"Tell me more about this…": An examination of the efficacy of follow-up open questions following an initial account

Journal:	*Applied Cognitive Psychology*
Manuscript ID	ACP-19-0165.R1
Wiley - Manuscript type:	Research Article
Keywords:	follow-up questions, eliciting information, timeline technique, accuracy-informativeness trade-off

Title: "Tell me more about this…": An examination of the efficacy of follow-up open questions following an initial account

Short title: ***TESTING THE EFFICACY OF FOLLOW-UP QUESTIONS***

Abstract

In information gathering interviews, follow-up questions are asked to clarify and extend initial witness accounts. Across two experiments, we examined the efficacy of open-ended questions following an account about a multi-perpetrator event. In Experiment 1, 50 mock-witnesses used the timeline technique or a free recall format to provide an initial account. Although follow-up questions elicited new information (18% to 22% of the total output) across conditions, the response accuracy (60%) was significantly lower than that of the initial account (83%). In Experiment 2 (N = 60), half of the participants received pre-questioning instructions to monitor accuracy when responding to follow-up questions. New information was reported (21% to 22% of the total output) across conditions, but despite using pre-questioning instructions, response accuracy (75%) was again lower than the spontaneously reported information (87.5%). Follow-up open-ended questions prompt additional reporting; however, practitioners should be cautious to corroborate the accuracy of new reported details.

Keywords: follow-up questions, eliciting information, timeline technique, accuracy-informativeness trade-off

In both intelligence and criminal investigation contexts, interviewers commonly ask follow-up questions to elicit additional information, and to clarify reported details and inconsistencies (Evans & Fisher, 2011; Shepherd & Griffiths, 2013). Spontaneously reported information can be highly accurate but witnesses often omit critical details in their reports that may be useful in an investigation, thus interviewers may need to use follow-up questions (Hope, Gabbert, & Fraser, 2013; Roberts & Higham, 2002; Smeets, Candel, & Merckelbach, 2004). The current experiments examine the efficacy, in terms of both quantity and accuracy, of follow-up, open-ended questions that prompt interviewees for further information based on their initial account.

Follow-up questions to extend and clarify witness accounts are recommended in evidence-based interviewing protocols such as the Cognitive Interview (CI; Fisher & Geiselman, 1992). After requesting an initial free narrative about the event, interviewers can prompt for further information by using various memory-enhancing techniques, including a focused-retrieval phase where open questions are used to expand on aspects of the initial account (Fisher, 1995; Fisher & Geiselman, 1992). Building on the principles of the CI, recommendations for practice have been made about the use of appropriate prompts such as questions that start with "Tell", "Explain", and "Describe" (TED questions; for a review see Oxburgh, Myklebust, & Grant, 2010). In this context, appropriate questions are open-ended, information-seeking questions that prompt the interviewee to elaborate in depth on what has been mentioned already (Gabbert et al., 2016). In fact, in their recent description of an effective evidence-based model of interviewing for practitioners, Brandon, Wells, and Seale (2018) discuss how interviewers might prompt the reporting of additional information using elements of the CI with broad and, if needed, more specific questions.

Even when interviewees are cooperative, they are likely to omit or provide inconsistent details, particularly when reporting complex events. Although both omissions and inconsistencies occur naturally during retrieval, both have important implications in applied contexts. Details may be omitted due to forgetting or because further retrieval support is needed to access the encoded information. It may also be the case that interviewees are unaware of what details interviewers consider to be relevant (Fisher & Geiselman, 1992). Prompting for specific omitted details after an interviewee provides a free report to an open invitation for information can enable interviewers to elicit more details directly related to investigative objectives (Brandon et al., 2018).

Witnessing complex incidents, such as events involving multiple perpetrators, may result in the reporting of inconsistent, or otherwise disjointed information. Given that both within and between-statement inconsistencies are perceived as diagnostic of the reliability of witness accounts, interviewers might use prompts to assess the accuracy of the reported detail by giving the interviewee the opportunity to clarify an inconsistency (Berman, Narby, & Cutler, 1995; Smeets et al., 2004). In sum, the use of follow-up prompts can serve a number of functions in the interviewing process, by encouraging the interviewee to retrieve more information and to elaborate on their initial account.

The notion that follow-up questions prompt retrieval is based, broadly, on the spreading activation theory, which posits that memory is represented as a network of traces that vary in strength (Anderson, 1983). With each retrieval attempt, a trace is activated and, as a result, it spreads activation throughout the associated elements in the network. Therefore, the use of additional prompts can trigger a search through the memory network, facilitating access to additional memories which were not readily available before (see also Bower 1967). When a memory is not accessible by a particular prompt, a different prompt might be useful

(see also Anderson & Pichert, 1978). The use of open-ended, non-leading prompts that do not introduce new information but build on a free narrative should effectively encourage retrieval, since the information included in the question can act as a cue for the interviewee (Ibabe & Sporer, 2004). Thus, additional prompts following an initial retrieval may cue more memories and elicit more information.

That asking follow-up questions can lead to the elicitation of more information is neither new nor surprising. Results from meta-analyses on the effects of the CI on memory reporting show that use of the CI, which includes various mnemonics and additional prompts, results in improved reporting of correct details compared to standard interviews. However, there is also sometimes an increase in erroneous reporting as overall reporting increases (Kohnken, Milne, Memon, & Bull, 1999; Memon, Meissner, & Fraser, 2010). One likely explanation for this increase in inaccurate reporting relates to how effectively (or not) interviewees regulate their memory outputs (Koriat & Goldsmith, 1996; Memon et al., 2010). When asked to report information from memory, interviewees face competing demands to be both informative and accurate (Goldsmith, Koriat, & Weinberg-Eliezer, 2002; Koriat & Goldsmith, 1996). To achieve a balance between the two, research suggests that they tend to strategically monitor the amount of information they report (Koriat & Goldsmith, 1996). Specifically, in a free narrative, interviewees can decide to withhold or volunteer information based on how confident they are about the accuracy of that information.

Interviewees avoid errors by metacognitively assessing how likely it is that an answer is correct and, if it exceeds a pre-set accuracy threshold (*the satisficing model*; Goldsmith et al., 2002), they volunteer the answer or withhold it instead (*control of report option*; Koriat & Goldsmith, 1996). Thus, by controlling their responses, interviewees can be highly accurate, even after a delay in reporting (Goldsmith, Koriat, & Pansky, 2005). However, by choosing

to report information that is certainly correct, there is a cost to the total amount of reported information, resulting in an accuracy-informativeness trade-off (Brewer, Vagadia, Hope, & Gabbert, 2018; Goldsmith et al., 2002; Koriat & Goldsmith, 1996). Conversely, if interviewees attempt to be more informative, they risk reporting details that they are not as confident about, resulting in an increase in erroneous reporting.

Although the increased reporting of errors in the context of elaborate memory reports can be attributed to metacognitive monitoring, we do not have a clear understanding of where errors are most likely to spontaneously occur within the interviewing process, assuming recommended practice (e.g., use of open-ended questions). Research on the benefits of the CI for recall has mostly focused on the effectiveness of the different mnemonics rather than on the use of prompts following an initial narrative (e.g., Brunel & Launay, 2013; Colomb & Ginet, 2012; Memon, Wark, Bull, & Koehnken, 1997; Paulo, Albuquerque, & Bull, 2013). Similar to the use of cues, asking follow-up questions can also further prompt interviewees to search through their memory (Fisher & Geiselman, 2010). Yet, systematic investigation into witness performance when additional prompts are applied is limited or only incidentally reported across research on the development of investigative interviewing techniques. Research frequently focuses on the reporting of an initial account when testing a specific technique or, when an interviewing protocol with mnemonics and prompts is used, the results refer to the total information output across the entire interview but not within each interviewing phase (although see Memon et al., 1997; Paulo et al., 2013; Paulo, Albuquerque, Vitorino, & Bull, 2017).

Across two experiments, the current research examined the efficacy of using open-ended questions following a self-administered account, provided with either the timeline technique, which uses a physical timeline format and interactive instructions to facilitate

memory for multi-perpetrator events (Hope, Mullis, & Gabbert, 2013), or a free recall format. Although the timeline technique facilitates retrieval compared to free recall (Hope et al., 2013; Hope et al., 2019), it has not been tested in conjunction with follow-up questions – which would likely be used in real settings. Specifically, we sought to examine the number of new details reported about a witnessed event in response to follow-up questions and the accuracy of any new information reported (Experiment 1). In Experiment 2, in an attempt to refine the questioning procedure, we tested the use of instructions designed to promote accuracy monitoring in responding.

The objectives of Experiment 1 were exploratory, in that we aimed to assess the quantity and the quality of additionally reported information. Given that there is not a strong rationale in the literature to inform a directional hypothesis, there were no specific expectations about the reporting of additional information in response to prompts following an initial account provided with either reporting format. However, it was expected that the use of the timeline technique would elicit more correct details compared to the free recall format at the initial reporting phase (Hope et al., 2013). Open-ended questions were used as invitations to elaborate on omitted information and gaps (e.g., "Tell me more about [detail already mentioned]"; "What else can you tell me about [detail already mentioned]"; Brubacher, 2007; Gabbert et al., 2016) or inconsistencies in the written account (e.g. "You mention four perpetrators arriving at the location but three leaving, can you explain in more detail what you mean about this part?"). To ensure that the questions matched the interviewee's retrieval pattern (*witness-compatible questioning;* Fisher & Geiselman, 1992; Wells, Memon, & Penrod, 2006), the participant's own words were used when formulating the questions (e.g., "You mentioned there was a leader of the group. Tell me more about this leader").

General Method

Participants

Participants who were fluent or native English speakers, and aged between 18-49 years old, were eligible to participate in both experiments. Participants were recruited through the department's student participation pool and through advertisements circulated across campus.

Materials

Timeline reporting format. The timeline format consists of three elements: (i) A physical cardboard (33 in. x 12 in.) which depicts a horizontal line running at mid-point from one end of the card to the other; (ii) Person Description cards (5 in. x 3 in.): blank, white and lined cards; (iii) Action cards (3 in. x 3 in.): blank and yellow cards (semi-adhesive strip on the back for easy removal and rearrangement on the timeline).

Follow-up open-ended questions. A question protocol was composed to prompt additional information based on the initial account, in relation to omitted information, gaps, and inconsistencies/need to clarify (see Table 1).

Insert Table 1 here

Coding

Coding of the interviews in both experiments was conducted by the first author following the scoring template used in Kontogianni, Hope, Taylor, Vrij, & Gabbert (2018). Each detail reported about the witnessed events was identified as a Person (P), Action (A), Object (O), and Setting (S) detail. A detail was scored as correct if it was present in the event and described correctly. A detail was scored as incorrect if it was present in the event but described incorrectly or if it was not present in the event. Details that were subjective or vague were not coded. A secondary coding was conducted in Experiment 1 with respect to attributions of reported actions to specific actors. Person-action details were scored as correct

when an action was correctly attributed to a specific actor (e.g., Male with red shirt raises the crowbar). Sequencing errors were also noted when events were reported in the wrong order. For instance, if ABCD is correct, in ACBD, C would be coded as one sequence error as it should follow B, but B would not be counted as out of sequence too.

To assess inter-rater reliability across categories, 15% of the interviews in each experiment were randomly selected and coded by an independent rater. Given the use of different reporting formats in Experiment 1, coding was blind to hypotheses and research questions, while coding in Experiment 2 was also blind to experimental conditions. Inter-rater reliability was computed ICC based on the mean value of two raters, using an absolute agreement definition and a two-way mixed effects model, as the raters were fixed (McGraw & Wong, 1996). Inter-rater reliability was high, *ICC* = .99, 95% *CI* [.987, .993] (Exp. 1), *ICC* = .98, 95% *CI* [.965, .984] (Exp. 2).

Experiment 1

Method

Participants and Design

Fifty participants (37 Females, Age: M = 24.64, SD = 6.99, Range 18-47 years) were randomly allocated to a timeline (n = 25) or a free recall condition (n = 25). The dependent variables were the number of correct and incorrect details reported in each interview phase (initial report and follow-up questioning), the number of correct person to action details provided in the initial report, and the accuracy rates for both types of details. Accuracy rates were calculated by dividing the number of correct details reported by total details (correct and incorrect) reported to obtain the proportion of accurate responses.

Materials

Stimulus event. Participants witnessed a 1 min 20 s long film of a multi-perpetrator assault and robbery [see Hope et al. (2013) and Kontogianni et al., (2018) for previous use of this stimulus]. The film starts with three males talking next to a parked car. Two other males join them. A woman with a laptop bag is seen walking in their direction. As she tries to walk past, they surround her, and one male is seen threatening her with a crowbar. One male takes her bag, which is then passed between several perpetrators, while another male films the incident on his phone. The perpetrators run away with the bag.

Procedure

Participants were asked to take part in a study investigating factors that affect people's reports for witnessed events. Participants witnessed the stimulus event on a computer screen while wearing headphones. Although there was no audible dialogue, headphones were used to ensure that participants were not distracted by incidental background noises. Participants were instructed to pay attention because they would later be asked about the event. After watching the event, participants completed a filler task for ten minutes. In another room, the researcher then presented participants with either a physical timeline format or a free recall format to provide their account. In both conditions, participants were asked to report all the details they remembered about the event and the people involved in order to provide a complete and accurate account. All participants were instructed to not make guesses about things they did not remember. Participants in the timeline condition were instructed to use the person description cards to provide descriptive details about the people involved in the event, and the action cards to report any actions and sequence information and to show "who did what and when".

After reporting their initial account (Phase 1), all participants across conditions were asked between three to five follow-up open-ended questions about the event (Phase 2). Question topics were not pre-selected, instead the questions were based on what participants reported, using an open-ended question format, such as "You mentioned X. Tell me more/Tell me everything about X"; "What else can you tell me about X?"; "Explain in more detail what you mean about X"; and "Describe X part in more detail". For example, "You mentioned there was a man in a red jumper. Tell me more about this man in the red jumper" or "Explain in more detail what you mean about this part where they threatened her". This procedure allowed for the interviewer to maintain the same phrasing of questions but avoid using a scripted list of cued-recall questions that did not relate to the initial account. Although not explicitly stated, participants were not forced to respond and if they answered by saying "I don't know" or "I don't remember", the interviewer asked the next question. Similarly, if participants repeated the information they had already reported, and/or responded by saying that they had nothing else to report, the interviewer asked the next question. As a final question, all participants were asked, "Is there anything else you would like to report?". During the questioning phase in both conditions, the participant's written account remained on the table and the interviewer pointed to the specific part to which the prompt referred to when asking each question. The follow-up questioning phase was audio and video-recorded, with the camera focusing on the written account placed in front of the participant. For a visual description of the interview stages, see Figure 1 in the Supplemental materials section.

Results

Initial Reporting (Phase 1)

Participants in the Timeline condition reported significantly more correct details than participants in the Free Recall condition, $t(37.59) = 2.44$, $p = .020$, $d = 0.69$, 95% CI[0.12,

1.26]. There was no difference in the mean number of incorrect details between conditions, $t(48) = .09$, $p = .931$, $d = 0.03$, 95% *CI* [-0.53, 0.58]. With respect to accuracy rates for reported information, there was no difference between conditions, $t(48) = .17$, $p = .864$, $d = 0.05$, 95% *CI* [-0.51, 0.60]. Table 2 displays the Means and *SDs* of both correct and incorrect details, and accuracy rates across reporting phases.

Insert Table 2

An independent t-test analysis showed that participants who used the timeline reported a similar mean number of correct attributions of actions to persons ($M = 3.72$, $SD = 1.77$) relative to the participants who used the free recall format ($M = 3.36$, $SD = 1.87$), $t(48) = .70$, $p = .487$, $d = 0.20$, 95% *CI* [-0.36, 0.75]. Regarding the overall accuracy of the reported attributions, there was also no significant difference between conditions (Timeline: $M = .80$, $SD = .22$; Free Recall: $M = .81$, $SD = .22$), $t(48) = .10$, $p = .919$, $d = 0.06$, 95% *CI* [-0.58, 0.53].

There was a significant difference between conditions regarding the number of sequence errors, $t(48) = 2.70$, $p = .010$, $d = 0.76$, 95% *CI* [0.19, 1.34]. Participants who used the timeline reported fewer sequence errors ($M = .48$, $SD = .51$) compared to participants who used the free recall format ($M = 1.00$, $SD = .82$).

Follow-up questioning (Phase 2)

There was no difference in the number of follow-up questions that were asked between the Timeline ($M = 4.52$, $SD = .51$) and Free Recall ($M = 4.44$, $SD = .58$) conditions, $t(48) = .516$, $p = .608$.

For responses to follow-up questions, there was no difference between conditions for the number of reported correct details, $t(48) = .47$, $p = .638$, $d = 0.13$, 95% *CI* [-0.69, 0.42],

or incorrect details, $t(48) = .63$, $p = .532$, $d = -0.18$, 95% *CI* [-0.73, 0.38]. Nor was there any difference between conditions for the accuracy of details reported, $t(48) = .45$, $p = .657$, $d = 0.13$, 95% *CI* [-0.68, 0.43].

A paired samples t-test showed that the accuracy rate of the reported information in the follow-up questioning phase was significantly lower than in the initial reporting phase both in the Timeline condition, $t(24) = 7.34$, $p < .001$, $d = 1.89$, and in the Free Recall condition, $t(24) = 5.98$, $p < .001$, $d = 1.64$ (see Table 2).

Total Interview Output

Participants in the Timeline condition reported a significantly larger number of correct details overall, compared to participants in the Free Recall condition, $t(38.66) = 2.29$, $p = .028$, $d = 0.65$, 95% *CI* [0.08, 1.21]. There was no statistically significant difference between conditions for the total incorrect details reported across phases, $t(48) = .43$, $p = .671$, $d = 0.12$, 95% *CI* [-0.68, 0.44], or for the total accuracy rate across phases, $t(48) = .99$, $p = .328$, $d = 0.28$, 95% *CI* [-0.28, 0.84].

Discussion

The results of Experiment 1 show that a sizeable amount of additional information was elicited through follow-up questions, representing 18% and 22% of the total information reported in the timeline and free recall conditions, respectively. It is likely that the use of follow-up questions, used here as open prompts rather than directive cued-recall questions, led to further retrieval attempts focusing on different components of the witnessed event. Therefore, in line with the activation theory of memory (Anderson, 1983), the use of open-ended prompts further cued participants' memory for the event.

Despite the opportunity to provide more information in response to follow-up prompts, participants in the free recall condition still reported fewer correct details overall compared to those in the timeline condition. Consistent with previous research (e.g., Hope et al., 2013; Hope et al., 2019), more correct details were initially reported with the timeline technique than with the free recall format, without a cost to accuracy. Meanwhile, in the follow-up questioning phase, participants reported a similar amount of new information across conditions. It is possible that participants who used the timeline technique engaged in extensive retrieval due to the cues that are inherent to the format (see Hope & Gabbert, 2019), and which outweighed the benefits of additional recall through open prompts. The two groups also reported a similar amount of attributions of actions to persons, which is inconsistent with previous findings showing that the timeline technique facilitates the correct reporting of such attributions (cf. free recall; Hope et al., 2013; Hope et al., 2019).

Consistent with previous findings (e.g., Colomb & Ginet, 2012; Hope et al., 2013; Memon et al., 1997), accuracy rates for the initial phase of the interview comprising the witness's self-initiated report were relatively high (average 83%), and did not differ between the timeline and free recall conditions. The accuracy of additional information reported in response to follow-up questions, however, was significantly lower, with an average accuracy rate of approximately 60% (similar between conditions). A possible explanation for the lower accuracy of the new information is that, when reporting their initial accounts, the interviewees were more conservative about the likelihood that the information was correct, than when answering follow-up questions. When interviewees have the freedom to control their reporting, they decide what information to volunteer based on whether it exceeds a certain threshold of confidence in the likelihood that the information is correct (Goldsmith et al., 2002; Koriat & Goldsmith, 1996). Although participants in the current study were not required to answer all the questions, the use of follow-up prompts in the context of an

interview may have implicitly suggested an increased expectation to be informative (Grice, 1975). As a result, participants may have adopted a more liberal criterion for accuracy to still provide informative answers (Goldsmith et al., 2002). Therefore, the finding that the information provided in response to follow-up questions was not as accurate as the initially reported information may have been due to an accuracy-informativeness trade-off, in that participants were able to report new information, but were not as confident in its accuracy relative to their initial account. In order to satisfy an informativeness criterion, the interviewees likely volunteered more details while risking accuracy.

The current experiment served as a first step to examine the efficacy of follow-up open-ended questions based on a free narrative. Given that the new information was not as accurate as the spontaneously reported information and considering the potential implications for applied contexts, a second experiment was conducted to examine whether the follow-up questioning phase could be refined using pre-questioning instructions designed to encourage accuracy by emphasizing the use of metacognitive processes in reporting.

Experiment 2

Research on decision-making mechanisms that are involved when reporting information from memory suggests that rememberers try to achieve a balance between informativeness and accuracy (Goldsmith et al., 2002; Koriat & Goldsmith, 1996). To this end, rememberers control how much information they report based on how confident they are about the accuracy of their recollection (Ackerman & Goldsmith, 2008; Koriat & Goldsmith, 1996).

Rememberers also regulate their answers by adjusting the precision of the reported information (control over grain size; Ackerman & Goldsmith, 2008; Goldsmith et al., 2002). For instance, if asked to provide quantitative information, they may offer a coarse-grain

answer (i.e., broad), instead of a fine-grain answer (i.e., specific), such as reporting that an event occurred "between 17.00 to 18.00" instead of "at 17.15". According to the satisficing model of the minimum-confidence criterion (Goldsmith et al., 2002), rememberers start by retrieving a fine-grain answer, which they will volunteer if it is likely to be correct, otherwise they will provide a coarse-grain answer to preserve accuracy. Further to the satisficing model (Goldsmith et al., 2002), the dual-criterion model suggests that informativeness also mediates reporting, in that even if coarse-grain responses are more likely to be correct, they may be withheld if assessed as not sufficiently informative (Ackerman & Goldsmith, 2008; Yaniv & Foster, 1995). In an investigative context, an interviewee could report coarse details to maximise accuracy. However, if the reported details are thought of as too broad to progress the investigation, the interviewee might choose to offer more specific information, thus using both a confidence and an informativeness criterion to regulate reporting. Given the pattern of findings observed in Experiment 1, it may be that interviewees initially reported information that they assessed as probably correct but in response to follow-up questions they were more willing to risk accuracy to satisfy a demand for informativeness.

Experiment 2 examined whether instructions designed to promote the exercise of metacognitive monitoring would improve accuracy rates for reporting in response to prompting. Half of the participants were instructed that they could withhold from providing an answer (i.e. say "I don't know") and that they could regulate the precision of their answers by providing coarse-grain (e.g., he wore dark clothes) or fine-grain information (e.g., he wore a grey jumper and black jeans). Previous research applying the metacognitive monitoring framework to a forensic context has shown that by using conservative criteria, mock-witnesses can successfully maintain the accuracy of their reporting after a delay (Goldsmith et al., 2005), and after being exposed to misinformation by a co-witness (Wright, Gabbert, Memon, & London, 2008), and that they can balance informativeness and accuracy when

answering cued-recall questions (Weber & Brewer, 2008). Other research examining how interviewees regulate the output and precision of their reporting in various contexts (e.g., reporting in private vs. with an audience) suggests that interviewees would often rather provide informative (i.e., fine) details. However, this tendency is reduced in the presence of an evaluative audience or when they receive penalties for inaccurate responses, in which case they report more coarse details, which are more likely to be accurate (McCallum, Brewer, & Weber, 2016). More recently, Brewer et al. (2018) showed that interviewees can use coarse-grain responses to report on a wide range of topics, from a person's appearance (e.g., hair length and hair colour) to the description of objects and locations, and that they can be provided in response to cued-recall questions even if they were not initially volunteered in a free narrative. Therefore, based on previous research, interviewees should be able to maintain accuracy in reporting by following the instructions that promote monitoring of their memory output and of the type of details they report.

Participants were also reminded that they should not guess. The use of warnings to interviewees to not guess and to reply "I don't know" or "I don't remember" throughout the interview are recommended in the use of the CI to avoid erroneous reporting (Memon et al., 2010). Similar warnings to avoid guessing are also included in other interviewing tools, to encourage interviewees to only volunteer information they are certain about (e.g., Self-Administered Interview; Gabbert, Hope, & Fisher, 2009). There is also evidence that warnings can contribute to the interviewees controlling their reporting more carefully over time (Gawrylowicz, Memon, & Scoboria, 2013). Research by Koriat and Goldsmith (1996) also suggests that participants are more likely to maintain accurate reporting when instructed to not guess if they are uncertain about any details. Related research on metacognitive monitoring indicates that allowing "I don't know" responses and not forcing interviewees to respond to prompts, reduces guessing and increases accuracy when both answerable and

unanswerable questions are asked (Scoboria & Fisico, 2013; Scoboria, Mazzoni, & Kirsch, 2008). Therefore, there is evidence that the use of warnings and instructions to control monitoring of memory output can lead to increased accuracy.

To determine whether the results regarding the accuracy of the information reported in the follow-up questioning phase of the first experiment would replicate, the procedure largely remained the same. A different stimulus was used to increase the generalizability and the relevance of our findings for different interviewing contexts. In Experiment 2, participants witnessed a stimulus event that initially depicted a meeting of a terrorist group who then progressed to placing explosives in a target location. Given the promising results on using the self-generated cues in conjunction with the timeline technique in previous research (Kontogianni et al., 2018), a modified version of the timeline was used here to include use of the mnemonic. Self-generated cues are salient details of the witnessed event that are produced by the interviewees themselves and facilitate recall compared to interviewer-generated cues and no cues (Kontogianni et al., 2018; Wheeler & Gabbert, 2017). In keeping with the procedure of the previous experiment, the same follow-up open-ended questions were used, with the addition of specific pre-questioning instructions to encourage accurate reporting.

Confidence plays a key role in monitoring and controlling reporting (Koriat & Goldsmith, 1996) as well as in the regulation of precision in reporting (Goldsmith et al., 2002). For instance, mock-witnesses are more confident about accurate than inaccurate reported details (Fisher, 1995; Roberts & Higham, 2002), and are more likely to volunteer responses in which they are highly confident (Weber & Brewer, 2008), and withhold responses when they are not confident (Evans & Fisher, 2011). To explore whether retrospective confidence judgments correspond to the pattern of the accuracy rates for the reported information, at the end of the session, all participants were asked to rate how

confident they felt about their written and spoken accounts. Unlike related research on the relationship of confidence-accuracy, we only used two measures regarding the total output for each reporting phase. This is because we were interested in the trajectory of interviewees' confidence ratings relative to that of the accuracy rates for the reported information. For instance, if accuracy for the information provided in response to follow-up questions was lower than the accuracy of the initial account, we were interested to explore if confidence was also lower in the follow-up questioning phase relative to the initial reporting phase.

We predicted that, when interviewees received instructions to monitor the accuracy of their responses to follow-up questions, the accuracy rate of their responses would be higher than when interviewees received no additional instructions. As the current experiment focused on the efficacy of the instructions to support accurate reporting in the follow-up questioning phase, all participants used the timeline technique to provide their initial account.

Method

Participants and Design

An a priori G*Power statistical analysis (Faul et al., 2007) showed that a sample of 60 participants was required for an 80% chance of detecting a large effect size (Cohen, 1992) for the finding of improved accuracy after receiving instructions to monitor reporting based on previous related findings (e.g., Goldsmith et al., 2002; Koriat & Goldsmith, 1996; Scoboria & Fisico, 2013; Weber & Brewer, 2008). The dependent variables were the number of correct and incorrect details, and accuracy rates for both reporting phases, as well as the confidence ratings provided at the end of reporting. Sixty participants (50 Females, Age: M = 20.72, SD = 3.73, Range = 18-33) provided an initial account using the timeline technique. They were then randomly allocated to one of two experimental conditions before follow-up questioning.

Half of the participants received accuracy monitoring instructions prior to the follow-up questioning phase while the remaining half received no instructions.

Materials

Stimulus event. Participants witnessed a 4.28 min long scripted film that depicted a meeting between four perpetrators (three males, one female) who plot a terrorist attack and then carry out the plan. At the outset, three of the perpetrators are seen waiting in a room. The film is shot from a first-person perspective to give the impression of the viewer being in the room. Another individual, acting as the group leader, enters and delivers information about the target of the attack. The leader assigns roles to each member; overseeing the operation, placing the explosives, acting as a look out, and being the getaway driver. The perpetrators discuss the explosives to be used and how they are to be detonated and when. Next the three perpetrators visit the selected target, a park, and are walking down a pathway. One of the males walks around a café with a briefcase which allegedly contains the explosives. The other male takes photos of the park while the female looks at a map. After the first male returns without the briefcase, the female hands him a mobile phone in a covert interaction. All three are seen exiting the park. There is a brief dialogue from inside the car confirming that the explosives have been placed.

Accuracy Monitoring Instructions. Based on previous research, the instructions reminded participants to refrain from guessing (Gabbert et al., 2009; Gawrylowicz et al., 2013; Memon et al., 2010), to feel free to withhold an answer (Scoboria & Fisico, 2013; Scoboria et al., 2008), and to consider the level of detail they felt they could accurately report (Goldsmith et al., 2002; Koriat & Goldsmith, 1996; Weber & Brewer, 2008; see Supplemental materials for verbatim instructions). With respect to the level of detail in reporting, participants were asked to provide all the information they believed to be accurate

from the event, regardless of whether it was fine or coarse in nature. Participants were provided with examples of fine-grain and coarse-grain details, such as describing a car as "small and dark coloured" (coarse), or as "a Volkswagen Golf, British Racing Green, 5-door hatchback, with tinted windows, and a registration number" (fine). To make sure that the instructions were clear, participants were asked to answer the practice question "what can you remember about what footwear the researcher in the room with you is wearing?", by reporting coarse and/or fine details about what they remembered.

Procedure

Participants were invited to take part in research investigating factors that affect people's memory reports for witnessed events. Participants viewed the stimulus event on a computer screen using headphones. Participants were instructed to imagine that they are an undercover agent that infiltrated a terrorist group and to pay attention because they would later have to provide a report on the activities of the group that would be passed on to intelligence analysts. After watching the event, participants completed a filler task for 10 minutes. In another room, the researcher then presented the participants with a physical timeline reporting format to provide their account. Following Kontogianni et al. (2018), participants were given a self-generated cues instruction to write down the first six things that they remembered from the event, without thinking too hard, to think about each of the things they listed and think about whether that memory helped them remember other things about the event. All participants received the same timeline instructions as in the first experiment. After completing their account, half of the participants were provided with the Accuracy Monitoring Instructions. These instructions were presented in written format after participants provided their initial account and prior to being asked any follow-up questions. After they had the chance to ask any questions about the instructions, the instructions were removed and the follow-up questioning phase began.

All of the participants were reminded of their role as an undercover agent with valuable information, and they were asked follow-up open-ended questions about the event. As in Experiment 1, the interviewers asked between three to five questions, based on what participants reported. For instance, the interviewer would ask "You mentioned there was a leader of the group. Tell me more about this leader" or "Explain in more detail what you mean about this part where they discussed the explosives". During questioning, the participant's account on the timeline format was on the table and the interviewer would point to the part based on which the question was asked. At the end of the interview, all participants were asked if there was anything else they wished to report. The follow-up questioning phase was audio and video-recorded, but the camera focused only on the table and the timeline format. At the end of the session, participants were given two separate confidence scales, which ranged from 0% (not at all certain) to 100% (completely certain) with 10% increments. They were asked to indicate how confident they felt about the accuracy of their written account and about their responses to the follow-up questions. For a visual description of the interview stages, see Figure 2 in the Supplemental materials section.

Results

Initial Reporting (Phase 1)

There was no statistical difference between conditions for the number of correct details reported in initial reports, $t(58) = 1.11$, $p = .272$, $d = 0.29$, 95% *CI* [-0.22, 0.79], which was expected as all participants used the timeline technique to provide an initial account. There was no significant difference between conditions with respect to incorrect details, $t(58) = .87$, $p = .388$, $d = 0.23$, 95% *CI* [-0.73, 0.28], or for the accuracy rate of details reported, $t(58) = 1.33$, $p = .189$, $d = 0.34$, 95% *CI* [-0.17, 0.85]. Table 3 shows Means and SDs for correct

details, incorrect details, and accuracy rates reported in both conditions, across reporting phases.

Insert Table 3

Follow-up questioning (Phase 2)

There was no significant difference between the number of follow-up questions that were asked in the accuracy monitoring instructions condition (M = 4.67, SD = 0.55) and in the no instructions condition (M = 4.63, SD = 0.56), $t(58) = .23$, $p = .816$. With respect to the amount of information reported in the follow-up questioning phase, there was no difference between instruction conditions for the number of correct details, $t(58) = .04$, $p = .970$, d = 0.01, 95% *CI* [-0.50, 0.52], or for the number of incorrect details reported in response to follow-up questions, $t(58) = .47$, $p = .642$, $d = 0.12$, 95% *CI* [-0.63, 0.39]. Despite the use of monitoring instructions, there was no significant difference between the two conditions for the accuracy rate of the additional information, $t(58) = .67$, $p = .504$, $d = 0.17$, 95% *CI* [-0.33, 0.68].

A paired samples t-test showed that the accuracy rate of the reported information in the follow-up questioning phase was significantly lower than in the initial reporting phase both in the no instructions condition, $t(29) = 2.85$, $p = .008$, $d = 0.67$, and in the accuracy instructions condition, $t(29) = 3.95$, $p < .001$, $d = 0.73$.

Total Interview Output

There was no difference between the two conditions for the number of correct details reported overall, $t(58) = 1.01$, $p = .317$, $d = 0.26$, 95% *CI* [-0.25, 0.77]. There was no difference between conditions for the total number of incorrect details reported, $t(58) = .91$, p

= .368, d = 0.24, 95% *CI* [-0.74, 0.27], nor for the total accuracy rate across reporting phases, $t(58) = 1.35$, p = .184, d = 0.28, 95% *CI* [-0.16, 0.86].

Confidence Ratings

An independent t-test analysis showed that there was no significant difference between conditions with respect to confidence ratings for the information provided in the initial account, $t(57) = 1.42$, p = .160, d = 0.37, 95% *CI* [-0.15, 0.88], or in response to follow-up questions, $t(57) = .42$, p = .674, d = 0.11, 95% *CI* [-0.40, 0.62]. A paired samples t-test showed that there was no significant difference in participants' confidence ratings for their initial account and for their responses to follow-up questions across conditions, $t(58) = .14$, p = .888, d = 0.02, 95% *CI* [-0.24, 0.27]. Table 4 shows the mean confidence ratings with standard deviations across conditions. A separate exploratory examination of the results for confidence was conducted to more closely examine how the mean accuracy rates provided across reporting phases were distributed at each level of confidence, as in Brewer et al. (2018). The means and *SD*s are shown in Table 5. The results show that most participants expressed between 60% to 80% confidence in the accuracy of their accounts although some participants appear as overconfident and others as underconfident, given the actual accuracy rates reported.

Insert Table 4

Insert Table 5

Discussion

Contrary to our hypothesis, providing participants with instructions designed to encourage accurate reporting did not significantly increase the accuracy of the information provided in response to follow-up questions, relative to participants who received no instructions. With

respect to the efficacy of follow-up questioning, the current results follow the same pattern observed in Experiment 1. Participants reported additional information in the follow-up questioning phase: specifically, 22% (accuracy monitoring instructions condition) and 21% (no instructions condition) of the total information reported was provided in response to open-ended questions. In terms of overall accuracy, accuracy rates for the initial account were high (87.5%) and consistent with previous research (e.g. Colomb & Ginet, 2012; Evans & Fisher, 2011; Gabbert, Hope, & Fisher, 2009) but the accuracy rate observed in the questioning phase was lower (75%).

The fact that accuracy was impaired in follow-up questioning despite the use of open-ended questions and instructions that discouraged guessing, and uncertain responses, and encouraged participants to monitor the precision of their responses, is surprising. Previous research suggests that specific instructions can assist interviewees in balancing accuracy and informativeness demands when asked follow-up questions (e.g., Evans & Fisher, 2011). Based on the similar rates of accurate reporting between conditions and given that the use of open prompts allows interviewees to control their reporting (Ackerman & Goldsmith, 2008; Brewer & Weber, 2008), it may be the case that participants were already regulating their responses to preserve accuracy; thus, the instructions used in Experiment 2 did not further contribute to their strategic reporting. Alternatively, it is possible the instructions were not as helpful as expected because of the short interval between encoding and reporting of the event.

Importantly, the current results suggest that there might be limitations in how well interviewees can balance accuracy and informativeness in reporting – even with the use of appropriate questions and instructions emphasizing accuracy. One potential reason for this is because follow-up questions implicitly increase the demand for informativeness. Recent research suggests that interviewees often show a tendency to report informative details for a

variety of reasons. There is evidence, for example, that interviewees perceive informative reporting to be specific and valuable to the investigator. They also consider how reporting affects their perceived image so that they do not appear to be uncooperative (McCallum et al., 2016; McCallum, Brewer, & Weber, 2019). Future research should investigate to what extent the interviewees' perceptions of what is required with respect to accuracy, informativeness and precision, interact with how they strategically regulate their reporting during the interviewing phase.

Participants' confidence in the accuracy of their reports remained stable across the initial reporting and follow-up questioning phases in both conditions. On average, participants' confidence was approximately 70% and did not vary with declining accuracy rates in the questioning phase. Although the current study measured participants' confidence in a different manner to related research (e.g., Koriat & Goldsmith, 1996; Robert & Higham, 2002; Weber & Brewer, 2008), our aim was simply to explore whether participants' retrospective confidence ratings would provide an insight with respect to the confidence threshold for accurate reporting in follow-up questioning. However, there are limitations that do not allow us to conclude how participants' judgments might be indicative of their regulation in reporting. For instance, although the administration order of the confidence ratings was used to match the way that information was reported through the session and to indirectly encourage participants to compare their reports, it may have contributed to an anchoring effect, whereby confidence estimates for responses to follow-up questions were biased towards the initial report ratings (Tversky & Kahneman, 1974). However, it is also likely that participants appeared under-confident in the accuracy of their initial reports but over-confident in their responses to follow-up questioning due to accuracy rates declining from the initial report to the follow-up questioning phase while confidence remained stable. Furthermore, retrospective ratings may not be as useful in assessing accuracy for such

elaborate free reports compared to cued-recall (e.g., Gwyer & Clifford, 1997; Ibabe & Sporer, 2004). Further research could assess confidence ratings for each response provided to an open prompt, to more closely examine how interviewees consider the accuracy of their reporting.

General Discussion

Across two experiments, the results showed that follow-up, open-ended questions are effective for eliciting new details after an initial free report. However, the accuracy rate for responses to follow-up questions was significantly lower than the accuracy rate for spontaneously reported information. This general pattern of results was replicated across two experiments, using different stimuli which depicted multi-perpetrator events.

The results of both experiments highlight the need to better understand how interviewees' reporting might differ when asked follow-up questions, compared to when they spontaneously report information. Previous research shows that the use of various retrieval attempts, such as techniques included in the CI, can produce increased reporting of more correct details but can also result in a slight increase in the reporting of incorrect details (cf. standard interviews; Memon et al., 2010). Given that the current findings indicate a trade-off in favour of informativeness, despite the use of instructions designed to promote accurate reporting, more research is needed on the generation of errors when additional (open-ended) prompts are used and the role of monitoring processes when demands for informativeness increase. Thus, although research suggests that there can be accuracy trade-offs at the cost of overall reporting when open invitations and varied retrieval attempts are encouraged (Fisher & Geiselman, 2010), the potential costs and benefits of recall when follow-up questions are used have not been systematically examined. Future research should examine whether follow-up questioning increases the perceived need for informativeness, and results in interviewees using a less conservative criterion to balance demands, even if explicitly

instructed to monitor accuracy. Further research could also assess confidence for responses to open prompts as well as to specific "what?", "when?", "where?", "who?", "why?" and "how" probes (Oxburgh et al., 2010), to closely examine how interviewees assess their responses and to what extent that is reflected to the actual reported accuracy.

We already know that asking multiple-choice or repeated questions will likely increase the amount of erroneous reporting (Fisher, 1995; Fisher & Geiselman, 1992), and that open-ended questions are preferable and more efficient (Fisher, Milne, & Bull, 2011; Oxburgh et al., 2010) as they allow interviewees to strategically monitor their reports (Evans & Fisher, 2011). The current findings confirm that follow-up open-ended questions are efficient in gaining new information. However, they also suggest that such information might not be as accurate as an initial spontaneous report. Thus, practitioners should be cautious about the reliability of new information provided in response to follow-up questions and seek further corroboration. It is crucial that future research extends our understanding of the limitations of memory reporting, as there is a limited pool of accurate details that interviewees can recall but an unlimited pool of inaccurate details to report.

References

Ackerman, R., & Goldsmith, M. (2008). Control over grain size in memory reporting--With and without satisficing knowledge. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *34*(5), 1224-1245.

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, R. C., & Pichert, J. W. (1978). Recall of previously unrecallable information following a shift in perspective. *Journal of verbal learning and verbal behavior*, *17*(1), 1-12. https://doi.org/10.1016/S0022-5371(78)90485-1

Berman, G. L., Narby, D. J., & Cutler, B. L. (1995). Effects of inconsistent eyewitness statements on mock-jurors' evaluations of the eyewitness, perceptions of defendant culpability and verdicts. *Law and Human Behavior*, *19*(1), 79-88. https://doi.org/10.1007/BF01499074

Bower, G. (1967). A multicomponent theory of a memory trace. In K.W. Spence & J. Spence (Eds.), The psychology of learning and motivation: I (Vol. 1, pp. 230-325). Oxford: Academic Press.

Brandon, S. E., Wells, S., & Seale, C. (2018). Science-based interviewing: Information elicitation. *Journal of Investigative Psychology and Offender Profiling*, *15*(2), 133-148. https://doi.org/10.1002/jip.1496

Brewer, N., Vagadia, A. N., Hope, L., & Gabbert, F. (2018). Interviewing witnesses: Eliciting coarse-grain information. *Law and human behavior*. Advance online publication. http://dx.doi.org/10.1037/lhb0000294

Brubacher, S. P. (2007). "Interviewing Children about a Repeated Event: Does Prior Practice in Describing a Specific Instance of an Unrelated Repeated Event Improve the Amount and Quality of Elicited Information?". *Theses and Dissertations (Comprehensive)*. Paper 842. Retrieved from http://scholars.wlu.ca/etd/842

Brunel, M., Py, J., & Launay, C. (2013). Cost and benefit of a new instruction for the cognitive interview: the open depth instruction. *Psychology, Crime & Law*, *19*(10), 845-863. https://doi.org/10.1080/1068316X.2012.684058

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155-159. http://dx.doi.org/10.1037/0033-2909.112.1.155

Cohen, J. (1992). A power primer. *Psychological bulletin*, *112*(1), 155-159. http://dx.doi.org/10.1037/0033-2909.112.1.155

Colomb, C., & Ginet, M. (2012). The Cognitive Interview for use with adults: an empirical test of an alternative mnemonic and of a partial protocol. *Applied Cognitive Psychology*, *26*(1), 35-47. https://doi.org/10.1002/acp.1792

Evans, J. R., & Fisher, R. P. (2011). Eyewitness memory: Balancing the accuracy, precision and quantity of information through metacognitive monitoring and control. *Applied Cognitive Psychology*, *25*(3), 501-508. https://doi.org/10.1002/acp.1722

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior research methods*, *39*(2), 175-191. doi:10.3758/BF03193146

Fisher, R. P. (1995). Interviewing victims and witnesses of crime. *Psychology, Public Policy, & Law, 1,* 732–764. doi:10.1037/1076-8971.1.4.732

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

Fisher, R. P., Milne, R., & Bull, R. (2011). Interviewing cooperative witnesses. *Current Directions in Psychological Science*, *20*(1), 16-19. https://doi.org/10.1177/0963721410396826

Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. *Law and human behavior*, *33*(4), 298-307. https://doi.org/10.1007/s10979-008-9146-8

Gabbert, F., Hope, L., La Rooy, D., McGregor, A., Ellis, T. & Milne, R. (2016, June). *Introducing a PEACE-compliant 'Structured Interview Protocol' to enhance the quality of investigative interviews.* Paper presented at the 9th Annual Conference of the International Investigative Interviewing Research Group (IIIRG), London, UK.

Gawrylowicz, J., Memon, A., & Scoboria, A. (2014). Equipping witnesses with transferable skills: the Self-Administered Interview©. *Psychology, Crime & Law*, *20*(4), 315-325. https://doi.org/10.1080/1068316X.2013.777961

Goldsmith, M., Koriat, A., & Pansky, A. (2005). Strategic regulation of grain size in memory reporting over time. *Journal of Memory and Language*, *52*(4), 505-525. https://doi.org/10.1016/j.jml.2005.01.010

Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size memory reporting. *Journal of Experimental Psychology: General*, *131*(1), 73-95. http://dx.doi.org/10.1037/0096-3445.131.1.73

Grice, H. P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and semantics* (pp. 41–45). New York, NY: Academic Press.

Gwyer, P., & Clifford, B. R. (1997). The effects of the cognitive interview on recall, identification, confidence and the confidence/accuracy relationship. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *11*(2), 121-145. https://doi.org/10.1002/(SICI)1099-0720(199704)11:2<121::AID-ACP443>3.0.CO;2-L

Hope, L. & Gabbert, F. (2019). *Expanding the Legacy of the Cognitive Interview: Developments and Innovations in Evidence-based Investigative Interviewing.* In J. J. Dickinson, N. Schreiber Compo, R. Carol, B. L. Schwartz, & M. McCauley (Eds.), *Evidence-based Investigative Interviewing: Applying Cognitive Principles* (1st ed., pp. 42-55). Routledge.

Hope, L., Gabbert, F., & Fraser, J. (2013). Post-incident conferring by law enforcement officers: Determining the impact of team discussions on statement content, accuracy, and officer beliefs. *Law and Human Behavior*, *37*(2), 117-127. doi:10.1037/lhb0000019

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. https://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

Ibabe, I., & Sporer, S. L. (2004). How you ask is what you get: On the influence of question form on accuracy and confidence. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *18*(6), 711-726. https://doi.org/10.1002/acp.1025

Köhnken, G., Milne, R., Memon, A., & Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology, Crime and Law*, *5*(1-2), 3-27. https://doi.org/10.1080/10683169908414991

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

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103*(3), 490-517.

McCallum, N. A., Brewer, N., & Weber, N. (2016). Memorial Monitoring and Control: How Confidence and Social and Financial Consequences Affect Eyewitnesses' Reporting of Fine-Grain Information. *Applied Cognitive Psychology*, *30*(3), 375-386. https://doi.org/10.1002/acp.3212

McCallum, N., Brewer, N., & Weber, N. (2019). A Measure of Perceived Informativeness for Investigations of Eyewitness Memory Reporting. *Journal of Applied Research in Memory and Cognition*, *8*(2), 214-220. https://doi.org/10.1016/j.jarmac.2019.04.004

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.

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

Memon, A., Wark, L., Bull, R., & Koehnken, G. (1997). Isolating the effects of the cognitive interview techniques. *British Journal of Psychology*, *88*(2), 179-197. doi:10.1111/j.2044-8295.1997.tb02629.x

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.

Paulo, R. M., Albuquerque, P. B., & Bull, R. (2013). The enhanced cognitive interview: Towards a better use and understanding of this procedure. *International Journal of Police Science & Management*, *15*(3), 190-199. https://doi.org/10.1350/ijps.2013.15.3.311

Paulo, R. M., Albuquerque, P. B., Vitorino, F., & Bull, R. (2017). Enhancing the cognitive interview with an alternative procedure to witness-compatible questioning: category clustering recall. *Psychology, Crime & Law*, *23*(10), 967-982. https://doi.org/10.1080/1068316X.2017.1351966

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

Scoboria, A., & Fisico, S. (2013). Encouraging and clarifying "don't know" responses enhances interview quality. *Journal of Experimental Psychology: Applied*, *19*(1), 72-82. http://dx.doi.org/10.1037/a0032067

Scoboria, A., Mazzoni, G., & Kirsch, I. (2008). "Don't know" responding to answerable and unanswerable questions during misleading and hypnotic interviews. *Journal of Experimental Psychology: Applied*, *14*(3), 255-265. http://dx.doi.org/10.1037/1076-898X.14.3.255

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

Smeets, T., Candel, I., & Merckelbach, H. (2004). Accuracy, completeness, and consistency of emotional memories. *American Journal of Psychology*, *117*(4), 595-609. doi:10.2307/4148994

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science*, *185*(4157), 1124-1131. doi: 10.1126/science.185.4157.1124

Weber, N., & Brewer, N. (2008). Eyewitness recall: regulation of grain size and the role of confidence. *Journal of Experimental Psychology: Applied*, *14*(1), 50-60. http://dx.doi.org/10.1037/1076-898X.14.1.50

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological science in the public interest*, *7*(2), 45-75. https://doi.org/10.1111/j.1529-1006.2006.00027.x

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

Wright, D. B., Gabbert, F., Memon, A., & London, K. (2008). Changing the criterion for memory conformity in free recall and recognition. *Memory*, *16*(2), 137-148. https://doi.org/10.1080/09658210701836174

Yaniv, I., & Foster, D. P. (1995). Graininess of judgment under uncertainty: An accuracy-informativeness trade-off. *Journal of Experimental Psychology: General*, *124*(4), 424-432.

Tables

Table 1

Protocol of follow-up open-ended questions to extend and clarify on 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 2

Experiment 1: Means and SDs of correct and incorrect details (and accuracy rates) provided in the initial reporting phase and in response to follow-up questions.

		Timeline condition		Free recall condition	
		Mean	***SD***	**Mean**	***SD***
Initial Account	Correct details	67.32	19.27	56.56	10.74
	Errors	8.84	5.11	8.72	4.65
	Accuracy rate	0.83	0.08	0.83	0.06
Follow-up questions	Correct details	14.72	6.52	15.60	6.61
	Errors	4.00	2.96	4.84	5.98
	Accuracy rate	0.59	0.16	0.61	0.17
Total	Correct details	82.04	18.63	72.16	10.88
	Errors	12.84	4.63	13.56	7.05
	Accuracy rate	0.86	0.04	0.85	0.06

Table 3

Experiment 2: Means and SDs of correct and incorrect details (and accuracy rates) provided in the initial reporting phase and in response to follow-up questions.

		Accuracy monitoring instructions		No instructions	
		Mean	***SD***	**Mean**	***SD***
Initial report	Correct details	40.10	12.58	44.27	16.24
	Errors	6.57	4.71	5.67	5.67
	Accuracy rate	0.86	0.08	0.89	0.05

Follow-up questions	Correct details	11.50	7.29	11.57	6.31
	Errors	3.20	2.09	2.93	2.32
	Accuracy rate	0.73	0.21	0.77	0.23
Total	Correct details	51.60	15.75	55.83	16.72
	Errors	9.77	4.03	8.60	5.77
	Accuracy rate	0.84	0.08	0.86	0.06

Table 4

Experiment 2: Means and standard deviations of confidence ratings between conditions for the initial reports and in response to follow-up questions.

	Accuracy monitoring instructions		**No instructions**	
	Mean	***SD***	**Mean**	***SD***
Initial reports	66.55	14.95	72.00	14.48
Follow-up questions	68.28	18.34	70.00	12.59

Table 5

Experiment 2: Mean accuracy rates and standard deviations for both the initial and follow-up reporting phases. Rates are collapsed across conditions.

	Mean			**Mean**		
Confidence	**Accuracy**	***SD***	***n***	**Accuracy**	***SD***	***n***
	Initial report			**Follow-up**		
100	0	0	1	0.79	0	1

90	0.87	0.06	8	0.76	0.21	7
80	0.88	0.05	12	0.75	0.21	13
70	0.87	0.06	22	0.74	0.15	17
60	0.87	0.04	9	0.74	0.29	14
50	0.87	0.03	4	0.74	0.15	4
40	0.80	0.19	2	0.87	0	1
30	0.91	0	1	0.36	0	1
20	0.82	0	1	0	0	0
10	0	0	0	0.63	0	1
0	0	0	0	0	0	0

809