Facilitating recall and particularization of repeated events in adults using a multi-method interviewing format

Feni Kontogianni[a]*, Eva Rubinova[a], Lorraine Hope[a], Paul J. Taylor[b], Aldert Vrij[a], and Fiona Gabbert[c]

[a] *Department of Psychology, University of Portsmouth, Portsmouth, United Kingdom;*
[b]*Department of Psychology, Lancaster University, Lancaster, United Kingdom & University of Twente, Enschede, The Netherlands;*
[c] *Department of Psychology, Goldsmiths University of London, United Kingdom*

*Corresponding author[1]: Feni Kontogianni, Department of Psychology, University of Portsmouth, Portsmouth, PO1 2DY (email: feni.kontogianni@port.ac.uk)

Short title: ADULTS' RECALL OF REPEATED EVENTS

Word count: 7917

[1] The corresponding author's affiliation has changed: Department of Psychology, University of Winchester, Winchester, SO22 4NR (email: feni.kontogianni@winchester.ac.uk)

Facilitating recall and particularization of repeated events in adults using a multi-method interviewing format

Reports about repeated experiences tend to include more schematic information than information about specific instances. However, investigators in both forensic and intelligence settings typically seek specific over general information. We tested a multi-method interviewing format (MMIF) to facilitate recall and particularization of repeated events through the use of the self-generated cues mnemonic, the timeline technique, and follow-up questions. Over separate sessions, 150 adult participants watched four scripted films depicting a series of meetings in which a terrorist group planned attacks and planted explosive devices. For half of our sample, the third witnessed event included two deviations (one new detail and one changed detail). A week later, participants provided their account using the MMIF, the timeline technique with self-generated cues, or a free recall format followed by open-ended questions. As expected, more information was reported overall in the MMIF condition compared to the other format conditions, for two types of details, correct details, and correct gist details. The reporting of internal intrusions was comparable across format conditions. Contrary to hypotheses, the presence of deviations did not benefit recall or source monitoring. Our findings have implications for information elicitation in applied settings and for future research on adults' retrieval of repeated events.

Keywords: repeated events; memory retrieval; timeline technique; self-generated cues; deviation

Witnesses, victims, or sources may be interviewed about a series of events that have occurred repeatedly over a period of time, such as, domestic violence, sexual abuse, industrial accident investigations, or meetings of a criminal gang. In investigations involving repeated events, interviewees will likely need to retrieve and report details of a specific incident (i.e., particularization; see p. 203, Brubacher & Roy, 2014; p.67, Powell et al. 2007). To date, numerous studies have examined how this can be achieved when interviewing children about repeated events of child abuse (Brubacher et al., 2014; Woiwod et al., 2019). However, there is little forensic research on adults' memory of repeated events (e.g., MacLean et al., 2018) and, more particularly, on techniques that might be used to effectively elicit information for specific instances (e.g., Leins et al., 2014; Theunissen et al., 2017; Willen et al., 2015).

Memory for repeated experiences differs from memory for unique experiences in various ways (Price & Connolly, 2013). Across a series of repeated events, some details are recurring and thus form a routine (e.g., every meeting starts with the leader of the terrorist group describing a plan). Other details vary across instances (e.g., a different target is selected for each attack). Because of repeated exposure, memory for recurring details is stronger than memory for variable details. Similarly, memory for the overall routine across events is stronger than for a specific incident within the series (McNichol et al., 1999). Consequently, interviewees are likely to under-report information about specific instances.

Interviewees are often required to describe instances with precision, such as by reporting dates or times, or by identifying perpetrators of specific actions. However, discriminating between repeated events is a challenging task, as there is a high likelihood of memory interference, particularly when the events are similar (Farrar & Boyer-Pennington, 1999; Lindsay & Johnson, 1989). Interviewees might attribute a

detail to the wrong incident (i.e., source misattribution) or they might remember what occurred but not when it occurred (Johnson et al., 1993); both of which can negatively affect the interviewee's credibility and the efficacy of the investigation (Brubacher et al., 2014; Weinsheimer et al., 2017).

There are distinct challenges in remembering and reporting repeated events that are not fully addressed by current information gathering protocols. Drawing from a rich theoretical framework on the representation of repeated events in memory, the current study tests the effectiveness of a multi-method interviewing format (MMIF), which includes: i) the self-generated cues mnemonic (SGC; Kontogianni et al., 2018; Wheeler & Gabbert, 2017); ii) the timeline technique (Hope et al., 2013; Hope et al., 2019); and, iii) the use of follow-up open-ended questions (Kontogianni et al., 2020). These elements have been previously tested separately or in combination in research examining memory for single events. To examine their combined effectiveness, we test them against two comparison groups: one intermediate group where participants used the self-generated cues and the timeline technique (SGC-Timeline), and a control group where participants used a free recall format followed by open-ended questions.

Repeated events: Schema and Fuzzy-Trace Theory

Repeated events are thought to be represented in memory as parts of an overarching higher-order knowledge structure that is characterized by a common theme, referred to as a schema (Ahn et al., 1992; Brewer & Nakamura, 1984). Schema theory suggests that one experience suffices to begin building a script about what typically occurs in an event. The experience of additional similar occurrences informs a more elaborate schematic representation and shapes our expectations about future occurrences (Ahn et al., 1992; Farrar & Boyer-Pennington, 1999; Hudson et al., 1992). In line with the

spreading activation theory of memory (Anderson, 1983), repeated exposure to recurring details forms strong associated traces within the memory network which in turn increases the probability of their retrieval (e.g., Hudson et al., 1992). Over the course of a repeated event, some *variations* (i.e., predictable alternatives of certain recurring actions) are likely to occur (Abelson, 1981). Compared to recurring details that are stable and characterise the general routine of the events, variations are thought to be absorbed by the script over time and, thus, become less likely to be retrieved (Abelson, 1981; Schank & Abelson, 1977). This notion aligns with the idea that, over time, the content of repeated instances can become part of semantic memory in an abstracted form (e.g., Brewer & Nakamura, 1984).

However, there is evidence that *deviations* from the script, which are atypical and unpredictable details, can be particularly memorable (Abelson, 1981). To the extent that deviations are schema-inconsistent, research suggests that they are likely to be recalled because they attract attention and require increased resources to be integrated to the script, therefore comprising a strong trace in the memory network (Anderson, 1983; Brewer & Nakamura, 1984). In other words, deviations are more memorable because they violate the script and are thus distinctive and salient (Davidson, 2006; see also Cohen & Java, 1995; Means & Loftus, 1991). There is also evidence that deviations can improve source monitoring because they serve as tags for specific instances *("script pointer plus tag" hypothesis*; Graesser et al., 1980). Recent studies that have manipulated the presence of deviations in a target instance of repeated events show that deviations can improve recall for that instance (targeted effect; MacLean et al., 2018), or even for all the events (general effect; Connolly et al., 2016; MacLean et al., 2018). If deviations from the script improve recall and particularization of repeated events, their occurrence can have implications for information elicitation.

Fuzzy-Trace Theory (FTT; Brainerd & Reyna, 1990) offers a similar conceptualization to schema theory regarding the retrieval of repeated events. Fuzzy-Trace Theory suggests that experiences are encoded and stored in memory in the form of two traces: gist, which in the context of repeated events refers to an overall understanding of what typically occurs (e.g., general recurring details); and verbatim, which represent instance-specific details (e.g., sources of variations; deviations). Research shows that gist and verbatim traces are retrieved separately via different cues, and that verbatim traces are more sensitive to forgetting than gist traces, so that memory over time tends to rely on gist (*principle of retrieval dissociation*; Brainerd & Reyna, 2001; 2004). In sum, although they differ regarding the recall of deviations, both FTT and schema theory suggest that specific details of separate instances are less likely to be accessed over time compared to the general routine of repeated events. This is true either because specific details fade out from memory more rapidly, or because they become absorbed by the memory of the routine itself.

Interviewing techniques to facilitate recall and reporting

In line with the theoretical work reviewed above, research has tested various cues to facilitate recall about both the general routine and specific details of repeated events – although cues are usually tested discretely rather than in combination. Most methods capitalize on the thematical and temporal organisation of autobiographical memories, according to which specific events are thought to be hierarchically nested within summarised, and extended events (Conway & Pleydell-Pearce, 2000). For instance, research suggests that inquiring about the overall number of witnessed events first, before asking about specific instances leads to more elaborate reporting (Connolly & Gordon, 2014). Another recommendation is that open and directive prompts are used, in that order, to inquire about the frequency of specific actions and variations (Brubacher

et al., 2014). In survey methodology, visual timelines have been effectively used to elicit information about major thematic events (e.g., relationships, work) over prolonged time periods (*event history calendars*; Belli, 1998; Van der Vaart & Glasner, 2007; Yoshihama et al., 2005). With respect to more mundane events, Means and Loftus (1991) interviewed participants about repeated health care visits, and found that asking participants to think about specific elements of each visit (e.g., type of doctor, weather etc.) and then construct a personal timeline improved the amount of recall and dating accuracy of separate instances relative to reporting following a free recall. More recently, Leins et al. (2014) found that the use of a timeline can benefit the reporting of details of family gatherings when administered with various mnemonics, such as the 'family tree mnemonic' (cf. unaided recall), as participants may identify meetings in relation to other temporal markers. For instance, participants were prompted with derived cues that asked about normative events likely shared by members of the same social background (e.g., Thanksgiving). They were then asked to report why such gatherings occurred in their family, and these personally relevant reasons were used as cues to prompt retrieval. A similar cuing technique was used in Willen et al. (2015), where context-specific cues were derived from the most salient details that participants remembered from a series of dental visits. When other participants used the context-specific cues from a similar perspective to theirs, they reported more specific instance details than when they used general cues, such as times and dates.

With respect to facilitating recall and particularization of details, the current research investigated how the use of self-generated cues (Kontogianni et al., 2018; Wheeler & Gabbert, 2017), the timeline technique (Hope et al., 2013), and open-ended questions (Kontogianni et al., 2020) can be used in conjunction as a multi-method interviewing format. The use of the timeline should effectively cue the retrieval of

details that are temporally and thematically associated in the context of repeated events (Belli, 1998). To date, the timeline technique has been found to enhance the reporting of accurate details, sequential information, and information about attributions of actions to multiple perpetrators in single events; both as a stand-alone technique (Hope et al., 2013) and in conjunction with the self-generated cues mnemonic (Kontogianni et al., 2018). To promote 'top-down' retrieval, we used a modified timeline so that interviewees are first asked to outline the number of the witnessed events before being asked to use the self-generated cues and describe each instance on a separate format (see also Hope et al., 2019).

Self-generated cues prompt interviewees to list the most salient details from each instance, and thus facilitate the retrieval of closely associated memories (Wheeler & Gabbert, 2017). Previous research has shown that when participants are asked to generate their own cues during encoding of target items, cues with distinctive properties are more specific than experimenter-generated general descriptions (Mantyla & Nilsson, 1988; Tullis & Benjamin, 2015). Anderson and Conway (1993) have also shown that participants tend to list distinctive details about previously experienced events first, followed by other thematically related details. Thus, evidence shows that reliable cues both reinstate the context of the experienced event (Tulving & Thomson, 1973) and provide diagnostic information about specific memories (*principle of cue overload*; Goh & Lu, 2012; Nairne, 2002). Based on previous applied research, we expected that the use of self-generated cues would facilitate the discrimination of specific instances thereby improving recall while reducing source confusion (Brubacher et al., 2018; Willen et al., 2015).

Interference between instances might lead to source intrusions and the reporting of inconsistent details (e.g., confusing a perpetrator's actions in one instance based on

what was witnessed in another instance) and omission errors (see also Lindsay, 2014). Indeed, practitioners often follow-up on an interviewee's account to clarify what has been reported and to address information gaps (Shepherd & Griffiths, 2013). There is evidence that the use of open questions that prompt separate instances in depth, rather than in breadth, is associated with the reporting of more specific (cf. general routine) details (Brubacher et al., 2012). In addition, their use could prompt source-monitoring judgments about specific instances, thus reducing the reporting of internal intrusions (Lindsay et al., 2004; Oeberst & Blank, 2012). Given how follow-up open questions are related to improved particularization and instance discrimination, they were incorporated in the MMIF as prompts that are tailored to the interviewees' accounts rather than as pre-set questions (*witness-compatible questioning*; Fisher & Geiselman, 1992; for a review see Oxburgh et al., 2010). Given the use of evidence-based cues, more accurate information and fewer internal intrusions should be reported in the MMIF and in the SGC-Timeline condition than in the Free Recall condition. However, the use of follow-up prompts should further improve reporting and source monitoring in the MMIF condition.

Similar to previous research on repeated events, some target activities were manipulated to change from one instance to another, while some remained stable across events (e.g., Brubacher et al., 2012; McNichol et al., 1999). To examine if the presence of deviations at encoding facilitates recall and particularization, half of the participants witnessed four events on separate occasions, where the third instance included two deviations (one changed and one novel detail). We expected that participants in the 'deviation present' condition would recall more correct information both for the third instance (targeted effect; e.g., Mac Lean et al., 2018) and across all events (general

effect; e.g., Connolly et al., 2016), compared to participants in the 'deviation absent' condition where no deviations were introduced.

Method

Participants and Design

A total of 150 participants (121 Females, Age: M = 21.26, SD = 5.21, Range 18-44 years) were randomly allocated to a 3 (Reporting format: Multi-Method Interviewing Format (MMIF) vs. SGC-Timeline vs. Free recall) x 2 (Deviation: Present vs. Absent) between-subjects design. Participants were recruited through the student participation pool and advertisements circulated across campus and were granted course credit or a £7 honorarium for participating. Overall, 164 participants were recruited but 14 did not attend all the sessions and so were excluded from analyses. Dependent variables were the number of correct details, correct gist and verbatim details, accuracy rates for all types of reported details, and intrusion errors. Based on previous findings of the initial testing of the timeline technique on the reporting of attributions of actions and statements to people, we also included this dependent variable (i.e., number of reported attributions and accuracy rate) to examine reporting across repeated events.

Materials

Stimulus events. Five stimulus events were scripted and filmed. Each event was a short film, 4-5 minutes long, depicting a meeting between four perpetrators (three males, one female) who plot a terrorist attack and then proceed to carry out the plan. Each event was shot from a first-person perspective to facilitate the cover story that the participant is an undercover agent acting as a group member. In each film, the leader delivers information to the perpetrators about the target of the attack and assigns the

following roles to each member: one member will oversee the operation and provide the detonator (a mobile phone) to another member who will plant the explosives; the third member will act as a look out, and the participant will be the getaway driver. There is a discussion about the explosives, how they are to be detonated and when. The meetings, which constitute the first part of the event, take place indoors and were all shot in the same location. In the second part of the event, the perpetrators are seen arriving at the selected location, which differed in each film, and act according to their assigned roles before they leave in a getaway car. The first part of the indoor conversations was highly similar across events, although specific details of the content of the discussions varied. The second part of the outdoor activities was similar in the overall structure, and on the general level i.e., people involved, actions performed, but the location and direction of people's movements varied (see Table 1 for all variable details across events; see Table 2 in Supplemental materials for a visual representation of the events).

Insert Table 1

To implement the deviation, an alternative version of one of the four events was developed including two deviations: i) a 'role switch': one of the four perpetrators, who always has the role of the lookout during the operation, was also in charge of the meeting, while the perpetrator who always has the role of the leader simply attended the meeting with the other members; and ii) a 'new character': when carrying out the plan, the perpetrator in charge of planting the explosives gestured to the female overseeing the operation to convey that there was a problem with the explosives. The female was seen making a phone call, and a woman, who was not seen in the other events, briefly appeared and handed her an envelope. This alternative version was only presented in the Deviation Present condition and was always presented third to avoid primacy or recency effects on recall. The presentation order of all the other events in both the Deviation

Present and Deviation Absent conditions was counterbalanced across participants to avoid order effects. Analysis using a univariate analysis of variance showed that there was no effect of the counterbalancing stream of 16 iterations used on the reporting of correct details, $F(15, 149) = .92$, $p = .544$, $\eta^2 = .09$, nor on the reporting of correct gist details, $F(15, 149) = 1.37$, $p = .170$, $\eta^2 = .13$.

Timeline Reporting Format. The two-level timeline format for reporting repeated events consisted of: (i) a "Scoping" timeline (33 in. x 12 in.) which depicts a horizontal line running at mid-point from one end of the card to the other to provide an overview of all the experienced events; (ii) "Specific event" timelines of the same size and layout (33 in. x 12 in.) which depicts a horizontal line running at mid-point from one end to the other representing the temporal space for each event; (iii) Person Description cards (5 in. x 3 in. white lined cards); (iv) Action cards (3 in. x 3 in.): yellow cards (semi-adhesive strip on the back for easy removal and rearrangement on the timeline); and (v) Statement cards (5 in. x 3 in.): blank, pink cards.

Follow-up open-ended questions. A protocol of open-ended in-depth questions was composed to prompt additional information based on the initial account, in relation to omitted information, gaps, and inconsistencies/clarifications (see Table 2).

Insert Table 2

Procedure

Stimuli administration. Participants visited the lab to witness four events on four separate occasions over the span of seven days (minimum one day and maximum four days between visits). On each occasion participants were instructed to imagine that they were an undercover agent who had infiltrated a terrorist group. They were asked to pay attention because they would later have to provide a report on the activities of the

group, which would be passed on to intelligence analysts. Participants witnessed the events on a computer screen while wearing headphones. After witnessing the final event, participants were invited to return after a seven-day delay (*M* = 7.29 days, *SD* = 0.53) to provide an account of the events. All participants provided their account approximately two weeks after the first visit and one week after the last visit to the lab.

Interview. When participants returned to provide their account, they were either provided with instructions for the Multi-Method Interviewing Format (MMIF), the SGC - Timeline, or the Free recall format. All participants were reminded that they are in the role of an undercover agent who infiltrated a terrorist group and that they are in possession of valuable intelligence information about the activities of the group. When asked to describe each event, all participants across conditions received the same general instruction: 'report all the details you remember about the events and the people involved; report exactly what was said when possible'. Also, all participants were instructed to not guess about things they could not remember.

In the MMIF condition, participants were instructed to begin by outlining all the events in the order they witnessed them and to then focus on each instance. Participants could use a card to label each event and place them on the scoping timeline. Next, following Kontogianni et al. (2018), participants received the self-generated cue instructions: *'Without thinking too hard, write down the first six things that you remember seeing or thinking when witnessing each event. It doesn't matter what these things are. All that is important is that they immediately come to mind when thinking back to each event. Please list them on a piece of paper. Think about each of the things in your list one at a time and think about whether that memory helps you remember other things that also happened in the event'*. Participants were instructed to list the self-generated cues for each individual event, prior to using a timeline to describe each

event in detail. At that point, they were instructed to use the person description cards to report descriptive details about each person, and the action and statement cards to report action, sequence information and any statements they remembered. They were also instructed to link the cards to show "who did/said what and when".

After they finished providing their account in written form, participants were asked follow-up open-ended questions. The interviewer asked three to four questions per event, so that all participants across conditions were asked an equivalent number of questions. Consistent with best practice interviewing guidance, the questions topics were not pre-selected, instead participants were asked open questions based on what they had reported in their account. For instance, if the participant had mentioned the leader of the group, the interviewer asked "You mentioned there was a leader of the group. Tell me more about this leader." Or, if explosives had been mentioned in the initial account, the interviewer asked, "Explain in more detail what you mean about this part where they discussed the explosives". This procedure allowed for interviewers to maintain the same phrasing of questions but avoid using a scripted list of cued recall questions not related to the witness's initial account. Participants were not required to answer all the questions and if they replied by saying "I don't know" or "I don't remember", the interviewer moved on to the next question. Finally, all participants were asked if there was anything else they would like to report. During the questioning phase, the participant's written account remained in sight and the interviewer pointed to the specific part to which the prompt referred when asking each question. The follow-up questioning phase was audio and video-recorded, but the camera was only focused on the participant's written account. Participants in the SGC-Timeline condition followed the exact same procedure in reporting their initial account in written form, but they were not asked any follow-up questions. Participants in the Free recall condition were

instructed to outline the events prior to describing each in detail, but they did not receive specific instructions on how to outline the events. After providing their account in written form, they were asked follow-up questions according to the procedure outlined above. At the end, participants were debriefed and compensated for their time.

Coding

Following Hope et al. (2019), a coding protocol was developed for the stimuli events. Each detail reported was coded as a person, action, object or setting detail. Details were coded as accurate[2] if they were present in each corresponding stimulus event and described correctly. Particularisation refers to the reporting of (highly) specific details as well as to the increased reporting of details that are specific to each instance of repeated events (Brubacher & Roy, 2014; Powell et al. 2007). Use of the current coding scheme allowed for particularisation to be reflected in our measure of correct recall. Details that were specific to each instance (e.g., about the location and placement of explosives; the equipment used) were reported and coded as following: 'the female (1-P) walked (1-A) into the building (1-S), into the lobby (1-S) and to the right (1-S)'; 'carrying (1-A) a black (1-O) backpack (1-O) with red markings (1-O) and a mobile phone (1-O).' Each instance took place in a different location and setting, therefore this example refers to the only one occasion where the female perpetrator placed explosives in a building lobby. Details that were vague or subjective (e.g., "he was young", "he looked satisfied") were not scored for accuracy. As the events included a conversation among the perpetrators, interviews were also coded for gist and verbatim statements,

[2] We acknowledge the potential for two definitions of reported accuracy (narrow and broad) as described in the meta-analysis by Woiwod et al. (2019). Given the framing on particularization and the hypotheses of the current research, we formulated our instructions and developed our coding scheme to deliberately focus on the amount of correct information about each instance of the repeated events.

based on the script that was developed for the stimulus events. Gist details reflected the overall meaning of what was discussed (e.g., “the leader told everyone what their role was”) and were scored as one point for each correct gist unit (i.e., correct extraction of the conversation that was not reported verbatim) and one point for each incorrect gist unit (i.e., incorrect extraction). If the gist statement was reported in a vague manner, it was not scored for accuracy (e.g., “they talked about doing something”). Verbatim details reflected the precise language used in the original stimulus. Verbatim units were scored as correct for every three verbatim words reported correctly and as incorrect when two or fewer words corresponded to the script. In the initial account, additional coding was conducted for the accuracy of attributions of both actions and statements to a person. Attribution details were scored as correct when an action or statement was correctly attributed to a specific actor (e.g., Female handed the detonator over).

All the accounts were coded for internal intrusions (i.e., source monitoring errors) by noting the type of the reported detail (person, action, object, setting, gist, verbatim, location, target, time; see Table 1) and the source of the stimulus event where it was witnessed. For example, if a participant reported that the “target in the Spinnaker tower was a local activist”, that would be coded as an intrusion as this was the target in the event at Victoria park. Therefore, if the event at Victoria park was witnessed third, this would be scored under intrusions as “1-Target Event-3”. The same coding scheme was used to code the responses to the follow-up questions, apart from coding for the attribution details which were only relevant to the initial reporting phase.

Coding of the interviews was mostly conducted by the lead researcher and partly by two research assistants. Twenty-four interviews (i.e., 15% of all interviews) were randomly selected and independently scored by the second author, who was blind to experimental conditions for Deviation (to some extent also for Format; i.e., between the

MMIF and SGC-Timeline conditions which used the same reporting format). Inter-rater reliability, which was high across coding categories, *ICC* = .97, 95% CI [.967, .974], was computed based on the mean value of two raters, using an absolute agreement definition and a two-way mixed effects model (McGraw & Wong, 1996).

Statistical Analyses

To examine the extent to which the independent variables predicted correct reporting, the dependent variables were analysed using linear mixed models (LMMs) with fixed effects of reporting format (categorical: MMIF vs. SGC-Timeline vs. Free recall) and deviation (categorical: Absent vs Present), and random intercepts for events nested within participants[3] (Finch et al., 2014). We arrived at this model by comparing: i) a baseline model with fixed effects only; and ii) a model with fixed effects and random intercepts for events nested within participants. We found that the second model was a better fit for the data, by conducting a likelihood ratio test (LRT, function *anova*) comparing the log-likelihoods of both models (for all the statistical comparisons of the models see Supplemental Materials). The model included two-way interactions between deviation and reporting format. We used simple contrasts to code the reporting format and the deviation. Specifically, for format, contrasts compared reporting with Free recall (reference level) to reporting with SGC-Timeline, and Free recall to MMIF. A separate model was used with a contrast comparing MMIF to SGC-Timeline. For deviation, the contrasts compared Absent vs Present. The reported coefficients (*b* value estimates) show the degree to which the dependent variable changed relative to the reference level, while the 95% CI represent the plausible range of the value of the

[3] The only exception was for the analysis of the number of intrusions reported across events, where random intercepts for events were nested within type of intrusions within participants.

regression coefficients (Cumming, 2013). With respect to accuracy rates, we were interested in the overall accuracy reported across events rather than within each instance, and thus analysis was not conducted with LMMs, but with factorial ANOVAs. Accuracy rates were calculated by dividing the number of correct details reported by the total number of details (correct and incorrect) reported.

The analyses were run in R version 3.5.0 (R Core Team, 2017) using the `lme` function from the `nlme` package (Pinheiro et al., 2017). Datasets, and R markdown scripts are available on the Open Science Framework website (https://osf.io/4mcsa/?view_only=85498381b88e44a881f5014b27976b8b).

Results

We present the results of the analysis on the number of correct details, related accuracy rate, and number of gist details for total reporting[4] first, and then examine the initial reports and the reporting of internal intrusions. In the initial reporting phase, we also present secondary analyses with respect to the number of correct attributions of actions and statements to people, and accuracy rate of attribution details. In the interest of parsimony, results for verbatim details and accuracy rates for both gist and verbatim details compared for initial and total reporting, are presented in Supplemental Materials, as no significant results emerged across conditions.

Total interview output

Participants in the MMIF condition were asked a similar number of follow-up questions

[4] In line with the current research aims and hypotheses, a model was built to examine total reporting instead of reporting in the follow-up phase, in order to compare the effects of all reporting formats across events in examining the reporting of information per instance of repeated events, while testing the effectiveness of each reporting format as a whole. Meanwhile, the information gain in the follow-up phase is also evident with the use of this model.

(M = 13.78, SD = 1.58) to the participants in the Free recall condition (M = 13.16, SD = 1.60). An independent t-test analysis showed that there was no statistically significant difference in the number of follow-up questions asked between groups, $t(98) = 1.95$, $p = .054$.

Reporting Format had a significant main effect on the reporting of total correct details, with more correct details being reported in the MMIF condition, $b = 7.48$, 95% CI [0.98, 13.98], $t(144) = 4.42$, $p < .001$, and in the SGC-Timeline condition, $b = 5.48$, 95% CI [-1.02, 11.98], $t(144) = 3.24$, $p = .001$ than in the Free Recall condition. Reporting of total correct details did not differ between the SGC-Timeline and the MMIF condition, $b = 2.00$, 95% CI [-4.50, 8.50], $t(144) = 1.18$, $p = .240$. Neither Deviation, $b = 0.95$, 95% CI [-6.26, 4.36], $t(144) = 0.69$, $p = .491$, nor the interaction between Deviation and Reporting format were significant: i) MMIF vs Free Recall, $b = 0.46$, $t(144) = 0.14$, $p = .889$; ii) MMIF vs SGC-Timeline, $b = 0.18$, $t(144) = 0.05$, $p = .960$; iii) SGC-Timeline vs Free Recall, $b = 0.28$, $t(144) = 0.08$, $p = .936$. Therefore, the use of different formats only affected the reporting of correct details (see Figure 1, for results at initial and follow-up reporting).

Insert Figure 1

A separate model was built to examine if there is a targeted effect of Deviation on the reporting of correct details for the instance that contained the deviations. Deviation did not have a significant effect on the reporting of correct details, $b = 2.53$, 95% CI [-1.10, 6.17], $t(144) = 1.38$, $p = .170$. Therefore, the presence of deviations did not affect the reporting of details overall (general effect on recall) or for the specific instance (target effect on recall). Figure 2 shows the reporting of correct details across events within deviation conditions (total reporting).

Insert Figure 2

A factorial ANOVA showed that neither Reporting Format, $F(1, 96) = 0.21$, $p = .650$, $\omega^2 = {}^{-}.009$, nor Deviation, $F(1, 96) = 1.04$, $p = .311$, $\omega^2 = .010$, had a significant main effect on the overall accuracy of accounts. The interaction between Deviation and Reporting format was not significant, $F(1, 96) = 0.93$, $p = .337$, $\omega^2 = {}^{-}.000$.

A repeated-measures analysis of variance showed that the accuracy rate of the reported information in the follow-up questioning phase was significantly lower than the accuracy rate in the initial reporting phase, $F(1, 98) = 32.54$, $p < .001$, $\omega^2 = .245$. However, accuracy in the follow-up questioning phase was not affected by the format participants used to provide their initial report, $F(1, 98) = 1.38$, $p = .243$, $\omega^2 = .003$. Means and standard deviations for accuracy rates can be found in Table 3.

Insert Table 3

Reporting format had a significant effect on the reporting of correct gist details, with more correct gist details reported in the MMIF than in both the Free recall, $b = 0.78$, 95% *CI* [0.09, 1.46], $t(144) = 2.24$, $p = .026$, and the SGC-Timeline condition, $b = 1.47$, 95% *CI* [0.14, 2.79], $t(144) = 4.24$, $p < .001$. More correct gist details were reported in the Free recall than in the SGC-Timeline condition, $b = 0.69$, 95% *CI* [-1.37, -0.01], $t(144) = 2.00$, $p = .047$. Deviation did not have an effect on the reporting of correct gist details $b = 0.06$, 95% *CI* [-0.50, 0.32], $t(144) = 0.21$, $p = .832$. The interaction between Deviation and Reporting format was not significant: i) MMIF vs Free Recall, $b = 0.47$, $t(144) = 0.68$, $p = .498$; ii) MMIF vs SGC-Timeline, $b = 0.79$, $t(144) = 1.14$, $p = .256$; iii) SGC-Timeline vs Free Recall, $b = .32$, $t(144) = 0.46$, $p = .646$.

Figure 3 presents the reporting of correct gist details as a function of Reporting format within Deviation conditions. Figure 4 shows the reporting of correct gist details across events within Deviation conditions (total reporting).

Insert Figure 3

Insert Figure 4

Initial reporting phase

Reporting format had a significant main effect on the number of correct details reported, with more correct details reported initially in the MMIF condition, $b = 9.04$, 95% *CI* [2.89, 15.19], $t(144) = 5.65$, $p < .001$, and in the SGC-Timeline condition, $b = 12.47$, 95% *CI* [6.32, 18.62], $t(144) = 7.80$, $p < .001$ (cf. Free Recall). More correct details were reported in the SGC-Timeline than in the MMIF condition, $b = 3.43$, 95% *CI* [-2.72, 9.58], $t(144) = 2.14$, $p = .034$. Neither Deviation, $b = 1.03$, 95% *CI* [-6.05, 3.98], $t(144) = 0.79$, $p = .431$, nor an interaction between Deviation and Reporting format were significant: i) MMIF vs Free Recall, $b = 1.50$, $t(144) = 0.47$, $p = .640$; ii) MMIF vs SGC-Timeline, $b = 0.58$, $t(144) = 0.18$, $p = .857$; iii) SGC-Timeline vs Free Recall, $b = 0.92$, $t(144) = 0.29$, $p = .772$.

A factorial ANOVA showed a significant main effect of Reporting format on the accuracy rate of reported details, $F(2,144) = 3.43$, $p = .035$, $\omega^2 = .032$. Bonferroni-adjusted post-hoc pairwise comparisons showed that there was a significant difference ($p = .035$) between the MMIF ($M = 0.87$, $SD = 0.06$) and the Free recall condition ($M = 0.83$, $SD = 0.08$), but not between the SGC-Timeline and Free recall conditions ($p = .228$). There was not a significant difference between the MMIF and SGC-Timeline conditions ($p = 1.00$). There was no significant main effect of Deviation on the accuracy rate $F(1,144) = 0.14$, $p = .709$, $\omega^2 = -.006$, and the interaction between Reporting format and Deviation was not significant $F(2, 144) = 2.47$, $p = .088$, $\omega^2 = .020$.

Reporting format had a significant main effect on the reporting of gist details, with more correct gist details reported initially in the MMIF than the Free recall format, $b = 0.67$, 95% *CI* [0.02, 1.32], $t(144) = 2.03$, $p = .044$. Reporting of correct gist details did not differ between the MMIF and SGC-Timeline conditions, $b = 0.21$, 95% *CI* [-0.86, 0.44], $t(144) = 0.64$, $p = .523$, or between the SGC-Timeline and Free recall conditions, $b = 0.46$, 95% *CI* [-0.19, 1.11], $t(144) = 1.39$, $p = .167$. Deviation did not have a significant main effect on the reporting of gist details ($p = .758$) and the interaction between Reporting format and Deviation was not significant: i) MMIF vs Free Recall, $b = 0.22$, $t(144) = 0.33$, $p = .742$; ii) MMIF vs SGC-Timeline, $b = 0.70$, $t(144) = 1.06$, $p = .291$; iii) SGC-Timeline vs Free Recall, $b = 0.48$, $t(144) = 0.73$, $p = .467$.

Reporting format had a significant main effect on the reporting of correct attributions of actions and statements to persons, with more attributions reported in the MMIF, $b = 1.58$, 95% *CI* [0.72, 2.43], $t(144) = 3.63$, $p < .001$, and the SGC-Timeline conditions, $b = 2.38$, 95% *CI* [1.52, 3.24], $t(144) = 5.49$, $p < .001$, than in the Free recall condition. Reporting was not significantly different in SGC-Timeline and MMIF conditions, $b = 0.81$, 95% *CI* [-0.05, 1.66], $t(144) = 1.86$, $p = .065$. Deviation did not have a significant main effect on the correct reporting of attributions, $b = 0.23$, 95% *CI* [-0.93, 0.47], $t(144) = 0.66$, $p = .510$, and the interaction between Deviation and Reporting format was not significant, i) MMIF vs Free Recall, $b = 0.07$, $t(144) = 0.08$, $p = .936$; ii) MMIF vs SGC-Timeline, $b = 0.25$, $t(144) = 0.29$, $p = .773$; iii) SGC-Timeline vs Free Recall, $b = 0.18$, $t(144) = 0.21$, $p = .834$. Figure 5 shows the number of correct attributions made as function of Reporting format within Deviation conditions.

Insert Figure 5

A factorial ANOVA showed that Reporting format had a main effect on the accuracy rate of the reported attribution details, $F(2,144) = 4.72$, $p = .010$, $\omega^2 = .048$. Deviation did not have a significant effect, $F(1,144) = 0.18$, $p = .672$, $\omega^2 = -.315$, and the interaction between Reporting format and Deviation was not significant, $F(2,144) = 0.11$, $p = .897$, $\omega^2 = -.161$. Levene's test was significant ($p = .010$). Bonferroni-adjusted post-hoc pairwise comparisons showed that there were significantly higher accuracy rates of reported attributions in the MMIF ($M = 0.73$, $SD = 0.16$) than in the Free recall condition ($M = 0.61$, $SD = 0.29$), $p = .023$, and in the SGC-Timeline ($M = 0.73$, $SD = 0.19$) than in the Free recall condition ($M = 0.61$, $SD = 0.29$), $p = .028$, with no difference between the MMIF and SGC-Timeline conditions ($p = 1.00$).

Internal intrusions across instances

Table 4 shows the distribution of the mean number of internal intrusions in each event across conditions.

Insert Table 4

Neither Deviation, $b = 0.01$, 95% *CI* [-0.06, 0.05], $t(144) = 0.44$, $p = .661$, nor Reporting format had a significant main effect on the reporting of intrusions: i) MMIF vs Free Recall, $b = 0.02$, 95% *CI* [-0.04, 0.08], $t(144) = 1.23$, $p = .221$; ii) MMIF vs SGC-Timeline, $b = 0.02$, 95% *CI* [-0.04, 0.08], $t(144) = 1.23$, $p = .221$; iii) SGC-Timeline vs Free Recall, $b = 0.00$, 95% *CI* [-0.06, 0.06], $t(144) = 0.00$, $p = 1.00$. The interaction between Deviation and Reporting format was significant for i) SGC-Timeline vs Free Recall, $b = 0.08$, $t(144) = 2.31$, $p = .022$ but not for: ii) MMIF vs Free Recall, $b = 0.02$, $t(144) = 0.54$, $p = .590$; iii) MMIF vs SGC-Timeline, $b = 0.06$, $t(144) = 1.77$, $p = .079$.

Discussion

The current study examined the effectiveness of the timeline technique used in combination with self-generated cues and follow-up open-ended questions, as a multi-method interviewing format, for facilitating recall and particularization of repeated events (i.e., correct information specifically reported for each instance). The findings show that participants reported overall more correct information about specific instances when using the MMIF than a free recall format followed by open-ended questions, without a cost to accuracy. Findings for the MMIF and SGC-timeline conditions were comparable with respect to the reporting of information about the attacks, but as more information was reported in the MMIF condition about the conversations around planning the attack, reporting was overall higher in the MMIF (cf. SGC-Timeline) condition.[5]

Although the responses to follow-up questions were not as accurate as the information provided in the initial accounts, previous research suggests that an increase in the reporting of both correct details and errors is likely when output increases overall (Memon et al., 2010; Roberts & Higham, 2002; see also, Kontogianni et al., 2020 for a replication of this result). Importantly, the use of follow-up questions about each instance facilitated the reporting of more episodic information, and accuracy rates were similar across reporting format conditions. For instance, the use of follow-up questions allowed participants to elaborate on some details, thus facilitating further reporting, but more effectively so when participants had already provided a detailed account using

[5] Although there was a main effect of Reporting format on the time of reporting, $F(2, 142) = 76.96$, $p < .001$, Bonferroni post-hoc comparisons show that the reporting time in the Free recall condition was significantly lower than in the MMIF and SGC-Timeline conditions ($p < .001$), but that reporting time did not differ between the MMIF and SGC-Timeline conditions ($p = 1.00$). Therefore, the use of follow-up open questions in the MMIF condition did not result in a significantly longer interview, compared to the other condition where the timeline technique was used.

effective mnemonics as shown in the MMIF condition.

A closer analysis of the initial reporting phase shows that the initial accounts were more detailed about each instance when participants used the self-generated cues in conjunction with the timeline technique rather than a free recall format. Based on previous research comparing the separate individual components of the timeline technique (i.e., full timeline, record cards, temporal context instructions; Experiment 2; Hope et al., 2013) to a free recall format, it is likely that the mnemonic benefits inherent in the timeline format and the use of separate timelines per instance in the current study facilitated the retrieval of specific details (see also Hope et al., 2019). It is possible that particularization was further facilitated by the use of the self-generated cues, which were used to prompt the retrieval of the salient details prior to describing each instance in detail (Brubacher et al., 2011). However, based on the current method we cannot be certain about the contribution of the use of the self-generated cues to overall performance. Consistent with previous findings, participants who used the timeline technique also reported more correct attributions of actions and statements to perpetrators than those who used a free recall format (Hope et al., 2013; 2019). Regarding the use of free recall, and potential concerns that any additional intervention would facilitate retrieval, we note that this comparison allowed the examination of self-administered reporting and matching of the general instructions across conditions. Also, participants in the free recall condition still had the opportunity to report more information in response to follow-up questions compared to participants in the SGC-Timeline condition. The current results provide further evidence of the usefulness of the timeline technique in conjunction with self-generated cues for multi-actor events in the reporting of both single and repeated experiences.

The increased reporting and particularization of instances in both conditions where the timeline technique and self-generated cues were used (cf. free recall) occurred without an increased cost in the reporting of internal intrusions. Given evidence that the use of open in-depth questions per instance elicit more specific details and facilitate discrimination between instances (see also, Brubacher et al., 2012), fewer intrusions were expected in the MMIF relative to the SGC-Timeline and to the free recall conditions. However, the mean number of intrusions was low across conditions. Notably, the number of intrusions varied by detail type. Participants most commonly confused the equipment used across events and the perpetrators who alternated roles in planting the explosives, with the latter being a particularly pertinent detail for interviewers who are interested in knowing 'what happened when' or 'who did what when' (Roberts, 2003). Considering that internal intrusions are unlikely to be completely avoided, the current findings suggest that corroboration is necessary for details that interviewees have difficulty attributing to an instance. Inquiring about which instance interviewees remember the best, and why, could help us better understand how interviewees encode and process information about repeated events (Danby et al., 2017). Given the current null finding, further research is necessary to determine the precise impact of the use of the MMIF components on source monitoring.

Contrary to our hypotheses, the presence of deviations had no effect on either recall or on particularization of specific instances. Accurate reporting was improved when memory-enhancing techniques were used (cf. free recall format), yet recall did not further benefit from the presence of deviations. The current results are inconsistent with previous research, which shows that the presence of deviations facilitates recall for all the witnessed events and for the targeted instance itself (Connolly et al., 2016; MacLean et al., 2018; although see Rubinova et al., 2020). Also, witnessing an instance that

deviated from the script did not facilitate source monitoring as participants reported a similar number of internal intrusions in both deviation conditions. One explanation for our results may be that the current deviations were not sufficiently salient[6] to impact recall. To this end, we could have included a manipulation check to directly inquire if participants encoded the deviations (e.g., MacLean et al., 2018), as it is difficult to reach any conclusions based on what participants reported alone.

Research suggests that the effect of deviations on recall may depend on whether the deviation has any consequence to the sequence of events (i.e., *continuous vs discrete deviations*; Connolly et al., 2016; *obstacles, errors, and distractions*; Schank & Abelson, 1977). Although the comparison between specific types of deviations was beyond the focus of our study, exploring how deviations are implemented across studies and their differential effect on recall might be relevant for future research. Notably, in recent studies showing improved recall for deviations, both children and adult participants witnessed the repeated events over one or two days (e.g., Connolly et al., 2016; MacLean et al., 2018; although see Brubacher et al., 2012) and usually with recall occurring after shorter intervals than in the current study. It is possible, then, that after the one-week interval, our participants could not access the deviation details. Further research on the effects of deviations on memory is needed to determine if the lack of an observed effect depends more on the salience of the deviations, or the delay until reporting. An interaction between the two factors is also conceivable.

---

[6] We can report that although the 'changed' detail (i.e., a different person leading the group meeting on one occasion) was reported by participants, the 'le 1new' detail (i.e., a new person briefly entering the scene in the park, passing by one of the perpetrators to hand her an envelope) was not reported. Possibly, the 'new' detail was not conspicuous enough in a busy scene; or maybe participants used the existing script to notice the 'changed' compared to the 'new' detail, since the script for the former was likely stronger than the latter, which occurred during the second and more variable part of the events (*schema-confirmation-deployment hypothesis*; Farrar & Goodman, 1992).

Lastly, it is possible that a larger sample size may be required to detect an effect of the presence of deviations on recall. Key reasons, identified before conducting the research, prevented us from conducting an a priori analysis. First, in the repeated events literature, recall is often measured in a different way than in the current experiment (e.g., reporting of pre-determined fixed vs variable details). Second, power calculations for Linear Mixed Models require the input of parameter estimates that are often not reported in the literature and/or are tied to the stimuli used (Kumle et al., 2020). These estimates were not available to the current research.

The current findings suggest that self-generated cues, the timeline technique, and follow-up open-ended questions can be used together to elicit detailed and accurate information for repeated events. With respect to the manipulation of deviations, our results suggest that further research should investigate how deviations affect (or not) delayed recall to improve reporting and discrimination between instances. The current research contributes to the development of an adaptive information gathering 'toolbox' of techniques that can be used flexibly in applied settings. However, further research is needed to address the challenges that both interviewers and interviewees face when investigating repeated events, including the implications of interference between instances for reporting.

Acknowledgements

This work was funded by the Centre of Research and Evidence on Security Threats (ESRC Award: ES/N009614/1) and the University of Portsmouth as a postgraduate bursary for the doctoral work of the first author.

References

Abelson, R. P. (1981). Psychological status of the script concept. *American Pychologist*, *36*(7), 715-729. http://dx.doi.org/10.1037/0003-066X.36.7.715

Ahn, W. K., Brewer, W. F., & Mooney, R. J. (1992). Schema acquisition from a single example. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*(2), 391-412. http://dx.doi.org/10.1037/0278-7393.18.2.391

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior, 22*(3)*,* 261–295. doi:10.1016/S0022-5371(83)90201-3

Anderson, S. J., & Conway, M. A. (1993). Investigating the structure of autobiographical memories. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19*(5), 1178–1196. https://doi.org/10.1037/0278-7393.19.5.1178

Belli, R. F. (1998). The structure of autobiographical memory and the event history calendar: Potential improvements in the quality of retrospective reports in surveys. *Memory*, *6*(4), 383-406. doi:10.1080/741942610

Brainerd, C. J., & Reyna, V. F. (2004). Fuzzy-trace theory and memory development. *Developmental Review*, *24*(4), 396-439. https://doi.org/10.1016/j.dr.2004.08.005

Brainerd, C. J., & Reyna, V. F. (2001). *Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience.* In H. W. Reese & R. Kail (Eds.), *Advances in child development and behavior* (p. 41–100). Academic Press.

Brainerd, C. J., & Reyna, V. F. (1990). Gist is the grist: Fuzzy-trace theory and the new intuitionism. *Developmental Review*, *10*(1), 3-47. https://doi.org/10.1016/0273-2297(90)90003-M

Brewer, W. F., & Nakamura, G. V. (1984). The nature and functions of schemas. *Center for the Study of Reading Technical Report; no. 325*. Illinois, USA: University of Illinois at Urbana-Champaign, Center for the Study of Reading. Bolt Beranek and Newman, Inc. Retrieved from: http://hdl.handle.net/2142/17542

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*(5), 600-609. https://doi.org/10.1002/acp.3440

Brubacher, S. P., Glisic, U., Roberts, K. P., & Powell, M. (2011). Children's ability to recall unique aspects of one occurrence of a repeated event. *Applied Cognitive Psychology*, *25*(3), 351-358. https://doi.org/10.1002/acp.1696

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. http://dx.doi.org/10.1037/law0000011

Brubacher, S. P., Roberts, K. P., & Powell, M. (2012). Retrieval of episodic versus generic information: Does the order of recall affect the amount and accuracy of details reported by children about repeated events?. *Developmental Psychology*, *48*(1), 111-122. https://doi.org/10.1037/a0025864

Brubacher, S. P., & La Rooy, D. (2014). Witness recall across repeated interviews in a case of repeated abuse. *Child abuse & neglect*, *38*(2), 202-211. https://doi.org/10.1016/j.chiabu.2013.06.010

Cohen, G., & Java, R. (1995). Memory for medical history: Accuracy of recall. *Applied Cognitive Psychology*, *9*(4), 273-288. https://doi.org/10.1002/acp.2350090402

Connolly, D. A., & Gordon, H. M. (2014). Can order of general and specific memory prompts help children to recall an instance of a repeated event that was different from the others?. *Psychology, Crime & Law*, *20*(9), 852-864. https://doi.org/10.1080/1068316X.2014.885969

Connolly, D. A., Gordon, H. M., Woiwod, D. M., & Price, H. L. (2016). What children recall about a repeated event when one instance is different from the others. *Developmental Psychology, 52*(7), 1038-1051. https://doi.org/10.1037/dev0000137

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*(2), 205-223. https://doi.org/10.1002/1099-0720(200103/04)15:2<205::AID-ACP698>3.0.CO;2-F

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological review*, *107*(2), 261-288. https://doi.org/10.1037/0033-295X.107.2.261

Cumming, G. (2013). *Understanding the new statistics: Effect sizes, confidence intervals, and meta-analysis*. Routledge, USA.

Danby, M. C., Brubacher, S. P., Sharman, S. J., Powell, M. B., & Roberts, K. P. (2017). Children's reasoning about which episode of a repeated event is best remembered. *Applied Cognitive Psychology*, *31*(1), 99-108. https://doi.org/10.1002/acp.3306

Davidson, D. (2006). Memory for bizarre and other unusual events: Evidence from script research. In R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 157–179). Oxford University.

Farrar, M. J., & Boyer-Pennington, M. E. (1999). Remembering specific episodes of a scripted event. *Journal of Experimental Child Psychology*, *73*(4), 266-288. https://doi.org/10.1006/jecp.1999.2507

Finch, W. H., Bolin, J. E., & Kelley, K. (2014). *Multilevel modeling using R*. CRC Press.

Fisher, R., & Geiselman, R. (1992). *Memory-enhancing techniques for investigative Interviewing: The cognitive interview*. Charles C. Thomas.

Goh, W. D., & Lu, S. H. (2012). Testing the myth of the encoding–retrieval match. *Memory & cognition*, *40*(1), 28-39. https://doi.org/10.3758/s13421-011-0133-9

Graesser, A. C., Woll, S. B., Kowalski, D. J., & Smith, D. A. (1980). Memory for typical and atypical actions in scripted activities. *Journal of Experimental Psychology: Human Learning and Memory*, *6*(5), 503.

Hope, L., Gabbert, F., Kinninger, M., Kontogianni, F., Bracey, A., & Hanger, A. (2019). Who said what and when? A timeline approach to eliciting information and intelligence about conversations, plots and plans. *Law and Human Behavior, 43*(3), 263-277. http://dx.doi.org/10.1037/lhb0000329

Hope, L., Mullis, R., & Gabbert, F. (2013). Who? What? When? Using a timeline technique to facilitate recall of a complex event. *Journal of Applied Research in Memory and Cognition, 2*(1), 20-24. doi:10.1016/j.jarmac.2013.01.002

Hudson, J. A., Fivush, R., & Kuebli, J. (1992). Scripts and episodes: The development of event memory. *Applied Cognitive Psychology*, *6*(6), 483-505. 10.1002/acp.2350060604

Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*(1), 3-28.

Kontogianni, F., Hope, L., Taylor, P. J., Vrij, A., & Gabbert, F. (2018). The benefits of a self-generated cue mnemonic for timeline interviewing. *Journal of Applied Research in Memory and Cognition*. *7*(3), 454-461. 10.1016/j.jarmac.2018.03.006

Kontogianni, F., Hope, L., Taylor, P. J., Vrij, A., & Gabbert, F. (2020). "Tell me more about this…": An examination of the efficacy of follow-up open questions following an initial account. *Applied Cognitive Psychology*. https://doi.org/10.1002/acp.3675

Kumle, L., Vo, M. L. H., & Draschkow, D. (2020). *Estimating power in (generalized) linear mixed models: an open introduction and tutorial in R*. PsyArXiv, https://doi.org/10.31234/osf.io/vxfbh

Leins, D. A., Fisher, R. P., Pludwinski, L., Rivard, J., & Robertson, B. (2014). Interview protocols to facilitate human intelligence sources' recollections of meetings. *Applied Cognitive Psychology*, *28*(6), 926-935. https://doi.org/10.1002/acp.3041

Lindsay, D. S., (2014). Memory Source monitoring applied. In Perfect, T. J. &. Lindsay, D. S (Eds.), *The SAGE Handbook of Applied Memory* (59-75). SAGE Publications Inc.

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for source. *Memory & Cognition*, *17*(3), 349-358. https://doi.org/10.3758/BF03198473

MacLean, C. L., Coburn, P. I., Chong, K., & Connolly, D. L. (2018). Breaking script: Deviations and postevent information in adult memory for a repeated event. *Applied Cognitive Psychology*, *32*(4), 474-486. https://doi.org/10.1002/acp.3421

Mäntylä, T., & Nilsson, L.-G. (1988). Cue distinctiveness and forgetting: Effectiveness of self-generated retrieval cues in delayed recall. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14*(3), 502–509. https://doi.org/10.1037/0278-7393.14.3.502

McGraw, K. O., & Wong, S. P. (1996). "Forming inferences about some intraclass correlations coefficients": Correction. *Psychological Methods, 1*(4), 390. https://doi.org/10.1037/1082-989X.1.4.390

McNichol, S., Shute, R., & Tucker, A. (1999). Children's eyewitness memory for a repeated event. *Child Abuse & Neglect*, *23*(11), 1127-1139. https://doi.org/10.1016/S0145-2134(99)00084-8

Means, B. & Loftus, E. (1991). When personal history repeats itself: Decomposing memories for recurring events. *Applied Cognitive Psychology*, *5*(4), 297–318. https://doi.org/10.1002/acp.2350050402

Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law*, *16*(4), 340-372. https://doi.org/10.1037/a0020518

Nairne, J. S. (2002). The myth of the encoding-retrieval match. *Memory, 10*(5-6), 389-395. doi:10.1080/09658210244000216

Oeberst, A., & Blank, H. (2012). Undoing suggestive influence on memory: The reversibility of the eyewitness misinformation effect. *Cognition*, *125*(2), 141-159. https://doi.org/10.1016/j.cognition.2012.07.009

Oxburgh, G. E., Myklebust, T., & Grant, T. (2010). The question of question types in police interviews: a review of the literature from a psychological and linguistic perspective. *International Journal of Speech, Language & the Law*, *17*(1), 45-66. 10.1558/ijsll.v17i1.45

Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., & R Core Team (2017). *nlme: Linear and Nonlinear Mixed Effects Models. R package version 3.1-131*. Retrieved from https://CRAN.R-project.org/package=nlme.

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. https://doi.org/10.1080/10345329.2007.12036415

Price, H. L., & Connolly, D. A. (2013). Suggestibility effects persist after one year in children who experienced a single or repeated event. *Journal of Applied Research in Memory and Cognition*, *2*(2), 89-94. https://doi.org/10.1016/j.jarmac.2013.03.001

R Core Team. (2017). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Roberts, A. (2003). Questions of who was there and who did what: The application of code D when a suspect disputes participation but not presence. *Criminal Law Review*, 709–716.

Roberts, K. P., Brubacher, S. P., Drohan-Jennings, D., Glisic, U., Powell, M. B., & Friedman, W. J. (2015). Developmental differences in the ability to provide temporal information about repeated events. *Applied Cognitive Psychology*, *29*(3), 407-417. https://doi.org/10.1002/acp.3118

Roberts, W. T., & Higham, P. A. (2002). Selecting accurate statements from the cognitive interview using confidence ratings. *Journal of Experimental Psychology: Applied*, *8*(1), 33-43. http://dx.doi.org/10.1037/1076-898X.8.1.33

Rubínová, E., Blank, H., Koppel, J., & Ost, J. (2020). Schema and deviation effects in remembering repeated unfamiliar stories. *British Journal of Psychology*. Advance online publication. https://doi.org/10.1111/bjop.12449

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures.* Erlbaum.

Shepherd, E. & Griffiths, A. (2013). *Investigative Interviewing: The Conversation Management Approach* (second edition). Oxford University Press.

Theunissen, T. P., Meyer, T., Memon, A., & Weinsheimer, C. C. (2017). Adult eyewitness memory for single versus repeated traumatic events. *Applied cognitive psychology*, *31*(2), 164-174. https://doi.org/10.1002/acp.3314

Tullis, J. G., & Benjamin, A. S. (2015). Cueing others' memories. *Memory & cognition*, *43*(4), 634-646. https://doi.org/10.3758/s13421-014-0478-y

Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*(5), 352–373. https://doi.org/10.1037/h0020071

Van der Vaart, W., & Glasner, T. (2007). Applying a timeline as a recall aid in a telephone survey: A record check study. *Applied Cognitive Psychology*, *21*(2), 227–238. https://doi.org/10.1002/acp.1338

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

Wheeler, R. L., & Gabbert, F. (2017). Using self-generated cues to facilitate recall: a narrative review. *Frontiers in psychology*, *8*, 1830. https://doi.org/10.3389/fpsyg.2017.01830

Willén, R. M., Granhag, P. A., Strömwall, L. A., & Fisher, R. P. (2015). Facilitating particularization of repeated similar events with context-specific cues. *Scandinavian Journal of Psychology*, *56*(1), 28-37. https://doi.org/10.1111/sjop.12180

Woiwod, D. M., Fitzgerald, R. J., Sheahan, C. L., Price, H. L., & Connolly, D. A. (2019). A meta-analysis of differences in children's reports of single and repeated events. *Law and human behavior*, *43*(1), 99-116. https://doi.org/10.1037/lhb0000312

Yoshihama, M., Gillespie, B., Hammock, A. C., Belli, R. F., & Tolman, R. M. (2005). Does the life history calendar method facilitate the recall of intimate partner violence? Comparison of two methods of data collection. *Social Work Research*, *29*(3), 151-163. https://doi.org/10.1093/swr/29.3.151

Table 1. Overview of superordinate recurring details across events and variations within each instance (in both deviation conditions).

Type of detail	**Events**			
	1	**2**	**3**	**4**
Target	The head of the National reform party	Swedish left-wing sociologist	Local activist	Investigative journalist
Attack Location	Spinnaker Tower	Portland building	Victoria park	Portsmouth & Southsea train station
	Base of the tower	Road in front of the building	North-west end of park	(Train arriving) from London
	Back of the café	Space under the stairs	Café lodge	Café inside the station
	Construction equipment	-	In bushes	Next to the sofa
	Seafront hotel	-	Premier hotel	Ibis hotel
	Two blocks away	The Hard	Closest bus stop	Square
	Within 900m	Within 900m	Within 900m	Within 800m

	Hotel lobby	-	Hotel lobby	Parking
Object/ equipment	M48 Mortar rounds	3 kilos	Six packages	9 kilos
	-	C-4	Semtex	106 rounds
	Improvised Explosive Device	-	Improvised Explosive Device	Improvised Explosive Device
	Nokia 3210	Nokia 6210	Nokia 5210	Nokia 3310
	Backpack	Backpack	Briefcase	Sports rucksack
Time	Wednesday	Thursday	Saturday	Monday
	10 am	6 pm	12 o'clock	9 am
	10.30 am	6.10 am	Half an hour before	Two hours before
	6.30 am	-	Half an hour upon start (of the event)	10 minutes
	-	-	9 am	6 am
	Until the party starts	Until the reception starts	After the blast	After the blast
Operational roles				
Person planting bomb	Niko (male)	Nina (female)	Niko (male)	Nina (female)
Lookout	Taking pictures	Not taking pictures	Taking pictures	Not taking pictures
Passing the detonator	Nina hands detonator to Niko over discussion[7]	Niko hands detonator to Nina[8]	Nina hands detonator to Niko over discussion	Niko hands detonator to Nina

1

2

3

[7] Nina asks Niko for directions and uses a map to conceal the detonator which she passes over to him

[8] Niko hands the detonator to Nina without any interaction

Table 2. Protocol for follow-up open-ended questions to extend and clarify information provided in the initial account.

1	Tell me more about (the part when/ person/object/activity) …
2	(You mentioned)…Tell me everything/ every detail about the part when …
3	What else can you tell me about …?
4	Explain in more detail what you mean about (this part where…)
5	Describe in more detail (this part when…)

Table 3. Mean and SDs accuracy rates of the initial reports and of the responses to the follow-up questions across format conditions.

	Initial report		**Follow-up questions**	
Format	Mean Accuracy rate	*SD*	Mean Accuracy rate	*SD*
MMIF	0.87	0.06	0.78	0.10
SGC-Timeline	0.86	0.08	-	-
Free Recall	0.83	0.08	0.79	0.10

Table 4. Mean number and SDs of intrusions reported across format and deviation conditions.

	Deviation Absent			**Deviation Present**		
	MMIF	**SGC-Timeline**	**Free Recall**	**MMIF**	**SGC-Timeline**	**Free Recall**
Events	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*	*M (SD)*
First	.84 (1.28)	.96 (1.10)	.88 (1.20)	.84 (1.18)	.84 (.99)	.88 (1.01)
Second	.92 (1.08)	1.28 (1.17)	.92 (1.22)	1.00 (1.12)	.52 (.82)	.96 (.98)
Third	.96 (1.06)	.88 (.97)	.56 (.96)	1.52 (1.23)	.68 (.95)	.92 (1.26)
Fourth	1.08 (1.12)	.92 (.91)	.52 (.71)	.64 (.95)	.56 (.87)	.80 (.82)

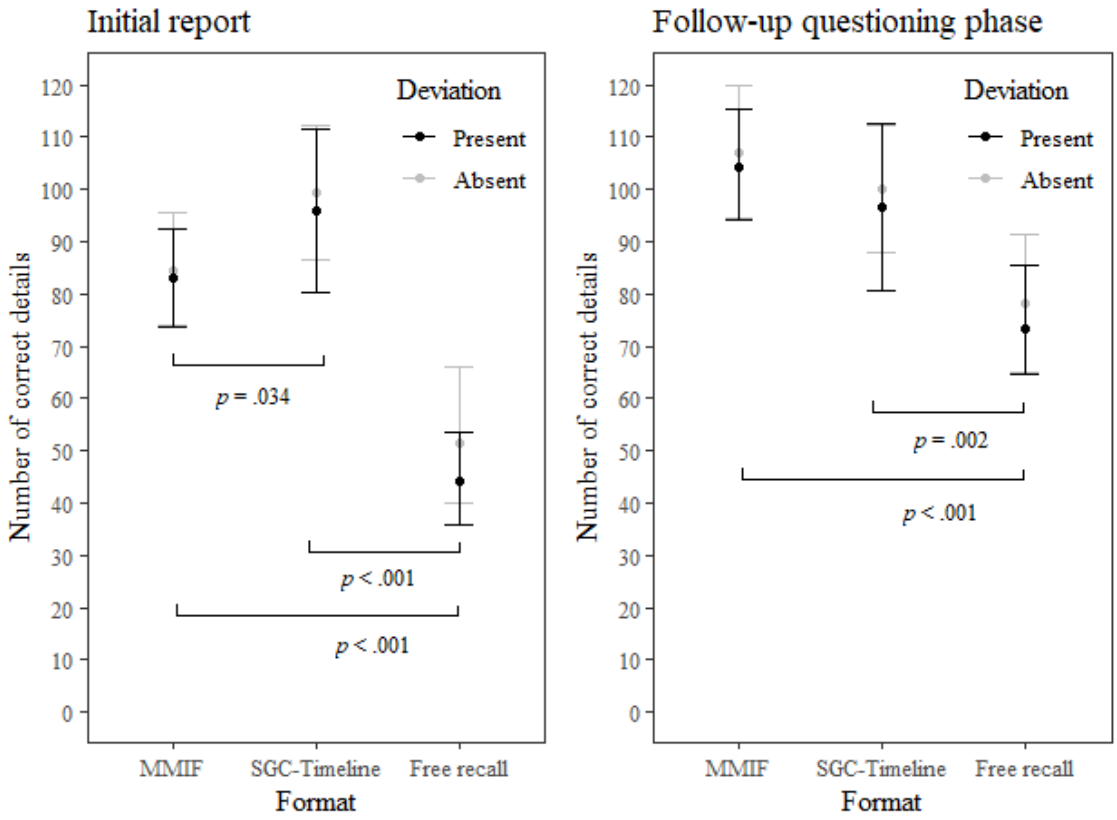

Figure 1. Number of correct details as a function of Reporting format within Present and Absent deviation conditions. Error bars represent 95% Confidence Intervals. Note: reporting in the SGC-Timeline condition remained stable (i.e., no follow-up questions were used).

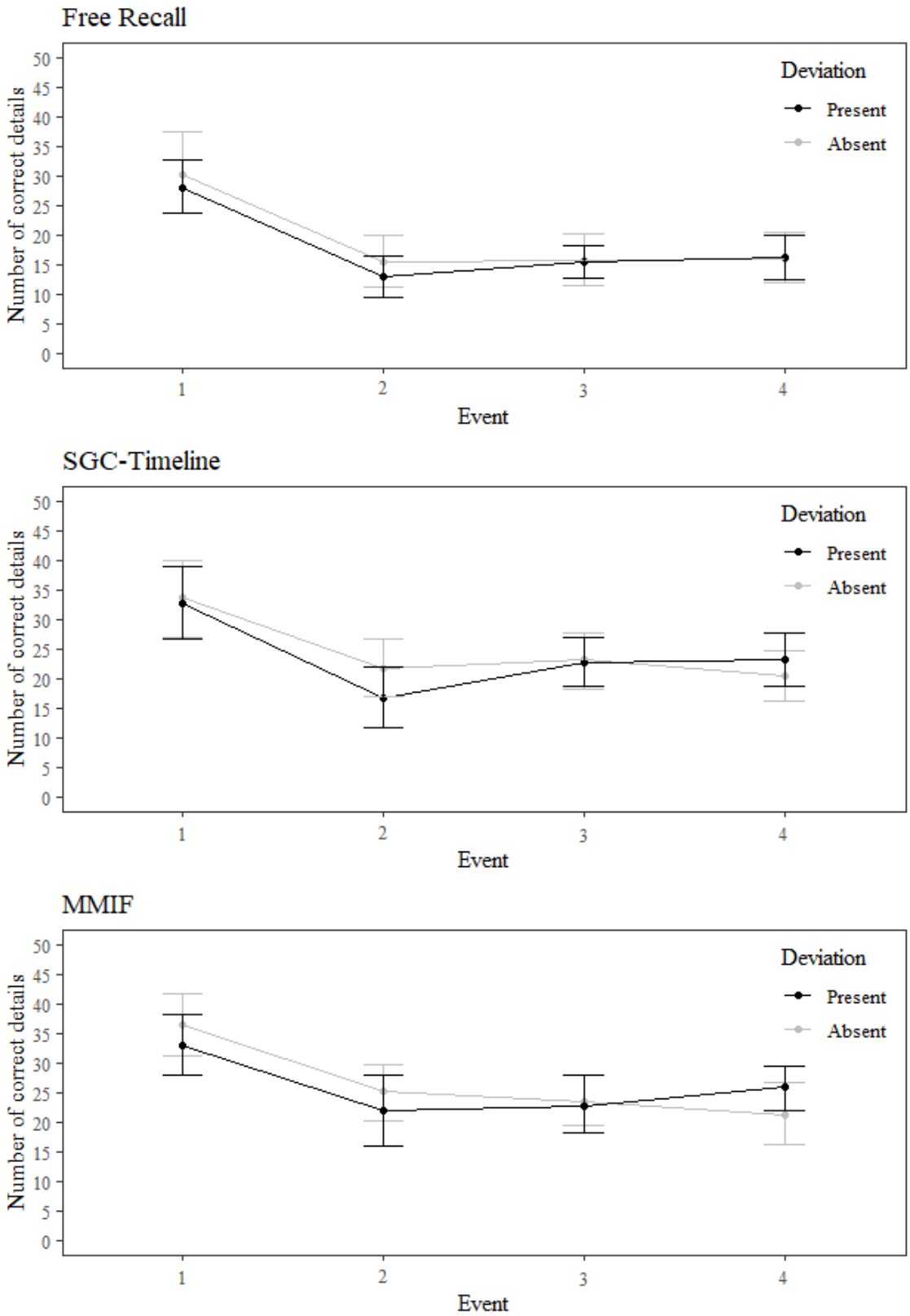

Figure 2. Number of correct details reported across events within Present and Absent deviation conditions for each format. Error bars represent 95% Confidence Intervals.

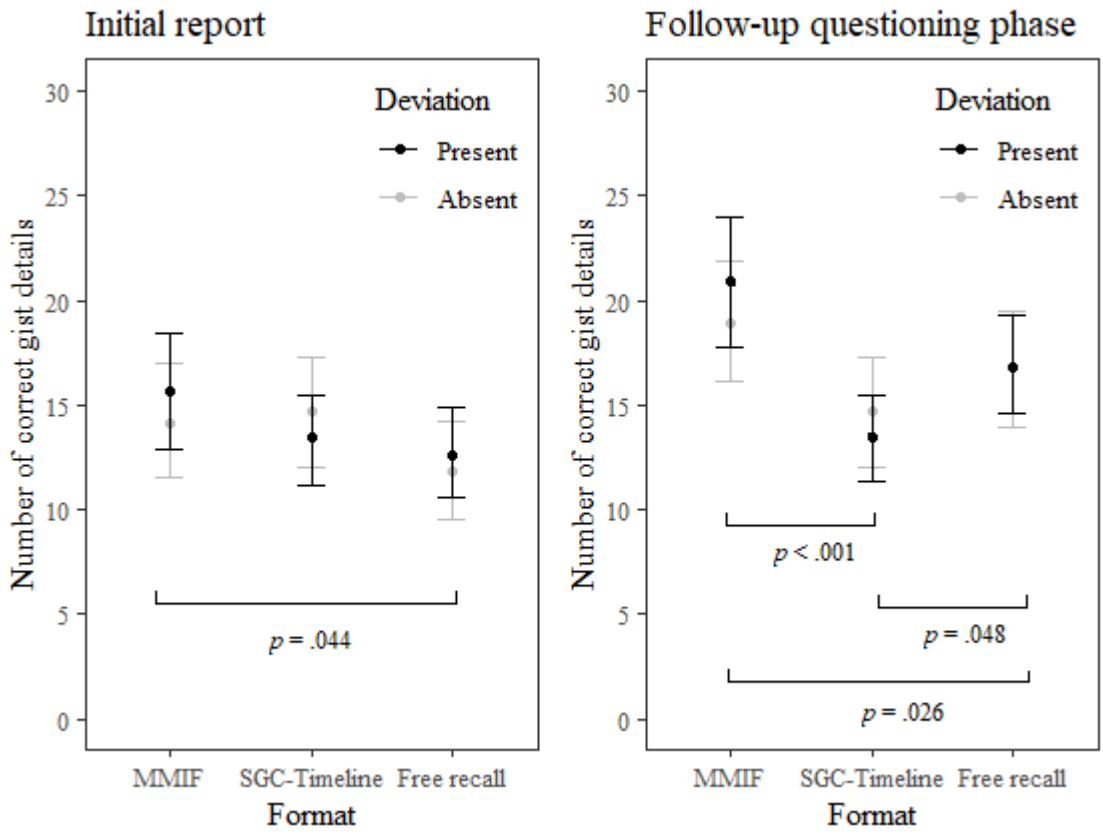

Figure 3. Number of correct gist details as a function of format within Present and Absent deviation conditions. Error bars represent 95% Confidence Intervals. Note: reporting in the SGC-Timeline condition remained stable (i.e., no follow-up questions were used).

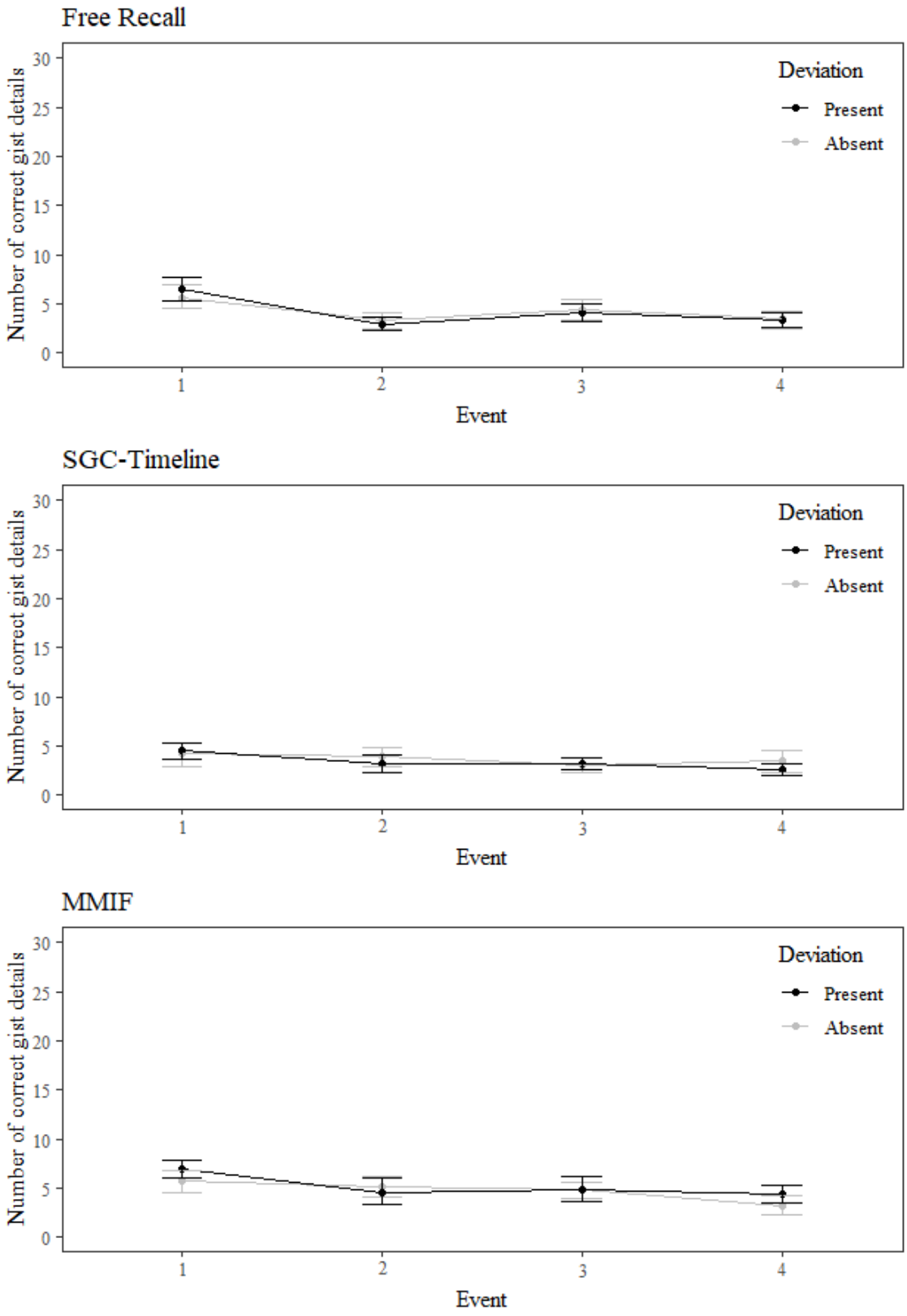

Figure 4. Number of correct gist details reported across events within Present and Absent deviation conditions for each format. Error bars represent 95% Confidence Intervals.

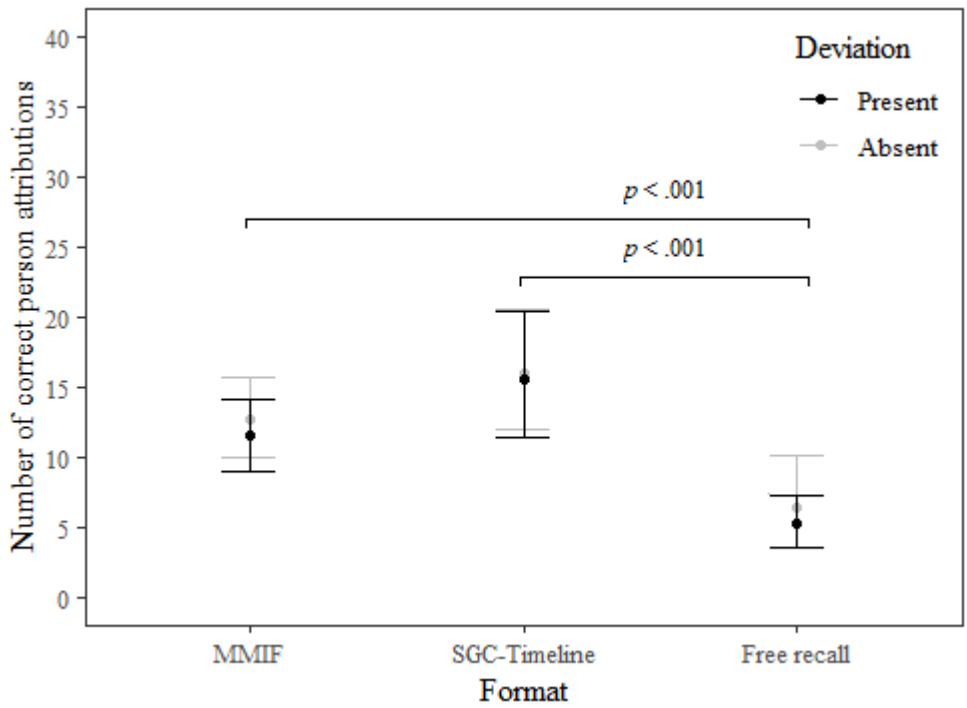

Figure 5. Number of correct attributions made as a function of format within Present and Absent deviation conditions. Error bars represent 95% Confidence Intervals.