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

Feni Kontogianni, Eva Rubinova, Lorraine Hope, Paul J. Taylor, Aldert Vrij & Fiona Gabbert


To link to this article: https://doi.org/10.1080/09658211.2021.1903508

© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group

Published online: 24 Mar 2021.

Article views: 194

View Crossmark data
Facilitating recall and particularisation of repeated events in adults using a multi-method interviewing format

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

a Department of Psychology, University of Portsmouth, Portsmouth, UK; b Department of Psychology, Lancaster University, Lancaster, United Kingdom; cUniversity of Twente, Enschede, Netherlands; dDepartment of Psychology, Goldsmiths University of London, London, UK

ABSTRACT

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 particularisation 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.

ARTICLE HISTORY
Received 7 August 2020
Accepted 10 March 2021

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., particularisation; see p. 203; Brubacher & La Rooy, 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; Willén 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 characterised 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 particularisation of repeated events, their occurrence can have implications for information elicitation.

Fuzzy-Trace Theory (FTT; Brainerd & Reyna, 1990) offers a similar conceptualisation 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 capitalise 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 pro-
longed time periods (event history calendars; Belli, 1998;
van der Vaart & Glasner, 2007; Yoshihama et al., 2005).
With respect to more mundane events, Means and Lofthus (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 time-
line 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 mnemon-
ics, 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 norm-
ative 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 Willén et al. (2015), where context-specific
cues were derived from the most salient details that par-
ticipants 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 particularisation
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 interview format. The use of the timeline should effec-
tively 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 attribu-
tions 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 sepa-
rate format (see also Hope et al., 2019).

Self-generated cues prompt interviewees to list the
most salient details from each instance, and thus facilit-
ate 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 (Mäntylä &
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; Willén 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 omis-
sion 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, 2014; Oeberst & Blank, 2012). Given
how follow-up open questions are related to improved
particularisation and instance discrimination, they were
incorporated in the MMIF as prompts that are tailored
to the interviewees’ accounts rather than as pre-set ques-
tions (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 facil-
tiates recall and particularisation, 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., MacLean 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) × 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 min 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).

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. × 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. × 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. × 3 in. white lined cards); (iv) Action cards (3 in. × 3 in.): yellow cards (semi-adhesive strip on the back for easy removal and rearrangement on the timeline); and (v) Statement cards (5 in. × 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 information gaps, and inconsistencies/clarifications (see 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 timing timeline. Next, following Kontogianni et al. (2018), participants received the self-generated cue instructions:

Without thinking too hard, write down the 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 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 & La Rooy, 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, 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 (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 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/).

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 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 (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).

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).

A factorial ANOVA showed that neither Reporting Format, F(1, 96) = 0.21, p = .650, ω² = .009, nor Deviation, F(1, 96) = 1.04, p = .311, ω² = .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, ω² = .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, ω² = .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, ω² = .003. Means and standard deviations for accuracy rates can be found in 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, $b = 0.32, t(144) = 0.46, p = .646$

Figure 4 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).

**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 ($p = .001$). 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

---

**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 3.** Number of correct details reported initially in the MMIF compared to the SGC-Timeline and Free recall conditions, 95% CI $[6.05, 1.91]$, $t(144) = 2.38, p = .020$. Deviation did not have a significant main effect ($p = .772$) and the interaction between Reporting format and Deviation was not significant ($p = .758$).

**Figure 4.** Number of correct details reported in the MMIF and SGC-Timeline conditions compared to the Free recall condition, 95% CI $[2.89, 15.19]$, $t(144) = 5.65, p < .001$. Reporting format had a significant main effect on the reporting of gist details, with more correct gist details reported in the MMIF than in the Free recall condition.
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 a function of Reporting format within Deviation conditions.

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, $F(2, 144) = 0.11, p = .897, \omega^2 = -.161$. Figure 2. Number of correct details reported across events within Present and Absent deviation conditions for each format. Error bars represent 95% Confidence Intervals.

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

<table>
<thead>
<tr>
<th>Format</th>
<th>Initial report</th>
<th>Follow-up questions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean accuracy rate</td>
<td>SD</td>
</tr>
<tr>
<td>MMIF</td>
<td>0.87</td>
<td>0.06</td>
</tr>
<tr>
<td>SGC-Timeline</td>
<td>0.86</td>
<td>0.08</td>
</tr>
<tr>
<td>Free Recall</td>
<td>0.83</td>
<td>0.08</td>
</tr>
</tbody>
</table>
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.

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) = 1.00, 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 particularisation of repeated events (i.e., correct information specific to 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.4

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 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 particularisation 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 particularisation 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 elicits more specific details and facilitates 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 particularisation 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

<table>
<thead>
<tr>
<th>Events</th>
<th>Deviation absent</th>
<th>Deviation present</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MMIF M (SD)</td>
<td>SGC-Timeline M (SD)</td>
</tr>
<tr>
<td>First</td>
<td>.84 (.128)</td>
<td>.96 (.10)</td>
</tr>
<tr>
<td>Second</td>
<td>.92 (.098)</td>
<td>1.28 (.17)</td>
</tr>
<tr>
<td>Third</td>
<td>.96 (.066)</td>
<td>.88 (.97)</td>
</tr>
<tr>
<td>Fourth</td>
<td>1.08 (.12)</td>
<td>.92 (.91)</td>
</tr>
</tbody>
</table>

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

Table 4. Mean number and SDs of intrusions reported across format and deviation conditions.
deviations facilitates recall for all the witnessed events and for the targeted instance itself (Connolly et al., 2016; MacLean et al., 2018; although see Rubinvová et al., 2021). 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 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.

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 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.

Notes
1. We acknowledge the potential for two definitions of reported accuracy (narrow and broad) as described in the meta-analysis by Woivod et al. (2019). Given the framing on particularisation 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.
2. 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.
3. To compare the differences between the Reporting formats as a whole regarding the information reported per instance, we built a model that examined total reporting rather than reporting in each interview phase. This approach was in line with the current research aims and hypotheses. The information gain in the follow-up phase is also evident with the use of this model (see Figures 1 and 3).
4. 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.
5. 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 ‘new’ 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).

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.

Disclosure statement
No potential conflict of interest was reported by the author(s).

Funding
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; Economic and Social Research Council.
References


---


