This is the published version


Available from Deakin Research Online

http://hdl.handle.net/10536/DRO/DU:30065406

Reproduced with the kind permission of the copyright owner

Copyright: 2014, Vathek Publishing
An examination of the prevalence of temporally leading questions in child witness interviews

Belinda L. Guadagno‡ and Martine B. Powell†
‡(Corresponding author) School of Psychology, Deakin University, 221 Burwood Hwy, Burwood, 3125, Victoria, Australia. Tel: +61 3 9244 6106; email: belinda.guadagno@deakin.edu.au
†School of Psychology, Deakin University, 221 Burwood Hwy, Burwood, 3125, Victoria, Australia
Submitted 28 October 2013, accepted 25 January 2014
Keywords: child witness, children’s memory, investigative interviewing, repeated events

Belinda Guadagno is a lecturer in psychology at the Faculty of Health, Deakin University, Melbourne, Australia.
Martine Powell is Professor Forensic Psychology at the Faculty of Health, Deakin University, Melbourne, Australia. Both authors conduct research primarily in the area of child witness testimony and investigative interviewer training, particularly children’s memory for repeated events.

ABSTRACT
Leading questions are generally defined as those that raise details not provided by the witness. Leading questions can raise content details (eg, actions, objects, persons) or can refer to the time when details occurred. The latter questions are referred to as temporally leading. Study 1 compared the incidence of content and temporally leading questions in field interviews conducted by police officers when eliciting accounts from children about repeated, or a single episode of, abuse. Study 2 extended the analysis to use standardised mock rather than field interviews, where there was a precise record of what events occurred. In both studies, temporally leading questions were more frequent than content-leading questions, but only in situations in which multiple occurrences of the event were being discussed. The implications of these results are discussed.

INTRODUCTION
When allegations of child abuse are reported, alleged victims are usually interviewed by police officers or child protection workers who specialise in child abuse investigations. The child’s statement during such interviews is of crucial importance; it informs the investigative process by unearthing leads for investigators to explore, and (in many jurisdictions) electronic recordings of the interview can be used as the child’s evidence-in-chief if the matter proceeds to court (Hoyano & Keenan, 2007). The accuracy of children’s statements about abuse is determined in large part by the quality of questions asked by the interviewer (Sternberg et al., 1996). The consensus is that interviewers should ask open-ended questions (ie, that do not dictate what specific information needs to be reported), and leading questions should be avoided. Leading questions are generally defined as those that presume certain information, suggest/imply a particular answer or raise specific details that have not previously been mentioned by the child witness (Powell & Snow, 2007).

Rates of interviewers’ use of leading questions have varied considerably across previous studies; from as low as 2 per cent to as high as 59 per cent (Cederborg, Orbach, Sternberg, & Lamb, 2000; Cyr, Dion, McDuff, &
Trotier-Sylvain, 2012; Davies, Westcott, & Horan, 2000; Hughes-Scholes & Powell, 2008; Korkman, Santtila, Westeråker, & Sandnabba, 2008; Orbach et al., 2000; Sternberg et al., 1996; Warren et al., 1999). The quality of training received by interviewers affects the incidence (Powell, 2008). So does the degree of precision in the way in which leading questions are measured. Although a single category has typically been used in prior evaluation research to represent all leading questions, researchers are now focusing on identifying the wide variety of subcategories of leading questions and their differential effects (Sharman & Powell, 2012). One type of leading question discussed is that which suggests details using yes/no or forced choice options (eg, ‘Was your brother home that day?’). Another type of leading question presumes information that the witness had not yet reported (eg, ‘What colour was the hat that Frank wore?’ when the witness had not previously mentioned that Frank wore a hat). Leading questions can be defined by whether they include few specific details (eg, ‘Was anyone else at your home that day?’) or an abundance of details that could be used by the child to construct an account of a false event (eg, ‘Was there a young man called Mr Briggs at your home that day?’). When comparing the differential effect of the question types, Sharman and Powell (2012) found that interview questions that contained highly specific details and that presumed information were the most harmful. Participants were more likely to agree with the misleading suggestions contained in these question structures and were more likely to falsely report those suggested details at a subsequent interview than misleading suggestions contained in other question structures.

Another way to categorise leading questions (which is the focus of this article) is whether the question includes temporally leading as opposed to content-leading information. Content-leading questions are defined as those that introduce, suggest or presume specific details (ie, actions, objects, persons) that occurred in the event. By contrast, temporally leading questions introduce, suggest or presume the time or temporal source of various content details. Content-leading and temporally leading question categories are not mutually exclusive. For example, consider the question ‘Your uncle touched your penis the last time you went to gymnastics training, didn’t he?’ when the child had not mentioned being touched on his penis. This question would be considered leading if the content detail suggested (ie, being touched on the penis by his uncle) was not previously mentioned by the child. The question could also be considered temporally leading if the child did not mention when the touching occurred.

Temporally leading questions can relate to both a one-time or repeated offence, but their incidence and detrimental impact may well be greater for interviews about repeated offences. In normal legal proceedings, for an alleged offender to be charged and convicted in relation to a repeated offence, at least one specific occurrence must be identified with reasonable precision with reference to place and time (S v R, 1989). For example, a child who has experienced repeated occurrences of abuse may be asked to remember contextual factors related to at least one of the occurrences, such as where the incident occurred, what clothing was worn, what the alleged perpetrator did and said during the occurrence of abuse and where members of the family were. The act of identifying an occurrence and distinguishing it from other occurrences is a difficult task for any witness after a delay in time, but it is especially difficult for child witnesses, whose knowledge and understanding of time are not as well developed as those of adults (Friedman, 1991).

The effect of event repetition on memory is such that details that are fixed (experienced exactly the same way during each
Leading questions in child witness interviews

Occurrence) are strengthened in memory and become highly resistant to suggestion (Connolly & Lindsay, 2001; Powell, Roberts, Ceci, & Hembrooke, 1999). Details that vary across occurrences, however, are easily confused and are subject to interference (Brubacher, Glisic, Roberts, & Powell, 2011). Memory of temporal-source information decays over time independently of content information, because the task involves identifying and distinguishing one act or occurrence of abuse from other similar acts or occurrences (Powell & Thomson, 1997). The detrimental effect of repeated experience on children's event memory is reflected by a high rate of migration of details from other occurrences of the event into the occurrence that is being recalled (Hudson, 1990; Powell & Thomson, 1996).

Consideration of the incidence of temporally leading (as distinct from content-leading) questions and examination of the context in which temporally leading questions are most likely to arise are important, because both errors can decrease the contribution of the child's statement and the outcome of the trial. For example, if a child accurately describes an incident of abuse, but confuses the time and place of the incident with another similar incident, the perpetrator may be able to provide an alibi and be exonerated, even though the child may be truthfully relaying abuse that had occurred at some point in the past. In this article, we examine the incidence of these questions in two studies which differed depending on the event paradigm. Study 1 explored the rate of these questions in field interviews about abuse, whereas Study 2 used a mock interview paradigm in which a record of the event and its structure were known, thus enabling examination of the effect of the questions on the accuracy of children's responses. Overall, we predicted that temporally leading questions would be more common in interviews about repeated occurrences of abuse than in those that focused solely on a single abusive act. We also expected higher error rates in response to temporally leading compared with content-leading questions.

STUDY 1

Method

A pseudo-random selection of 50 de-identified interview transcripts was obtained from child abuse investigation units within police organisations in three states of Australia. The only restraint for the selection of interviews was that they were conducted by an experienced police officer from a child abuse unit, the interviewee was a middle-school-aged child (approximately 8 years), the interviews constituted the first recorded interview with the child, and half the interviews focused on a single abusive act, while the other half questioned the child about repeated occurrences of abuse.

Each question asked in the interviews was first classified as leading or non-leading. Each non-leading question was assigned to one of the following mutually exclusive categories.

- Open-ended questions included any question designed to elicit an elaborate response without dictating what specific details the child needed to report (eg, ‘Tell me what happened the first time Mr Smith touched your rude spot?’).

- Specific questions included questions that focused the child on specific aspects of the abuse that the child had mentioned and requested particular information (eg, ‘You said Mr Smith took your clothes off. What clothes were you wearing?’).

Every identified leading question was classified into one of the following mutually exclusive categories:

- Specific closed leading questions that contain specific information that the child had not previously mentioned and could be used by the child to construct
an account of the alleged abuse (eg, ‘Did your Mum hit you with a stick?’ when the child had not reported what her mother hit her with).

- Presumptive leading questions presumed specific information that the child had not previously mentioned (eg, ‘Was the stick that your Mum hit you with long?’ or ‘Tell me about the stick your Mum hit you with’ when the child had not previously mentioned that her mother hit her with a stick).

For every identified leading question, it was noted whether the question was leading in terms of content and/or temporal details. In the case of temporally leading questions, the context in which each question arose within the interviews was also noted.

All the transcripts were coded by the principal researcher, and another researcher who was not otherwise involved in the study coded a random selection of 20 per cent of the transcripts. Inter-rater reliability, calculated as agreements/(agreements + disagreements), was at least 97 per cent for each of the categories listed above.

Results and discussion

The proportion of leading and open-ended questions was $M = 0.23$ ($SD = 0.09$) and $M = 0.17$ ($SD = 0.07$), respectively. Table 1 provides a description of the mean number and proportion of leading questions asked by the interviewers. The questions are presented separately according to whether they were content-leading or temporally leading, and whether they took the form of specific closed or presumptive leading questions. The data are also presented separately for those interviews that involved discussion about repeated occurrences of abuse, and those that focused solely on a single abusive act. It should be

<table>
<thead>
<tr>
<th>Leading questions</th>
<th>Interviewers (N = 25) who questioned about repeated abuse</th>
<th>Interviewers (N = 25) who focused solely on one abusive act</th>
<th>All interviewers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean questions</td>
<td>Proportion</td>
<td>Mean questions</td>
</tr>
<tr>
<td>Highly specific closed*</td>
<td>21.71 (13.13)</td>
<td>0.15 (0.07)</td>
<td>21.88 (12.35)</td>
</tr>
<tr>
<td>Content leading</td>
<td>13.71 (9.27)</td>
<td>0.10 (0.05)</td>
<td>17.50 (11.24)</td>
</tr>
<tr>
<td>Temporally leading</td>
<td>18.57 (10.95)</td>
<td>0.14 (0.06)</td>
<td>7.25 (5.34)</td>
</tr>
<tr>
<td>Presumptive*</td>
<td>14.29 (9.58)</td>
<td>0.10 (0.06)</td>
<td>2.13 (1.25)</td>
</tr>
<tr>
<td>Content leading</td>
<td>1.67 (1.65)</td>
<td>0.01 (0.01)</td>
<td>1.13 (1.25)</td>
</tr>
<tr>
<td>Temporally leading</td>
<td>13.90 (9.49)</td>
<td>0.10 (0.06)</td>
<td>1.00 (1.20)</td>
</tr>
<tr>
<td>TOTAL*</td>
<td>36.00 (20.42)</td>
<td>0.26 (0.09)</td>
<td>24.00 (12.74)</td>
</tr>
<tr>
<td>Content leading</td>
<td>15.38 (9.85)</td>
<td>0.11 (0.05)</td>
<td>18.63 (12.05)</td>
</tr>
<tr>
<td>Temporally leading</td>
<td>32.48 (18.17)</td>
<td>0.23 (0.10)</td>
<td>8.25 (5.65)</td>
</tr>
</tbody>
</table>

Note:
Standard deviations appear in parentheses.
*Total scores do not merely reflect the addition of the content-leading and temporally leading categories, because these categories were not mutually exclusive.
noted that the total scores do not merely reflect the addition of the content-leading and temporally leading question categories because these categories were not mutually exclusive.

The proportion of leading questions asked in interviews about repeated abuse was significantly higher than that observed in interviews that focused on a single abusive act, \( t(48) = 3.10, p < 0.01 \). Although there was no difference in the proportion of content-leading questions across the two types of interviews \( t(48) = 0.17, p = 0.87 \), interviews about repeated abuse contained a greater proportion of temporally leading questions than interviews that focused on just one abusive act, \( t(48) = 4.91, p < 0.01 \). The leading questions asked in interviews about repeated abuse were largely temporally leading; the rate of temporally leading questions for this group was higher than the rate of content-leading questions, irrespective of whether the questions were asked in the form of specific closed \( t(24) = 4.22, p < 0.01 \) or presumptive leading questions \( t(24) = 6.83, p < 0.01 \). Interviews that focused solely on one abusive act consisted of largely content-leading questions; these interviews contained a greater proportion of content-leading as opposed to temporally leading questions \( t(24) = 3.31, p < 0.01 \).

Temporally leading questions tended to be asked in four distinct contexts: to establish when the abuse or an occurrence of the abuse occurred; to establish whether the abuse was repeated; to generate labels to identify different occurrences of the abuse; and to determine how details varied across the reported occurrences of abuse. Interestingly, not all the temporally leading questions contained false details in the questions. Many of these questions arose when interviewers shifted their focus between occurrences of abuse without providing appropriate verbal cues to the child witness to indicate that they had done so. Interviewers who questioned about repeated abuse asked temporally leading questions in all four contexts outlined above, whereas temporally leading questions in interviews about a single abusive act were confined to the first two contexts only.

In summary, temporally leading questions tended to occur in interviews about repeated abuse, where one or more occurrences need to be distinguished with respect to time, place or some other unique contextual detail. The questions occurred because of false presumptions on the part of the interviewer about the structure of the event, and because of a lack of specificity (on the part of the interviewer and the child) about which occurrence of the event was being referred to in the question.

**STUDY 2**

The current study extended the prior findings using a different interview paradigm: a mock interview with school children recalled an event that was staged in their school. There were two benefits to incorporating this laboratory design. First, having a record of the event enabled us to demonstrate the effect of temporally leading versus content-leading questions on the accuracy of children’s responses. Further, a staged event (where we had full control over the event details and prior information given to interviewers) allowed us to ascertain whether the results in Study 1 were robust and not merely due to inherent differences in the nature of events that resulted in disclosures of isolated or repeated acts of abuse (apart from event repetition per se).

**Method**

**Participants**

Thirty-eight children (24 males and 14 females) participated in the interviews conducted for this study \( M \) age = 119.37 months; \( SD = 12.4 \) months; age range = 7 years, 3 months to 11 years, 9 months). All
children were located in the metropolitan area and were recruited through letters to their carers, outlining the nature of the project and seeking consent for the children's participation. While 45 parents gave consent for their child to be involved, seven children were excluded from the final sample because they were absent from school during at least one of the occurrences of the event or the interview.

The interviewers were 38 Australian police officers (14 males and 24 females); one for each child who had attended all four occurrences of the staged event. All the police officers were employed in a child abuse investigation unit and were recruited via letters distributed by a senior member of their training unit. The officers and children came from a range of areas and their participation was voluntary. All officers had completed internal training in child abuse interviewing, and had been conducting field interviews on a regular basis.

The event
All children participated in a 30-minute event referred to as the ‘Deakin Activities’. A research assistant (named ‘Sarah’) administered the event in the presence of the children's classroom teachers. This event was selected because it has been used successfully in many prior studies with similar-aged children (Guadagno & Powell, 2009; Powell & Thomson, 1996). It was administered on four separate occasions (twice a week for two weeks) in the children's regular classroom and consisted of 17 items, which were administered in the same temporal order and centred around six activities: meeting a koala; listening to a story; doing a puzzle; having a rest; getting refreshed; and receiving a surprise. The items represented various types of information (e.g., verbalisations, actions, objects, persons) and were repeated in different ways across the occurrences. Although all occurrences of the event adhered to the same structure, many of the specific items had different instantiations across the occurrences. For example, the story that was read to the children changed in every occurrence of the event (see Powell & Thomson, 1996, for a full description of the event paradigm). All teachers were instructed not to talk with the children about the event or to inform them that the research assistant would return to administer subsequent occurrences of the event. The teachers were also instructed not to inform the children that they were to be interviewed by a police officer about the event until the morning of the interview.

Procedure
Each police officer individually interviewed one child each within two weeks of the final occurrence of the event. Before commencing the interview, each of the officers was told that a lady called Sarah attended the child's school to administer an event called the ‘Deakin Activities’, and that their job was to try to elicit an accurate and detailed account of this event ‘in a manner they would normally do when interviewing a child in the field’. Although the officers were not told any specific details about the event, they were advised to keep an open mind regarding what may have occurred. The officers were informed that the event may or may not have been repeated.

The interviews were conducted in an isolated room at the child's school (not the room where occurrences of the event had taken place). Given that the interviews were conducted during school time, a time limit needed to be set for each interview. The police officers were allocated a maximum of 12 minutes to conduct each interview; 10 minutes to elicit information about the event with 2 minutes for rapport building at the commencement of the interview.

Limitations were also imposed on the way in which the police officers could commence the substantive component of the interview. Specifically, the officers were instructed to
use the following leading statement to introduce the topic of concern: ‘I heard that a lady called Sarah came to your school to do the “Deakin Activities” event. Tell me everything you can remember about that.’ The reason for commencing with this statement is that the focus of the study was on how the interviewers elicited an account of the event as opposed to the questions used to elicit a non-leading disclosure. The use of the opening leading statement ensured that each interviewer had the maximum time available to elicit a free narrative account (the cue was effective in eliciting event-related details from all children). All children were briefed both prior to and subsequent to the interviews.

**Data management**

The interviews were audiotaped, transcribed verbatim and checked for accuracy. The coding procedure used to identify the questioning technique adopted by the interviewers was identical to that used in Study 1. The number of errors in the children’s responses to interviewers’ questions was also noted. Any incorrect response was defined as an error, irrespective of whether it referred to a detail that occurred in the Deakin Activities, but not the particular occurrence being recalled by the child or a detail that did not occur in the event. For each error type, it was noted what type of question elicited the error (specific closed leading or presumptive leading questions). The first author coded all the transcripts (interviewer and child responses), and a person not otherwise involved coded a random selection of 20 per cent of the transcripts. Inter-rater reliability, calculated as agreements/(agreements + disagreements), was at least 94 per cent for each of the categories.

**Results**

The proportion of leading and open-ended questions was $M = 0.26$ ($SD = 0.14$) and $M = 0.22$ ($SD = 0.13$) respectively. Table 2 outlines the mean number and proportion of

<table>
<thead>
<tr>
<th>Leading questions</th>
<th>Interviewers ($N = 16$) who distinguished between occurrences</th>
<th>Interviewers ($N = 22$) who focused on one occurrence of the event</th>
<th>All interviewers ($N = 38$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean questions</td>
<td>Proportion</td>
<td>Mean questions</td>
</tr>
<tr>
<td>Highly specific closed*</td>
<td>15.63 (10.13)</td>
<td>0.21 (0.10)</td>
<td>8.91 (7.66)</td>
</tr>
<tr>
<td>Content leading</td>
<td>6.63 (5.46)</td>
<td>0.08 (0.05)</td>
<td>5.27 (5.68)</td>
</tr>
<tr>
<td>Temporally leading</td>
<td>13.13 (9.37)</td>
<td>0.18 (0.10)</td>
<td>4.55 (6.34)</td>
</tr>
<tr>
<td>Presumptive*</td>
<td>8.63 (5.24)</td>
<td>0.13 (0.07)</td>
<td>3.73 (3.33)</td>
</tr>
<tr>
<td>Content leading</td>
<td>1.38 (1.75)</td>
<td>0.02 (0.02)</td>
<td>1.05 (1.53)</td>
</tr>
<tr>
<td>Temporally leading</td>
<td>8.19 (5.42)</td>
<td>0.13 (0.07)</td>
<td>2.77 (3.25)</td>
</tr>
<tr>
<td>TOTAL*</td>
<td>24.26 (13.83)</td>
<td>0.34 (0.13)</td>
<td>12.64 (9.06)</td>
</tr>
<tr>
<td>Content leading</td>
<td>8.00 (6.27)</td>
<td>0.10 (0.06)</td>
<td>6.31 (6.24)</td>
</tr>
<tr>
<td>Temporally leading</td>
<td>21.32 (13.25)</td>
<td>0.30 (0.14)</td>
<td>7.32 (8.81)</td>
</tr>
</tbody>
</table>

**Note:**

Standard deviations appear in parentheses.

*Total scores do not merely reflect the addition of the content leading and temporally leading categories, because these categories were not mutually exclusive.
leading questions asked across the categories. The results are reported separately for officers who established that there were repeated episodes of the event (and attempted to distinguish between them) and those who focused on only one occurrence. The results are also separated according to whether questions were content-leading or temporally leading, and asked in the form of specific closed or presumptive leading questions.

The high rate of temporally leading questions was evident only when interviewers were trying to establish the nature of the repeated event. Interviews in which officers attempted to distinguish between one or more occurrences of the repeated event comprised a greater proportion of temporally leading questions, $t(36) = 3.77, p < 0.01$, and a greater overall proportion of leading questions, $t(36) = 4.03, p < 0.01$, than interviews that focused on a single occurrence of abuse only. There was no difference in the proportion of content-leading questions across groups, $t(36) = 0.74, p = 0.47$. Furthermore, officers who sought to distinguish one or more occurrences in the series, asked predominantly temporally leading questions; the proportion of temporally leading questions for this group was higher than the proportion of content-leading questions, irrespective of whether the questions were asked in the form of specific closed, $t(15) = 3.68, p < 0.01$, or presumptive leading questions, $t(15) = 5.90, p < 0.01$. In the case of interviews that focused on just one occurrence of the event, equal proportions of content-leading and temporally leading questions were observed, $t(21) = -1.56, p = 0.13$.

Overall, the proportion of leading questions that elicited an error was 19 per cent. In response to these questions, it was more common for the child to indicate confusion, provide an ambiguous response or say that they did not know the answer. For those leading questions where a clear but incorrect response was given, a greater proportion of errors was provided in response to specific closed leading, as opposed to presumptive leading questions, $t(37) = 6.10, p < 0.01$. Further, a greater proportion of errors arose in response to temporally leading as opposed to content-leading questions when they were asked in the form of specific closed questions, $t(37) = 4.28, p < 0.01$. No difference in error rate was observed between temporally leading and content-leading questions when they were phrased as presuppositions (i.e., where the interviewer presumed specific information that the child had not previously mentioned), $t(37) = 0.61, p = 0.55$.

**GENERAL CONCLUSION**

This article contributes to the literature on investigative interviewer performance by highlighting a type of leading question that has received relatively little discussion to date. The question, referred to as temporally leading, is that which introduces, suggests or presumes the time or temporal source of details when such details have not previously been mentioned by the witness. Temporally leading questions can occur in interviews about a single act of abuse or multiple occurrences; however, this study showed that such questions are quite prevalent in interviews about repeated abuse. In fact, we showed that interviewers are particularly susceptible to using these questions when the offence being elicited was repeated and the witness is being asked to distinguish individual occurrences from one another. The detrimental effect of the temporally leading questions was illustrated in Study 2. Consistent with prior research on the impact of repeated experience on children’s suggestibility, the likelihood of error in response to specific closed leading questions was found to be higher for temporally leading as opposed to content-leading questions (Powell & Thomson, 1997).

The temporally leading questions were not isolated to a particularly section of the interview; they were widespread. For example, such questions were asked in the context
of: establishing when an act or occurrence of the abuse occurred; whether the abuse was repeated; when trying to generate labels to identify different occurrences of the abuse; and to determine how details varied across the reported occurrences of abuse. On the positive side, laboratory research has now started to articulate practical ways to establish these components with child witnesses of abuse while minimising the risk of error in their accounts (Brubacher et al., 2011; Brubacher, Earhart, Roberts, & Powell, 2013; Brubacher, Roberts, & Powell, 2012; Drohan-Jennings, Roberts, & Powell, 2010; Powell & Thomson, 2003). The current research provides impetus for experts to address these strategies in interviewer training programmes, to monitor their inclusion in field interviews through careful observation and assessment of their impact, and to provide adequate trainee interviewer supervision and feedback to ensure that children are given the best opportunity to report what they remember with minimal errors (Orbach & Lamb, 2007). Overall, this work showed that a major focus of training should be to assist interviewers to avoid making presumptions about the timing of event details, to understand the limitations associated with children’s understanding of time, and to ensure that they exhaust a narrative account of one occurrence before moving the child’s focus to another occurrence.

Acknowledgements

The authors would like to thank the children and police officers for their time as well as the police managers who assisted with the recruitment of participants.

References


