the 46 labels generated during the police interviews, 36 were eventually replaced with a different label. Most commonly, replacements happened in cross-examination ($n = 12$; 26% of police interview label replacements). Nine labels generated during the police interview were replaced during the police interview, six were replaced during the prosecutor's opening statements, one during the defence lawyer's opening statements, seven during the complainant's evidence-in-chief, and one during the judge's closing statements.

Reusing labels

To track labels' continuity throughout the legal process, we explored how labels were used after they had been introduced. Of the 177 labels, only 53 were reused in a new stage or by a different legal professional. These 53 labels were reused between one and six times ($M = 1.87$, $SD = 1.36$), for a total of 99 instances of a previously established label being reused. Most were reused during the latter stages of the trial: the cross-examination of the complainant ($n = 32$, 31%), re-examination of the complainant ($n = 17$, 17%), and the judge's closing statement ($n = 23$, 24%). The remaining reuses occurred during the police interview (i.e., the interviewer repeated a child-created label or vice versa, $n = 10$, 11%), prosecutor and defence lawyers' opening statements ($n = 8$, 8% and $n = 4$, 4%, respectively), and the complainant's evidence-in-chief ($n = 5$, 5%).

A 5 (type of label: temporal, locational, situational, abuse-related, mixed) \times 2 (reused or not) chi-square analysis was conducted to explore whether certain types of labels were more often reused. Temporal labels were more likely than expected values to be reused later in the process $\chi^2(4, N = 177) = 34.05$, $p < .001$, Cramer's $V = .44$, while mixed labels showed the opposite pattern.

We examined the trajectory of labels throughout each stage of the criminal justice process to assess whether original labels were reused by other parties or at other stages of the trial or were replaced by new labels. Proportions of label replacements versus label repeats at each stage were examined (see Table 4). A paired-samples t -test confirmed that the overall replacements and repeats proportions across all stages of the criminal justice process were not significantly different from each other ($M_{\text{replacements}} = 0.53$, $SD = 0.20$; $M_{\text{repeats}} = 0.47$, $SD = 0.20$), $t(22) = 0.64$, $p = .53$, Cohen's $d = 0.30$. It was equally likely that a new label would be created to refer to a specific occurrence as that a previously established label would be reused.

Table 4. Average proportion of label replacements and reuses at each stage of the criminal justice process

Stage	Proportion of labels replaced	Proportion of labels reused
Police interview	0.38 (0.43)	0.62 (0.43)
Prosecution opening	0.63 (0.41)	0.37 (0.41)
Defence opening	0.56 (0.52)	0.44 (0.52)
Evidence-in-chief	0.74 (0.40)	0.26 (0.40)
Cross-examination	0.50 (0.34)	0.50 (0.34)
Re-examination	0.27 (0.34)	0.73 (0.34)
Judge's closing	0.36 (0.37)	0.64 (0.37)

Note. Standard deviations are in parentheses.

Discussion

The present study is the first to examine the use of labels during CSA trials and provides a rich description of the way that labels are used for particularized occurrences from the police interview to the trial. Previous research on children's memory and general cognitive development supports the recommendation that labels should be generated by the child (or at least using the child's words) in the police interview so that he/she knows precisely which occurrences are being discussed, and should then be used consistently thereafter by all parties (Brubacher *et al.*, 2014; Powell & McMeeken, 1998). Overall, this study showed that labels were not used in accordance with these guidelines.

Analyses revealed that many particularized occurrences of abuse were labelled during the police interview (62%), and a large majority were labelled by the end of the prosecutor's opening statement (82%). Although it is encouraging that most occurrences of abuse are labelled early in the legal process, recommendations for forensic interviewers specify that all occurrences of child abuse should be labelled during a child's initial interview with authorities to promote clarity (Powell & McMeeken, 1998). There remained much room for improvement in adherence to recommendations regarding early labelling during police interviews in the current sample.

Providing scaffolding for children's recall of autobiographical events using questioning techniques has been shown to facilitate accurate recall (Danby, Sharman, Brubacher, Powell, & Roberts, 2017) and recall of additional details (Klemfuss, Cleveland, Quas, & Lyon, 2016). Clear and consistent labelling is one means of scaffolding children's recall so that children know precisely which occurrences are being discussed. However, labels were frequently replaced with alternate labels for the same occurrence, with on average three different labels per occurrence. In fact, after original labels had been established for occurrences, labels were just as likely to be replaced as they were to be reused. While some replacements occurred during lawyers' opening statements (and thus not in the presence of the child witness), these changes are still detrimental to a child's evidence unless the label is changed back when the child provides evidence, and further serve to potentially confuse others' (e.g., the jury) ability to follow the evidence.

Young children have difficulty distinguishing repeated events and tend to confuse details across occurrences (Powell & Thomson, 1996; Powell *et al.*, 1999). Replacing labels during a child's testimony may add to this confusion, so it is of concern that label replacements happened so frequently in this sample. The high rates of label replacement may be an especially important issue given that the child is the only one who knows if a

label is unique to one occurrence. For example, ‘the time he kissed you’ is not an ideal label if the child was kissed on multiple occurrences. Interviewer-generated labels are less likely to be specific to one occurrence because the interviewer is naïve about the events in question (Brubacher, Earhart, Roberts, & Powell, 2018), and non-unique labels may contribute to greater confusion between events. Interviewers and legal professionals may benefit from guidance about clearly and consistently labelling occurrences of repeated events in order to maximize children’s opportunities to provide accurate testimony.

We hypothesized that children and police interviewers would create approximately equal numbers of labels during the police interview. Contrary to this hypothesis, children infrequently created their own labels for the occurrences they discussed. Only 13% of all labels were generated by children; a low figure compared to that observed by Brubacher *et al.* (2013), who found that children created 48% of all labels during police interviews. Whereas Brubacher *et al.* studied interviews conducted in accordance with best-practice guidelines (i.e., using a high proportion of open-ended questions), both the police interviews and court testimony in the present study were dominated by closed questions (e.g., yes–no questions; Powell *et al.*, 2016). Open-ended questioning has reliably been shown to facilitate narrative recall from children (Snow, Powell, & Murfett, 2009), and to promote reporting of additional event details (Brown *et al.*, 2013), both of which may provide children more opportunity to label occurrences. Although the closed questioning in the current sample is common practice around the world, it may have contributed to the lower rates of spontaneous labelling by children observed in this study and therefore reduced the possibility for label replacements to take place. Although spontaneous labels by children were rare, only one analogue study has provided insight into children’s capabilities to provide effective labels when explicitly instructed to do so. This study found that children aged 6–8 years old were quite good at producing effective labels with appropriate interviewer scaffolding (Brubacher *et al.*, 2018). This is an important area for future research.

In the present study, police interviewers most frequently created original labels for occurrences, whereas defence lawyers replaced labels most often; they replaced both their own labels and labels created by children more often than would be expected by chance. Cross-examination procedures with children have already been criticized for a variety of reasons (Kebbell, O’Kelly, & Gilchrist, 2007; Zajac, Gross, & Hayne, 2003; Zydervelt, Zajac, Kaladelfos, & Westera, 2017), and the current study has identified another way in which defence lawyers may impede children’s reporting. Irrespective of whether defence lawyers are aware that they are replacing labels, doing so may elicit inconsistencies in children’s accounts if label replacement contributes to further confusions between occurrences of repeated events.

Labels using temporal, locational, or mixed types of information were the most frequent, whereas abuse-related and situational information were used less often. The types of labels generated varied as a function of who created them, with children and prosecutors using temporal labels, whereas defence lawyers tended to apply mixed labels. It is well documented that appropriate use of time-related words is late to develop in childhood (see Friedman, 2007, for a review). As such, temporal labels may be more difficult for young children to understand, and these labels may not serve their intended purpose of clarifying which occurrence is being discussed. Nevertheless, children did generate temporal labels more frequently than other types of labels. Although interviewers used more temporal labels than did children in the study by Brubacher *et al.* (2013), of the labels generated by children in that study, temporal labels were also the most prevalent type.

Limitations and future directions

Generalizability of the present findings may be limited, given that the sample was drawn from a single jurisdiction in Australia, and that the sample was small due to coding procedures that involved reading entire interview and trial transcripts within a limited timeframe to complete the study. The small sample size may have impacted the ability to detect significant differences in some cases (e.g., the comparison of the proportions of label repeats and label replacements). Children's ability to label and recall occurrences of abuse is expected to increase with age (Roberts, 2002); however, due to the small sample size, we were unable to assess developmental differences in any of the variables. Lengthy trial delays or delays in reporting can occur in some cases, which could further impact children's labelling and particularization of occurrences.

It was beyond the scope of the current study to assess the source of the terms adult professionals adopted into their labels. Given recommendations that adults use children's words to create labels (Brubacher *et al.*, 2013; Powell & McMeeken, 1998), research into the frequency with which adults adopt children's words as labels and/or change labels based on children's disclosure of new information (e.g., a label was changed due to the child disclosing new unique details about an occurrence) should be considered in future. Given the extensive coding procedures in the current study, it was not feasible to consider the amount of information that children reported about each occurrence. The accuracy of their reports also could not be assessed given that the data came from the field. Future laboratory research should directly investigate the impact of replacing labels on the amount and accuracy of the information that children report.

Conclusions

Understanding the use of labels throughout investigations of CSA provides insight into how investigators, lawyers, and judges might help children overcome the cognitive challenges inherent in particularization. This study identified an important area in which practice does not align with recommendations: while most occurrences of CSA were labelled early in the process, these labels were frequently replaced between the police interview and the end of the trial. Professionals involved in eliciting or presenting children's evidence should consider allowing children to create labels wherever possible, and use those labels consistently throughout the entire criminal justice process in order to maximize children's capacity to describe occurrences of repeated abuse.

Acknowledgements

This research was supported by funding from the Royal Commission into Institutional Responses to Child Sexual Abuse. The views and findings expressed in this report are those of the authors and do not necessarily reflect those of the Royal Commission.

References

- Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review, 10*, 3–47. [https://doi.org/10.1016/0273-2297\(90\)90003m](https://doi.org/10.1016/0273-2297(90)90003m)

- Brown, D. A., Lamb, M. E., Lewis, C., Pipe, M.-E., Orbach, Y., & Wolfman, M. (2013). The NICHD Investigative Interview Protocol: An analogue study. *Journal of Experimental Psychology: Applied*, *19*, 367–382. <https://doi.org/10.1037/a0035143>
- Brubacher, S. P., Earhart, B., Roberts, K. P., & Powell, M. B. (2018). Effects of label training and recall order on children's reports of a repeated event. *Applied Cognitive Psychology*, *32*, 600–609. <https://doi.org/10.1002/acp.3440>
- Brubacher, S. P., Malloy, L. C., Lamb, M. E., & Roberts, K. P. (2013). How do interviewers and children discuss individual occurrences of alleged repeated abuse in forensic interviews? *Applied Cognitive Psychology*, *27*, 443–450. <https://doi.org/10.1002/acp.292>
- Brubacher, S. P., Powell, M. B., & Roberts, K. P. (2014). Recommendations for interviewing children about repeated experiences. *Psychology, Public Policy, and Law*, *20*, 325–335. <https://doi.org/10.1037/law0000011>
- Burrows, K. S., & Powell, M. (2013). Prosecutors' recommendations for improving child witness statements about sexual abuse. *Policing and Society*, *24*, 189–207. <https://doi.org/10.1080/10439463.2013.784305>
- Collin-Vézina, D., De La Sablonnière-Griffin, M., Palmer, A. M., & Milne, L. (2015). A preliminary mapping of individual, relational, and social factors that impede disclosure of childhood sexual abuse. *Child Abuse & Neglect*, *43*, 123–134. <https://doi.org/10.1016/j.chiabu.2015.03.010>
- Connolly, D. A., & Lindsay, D. S. (2001). The influence of suggestions on children's reports of a unique experience versus an instance of a repeated experience. *Applied Cognitive Psychology*, *15*, 205–223. [https://doi.org/10.1002/10990720\(200103/04\)15:2<205:AID-ACP698>3.0.CO;2-F](https://doi.org/10.1002/10990720(200103/04)15:2<205:AID-ACP698>3.0.CO;2-F)
- Conover, W. J. (1980). *Practical nonparametric statistics* (2nd ed.). New York, NY: John Wiley & Sons.
- Criminal Code* (Qld) s 229B (Austl.).
- Criminal Legislation Amendment [Child Sexual Abuse] Act 2018 No 33* (NSW) s 66EA (Austl.).
- Danby, M. C., Sharman, S. J., Brubacher, S. P., Powell, M. B., & Roberts, K. P. (2017). Differential effects of general versus cued invitations on children's reports of a repeated event episode. *Psychology, Crime and Law*, *23*, 794–811. <https://doi.org/10.1080/1068316x.2017.1324028>
- Dietze, P. M., & Thomson, D. M. (1993). Mental reinstatement of context: A technique for interviewing child witnesses. *Applied Cognitive Psychology*, *7*, 97–108. <https://doi.org/10.1002/acp.2350070203>
- Fisher, R. P., & Geiselman, R. E. (1992). *Memory enhancing techniques for investigative interviewing: The cognitive interview*. Springfield, IL: Thomas.
- Friedman, W. J. (2007). The development of temporal metamemory. *Child Development*, *78*, 1472–1491. <https://doi.org/10.1111/j.1467-8624.2007.01078.x>
- Goodman-Delahunty, J., Nolan, M., & van Gijn-Grosvenor, E. (2017). *Empirical guidance on the effects of child sexual abuse on memory and complainants' evidence*. Sydney, NSW: Royal Commission into Institutional Responses to Child Sexual Abuse.
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3–28. <https://doi.org/10.1098/rstb.1997.0156>
- Kebbell, M. R., O'Kelly, C. M. E., & Gilchrist, E. L. (2007). Rape victims' experiences of giving evidence in English courts: A survey. *Psychiatry, Psychology & Law*, *14*, 111–119. <https://doi.org/10.1375/pplt.14.1.111>
- Klemfuss, J. Z., Cleveland, K. C., Quas, J. A., & Lyon, T. D. (2016). Relations between attorney temporal structure and children's response productivity in cases of alleged child sexual abuse. *Legal and Criminological Psychology*, *22*, 228–241. <https://doi.org/10.1111/lcrp.12096>
- Lindsay, S. D., Allen, B. P., Chan, J. C. K., & Dahl, L. C. (2004). Eyewitness suggestibility and source similarity: Intrusions of details from one event into memory reports of another event. *Journal of Memory and Language*, *50*, 96–111. <https://doi.org/10.1016/j.jml.2003.08.007>
- Orbach, Y., & Pipe, M.-E. (2011). Investigating substantive issues. In M. E. Lamb, D. J. La Rooy, L. C. Malloy & C. Katz (Eds.), *Children's testimony: A handbook of psychological research and forensic practice* (pp. 147–163). Chichester, UK: John Wiley & Sons. <https://doi.org/10.1002/9781119998495>

- Parkinson, P. N., Shrimpton, S., Swanston, H. Y., O'Toole, B. I., & Oates, R. K. (2002). The process of Attrition in child sexual assault cases: A case flow analysis of criminal investigations and prosecutions. *Australian and New Zealand Journal of Criminology*, *35*, 347–362. <https://doi.org/10.1375/000486502320814368>
- Pearse, S. L., Powell, M. B., & Thomson, D. M. (2003). The effect of contextual cues on children's ability to remember an occurrence of a repeated event. *Legal and Criminological Psychology*, *8*, 39–50. <https://doi.org/10.1348/135532503762871228>
- Powell, M. B., & McMeeken, L. (1998). "Tell me about the time when...": 9 golden rules for interviewing a child about a multiple offence. *Australian Police Journal*, *52*, 104–108.
- Powell, M. B., Roberts, K. P., Ceci, S. J., & Hembrooke, H. (1999). The effects of repeated experience on children's suggestibility. *Developmental Psychology*, *35*, 1462–1477. <https://doi.org/10.1037/0012-1649.35.6.1462>
- Powell, M., Roberts, K., & Guadagno, B. (2007). Particularisation of child abuse offences: Common problems when questioning child witnesses. *Current Issues in Criminal Justice*, *19*(1), 64–74.
- 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*, 1988–2004. <https://doi.org/10.2307/1131605>
- Powell, M. B., Westera, N. J., Goodman-Delahunty, J., & Pichler, A. S. (2016). An evaluation of how evidence is elicited from complainants of child sexual abuse. Report for the Royal Commission into Institutional Responses to Child Sexual Abuse.
- 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*, 403–435. [https://doi.org/10.1016/s0273-2297\(02\)00005-9](https://doi.org/10.1016/s0273-2297(02)00005-9)
- Roberts, K. P., & Blades, M. (1999). Children's memory and source monitoring of real-life and televised events. *Journal of Applied Developmental Psychology*, *20*, 575–596. [https://doi.org/10.1016/S0193-3973\(99\)00030-1](https://doi.org/10.1016/S0193-3973(99)00030-1)
- Royal Commission into Institutional Responses to Child Sexual Abuse (2017). *Criminal justice report: Executive summary and parts I–II*. Retrieved from https://www.childabuseroyalcommission.gov.au/sites/default/files/file-list/final_report_-_criminal_justice_report_-_executive_summary_and_parts_i_to_ii.pdf
- Smith, K., & Milne, R. (2011). Planning the interview. In M. E. Lamb, D. J. LaRooy, C. Malloy & C. Katz (Eds.), *Children's testimony: A handbook of psychological research and forensic practice* (2nd ed., pp. 87–108). Sussex, UK: John Wiley & Sons, Ltd.
- 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*, 555–568. <https://doi.org/10.1080/10683160802409347>
- Trocmé, 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: Ottawa, ON.
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352–373. <https://doi.org/10.1037/h0020071>
- Zajac, R., Gross, J., & Hayne, H. (2003). Asked and answered: Questioning children in the courtroom. *Psychiatry, Psychology & Law*, *10*, 199–209. <https://doi.org/10.1375/pplt.2003.10.1.199>
- Zajac, R., & Hayne, H. (2003). I don't think that's what really happened: The effect of cross-examination on the accuracy of children's reports. *Journal of Experimental Psychology*, *9*, 187–195. <https://doi.org/10.1037/1076-898X.9.3.187>
- Zydervelt, S., Zajac, R., Kaladelfos, A., & Westera, N. (2017). Lawyers' strategies for cross-examining rape complainants: Have we moved beyond the 1950s? *British Journal of Criminology*, *57*, 551–569. <https://doi.org/10.1093/bjc/azw023>