UNIVERSITYOF
PORTSMOUTH

Eliciting information from cooperative sources about single & repeated multi-actor events

Feni Kontogianni

September 2018

This thesis is submitted in partial fulfilment of the requirements for the award of the degree of

Doctor of Philosophy of The University of Portsmouth

CREST CENTRE FOR RESEARCH AND
EVIDENCE ON SECURITY THREATS

Abstract

Successful investigations in forensic and security contexts depend on eliciting reliable and detailed information from sources. Although research has developed effective interviewing protocols to improve recall of witnessed events in criminal investigations, there is only a small body of research on information elicitation tools for intelligence gathering. The overarching aim of this thesis is to contribute to the development of interviewing techniques for use with cooperative sources in security settings. Specifically, we examined the effectiveness of mnemonics, reporting formats and prompts to facilitate recall for multi-actor events witnessed on both single and repeated occasions. In **Experiment 1**, we introduced a new mnemonic to the timeline technique. Participants witnessed a multi-actor crime-event under full or divided attention and provided an account using self-generated cues, other-generated cues or no additional cues across timeline reporting conditions. The results showed that use of the self-generated cues increased the reporting of correct details (cf. other-generated and no cues) under full but not under divided attention conditions. In **Experiments 2 and 3**, we examined the efficacy of open-ended questions to follow-up on an initial report. In Experiment 2, participants witnessed a multi-actor crime-event and used the timeline or a free recall format to provide an initial report. In Experiment 3, participants used the timeline to provide their recall of a video depicting a group planning and carrying out an attack. Before being asked follow-up questions, half of the participants were instructed to avoid guessing, to feel free to withhold an answer and to consider the precision of their answers (i.e. provide general or specific details). The results showed that follow-up questions elicited new information across conditions. However, the accuracy of the responses to the follow-up questions was not as high as the initially reported information (Expt. 2 & 3), even after participants were instructed to monitor the accuracy of their

responses (Expt. 3). In **Experiment 4**, we tested the effectiveness of self-generated cues, the timeline technique and follow-up open-ended questions, as part of a Multi-Method Interviewing Format (MMIF) to facilitate the retrieval and particularization of repeated events. Over the course of a week, participants witnessed four videos of a group planning and carrying out an attack, where either all four videos were highly similar, or where three videos were highly similar, and one video included a new and a changed critical detail to introduce a deviation to the script. After a week, participants provided an account using the MMIF, the timeline technique or a free recall format. The results showed that use of the MMIF elicited more correct information and increased particularization of specific instances of the repeated events (cf., timeline and free recall format). There was no additional benefit for recall when deviations were present in a specific instance. This set of experiments successfully extended the timeline technique into a format that can be used flexibly for the reporting of complex and repeated events. In the discussion of our results, we suggest avenues for future research focused on retrieval techniques for use in applied information elicitation contexts.

Table of Contents

Abbreviations

ANOVA: Analysis of variance

CI: Cognitive Interview

GLMM: Generalized Linear Mixed Models

FTT: Fuzzy-Trace-Theory

ICC: Intra-class Correlation

MMIF: Multi-Method Interviewing Format

MRC: Mental Reinstatement of Context

SM: Source-monitoring framework

Declaration

Whilst registered as a candidate for the above degree, I have not been registered for any other research award. The results and conclusions embodied in this thesis are the work of the named candidate and have not been submitted for any other academic award.

Word count: 38.400

Table of Figures and Tables

Figures

Tables

Acknowledgments

I would like to extend my thanks to the many people who have helped to make this work possible, and this PhD journey a memorable experience.

I am grateful to my supervisors who have guided me throughout this process. Thank you all for your generosity and for challenging me to push myself. Lorraine, you taught me to attend to the details but to never lose sight of the wider implications of my research, to be confident in my work but to never stop trying to improve it. Thank you for ceaselessly mentoring me these years while sharing a laugh through it all. Paul, thank you for always introducing a new perspective to my work and for making me consider the potential connections between different methods and ideas. Aldert, thank you for your honest feedback and for always offering your support amidst your travels and busy schedule. A special thank you to Prof. Fiona Gabbert. I am thankful for your valuable advice and constructive comments which always improved my work.

Thank you to my friends, colleagues and office-mates for the support, the laughs, and the endless soundboard sessions. I have enjoyed sharing this experience with you all. Marc and Gary, thank you for being there during both the good and the somewhat rocky times over these three years.

My deepest gratitude goes out to my family, particularly my parents for their constant support. Thank you for having trust in me, and for always encouraging me to pursue my ambitions, wherever they might take me.

I would like to thank all of the technical staff for their help over these past three years. Also, thank you to all of the participants who offered their time for my research and without whom this thesis would not have been possible.

Dissemination

Conference presentations

Experiment 1 was presented at the 6th International Conference on Memory in July 2016, at the annual meeting of the Centre for Research and Evidence on Security Threats in September 2016, and at the 12th conference of the Society for Applied Research in Memory and Cognition in January 2017:

Kontogianni, F., Hope, L., Taylor, P., Vrij, A., & Gabbert, F. (2016, July). *Introducing a new mnemonic to the Timeline Technique: the use of Self-Generated Cues*. Poster presented at the 6th International Conference on Memory (ICOM) conference, Budapest, Hungary.

Kontogianni, F., Hope, L., Taylor, P., Vrij, A., & Gabbert, F. (2016, September). *The Benefits of a Self-Generated Cue mnemonic for Timeline Interviewing*. Poster presented at the annual meeting of the Centre for Research and Evidence on Security Threats (CREST), London, UK.

Kontogianni, F., Hope, L., Taylor, P., Vrij, A., & Gabbert, F. (2017, January). *Introducing a new mnemonic to the Timeline Technique: the use of Self-Generated Cues*. Presentation at the 12th conference of the Society for Applied Research in Memory and Cognition (SARMAC), Sydney, Australia.

Experiments 1, 2 and 4 were presented at the annual meeting of the Centre for Research and Evidence on Security Threats in September 2017:

Kontogianni, F., Hope, L., Taylor, P., Vrij, A. (2017, September). *Developing innovative techniques to elicit complex and detailed intelligence.* Paper presentation at the annual meeting of the Centre for Research and Evidence on Security Threats (CREST), Reading, UK.

Experiment 4 was presented at the 11th annual conference of the International Investigative Interviewing Research Group conference in July, 2018 and at the 1st International Conference on Behavioural and Social Sciences in Security in July, 2018:

Kontogianni, F., Rubinova, E., Hope, L., Taylor, P., Vrij, A. (2018, July). *Testing a Multi-Modal Interviewing Format for recall of repeated events.* Paper presentation at the 11th annual conference of the International Investigative Interviewing Research Group (IIIRG), Porto, Portugal.

Kontogianni, F., Rubinova, E., Hope, L., Taylor, P., Vrij, A. (2018, July). *Testing a Multi-Modal Interviewing Format for recall of repeated events.* Paper presentation at the 1st International Conference on Behavioural and Social Sciences in Security (BASS), Lancaster, UK.

Publications

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. doi: 10.1016/j.jarmac.2018.03.006

Chapter 1: Thesis General Introduction

The successful progress of both criminal and human intelligence (HUMINT) investigations depends on eliciting reliable and detailed information from suspects, witnesses, victims, and informants (DeClue, 2010). Although cooperative sources are typically expected to report all the information they possess, the use of ineffective communication practices and failure to support the retrieval of information from memory can restrict reporting (Evans, Meissner, Russano, & Kleinman, 2010; Vrij, Hope, & Fisher, 2014). Further, memory for past experiences is malleable and often prone to errors of distortion, confabulation and omission (Bartlett, 1932; Loftus, 1979; Loftus, 2003). To facilitate the reporting of reliable and detailed accounts, the use of evidence-based information elicitation techniques is necessary. Although psychological research has contributed to the development of established interviewing protocols such as the Cognitive Interview (CI; Fisher & Geiselman, 1992) for use in police investigations, there is only a small body of research on empirically-based information elicitation tools tailored for security contexts. Following recent calls for more focused research in the area (Granhag, Vrij, & Meissner, 2014), the overarching aim of the current programme of doctoral research is to contribute to the development of information elicitation techniques for use in security settings. To this end, the research presented in this thesis aimed to improve currently used techniques and test adaptive and effective tools to use with cooperative sources.

Group, gang or cell involvement is common in both criminal and terrorist activities thus, reporting of multi-actor events is relevant in both forensic and security contexts, as it can lead to crucial information about a criminal network (Taylor, Snook, Bennell, & Porter, 2015; Ozgul, 2016). Additionally, HUMINT interviewers are often interested in non-recent incidents that might have occurred between other similar events (e.g., meetings related to the planning of

terrorist attacks) and in persons that might not have been of particular interest to the interviewee at the time (Evans et al., 2010; Rivard, Fisher, Robertson, & Hirn Mueller, 2014). However, obtaining extensive and high-quality information can be particularly challenging in cases of complex multi-actor events. Across four experiments, this thesis tested the use of a cognitive mnemonic in conjunction with an innovative reporting format to facilitate recall of a unique multi-actor event witnessed under both optimal and sub-optimal conditions (Experiment 1); examined the efficacy of follow-up open-ended questions based on a self-report for a unique multi-actor event (Experiments 2 and 3); and tested the effectiveness of cues, retrieval techniques and follow-up prompts, combined in a multi-method interviewing format, to facilitate the recall of multi-actor events that were witnessed on repeated occasions (Experiment 4).

The current research draws on a rich theoretical background describing the role of encoding, retrieval and metacognitive processes in memory reporting. The purpose of this chapter is to review the literature on memorial processes that are involved in the reporting of witnessed events and to outline key findings from eyewitness memory research relevant for the focus of this thesis.

The associative structure of memory

It has been argued that episodic memory is formed by multiple features bound together to represent a cohesive representation of an event (*multi-component view of memory;* Bower, 1967; see also, Thomson & Tulving, 1970). According to this view, each event consists of many components which are interconnected in memory. Spreading activation theory (Anderson, 1983), one of the most comprehensive theoretical models of episodic memory, proposes that an event is represented in memory as a network of associated traces. When one witnesses an event, details of

the event are encoded and stored in memory as nodes which are connected to each other through links that vary in strength. How strongly a trace is represented in the memory network depends to a great extent on the amount of attention one was paying at the time of encoding. Imagine that you are a witness to a bank robbery. It is likely that you have to pay attention to multiple perpetrators performing many actions at once, while experiencing feelings of stress and physiological arousal that cause your attention to divert to internal thoughts and fears (Lane, 2006). As a result, some components of the event will be poorly encoded in memory, making them less accessible at retrieval.

Research shows that the division of attentional resources affects encoding and retrieval processes in different ways. To examine the effect of attention on memory performance, participants in experimental research are usually asked to perform two different tasks at the same time (*dual-task performance*). For instance, across four experiments, Craik, Govoni, Naveh-Benjamin, and Anderson (1996), instructed participants to track when an asterisk appeared on the computer screen during a computer task and respond by pressing a key on the keyboard as quickly as possible. While performing the task, participants had to either study a list of word pairs presented auditory, or to recall the studied word. Therefore, their attention was divided either when they encoded or when they retrieved the words. Memory performance was measured across free recall (i.e. recalling as many words as possible), cued recall (i.e. recalling the paired associate of the presented word), and recognition tasks (i.e. indicating whether a word presented at retrieval was present during encoding or not). The division of attention during encoding of the word lists consistently reduced participants' reaction time to the visual task, and negatively affected their memory performance both for free recall and recognition tasks. When attention was instead divided at retrieval, there was a small effect on free memory recall and no effect at

all on recognition memory. Therefore, it appears that there is an asymmetry in the way that divided attention affects memory. Further findings of impaired memory performance when attention is divided at encoding have been demonstrated via the use of auditory concurrent tasks (e.g., Naveh-Benjamin, Craik, Perretta, & Tonev, 2000; Naveh-Benjamin, Kilb, & Fisher, 2006) and across various stimuli (e.g., actions, pictures etc., for a review see Mulligan, 2014). Overall, research suggests that divided attention affects encoding but not retrieval processes because encoding is a controlled process that requires deliberate effort, whereas retrieval appears to be an automatic process (Craik et al., 1996).

Although there is a wealth of research examining the effects of divided attention with basic experimental paradigms, applied research is rather limited. Recent findings from eyewitness memory research indicate that divided attention at encoding can have serious consequences for the reporting of crime events. In a study using a mock-witness paradigm, participants viewed a series of slides depicting a theft, under full or divided attention, and then answered questions about the event (some of which included misleading information which was never presented). Mock-witnesses whose attention was divided were less accurate and more suggestible in their responses, compared to participants who fully focused on the event (Lane, 2006). More recently in Marsh et al. (2017), mock-witnesses were instructed to ignore distractions of cell-phone conversations while witnessing a staged crime-event. Despite this instruction, later retrieval of the witnessed perpetrators' faces was impaired. Hence divided attention at encoding affects both the amount of recalled information, accurate reporting and identification decisions.

Facilitating the retrieval of encoded information. According to the spreading activation theory, the stronger a memory trace is, the more likely it is that it will be remembered

(Anderson, 1983). Every retrieval attempt causes the activation of a trace, as if a spotlight focuses on that memory, thus further increasing its strength. As the trace is activated it also spreads activation to closely associated traces within the memory network. Therefore, remembering one part of a witnessed event increases the likelihood of remembering other related parts, suggesting that there is a degree of dependency between the different traces that represent an event (Horner & Burgess, 2013; Naveh-Benjamin et al., 2000). Using the example of the bank robbery, remembering what a perpetrator looked like increases the likelihood of remembering what they sounded like and what they said to the customers who were in the bank. Below we present an overview of key research findings that demonstrate how the use of mnemonics capitalize on the associative structure of memory to facilitate retrieval.

Successful retrieval depends on the interaction between stored information and conditions at remembering. Early research by Tulving and Thomson (1973) demonstrated that event details that are not spontaneously recalled can be retrieved with the use of appropriate cues, i.e., cues that include information that was associated with the to-be-remembered trace at the time of encoding and storage in memory (*encoding specificity principle*). In a series of experiments, participants were asked to memorize word-pairs of a target (to-be-remembered) word accompanied by weakly associated cues (e.g., BLACK – train). At test, they had to recall the target words in capital letters, in the presence of the studied weak cue words or semantically strong cues, which were not present at the study phase (e.g., BLACK – white), or without any accompanying cues. Results showed that weak cues facilitated retrieval more than strong cues, because the target words were encoded in the context of the weak cues and were thus associated in memory (Thomson & Tulving, 1970; Tulving & Osler, 1968). In other words, since the otherwise weak cues were stored in memory together with the target words during the study

phase, their presence at retrieval improved recall due to an overlap between the conditions at encoding and at retrieval.

Other research on the effect of divided attention on memory performance shows that even poorly encoded traces might be accessible under cued recall. In a study by Backman and Nilsson (1991), participants' attention was divided while they studied a set of words that were listed in five different semantic categories. Participants had to report all the words they remembered without any support in a free recall task and then they recalled the words again with the use of the names of the lists' categories as cues. The results showed that participants reported more information under cued-recall, suggesting that the cues prompted retrieval further to the free recall task even under divided attention conditions, because they were encoded at the same time as the target words. Therefore, although there is limited research on the use of mnemonics to facilitate recall under divided attention at encoding, evidence shows that contextual cues can increase access to memory for poorly encoded events relative to a lack of support at retrieval.

The *encoding-specificity principle* suggests that a cue is effective, if the probability for retrieving the to-be-remembered item increases in the presence of the cue, relative to a free recall where no support at retrieval is provided (Tulving & Thomson, 1973). Tulving and colleagues (1968; 1973) demonstrated that cues effectively improved recall by maximizing the overlap between the conditions that were present at encoding and at retrieval. The encoding-specificity principle serves as the rationale behind context reinstatement techniques (Tulving & Thomson, 1973; for a review, see Pansky, Koriat, & Goldsmith, 2005). Godden and Baddeley (1975) were among the first to demonstrate a context-dependent effect on memory outside of the laboratory, where free recall of word lists was higher when retrieved in the same learning environment compared to a different environment (e.g., underwater versus on land). Subsequent research

suggests that the benefits of context-dependent retrieval extend from physical reinstatement to both mental and physical reinstatement. In a review and meta-analysis of the literature, Smith and Vela (2001) found that the use of mental reinstatement of the encoding context facilitates memory recall, even if the physical environment is different to that during encoding. The Mental Reinstatement of Context (MRC) technique, which is part of the Cognitive Interview, is a well-known example of such a guided retrieval mnemonic that is used to facilitate retrieval by instructing interviewees to think back to when they witnessed the event, to think about what they could see, what they could hear, what the surroundings were, and what they were thinking and feeling at the time (Fisher & Geiselman, 1992). Mock-witnesses tend to report more correct details when Mental Reinstatement of Context is used compared to a free recall format without any cues at retrieval (e.g., Dando, Wilock, & Milne, 2009a). However, there is evidence that suggests that context reinstatement might not always be effective. Rosenbluth-Mor (2001 as cited in Pansky et al., 2005) replicated the procedure used in studies by Tulving and colleagues that examined the effectiveness of weak-associate cues in facilitating retrieval for target words (Tulving & Osler, 1968; Tulving & Thomson, 1973). The results showed that using the same weakly associated contextual cue at encoding and at retrieval increased the amount of information recalled (cf. no cues at retrieval). However, using a different weak cue at retrieval to the one that was present at encoding impaired recall with respect to both the number and accuracy of the recalled words. Therefore, it may not be the overlap between conditions at encoding and retrieval that facilitates recall, but the mismatch between conditions that impairs recall.

Other research suggests that although the encoding-specificity principle is necessary to improve recall, it is not sufficient. Rather, it may be the quality of cues that moderates the extent

to which retrieval improves than the presence of the cue alone. As the number of encoded memory traces that are prompted by a cue increases, the efficacy of the cue to facilitate retrieval decreases (*principle of cue overload*; Nairne, 2002). Instead, a cue is effective when it is distinctive, that is when it uniquely matches a memory to the exclusion of other related memories. Watkins and Watkins (1975), and more recently Goh and Lu (2012) tested the diagnostic value of a retrieval cue by manipulating the number of items it was associated with. Participants studied target words that were listed in various categories. Each target was paired with a semantically unrelated word. The results show that recall was improved when the unrelated cue to the target word was used, compared to the use of the list's category name. Therefore, cues effectively facilitate retrieval when they are distinctive in addition to satisfying the encoding-retrieval match (Tullis & Benjamin, 2015; Watkins & Watkins, 1975). If a cue is distinctive, in that it matches a single memory trace in the network, retrieving that memory will be easier than if the cue matches multiple traces (Kahana, 2012; Nairne, 2002).

Mantyla and Nilsson (1988) demonstrated the efficacy of distinctive self-generated cues over a series of experiments. In a first experiment, participants were instructed to study a list of nouns and generate three properties that either described or were distinctive to each item (i.e., details that they would generate for this particular item on other occasions). Participants were later invited to describe the same target items: distinctive properties were more similar and specific than the general descriptions, suggesting that they can serve as reliable cues. Further experiments showed that recall was enhanced even after a six-week interval when participants used self-generated distinctive cues at test. Tullis and Benjamin (2015) used a similar procedure and found that cues that focused on specific properties of the target items were associated to fewer possible targets (i.e., they were a better match for specific traces) than the item

descriptions. A second experiment showed that self-generated cues were more effective than cues generated by other participants. Therefore, the findings suggest that retrieval was most improved when participants generated their own cues during encoding of the target items. Similarly, Anderson and Conway (1993) demonstrated that, when asked to list details about specific events they had experienced in the past, participants first listed "distinctive details" (i.e., "details that really stand out and make that memory what it is", p. 1188). They then listed other details, highly associated with those distinctive details. Thus, self-generation of distinctive cues can trigger thematically related memories from a witnessed event (Anderson & Conway, 1993). Therefore, to be effective, cues need to be encoded within the context of the witnessed event, and to offer diagnostic information identifying a single target to the exclusion of others, rather than matching multiple related targets (i.e., matching but not distinctive) (Goh & Lu, 2012; Nairne, 2002). Given that accurate information is crucial to applied information elicitation contexts, it is important to use distinctive cues that are compatible to what the witness encoded. To this end, recent research has focused on the use of self-generated cues which are defined as salient details that are actively generated by the individuals themselves and facilitate retrieval of a target memory (Wheeler & Gabbert, 2017).

Taken together, the research reviewed here indicates that use of distinctive self-generated cues may improve recall relative to other-generated cues. Further, there is evidence suggesting that cued-recall may facilitate access to even poorly encoded information. Therefore, the current thesis aimed to examine the effectiveness of self-generated cues to facilitate recall of multi-actor events under both optimal and sub-optimal encoding conditions.

Retrieving information about multi-actor events

Eliciting information about multi-actor events in security contexts could be critical in the aftermath of an event such as the London bridge terror attack (https://www.bbc.co.uk/news/uk-england-london-40147164). In such cases, reliable information could lead to the identification of relevant targets within a terrorist network. Critically, correctly identifying what actions were performed by each perpetrator ("who did what and when") is likely to be critical for successful prosecution and appropriate court-sentencing (Roberts, 2003). However, there is little empirical research on memory-enhancing techniques that can be used to increase recall of multi-actor events.

Tulving (1983) argued that information in episodic memory is organized according to the temporal and spatial perception of the witnessed event. This representation of events is consistent with the encoding-specificity principle, in that event-details are associated in memory with the temporal and spatial context that was present at encoding (Tulving & Thomson, 1973). There is evidence for the role of temporal context at retrieval in research examining the order in which previously studied items are reported in free recall (Kahana, 1996). For instance, in Howard and Kahana (1999), participants across two experiments studied twelve-item word lists and then completed a free recall task either immediately, after performing an arithmetic task for several seconds before recall, or after performing a distractor task in between studying the last item in the list and starting with the recall task. Results overall showed that for every reported item, the next item to be recalled tended to be from a close serial position (*lag recency effect*). In other words, items that are temporally associated at encoding tend to be recalled in close proximity (*temporal contiguity effect*; Polyn, Norman, & Kahana, 2009) and so, the temporal context of an item can be used to cue the next item that is closely associated in memory (Howard & Kahana, 1999;

Unsworth, 2008). Translating this to the context of multi-actor events, the actions of a perpetrator would therefore be temporally associated at encoding with the perpetrators themselves and with the context in which they occurred (Hope, Mullis, & Gabbert, 2013b).

Except for the Cognitive Interview (Fisher & Geiselman, 1992), which recommends that interviewees are allowed to provide an uninterrupted free narrative about the witnessed event, most interviewing approaches invite interviewees to "start at the beginning". However, reporting complex events, such as the actions of multiple perpetrators, in a linear narrative limits the potential for temporal-context retrieval as interviewees are forced to report events in a specific sequence. Moreover, withholding information about related components of the event to preserve the linear structure at reporting likely disrupts the flow of retrieval and increases cognitive load, thus restricting the information output (Oberauer & Bialkova, 2009). Hope and colleagues (2013b), tested the use of a timeline technique, which uses a physical timeline format, to facilitate recall of events with multiple perpetrators. Mock-witnesses who used the timeline technique reported more correct details and attributions of actions to actors, compared to a free recall format at both immediate testing and after a two-week delay. The timeline format is thought to represent the temporal space of the incident, where interviewees can place information about actions, the people involved and the sequence of events as they occurred, thus reinstating the temporal and spatial context to facilitate recall. Further, the structure of the timeline format encourages witness-compatible recall whereby interviewees can report events as they remember them, at any point of the timeline, and re-arrange details if necessary, without interrupting the conversational narrative flow (Grice, 1975). Therefore, the use of the timeline technique likely prompts interviewees to thoroughly search through their memory, by reinstating the temporal context at encoding and by allowing for the flexible reporting of events and of the perpetrators' actions.

Recalling multi-actor repeated events. Obtaining high-quality information in HUMINT contexts can also be crucial for multi-actor events that have occurred repeatedly over time, such as meetings between groups planning future terrorist activities (Rivard et al., 2014). In such cases, sources will likely need to retrieve memories of specific incidents across these repeated events. However, reporting information about specific details that occurred within a series of repeated events poses additional challenges to remembering an event that occurred on a single occasion. There is a growing body of developmental research which shows that memory for repeated experiences differs from memory of unique experiences, mainly with respect to the type of remembered details and the focus of recall prompt (Price & Connolly, 2013). When experiencing a series of repeated similar events, certain details remain constant across events, whereas others will change in each incident. Compared to a unique experience, research shows that memory for details that are constant across repeated events is stronger, because of exposure to repetition. Instead memory for variable details which change from one incident to the next is weaker than memory for both a unique experience and for constant details across events (Connolly & Lindsay, 2001; Fivush, Hudson, & Nelson, 1984).

When inquiring about repeated events, interviewers may focus on the general routine or on specific details of individual incidents. However, eliciting information about specific details (i.e., particularization of specific *instances*) might be challenging given that the general memory of the repeated events is stronger (Roberts, 2002). To date, research on interviewing strategies for recalling repeated events has focused on children in the context of abuse investigations where successful prosecutions rely on evidence about specific incidents of recurring events (for a review see Brubacher, Powell, & Roberts, 2014; Schneider, Price, Roberts, & Hedrick, 2011). However, there is a lack of research on adults' retrieval and on the use of techniques that might facilitate the

particularization of specific incidents (e.g., Cohen & Java, 1995; Leins, Fisher, Pludiwinski, Rivard, & Robertson, 2014; MacLean, Coburn, Chong, & Connolly, 2018; Means & Loftus, 1991; Rivard et al., 2014; Theunissen, Meyer, Memon, & Weinsheimer, 2017; Willen, Granhag, Stromwall, & Fisher, 2015).

Much of our current knowledge about memory for repeated events is based on Bartlett's (1932) seminal work on the dynamic and reconstructive nature of memory for past experiences. Bartlett asked participants to study "*The War of the Ghosts*" story in the laboratory and observed how their retellings of the story changed over time. He inferred that although their memories derived from elements of the story, their account was a reconstruction and interpretation of the original input based on their general knowledge representations or *schemas.* Schemas are higher-order knowledge structures that include general representations of complex concepts - such as sequences of multiple activities – and are thus considered a part of semantic memory that also interact with new episodic information (Ahn, Brewer, & Mooney, 1992; Brewer & Nakamura, 1984). A script represents a simple type of schema, including knowledge of what typically occurs over a series of events, such as the typical actions involved when ordering food in a restaurant (Abelson, 1981; Ahn et al., 1992). According to schema theory, details that recur across repeated events (i.e., typical actions) are schema-consistent because they preserve the general representation of the events, whereas variations are predictable alternatives that change from one instance to the next (e.g., ordering a different meal at a restaurant every time; Brewer & Nakamura, 1984). Although a schematic representation of events in memory can lead to improved recall for the general routine of what usually occurred, it can have the opposite effect for recall of specific details or variations (e.g., Brewer & Nakamura, 1984). In a study by Kuebli (1990), 4- and 7-year-old children took part in a series of play activities on four occasions. Over the course of the events,

some details remained fixed across events and some varied either on one or on every instance. On a fifth session, children were asked to freely recall what they remembered and were then questioned specifically about details that varied across events. On free recall, children tended to describe the general routine and not mention the specific details. It was only when directly asked, that they were able to describe the alternative activities, but that was increasingly difficult for children who witnessed the alternative activities only on one instance (cf. every instance) (see also Brubacher, Glisic, Roberts, & Powell, 2011). Consistent with schema theory, these findings suggest that variations tend to be integrated into the script to preserve the consistency of the overall schema about the repeated events (Brewer & Nakamura, 1984).

Fuzzy-Trace-Theory proposes a similar conceptualization to schema theory with respect to the processing, storage and retrieval of events (FTT; Brainerd & Reyna, 2002; 2004). According to FTT, experiences are encoded and stored in memory in the form of gist, and of verbatim traces. Gist traces represent semantic information or in other words, one's understanding of the event, whereas verbatim traces represent specific details of the experience. Evidence for this differential perception of events is based on experimental research where participants need to respond as quickly as possible when deciding on the gender of a series of names displayed on the screen during a computer-task. Before seeing the target name, participants are primed with a name briefly flashing on the screen. The prime either matches the gender of the target or not. Results show that once primed with the name that matches the target's gender, participants make a decision, before they even have time to process the actual name on the screen. These results suggest that we tend to access memory for meaning as soon as the encoding of the actual event begins. In other words, we tend to extract the general pattern of an event before specific details are stored in memory (see Brainerd & Reyna, 2004, for a review). According to FTT, the dissociated processing of gist and

verbatim details further dictates that they are stored separately in memory and are also dissociated at retrieval. Specifically, verbatim traces are thought to be more vulnerable to forgetting than gist traces (*principle of retrieval dissociation*; Brainerd & Reyna, 2002; 2004). In other words, over time, memory tends to rely more on gist information than specific details about an experience. Also, the parallel processing of gist and verbatim traces is consistent with the distinction between episodic and semantic memory. Tulving (1983) proposed that access to semantic memory is immediate whereas, access to episodic memory requires active search to retrieve information. In the context of memory for repeated events and similar to schema theory, the gist or the general pattern of events is more accessible in memory over time, compared to details specific to individual instances.

In addition to increased recall of the general routine of repeated events, schema-processing also increases the likelihood that details from one instance will be confused with details from other instances. McNichol, Shute, and Tucker (1999), examined children's recall for events that occurred on three occasions or only once. The repeated events consisted of both fixed and variable details across instances. Although children who experienced the repeated events were accurate in recalling the fixed details, they were more inaccurate in remembering variable details than the children who only experienced the event once. The results showed that inaccuracies regarding the variable details were driven by intrusions, whereby children confused which instance a detail occurred in (see also, Lindsay, Allen, Chan, & Dahl, 2004). Such confusion errors between similar events are explained within the Source-Monitoring framework (Johnson, Hashtroudi, & Lindsay, 1993). According to the SM framework, we usually automatically attribute a memory to its source based on perceptual and contextual characteristics (e.g., "I heard this on the news yesterday"). However, making accurate source attributions of a memory with high precision might require deliberate

effort (e.g., "I heard this on the BBC news while driving to work around 8.30am"). Research shows that source misattributions are more frequent when witnessing repeated similar events. For instance, in a series of experiments by Lindsay and colleagues (2004), participants listened to a narrative description of an event and then viewed a video that was either similar to the narrative or not. They were asked to report only what they witnessed in the video. The results showed that intrusions between the witnessed event and the narrative were sometimes reported, but they increased when the narrative was similar to the witnessed event. Notably, in the context of the FTT, information about the source of an event is processed and stored as a verbatim trace, as it is specific to the actual experience. Therefore, as verbatim information, it is more susceptible to forgetting (cf. gist traces) and might not be recalled at retrieval (Brainerd & Reyna, 2004). Instead, schema theory explains errors of intrusion between events as a result of increased reliance of memory on schema-consistent details over time, since variations are absorbed by the script (Brewer & Nakamura, 1984; Hudson, Fivush, & Kuebli, 1992). However, schema theory suggests that there is one exception where source monitoring and hence, the particularization of specific instances is improved, that is when details deviate from the script.

Deviations from the script, are similar to variations, in that they also change from one instance to another, but they are considered as atypical and unpredictable (e.g. a mistake in one's order at the restaurant; e.g., Connolly, Gordon, Woiwood, & Price, 2016). Unlike variations, deviations are likely to be recalled because they attract attention at encoding and require increased resources to be integrated to the script, thus resulting in a strong memory trace (Anderson, 1983; Brewer & Nakamura, 1984). In a study by Graesser, Gordon, and Sawyer (1979), participants were asked to describe several actions that typically occur in a series of activities, such as going to a restaurant or a birthday party. Based on these descriptions, researchers composed several stories

that included i) only typical actions; ii) typical and atypical but related to the story actions; iii) and typical and atypical unrelated actions. A new group of participants listened to two scripts of each version and completed a recognition test where they had to decide whether actions were presented in the story or not. The results showed that participants were accurate in judging if atypical actions were presented, regardless of whether they were related to the story or not. However, they were unable to discriminate whether typical actions were presented or not. The findings suggest that deviations were easier to discriminate because they served as "tags" for specific instances (*script pointer plus tag;* see also Abelson, 1981). Research by Farrar and Goodman (1992) further tested the "script pointer plus tag hypothesis" by examining children's memory for deviations during interactive events. Four- and seven-year-old children participated in a series of play activities on two or four occasions. At the last visit, children were presented with new activities that deviated from the typical routine. When interviewed under free and cued-recall, seven-year-old children were more accurate than four-year-olds in discriminating details from the deviation and the typical routine (see also Farrar & Boyer-Pennington, 1999). More recently, research with both children and adults, shows that the presence of deviations in a specific instance benefits recall overall for all instances (general effect; Connolly et al., 2016; MacLean et al., 2018) or only for that specific instance in which they were presented (targeted effect; MacLean et al., 2018). In other words, deviations are more memorable because they violate the script and are therefore distinctive and salient in the memory network (Davidson, 2006; Means & Loftus, 1991). Thus, information elicitation techniques could capitalize on any positive effects that deviations may have on the encoding of repeated events to improve recall. To this end, this thesis aimed to examine whether the presence of deviations in an instance of witnessed repeated events would benefit recall, with

respect to the reporting of more correct information for the specific instance and for the whole series of events.

Previous research examining adults' recall of medical visits over prolonged periods of time found that use of various retrieval techniques, such as cued-recall and placing events on a visual timeline increased the likelihood of particularization and accurate source attributions (Means & Loftus, 1991). More recently, Leins et al. (2014), extended the Cognitive Interview with the use of various mnemonics including a timeline adapted after Hope et al. (2013b), to facilitate retrieval for multiple family meetings. The results showed that participants reported twice as many meetings and twice as much information about the events with the extended Cognitive Interview, compared to a free recall format. Also, timelines and event calendars have been successfully used in survey methodology to elicit information about long life periods for medical records, educational background, and intimate partner violence incidents (Belli, 1998; Van der Vaart & Glasner, 2007; Yoshihama & Bybee, 2011). The findings of survey research suggest that the use of event history calendars elicited more information compared to a standard interview because they offer a temporal reference point to interviewees (Belli, Stafford, & Alwin, 2009). Recent research on the timeline technique shows that mock-witnesses reported more correct details and more correct attributions of actions and statements to perpetrators (cf. free recall format), after witnessing conversations between multiple speakers both on a single and on three occasions (e.g. hostage scenario, gang meetings; Hope, et al., 2018). Therefore, there is some evidence which suggests that the timeline technique could facilitate recall for repeated witnessed events, by capitalizing on the temporal organization of repeated events in memory (Conway & Pleydell-Pearce, 2000; Tulving, 1983). One of the main aims of the current thesis was to test the effectiveness of the

timeline technique in improving recall for both single and repeated multi-actor events relative to a free request for information.

Multiple attempts at retrieval and strategic monitoring over reporting

As outlined so far in this review, episodic memory is a coherent representation of an event formed by multiple traces that are part of an associative structure (e.g., Anderson, 1983; Bower, 1967; Tulving, 1983). For various psychological reasons, not all encoded details of an event are always accessible. For instance, in a series of experiments, Anderson and Pitchert (1978) asked participants to study a story about two boys skipping school and spending the day at one of the boys' homes. The story consisted of details that would be relevant to the point of view of both a burglar and a homebuyer. Participants recalled the story twice, from each different perspective. Results showed that participants recalled additional information the second time, which was initially unimportant but became relevant when they shifted perspectives. The findings suggest that participants relied on a different schematic view of the event that guided an additional memory search for important information. Fisher and Geiselman (1992) were the first to apply the use of multiple and various retrieval techniques in the Cognitive Interview, to increase access to different aspects of memory and therefore, to improve recall for eyewitnesses and victims. To this end, the Cognitive Interview begins with a "report everything" instruction to encourage the interviewee to provide a free narrative about the witnessed event, which is then followed by various memory-enhancing strategies to facilitate access to different memorial information. Among other techniques, interviewers can use open-ended questions following the free report to prompt for depth of information ("Tell me more about x, y, z"), in a manner that does not interfere with the interviewee's own idiosyncratic retrieval processes (*witness-compatible questioning*; Fisher & Geiselman, 1992). For instance, if the interviewee is describing a specific perpetrator, the

interviewer should not inquire about a perpetrator that was mentioned earlier, as that would interrupt the spontaneous retrieval of information (Fisher, 1995). Notably, the strategy of witness-compatible questioning contrasts with common interviewing approaches were interviewers adhere to a checklist of pre-set questions, rather than follow the interviewee's pattern of retrieval (Wells, Memon, & Penrod, 2006).

As with the use of different cues, follow-up questions can prompt an additional memory search, because the information included in the question can act as a cue for a specific part of the witnessed event (Gabbert et al., 2016; Ibabe & Sporer, 2004). Research on practitioners' use of the Cognitive Interview, shows that witness compatible questioning is perceived to be one of the most useful components (Kebbell, Milne, & Wagstaff, 2008). Certainly, it is likely to be necessary to use follow-up questions to elicit further information following an initial report. For instance, Hope, Gabbert and Fraser (2013a), and Smeets, Candel and Merckeclbach (2004) consulted with legal professionals to compose a list of forensically relevant information about a witnessed event and examined police officers' and mock-witnesses' recollections, respectively. Both studies showed that although the interviewees were accurate, they also omitted a significant amount of critical information, which interviewers could prompt for with additional questions. In both intelligence and criminal investigation contexts, interviewers commonly ask follow-up questions to prompt for depth, and to clarify reported details (Evans & Fisher, 2011; Shepherd & Griffiths, 2013). However, although numerous studies have examined the mnemonics used in the Cognitive Interview, there is limited empirical research on the efficacy of follow-up open-ended questions based on a free narrative (Brunel, Py, & Launay, 2013; Colomb & Ginet, 2012; Dando, Wilcock, Milne, & Henry, 2009b; Davis, McMahon, & Greenwood, 2005; Memon, Wark, Bull, & Koehnken, 1997; Paulo, Albuquerque, & Bull, 2013). One of the aims of this thesis was to

systematically test the efficacy of follow-up questions to an initial report provided with the timeline technique, by examining the amount and the accuracy of the new information.

Follow-up questioning is likely to lead to the reporting of new information by encouraging additional memory searches within the memory network (Anderson, 1983; Gilbert & Fisher, 2006). However, based on previous research, we cannot predict an equally direct outcome with respect to the accuracy of the new reported information. Meta-analyses on the effects of the Cognitive Interview on reporting show that this approach elicits more information compared to standard interviewing approaches. Although most of the reported information is correct, there is also a small increase in erroneous details (Kohnken, Milne, Memon, & Bull, 1999; Memon, Meissner, & Fraser, 2010). It is suggested that increased erroneous reporting may be a result of how interviewees strategically regulate their memory output in an informativeness-accuracy trade-off (Goldsmith, Koriat, & Weinberg-Eliezer, 2002; Koriat & Goldsmith, 1996; Memon et al., 2010; Roberts & Higham, 2002; Yaniv & Foster, 1995). That is, when asked to report their recollection of an event interviewees are faced with two competing demands, to be informative and accurate (Goldsmith et al., 2002). Although, retrieval techniques can facilitate recall, the amount, type and quality of information reported is mediated by the interviewee's use of metacognitive regulation strategies. Informativeness relates to the amount of memory output but also to the level of a detail's precision or *granularity*. Accuracy on the other hand relates to whether the information is correct or incorrect. For instance, coarse-grain details (i.e., broad or imprecise) about a perpetrator's description (e.g., "he was wearing a dark colour jacket") are less informative compared to fine-grain details (i.e., specific or precise; e.g., "he was wearing a navy-blue jacket"). Naturally, informative answers are preferable. However, if an interviewee is uncertain about their

recollection, fine-grain details are less likely to be correct (cf. coarse-grain details) (Koriat & Goldsmith, 1996; Wells & Brewer, 2008; Yaniv & Foster, 1995).

Importantly, regulation of the quantity and quality in reporting differs when information is spontaneously provided in a free narrative compared to information provided in response to specific questions. Koriat and Goldsmith (1996) administered a general-knowledge test to participants under both forced-report and free-report conditions. That is, initially participants were required to answer all the questions and assess how likely they thought it was that each answer was correct on a scale of 0-100%, whereas subsequently they were free to answer a question or not. In the latter reporting condition, they were provided with monetary incentives for accurate answers. They were provided and penalized with the same amount of money for correct and incorrect answers, or they were penalized for incorrect answers with ten times the amount they were rewarded with for correct answers. There was no penalty for not answering a question. The results showed that when participants were free to choose whether to report an answer or not, they answered questions that they were certain about, with participants that received high penalties for incorrect answers exercising a stricter criterion at reporting (cf. participants with moderate penalties). Also, participants under free-report conditions were able to monitor the accuracy of their answers to some cost relative to how many questions they answered, but the cost was greater when they were highly penalized for inaccurate responses. In other words, participants monitored the accuracy of their responses by volunteering answers that exceeded a pre-set criterion of confidence that the answer was correct (*satisficing model;* Ackerman & Goldsmith, 2008; see also Goldsmith et al., 2002). Also, participants were more effective at monitoring their accuracy when they were able to control whether to volunteer or withhold an answer (*control of report option*). Finally, participants were able to increase the accuracy of their answers when provided with

accuracy incentives, but that resulted in an accuracy informativeness trade-off, where fewer responses were certainly accurate.

Further research by Goldsmith and colleagues (2002) used the same procedure as Koriat and Goldsmith (1996) with the difference that when providing an answer to a question, participants were asked to provide a coarse-grain and a fine-grain answer. Specifically, the questions referred to quantitative information (e.g., "When did Boris Becker last win the Wimbledon men's tennis finals?"), and therefore a coarse-grain answer included a wide interval (e.g., ten years), whereas a fine-grain answer included a narrow interval (e.g., three years) or a specific value. In a second phase, participants were free to choose whether to provide a coarse-grain or fine-grain answer. Across three experiments, the procedure further included i) an instruction to participants to provide a confidence criterion for both coarse and fine-grain answers on a 0-100% scale, assessing the likelihood that the answer was correct, and ii) different monetary incentives, where fine-grain answers were rewarded either five times or twice as much as coarse-grain answers, similar to Koriat and Goldsmith (1996). The results showed that participants strategically regulated the precision level in their answers. Although, in general they showed a preference for reporting informative answers, they only did so when they assessed these answers as probably correct and reported more fine-grain answers when incentivized to do so by accordingly reducing their confidence criterion for accuracy (see also, McCallum, Brewer, & Weber, 2016). Therefore, in addition to monitoring accurate reporting via opting to report information or not, participants are able to regulate the level of granularity to balance both informativeness and accuracy (see also, Evans & Fisher, 2011; Weber & Brewer, 2008).

Based on the findings of Koriat and Goldsmith (1996) and of Goldsmith and colleagues (2002), researchers suggest that interviewers do not require or force interviewees to report

information and encourage them not to guess when uncertain about their memory to promote accurate reporting (Fisher & Geiselman, 1992; Memon et al., 2010). More recent research suggests that witnesses can effectively monitor the accuracy of their responses by responding "I don't know" if uncertain (Ackerman & Goldsmith, 2008; Scoboria & Fisico, 2013). Similarly, the use of reporting formats that invite a free report are more likely to facilitate effective monitoring of accuracy as interviewees can decide on what information to volunteer (e.g., Evans & Fisher, 2011). Importantly, although the use of open prompts also allows interviewees to control their information output to a larger degree than closed and multiple-choice questions (Fisher, 1995), follow-up questions by definition, prompt for additional details to what is already reported. Therefore, it is less clear how interviewees can balance informativeness and accuracy. To this end, one of the aims of this thesis was to examine whether warnings about accurate reporting can improve monitoring of accuracy in response to follow-up questions.

Outline of the chapters in this Thesis

The aim of the current research is to test the use of memory-enhancing techniques and formats to improve recall of multi-actor events witnessed on both single and repeated occasions. Based on the literature reviewed here, there is evidence that various cues and techniques, which are grounded in the notion of the associative structure of memory, can effectively facilitate the retrieval and reporting for past experiences. However, there is limited research on methods to facilitate recall of multi-actor events (Hope et al., 2013b). In this thesis, we report four experiments in which we tested the effectiveness of a self-generated cue mnemonic in conjunction with the timeline technique and the use of follow-up open-ended questions for multi-actor events witnessed under different circumstances.

In Experiment 1, we tested a theory-driven mnemonic to facilitate recall for an event witnessed under sub-optimal conditions. Although the effects of divided attention on memory performance are well documented in the literature (e.g., Craik et al., 1996; Naveh-Benjamin, 2000), there is a lack of research on retrieval strategies to facilitate access to poorly encoded memories. Based on research on effective cued-recall retrieval, we tested a self-generated cue in conjunction with the timeline technique (cf. other-generated cues and no-cues) for an event encoded under full and divided attention. This experiment is presented in Chapter 2.

In Experiments 2 and 3, we systematically examined the efficacy of open-ended questions to probe on gaps and inconsistencies following up on an initial self-report, provided with either the timeline technique or a free recall reporting format. Although the use of additional prompts may lead to further reporting, the quantity and accuracy of the new information is likely to be mediated by interviewees' metacognitive monitoring strategies (Goldsmith et al., 2002). We draw from research on metacognitive monitoring and precision over reporting to interpret interviewees' responses to follow-up prompts, with respect to the quantity and quality of the additionally reported information. These experiments are presented in Chapter 3.

In Experiment 4, we tested the effectiveness of cues, the timeline format and follow-up prompts, as part of a multi-method interviewing format to facilitate the retrieval and particularization of repeated events (Experiment 4). Despite a wealth of research on the use of interviewing strategies to improve children's recall of repeated events (e.g., Brubacher et al., 2014), research on adults' recall is limited. We draw from schema theory, and the Fuzzy-Trace-Theory to examine recall for specific instances and deviations from the routine of repeated events. This experiment is presented in Chapter 4.

In the general discussion we provide an overview of the key findings. We also consider the theoretical implications of our results for research on memory retrieval of witnessed events, and the practical implications for applied information elicitation contexts. Finally, we discuss some of the limitations of the current experiments together with suggestions for future research. This discussion is presented in Chapter 5.

Chapter 2: The Benefits of a Self-Generated Cue Mnemonic for Timeline Interviewing

Abstract

Obtaining detailed accounts from individuals who have witnessed complex events under challenging encoding conditions presents a difficulty for investigators. In the present research, participants ($N = 132$) reported their recall of an event witnessed under full or divided attention using a timeline reporting format. Extending the timeline technique to assess the relative performance of two additional mnemonics, self-generated cues and other-generated cues, participants provided an account across three timeline reporting conditions comparing the efficacy of self-generated cues, other-generated cues, and no cues (control). Mock-witnesses using self-generated cues provided more correct details than mock-witnesses in the other-generated cues or no cues conditions, under full but not under divided attention conditions. There was no difference between cue conditions with respect to the number of errors reported across attention conditions. Findings show self-generated cues to be a promising addition to interviewing techniques as a retrieval support mnemonic with implications for applied contexts.

Introduction

Successful criminal and intelligence investigations rely on detailed and accurate information from suspects, witnesses, victims, and informants (Borum, Gelles, & Kleinman, 2009; DeClue, 2010). However, memory for experienced events is fallible and hence, sometimes inaccurate and often incomplete (Frenda, Nichols, & Loftus, 2011; Loftus, 2003). Obtaining high-quality information can become even more difficult in cases of complex multi-actor events witnessed under challenging conditions. Given that 25% of violent crimes committed by strangers involve four or more perpetrators (Office for National Statistics, 2015), and that group involvement is common in terrorist activities (Ozgul, 2016), reporting of multi-actor events is relevant in both forensic and security contexts. To date, only a small body of empirical research has examined ways to improve intelligence gathering practices with calls for more focused contributions in this area (Granhag et al., 2014). The current research extends the timeline technique (Hope et al., 2013b), which uses an innovative reporting format to enhance retrieval of complex events, by testing the introduction of a new mnemonic, self-generated cues, to facilitate recall for multi-actor events witnessed under optimal (full attention) and sub-optimal conditions (divided attention).

Use of Cognitive Mnemonics in Interviewing

The use of mnemonics is already embedded in gold standard investigative interviewing practices. One example is the Mental Reinstatement of Context (MRC) of the Cognitive Interview (CI; Fisher & Geiselman, 1992). 'Context reinstatement' capitalizes on the notion that recall increases when there is an overlap between the conditions present at encoding and at retrieval (*encoding-specificity principle*; Tulving & Thomson, 1973; for a review, see Pansky et al., 2005). The

administration of the MRC mnemonic, which typically elicits more correct information than free recall (e.g., Dando et al., 2009a), involves directing interviewees to think back to the surroundings, their emotional state, and their thoughts around the time of the event (Memon et al., 1997) using pre-defined generic instructions.

Although the encoding-retrieval match appears to aid memory, it is the quality of cues that moderates the extent to which retrieval improves (Nairne, 2002). Cues effectively facilitate retrieval when they are distinctive in addition to satisfying the encoding-retrieval match (Tullis & Benjamin, 2015; Watkins & Watkins, 1975). A distinctive cue uniquely matches a memory to the exclusion of other related memories (*principle of cue overload*; Nairne, 2002). Therefore, to be effective, cues need to be encoded within the context of the witnessed event (*encoding-specificity principle*), and to offer diagnostic information identifying a single target to the exclusion of others, rather than matching multiple related targets (i.e., matching but not distinctive) (Goh & Lu, 2012; Nairne, 2002). To date, research on the efficacy of cues in interviewing has mainly focused on cues generated by an interviewer, such as in the administration of context reinstatement techniques. However, recent work (Wheeler, Gabbert, Hope, Jones, & Valentine, 2017) examined a new mnemonic, self-generated cues and found, across two studies, that self-generated cue techniques increased reporting, with no cost to accuracy, in comparison to cues generated by another witness (other-generated cues), or free recall.

Self-generated cues are salient details that are actively generated by the individuals themselves and facilitate retrieval of a target memory (Wheeler & Gabbert, 2017). When episodic information is recalled, stored traces are activated and these prompt related details, thereby "spreading activation" throughout an associative network (Activation Theory; Anderson, 1983). Every attempt to remember a detail strengthens the memory trace. The stronger the

memory, the more likely it is that it will be recalled later and that it will activate associated

memories (Anderson, 1983). Similarly, Anderson and Conway (1993) showed that, when asked

to list event-details in free recall, participants first listed "distinctive details" (i.e., "details that

really stand out and make that memory what it is", p. 1188). Then they listed other details, highly

associated with those distinctive details. Thus, self-generation of distinctive cues can trigger

related memories by tapping on a common theme (Anderson & Conway, 1993; Belli, 1998).

More recently, Berntsen, Staugaard, and Sørensen (2013) showed that it is possible to activate

specific involuntary autobiographical memories in the lab, by manipulating the unique match

between cue and item.

In light of Anderson and Conway's (1993) findings, use of self-generated cues (i.e., the

most memorable details), should trigger the retrieval of related event-details while excluding

unrelated details, thus satisfying both the encoding-specificity principle (Tulving & Thomson,

1973), and the principle of cue overload (Nairne, 2002). Therefore, the present study tests the

effectiveness of self-generated cues in comparison to other-generated cues and no cues (control)

across timeline reporting conditions. To maximize our test of the efficacy of self-generated cues,

in the other-generated cues condition, we administered standard MRC instructions as a generic

mnemonic (i.e. not generated by the witness). Although MRC instructions do not provide

directive cues to specific aspects of an event, they suggest aspects the rememberer might focus

on during retrieval. Following Wheeler et al. (2017), we predicted that use of SGC would

activate unique associated memories, thus facilitating higher rates of correct recall. To examine

the effectiveness of cues, and given previous research showing that accounts can be incomplete

despite being accurate (Hope et al., 2013a; Smeets et al., 2004), we also explored how the use of

mnemonics affects account completeness for critical details.

Obtaining information using the Timeline Technique

The timeline technique (Hope et al., 2013b) uses a reporting format with a physical timeline to facilitate retrieval of multi-actor events. In Hope et al. (2013b), the timeline technique elicited more accurate information than free recall for a multi-actor event and enhanced the reporting of connections between perpetrators and actions, at immediate testing and after a two weeks' delay. Importantly, instead of asking for a linear narrative of the events, the timeline format encourages witness-compatible reporting whereby interviewees can report events as they remember them, at any point of the timeline, and re-arrange details if necessary. The current study combines this reporting format with the distinctiveness of self-generated cues to extend the timeline technique and evaluate a novel mnemonic.

Attention and eyewitness memory

Given the role of attention for successful encoding of witnessed events (for a review, see Pansky et al., 2005), a secondary aim was to examine recall under different encoding conditions. When witnessing a real crime, the experience of stress or physiological arousal can divert attention to aspects of the scene and/or to internal thoughts (Lane, 2006). However, laboratory studies typically use optimal conditions where participants pay full attention to events, thus possibly overestimating witnesses' memory performance (Ihlebaek, Løve, Eilertsen, & Magnussen, 2003). Although there is some evidence of enhanced recall using cued versus free recall when attention is divided at encoding (Backman & Nilsson, 1991), many studies have shown that divided attention has a robust negative effect on later remembering across stimuli (e.g., word lists, actions, pictures etc.; e.g. Craik et al., 1996; Mulligan, 2014; Naveh-Benjamin et al., 2006).

Using a mock-witness paradigm, Lane (2006) also found that divided attention at encoding resulted in lower accuracy and greater suggestibility to misinformation.

Based on Activation Theory (Anderson, 1983) and given previous positive results for cued versus free recall under divided attention (Backman & Nilsson, 1991), we predicted that use of self-generated cues should enhance retrieval of even weakly encoded traces through the activation of memorable and associated details. Although witnesses under divided attention conditions were expected to provide less information overall, indicating poorer episodic memory, we hypothesised that witnesses in the self-generated cues condition would provide more correct information (cf. other-generated cues and no cue conditions) under both encoding conditions.

Method

Participants and Design

A G*Power statistical analysis (Faul, Erdfelder, Lang, & Buchner, 2007) indicated that a sample of 132 participants was required for a 95% chance of detecting a large effect size (Cohen, 1992) for the main effects of Attention and Cues on recall, based on previous findings (e.g., Craik et al., 1996; Naveh-Benjamin et al., 2000; Wheeler et al., 2017). A total of 135 participants were recruited through the department's participation pool and through advertisements on the university campus. Participants were randomly allocated to a 3 (Cues: self-generated cues vs other-generated cues vs no cues) x 2 (Attention at encoding: divided attention vs full attention) between-subjects design. Data were excluded for three participants who, respectively, did not meet the English fluency criterion, did not follow the instructions in the divided attention task, and experienced an unanticipated interruption during reporting. The reported analyses are based on the data for the remaining 132 (85 females; 18-59 years of age; $M_{age} = 25$ years, $SD = 8.91$)

participants, with 22 participants allocated per group cell (self-generated cues x full attention; self-generated cues x divided attention etc.).

Materials

     Stimulus event. Consistent with Hope et al. (2013b), the stimulus event was a multi-perpetrator short film lasting 1 min 20s. The event showed an assault and robbery by five male perpetrators against a female victim. The film starts with three males loitering by a parked car. Two other males join them. A woman walks toward the group carrying a laptop computer bag and tries to walk past them. They surround her, and one male is seen threatening her with a crowbar. Her bag is taken from her and passed between several perpetrators, while another perpetrator films the incident on his cell phone. At the end of the event, the perpetrators run away with the bag. Although there was an audio component to the video stimulus, this was mainly background traffic / outdoor noise. The content of what was said by the gang members was inaudible (in all conditions) and, as such, would not offer any additional information about the incident or actions performed.

     Divided attention task. Participants allocated to the divided attention condition listened to an audio recording of a series of numbers and were instructed to respond by pressing a key when an even number was heard (adapted from Naveh-Benjamin et al., 2006) while they watched the stimulus event. The number of correct responses (hits) and reaction times to the auditory task were recorded to verify that participants attended to the distraction task as instructed. Participants who performed at lower than 50% success at the task (from a total of 18 hits) were to be excluded from analysis, however no participants had to be excluded on this basis. As noted, one participant was excluded for not following the instructions (i.e. pressing a key to every number and not to even numbers only).

Timeline Technique. The timeline technique consists of three elements: (1) a physical cardboard timeline (33 in. x 12 in.) that has a horizontal line running at mid-point from one end of the card to the other representing the temporal context during which the event occurred; (2) blank, white, lined person description cards (5 in. x 3 in.); (3) blank yellow action cards with a semi-adhesive strip on the back (3 in. x 3 in.) for easy removal and rearrangement on the cardboard timeline.

Other-Generated Cues Instructions. Participants in the other-generated cues condition were administered a version of Mental Reinstatement of Context (MRC) instructions. Consistent with the standard administration of MRC, participants were instructed to think back to when they witnessed the event, to think about what they could see, what they could hear, what the surroundings were, and what they were thinking and feeling at the time. Participants were encouraged to consider whether each prompt helped them remember other things that occurred in the event. Participants were also invited to close their eyes or look at a blank wall if it helped them concentrate (Dando et al., 2009a).

Self-Generated Cues instructions. The instruction in the self-generated cues condition was adapted from Gabbert, MacPherson, and Hope (2014). Participants were instructed to write down the first six things that they remembered seeing or thinking when viewing the event and to then focus on each of these things one at a time, considering for each whether or not that memory helped them remember other parts of the event. Participants were also encouraged to close their eyes or look towards the wall to focus.

Procedure

Half of the participants watched the stimulus event while the other half watched the stimulus event and simultaneously performed the auditory distraction task. All participants were given the following instruction prior to watching the stimulus: "During the study, you will watch a video of a crime event. Please pay attention because later you will be asked to provide an account of the event." Participants in the divided attention condition also received the following instruction: "While you watch the video you will also listen to an audio recording of a series of numbers through the headphones. Please press the "enter" key on the keyboard every time you hear an even number".

After witnessing the event, all participants completed a 10-minute filler task (Sudoku puzzle). They were then moved to a different room and were given instructions for reporting their account of what happened in the event using the timeline reporting format and the instructions used in Hope et al. (2013b). Participants in all conditions were told to report all the details about the event and the people involved that they remember, without guessing. Participants were instructed on how to use the person description cards to provide information about the people involved by using a new card per each individual. They were also instructed to use action cards to describe any actions and information about the sequence of the events. The instructions further advised that they should place all the cards on the timeline format in order, with links between the individuals reported and each action to show "who did what and when". Depending on condition, participants also received instructions to use Mental Reinstatement of Context, or the self-generated cues. Participants in the no cues (control) condition did not receive any further instructions and simply reported their account using the original timeline technique reporting instructions. Participants were left alone in the room while providing their account by

completing the timeline format, although the researcher was available nearby to answer any questions if necessary. Participants were not asked any questions about the witnessed event by the interviewer. All participants were video-recorded while generating their accounts. After participants finished providing their account, they were thanked and debriefed. We should note that since participants in the self-generated cues condition were provided with a draft piece of paper to list the first six things they remembered from the witnessed event, it was evident at completion of the experiment that all participants had indeed written down their own cues.

Coding

The details reported by the participants on the person and action cards and placed on the timeline format were then coded according to the scoring template used in Hope et al. (2013b). Briefly, each detail reported was identified as a Person (P), Action (A), Object (O) and Setting (S) detail. A detail was scored as accurate if it was present in the stimulus event and described correctly. Details that were subjective or vague were not coded for accuracy. A secondary coding was conducted regarding the accuracy of attributions of the 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 3 raises the crowbar). Moreover, sequencing errors were 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. Therefore, this example reflects a total of one sequence error.

Finally, the reporting of critical details was coded according to the process described in Smeets et al. (2004). Six legal professionals were separately asked to view the stimulus event and list the details that they considered to be relevant to the investigation and legal proceedings of an assault. All details mentioned by at least four of the six raters were included in a list of 24

critical details, which was used to code participants' reports [a detailed description of the coding is provided in Supplemental materials; see Appendix A].

To assess overall inter-rater reliability, 20 interviews (i.e. 15% of all interviews) were randomly selected and coded independently by a rater who was blind to experimental conditions. Inter-rater reliability was high, *ICC* = .98, 95% CI [.967, .988] across coding categories.

Results

Bonferroni-corrected pairwise comparisons were conducted for all interactions. In the interests of parsimony, we only report pairwise comparisons where they indicate significant differences (even for non-significant interactions). Where Bonferroni-corrected pairwise comparisons are not significant (and therefore do not aid interpretation beyond the non-significant interactions), they are not reported.

Reporting of Correct Details

A between-subjects ANOVA showed a significant main effect of Cues, $F(2,126) = 4.39$, $p = .014$, $\omega^2 = .049$, for the number of correct details reported. Post hoc tests showed that, across attention conditions, more correct details were reported in the Self-Generated Cues condition than in the No Cues condition ($p = .012$). The number of correct details reported in the Other-Generated Cues condition did not differ from the number of correct details reported in the Self-Generated Cues ($p = .241$) and No Cues ($p = .718$) conditions. There was also a main effect of Attention, $F(1,126) = 24.78$, $p < .001$, $\omega^2 = .156$, with significantly more correct details reported in the Full attention condition than in the Divided attention condition. The interaction between Attention and Cues was not significant, $F(2,126) = 2.23$, $p = .111$, $\omega^2 = .018$. Bonferroni-corrected pairwise comparisons showed that more correct details were reported in the Self-Generated Cues

condition than in either the Other-Generated Cues ($p = .046$) or No Cues ($p = .002$) condition, under full attention, while there was no difference between conditions under divided attention ($p = 1.00$). Results for the number of incorrect details are reported in Supplemental materials (see Appendix A).

The effect of cues on the mean number of correct details reported within Full and Divided attention conditions are presented in Figure 2.1.
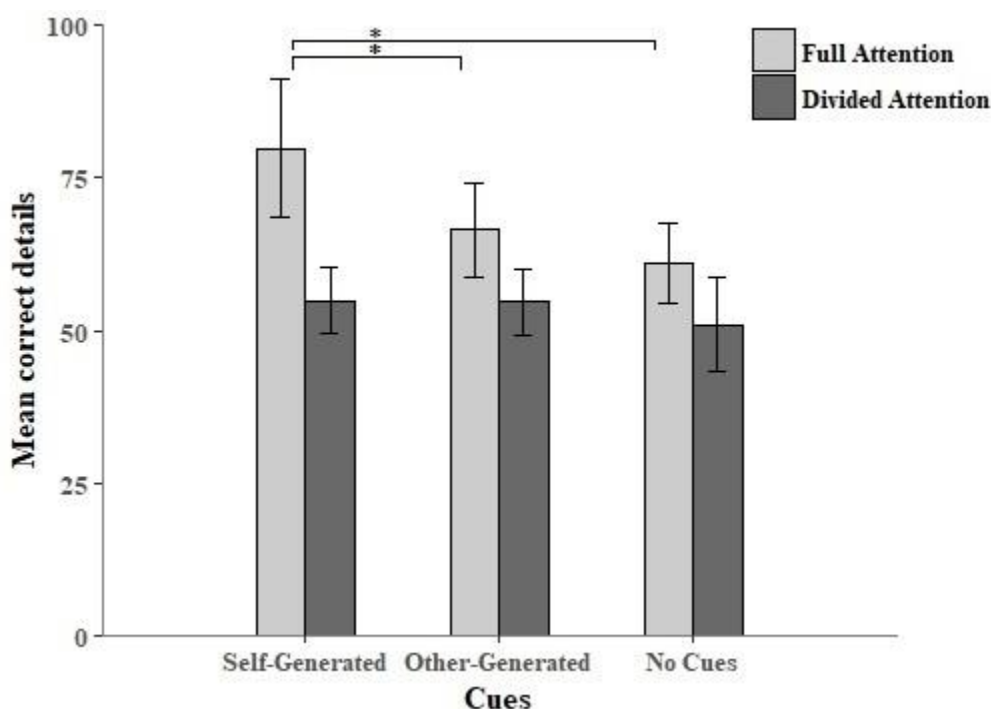
Figure 2.1. Mean number of correct details reported as a function of cues (Self-Generated Cues vs Other-Generated Cues vs No Cues) within Full and Divided attention conditions. Error bars represent ± 1.96 standard errors (95% confidence intervals). Asterisks indicate significant differences between cue conditions, *p < .05.

Accuracy Rate of Reported Details

Accuracy rate was calculated by dividing the number of correct details by the sum of both correct and incorrect details (total number of items) to obtain the proportion of accurate reported information. Levene's test was significant ($p = .004$). A boxplot showed that the distribution was

not symmetrical but negatively skewed with two outliers who had particularly low scores. However, given the overall robustness of the test, no action was taken. Analysis revealed a significant main effect of Attention, $F(1,126) = 10.37$, $p = .002$, $\omega^2 = .068$, with higher accuracy rates in the Full (cf. Divided) attention condition. There was also a main effect of Cues, $F(2,126) = 3.43$, $p = .035$, $\omega^2 = .036$, on accuracy rates. Post hoc tests with a Bonferroni adjustment showed that across attention conditions, there was no significant difference between the accuracy rate in the Self-Generated Cues condition and the accuracy rate in the Other-Generated Cues ($p = 1.00$) or No Cues conditions ($p = .188$). However, the accuracy rate in the Other-Generated Cues condition was significantly higher than the rate in the No Cues ($p = .039$) condition. The interaction was not significant, $F(2,126) = .63$, $p = .536$, $\omega^2 = -.005$. Bonferroni-corrected pairwise comparisons showed that there was no significant difference in accuracy rates between Self-Generated Cues and Other-Generated Cues conditions ($p = 1.00$), Self-Generated Cues and No Cues conditions ($p = .783$) or Other-Generated Cues and No Cues conditions ($p = .932$) under full attention. Under divided attention, there was a significantly higher accuracy rate in the Other-Generated Cues condition compared to the No Cues condition ($p = .036$), however there was no significant difference between accuracy rates in the Self-Generated Cues and Other-Generated Cues conditions ($p = .388$). Means for accuracy rates of reported details as a function of cues within attention conditions are presented in Table 2.1.

Table 2.1

Mean accuracy rates (SE) of reported details by cues (Self-Generated Cues, Other-Generated cues, No Cues) within Full and Divided attention conditions.

	Accuracy rates of reported details					
	SGC		**OGC**		**NC**	
Attention	*M (SE)*	95%CI	*M (SE)*	95%CI	*M (SE)*	95%CI
Full	0.89 (0.01)	[0.86, 0.92]	0.89 (0.01)	[0.86, 0.91]	0.86 (0.01)	[0.84, 0.89]
Divided	0.84 (0.01)	[0.82, 0.87]	0.87 (0.01)	[0.84, 0.89]	0.81 (0.02)	[0.78, 0.84]

Attribution of Actions

With respect to correct person-action details, there was a significant main effect of Attention, $F(1,126) = 8.94$, $p = .003$, $\omega^2 = .058$, but not of Cues, $F(2,126) = .003$, $p = .997$, $\omega^2 = -.007$. The interaction between Attention and Cues was not significant, $F(2,126) = .21$, $p = .814$, $\omega^2 = -.012$. Results for incorrect person-action details are reported in supplementary materials. The main effects for correct person-action details are presented in Figure 2.2.
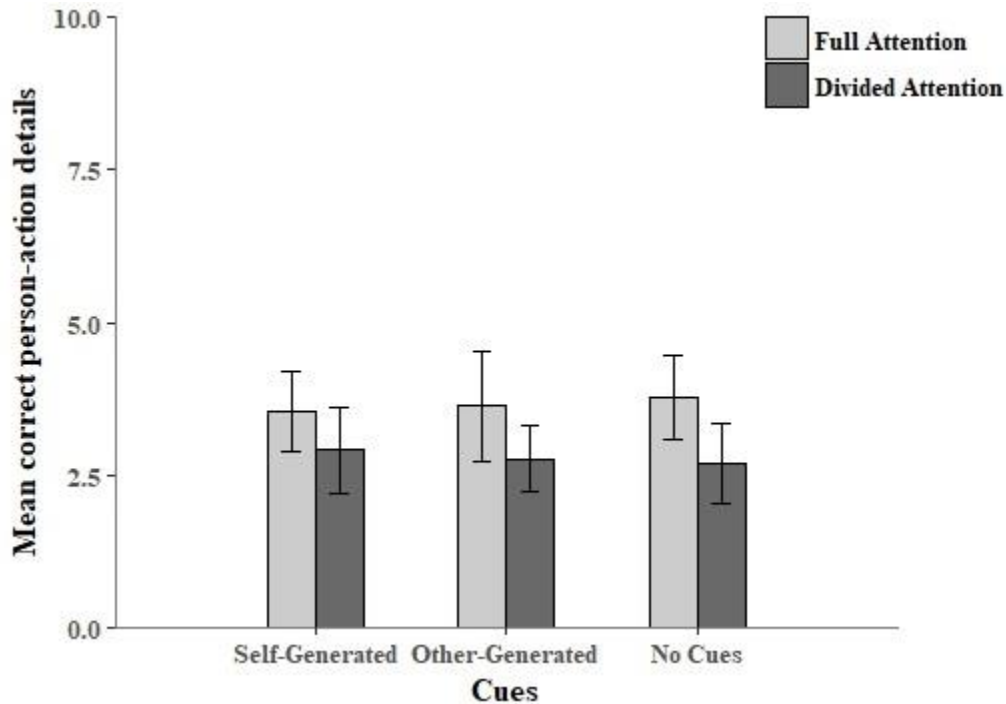
Figure 2.2. Mean number of correct person-action details as a function of cues (Self-Generated Cues vs Other-Generated Cues vs No Cues) and attention (Full vs Divided attention). Error bars represent ± 1.96 standard errors (95% confidence intervals).

Accuracy Rate of Person-Action Details

With respect to the accuracy rate of person-action details, there was no significant main effect of Attention, $F(1,126) = 2.08$, $p = .152$, $\omega^2 = .008$, or Cues, $F(2,126) = .10$, $p = .910$, $\omega^2 = -.014$. The interaction was also not significant, $F(2,126) = 2.77$, $p = .066$, $\omega^2 = .026$.

Sequence errors

There was a main effect of Attention $F(1,126) = 4.19$, $p = .043$, $\omega^2 = .024$, but not of Cues, $F(2,126) = .029$, $p = .971$, $\omega^2 = -.015$ on the total number of sequence errors reported by participants. The interaction between Attention and Cues for the total number of sequence errors reported by participants was significant, $F(2,126) = 3.75$, $p = .026$, $\omega^2 = .040$. Pairwise

comparisons showed that there were significantly more sequence errors made with the use of Other-Generated Cues under Full attention ($M = .55$, SE $= .05$) compared to the Divided attention condition ($M = .05$, SE $= .02$) ($p = .001$). However, there was no difference between attention conditions for the number of sequence errors made in the Self-Generated Cues ($p = .377$) and No Cues ($p = .556$) conditions. Levene's test was significant for the analysis of sequence errors ($p < .001$). Since the values in the reporting of sequence errors were overall very low ($M = .30$, SD $= .52$), no action was taken to recover the assumptions violation. Instead, emphasis was given to the fact that the overall mean number of sequence errors was low.

Results for the effects of Cues and Attention on the reporting of critical details and detail type (person, action, object, setting) are reported in Supplemental materials (see Appendix A).

Discussion

We tested the effectiveness of cognitive mnemonics used in conjunction with the timeline technique under full and divided attention. As predicted, mock-witnesses who used self-generated cues reported more correct details than mock-witnesses in other-generated and no cue conditions, at no cost to accuracy. However, this enhanced performance with self-generated cues was only observed under full attention. Participants under divided attention consistently reported less correct information than those under full attention, and there was no effect of cues under divided attention.

The apparent lack of benefit of self-generated cues under divided attention is noteworthy. The sizeable main effect of the divided attention task across cue conditions suggests that performing a secondary task significantly challenged attentional processes and likely drew participants' attention away from the target event, thus restricting encoding and retrieval (see

also Marsh et al., 2017, for a similar divided attention effect when participants were instructed to ignore distractions). These findings are consistent with literature on the powerful effect of divided attention on remembering (e.g., Craik et al., 1996) and, although it is not surprising that our task restricted encoding (as intended), it is possible that the to-be-remembered information was not stored from the outset, thus hindering retrieval despite the additional support of cues. Another possibility is that the self-generated cues manipulation was simply not powerful enough to access weakly encoded memories. Given that research on the effectiveness of memory-enhancing techniques under sub-optimal encoding conditions is limited, more research is needed to determine the most likely explanation. Research should also examine the effectiveness of self-generated cues possibly with more naturalistic divided attention measures, such as using a smartphone or conversing (e.g. Marsh et al., 2017), to delineate the limitations of the use of cues.

Nevertheless, mock-witnesses reported more correct information under full attention with self-generated cues than with other-generated cues. Possibly, the use of self-generated cues facilitated retrieval more effectively across the whole event by activating the "stronger" memories (Anderson, 1983) that distinctively identify associated targets (Nairne, 2002). It is also possible that initially identifying six event-details and processing them further might contribute to the self-generated cues advantage. By comparison, other-generated cues, administered here in the form of generic context-retrieval cues, failed to activate as many event-details. Further research is needed to increase understanding about the underlying mechanisms of self-generated cues relative to more generic cues (e.g., other-generated cues).

Another caveat to our finding of superior performance by self-generated cues is that there was no effect of cues on the reporting of critical details. Overall, only 50% of the critical details identified by legal professionals were reported across conditions, suggesting that even highly

accurate and detailed accounts can be lacking in information relevant to investigators (see Hope et al., 2013a; Smeets et al., 2004). Notably, most of these critical details related to specific details of the assault. It is possible that mock-witnesses did not appreciate the level of detail required or that, given the brevity of the event, such details were poorly encoded or simply not salient for participants and, therefore, not prompted by the self-generated cues. Future research might examine whether follow-up questioning facilitates the reporting of such details.

Regarding person-action links, there was no effect of cues on the number of correct attributions of actions. Accounts of witnesses using self-generated cues or other-generated cues did not include more person-action details than accounts of witnesses in the control condition, who only used the timeline technique. Therefore, the use of mnemonics did not increase the reporting of person-action details. Thus, features of the timeline technique (likely the use of different person and action cards and the instruction to show "who did what when") possibly drove the reporting of person-actions details. Indeed, in Hope et al. (2013b) reporting of person-action details did not differ between participants when using the timeline technique to participants using person and action cards only (Experiment 2). Given that self-generated cues increased retrieval of correct information overall but did not improve the reporting of person-action details compared to use of the timeline alone, it may be worth exploring whether self-generated cues and timeline capitalize on different retrieval processes to access different types of information.

Although our expectations about the benefit of self-generated cues across encoding conditions were not fully met, the results of self-generated cues in the full attention condition are promising. Notably for applied contexts where person descriptions are valuable in investigations (Brown, Lloyd-Jones, & Robinson, 2008; Gabbert & Brown, 2015), witnesses in the self-

generated cues condition reported more person details in their account provided with the timeline format, compared to witnesses in the other conditions, with person details being reported to a greater extent than any other details.

Current findings suggest that, when attention at encoding has not been compromised, self-generated cues may be a useful addition to interviewing techniques as a retrieval support mnemonic that promotes witness-led interviewing. In intelligence gathering, interviewers may be unaware of what information interviewees possess and what is memorable to each interviewee. Accordingly, the use of self-generated cues may support the interviewing process by facilitating an open-ended, largely self-administered report. Not only does this approach allow witnesses to report event-details in their own words; it also limits the potential for use of inappropriate or leading questions.

Chapter 3: A close examination of the efficacy of follow-up questions based on a self-report

Abstract

In information gathering contexts, after obtaining an initial report about an event, interviewers often ask follow-up questions to clarify or elicit further details. However, the information provided in response to such questions may not be as accurate as information that is spontaneously reported as part of an initial account. Across two experiments, we tested the efficacy of using open-ended questions to follow-up on an initial report provided with the timeline technique about a multi-actor event by examining the accuracy and the number of details in participants' responses. In the first experiment (N = 50), mock-witnesses used the timeline technique or a free recall format to provide an initial report. The use of follow-up questions elicited new information (18% to 22% of the total output) in both conditions. However, the accuracy of the responses to the follow-up questions was not as high as the initially reported information (60% vs 83%). In the second experiment (N = 60), we examined the use of pre-questioning instructions to improve accuracy for responses to follow-up questions based on an initial report provided with the timeline technique. Half of the participants were reminded to avoid guessing and were encouraged to feel free to withhold an answer and to consider the level of detail in their answers (i.e. provide general or specific details). Despite the use of pre-questioning instructions, the accuracy of responses to follow-up questions did not improve relative to a control group. New information was elicited (21% to 22% of the total output) across conditions, but as in the first experiment, accuracy in the follow-up phase was not as high as in the initial reporting phase (75% vs 87.5%). Results are discussed in relation to the role of metacognitive strategies in reporting and the use of follow-up questions in applied settings.

Experiments 2 and 3 are presented together in Chapter3 because these experiments are being prepared for publication together.

Introduction

In both intelligence and criminal investigation contexts, interviewers commonly ask follow-up questions to elicit specific details, and to clarify reported details and inconsistencies (Evans & Fisher, 2011; Shepherd & Griffiths, 2013). There is evidence that mock-witness accounts can be accurate but incomplete with regards to critical details that may be useful in an investigation, thus interviewers may need to prompt for further information (Hope et al., 2013a; Smeets et al., 2004; Roberts & Higham, 2002). However, because these prompts probe beyond what was initially reported, the additional information might not be as accurate as the information that was spontaneously provided in the initial reporting phase. The current experiments examine the efficacy of follow-up questions based on an initial free narrative.

Additional prompts or follow-up questions on witness accounts are used 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 probe for further information by using various memory-enhancing techniques, including a focused-retrieval phase where open questions are allowed to follow-up on parts of the 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, Mycklebust, & Grant, 2010). These questions are open-ended, information-seeking questions that prompt the interviewee to elaborate in more depth on what has previously been mentioned, followed by more specific prompts if necessary (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, after a free narrative is reported, the interviewer can follow-up using elements of the CI with broad and more specific questions to elicit more information.

Even when interviewees are cooperative, they are likely to omit details and to provide reports that include inconsistencies, particularly when reporting complex multi-actor events. Both errors of omission and inconsistencies naturally occur during the process of retrieval, but they have important implications in applied contexts. Details may be omitted because of forgetting or because further retrieval support is needed to access the encoded information, or because interviewees are unaware of what details interviewers consider to be relevant (Fisher & Geiselman, 1992). Prompting for specific omitted information can enable interviewers to gain more information that is related to investigative purposes (Brandon et al., 2018).

In the context of complex multi-actor events, some components of an event might be poorly encoded compared to other components of the same event (Fisher, Brewer, & Mitchell, 2009). This may result in the reporting of inconsistent details. For instance, one may report that four perpetrators were initially present but only describe three perpetrators at a later point. Given that both within and between-statement inconsistencies are perceived to be predictors of the reliability of one's account, 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 various functions in the interviewing process, by encouraging the interviewee to retrieve more information from memory and elaborate on an initial account.

The rationale behind the idea that follow-up questions can prompt further retrieval is based 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 serve as an opportunity to encourage a search through the memory network, facilitating access to additional memories. Similarly, based on a multicomponent view of memory (Bower, 1967), a memory is not a single representation of an event but rather comprises a network of many features. According to this notion, which also serves as the rationale behind certain memory-enhancing techniques used in the CI (Fisher & Geiselman, 1992), not all features are constantly available. Therefore, when a memory is not accessible by a particular probe or prompt, a different probe might be of use (see also Anderson & Pichert, 1978). The use of open-ended, non-leading prompts that do not introduce any new information but follow-up on a free narrative should be an effective means to 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.

The idea 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, consistently results in improved reporting of correct details compared to standard interviews, which do not include any retrieval techniques. However, results show that sometimes there is also a slight increase in incorrect details, as overall reporting increases (Kohnken et al., 1999; Memon et al., 2010). This increase in erroneous reporting is considered to depend on 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 et al., 2002; Koriat & Goldsmith, 1996). To achieve a balance between the two, research shows that they tend to strategically regulate the amount of information they report (Koriat & Goldsmith, 1996). Specifically, in a free narrative, interviewees can decide when to withhold or volunteer information based on how confident they feel 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 threshold for accuracy (*the satisficing model*; Goldsmith et al., 2002), then they volunteer the answer or withhold it instead (*control of report option*; Koriat & Goldsmith, 1996). Evidence shows that 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 (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, and as a result an increase in erroneous reporting is likely.

Although, the increased reporting of errors in the context of elaborate memory reports is attributed to metacognitive monitoring, we do not have a clear understanding of how this increase occurs throughout the interviewing process. Research on the benefits of the CI for memory recall has mostly focused on the effectiveness of the different mnemonics rather than on the use of additional prompts following an initial narrative (e.g., Brunel et al., 2013; Colomb & Ginet, 2012; Memon et al., 1997; Paulo et al., 2013). Similarly to the use of retrieval cues, asking

follow-up questions can also be a further prompt for the interviewee to search through their

memory (Fisher & Geiselman, 2010), yet systematic investigation into one's performance when

asked additional prompts is limited or incidentally reported without being the main research

focus. Therefore, the current research aimed to examine the efficacy of using open-ended

questions following a self-administered account provided with the timeline technique (Hope et

al., 2013b) or a free recall format. Also, it sought to explore the necessity of further prompting

for additional information from memory after an initial self-report (Experiment 1). In

Experiment 2, we aimed to refine the 'questioning procedure' by testing the use of instructions

that enhance accuracy in responding (Experiment 2) for a single witnessed event.

The nature of the first experiment was mainly exploratory in terms of the quantity of

additionally reported information (i.e., how much new information is elicited) and the quality of

the information (i.e., how accurate is the new information) in response to follow-up questions.

However, it was expected that the use of the timeline technique, which uses a physical timeline

format and interactive instructions to facilitate memory for multi-actor events, would elicit more

correct details for the witnessed event compared to the free recall format, as shown in Hope et al.

(2013b). Open-ended questions were used as invitations to elaborate on omitted information and

gaps (e.g., "Tell me more about [detail already mentioned]" and "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").

Experiment 1

Method

Participants and design

Fifty participants (37 Females, Age: M = 24.64, SD = 6.99, Range 18-47 years) were recruited and randomly allocated to a timeline ($n = 25$) or a free recall condition ($n = 25$). A post-hoc G* Power statistical analysis (Faul et al., 2007) using the effect size ($d = 0.69$) of our main hypothesized finding (i.e., the reporting of correct information using the timeline technique vs free recall format), showed that the achieved power was 0.78. Participants were recruited through the student participation pool and through advertisements circulated across campus. They were granted course credit or a £5 honorarium for participating. The dependent variables were the number of correct details, the number of correct person to action details, and the accuracy rates for both types of details. We also report the number of errors across phases.

Materials

Stimulus event. Participants witnessed a 1min20s long film of a multi-perpetrator crime event depicting an assault and robbery that was used in Hope et al. (2013b) and in Kontogianni et al. (2018). The film starts with three males loitering by a parked car. Two other males join them. A woman walks toward the group carrying a laptop computer bag and tries to walk past them. They surround her and one male is seen threatening her with a crowbar. Her bag is taken from her and passed between several perpetrators, while another perpetrator films the incident on his cell phone. At the end of the event, the perpetrators run away with the bag.

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 cardboard).

Follow-up open-ended questions. A list of five open-ended questions was composed to probe about additional information based on the initial account, in relation to omitted information, gaps, and inconsistencies/need to clarify:

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…)

Free recall reporting format. The free recall format consists of an A4 5-page long lined booklet.

Procedure

Participants were asked to take part in a study investigating factors that affect people's memory reports for witnessed events. Participants witnessed the stimulus event on a computer screen while wearing headphones. Although there was no audible dialogue in the stimulus (mostly background traffic noise), headphones were used to ensure that participants were not distracted

by any incidental surrounding noise. 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, participants were then presented with either the timeline format or the free recall format to provide their account. Participants in both reporting conditions were asked to provide all the details they remembered about the event and the people involved and to not make any guesses about things they did not remember. Participants in the timeline condition were instructed to use the person description and the action cards to provide their account and to show "who did what and when".

After providing their account with the use of either the timeline or the free recall reporting format, all the participants were asked follow-up open-ended questions about the event. The interviewer selected between three to five questions from a list, so that all participants across conditions were asked an equal number of questions. The topics were not pre-selected, instead the questions were asked based on what participants reported, using questions such as "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 instance, the interviewer would ask "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 interviewers to maintain the same phrasing of questions but avoid using a scripted list of cued recall questions. We should note that, although it was not explicitly stated, participants were not required to answer all the questions and if they answered by saying "I don't know" or "I don't remember", the interviewer moved on to the next question. All participants were finally asked if there was anything else they would like to report. During the questioning phase in both conditions, the participant's account remained on the table

and the interviewer would point to the specific part to which the prompt referred to when asking each question. The follow-up questioning phase was audio and video-recorded, but the camera was only focused on the table and the format with the participant's account. On completion of all of the questions, participants were debriefed and compensated for their time.

Coding

Coding of the interviews was conducted by the researcher according to the scoring template that was used in Kontogianni et al. (2018). Each detail reported was identified as a Person (P), Action (A), Object (O) and Setting (S) detail. A detail was scored as accurate if it was present in the stimulus event and described correctly. Details that were subjective or vague were not coded for accuracy. A secondary coding was conducted regarding the accuracy of attributions of the 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 3 raises the crowbar). The same coding scheme was used to code the responses of participants to the follow-up questions for type of detail and accuracy. Only new information was coded. 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. Therefore, this example reflects a total of one sequence error.

To assess inter-rater reliability across categories, 8 interviews (i.e. 15% of all interviews) were randomly selected and coded by an independent rater. Given the use of different reporting formats, coding was blind to hypotheses and research questions but not to the experimental conditions. Inter-rater reliability was high, *ICC* = .99, 95% CI [.987, .993].

Results

Initial Reporting phase

Participants in the Timeline condition reported significantly more correct details than participants in the Free Recall condition, $t(37.588) = 2.44$, $p = .020$, $d = 0.69$, $CI[0.12, 1.26]$. There was no difference in the mean number of errors between conditions, $t(48) = .087$, $p = .931$, $d = 0.03$, CI [-0.53, 0.58]. With respect to accuracy rates for reported information, there was no difference between conditions, $t(48) = .173$, $p = .864$, $d = 0.05$, CI [-0.51, 0.60]. Table 3.1 displays the Means and *SDs* of both correct and incorrect details, and accuracy rates across reporting phases.

Table 3.1

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 Report	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

An independent t-test analysis was conducted to examine the mean number of correctly reported attributions of actions to persons. The analysis showed that participants who used the Timeline reported a similar mean number of correct attributions ($M = 3.72$, $SD = 1.77$) to the participants who used the Free Recall format ($M = 3.36$, $SD = 1.87$), $t(48) = .700$, $p = .487$, $d = 0.20$, CI [-0.36, 0.75]. Regarding the overall accuracy of the reported attributions, there was also no significant difference between the two conditions. Participants who used the Timeline reported equally accurate information ($M = .80$, $SD = .22$) as participants who used the Free Recall format ($M = .81$, $SD = .22$), $t(48) = .103$, $p = .919$, $d = 0.06$, CI [-0.58, 0.53].

There was a significant difference between conditions regarding the number of sequence errors, $t(48) = 2.701$, $p = .010$, $d = 0.76$, 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

There was no difference in the number of follow-up questions that were asked across 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 $t(48) = .474$, $p = .638$, $d = 0.13$, CI [-0.69, 0.42], or incorrect details $t(48) = .630$, $p = .532$, $d = -0.18$, CI [-0.73, 0.38]. Nor was there any difference between conditions for the accuracy of details reported $t(48) = .447$, $p = .657$, $d = 0.13$, CI [-0.68, 0.43].

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, $F(1,48) = 87.57$, $p < .001$, $\omega^2 = .634$.

Total Reporting across Initial Report and Follow-up Questioning Phase

Participants in the Timeline condition reported a significantly larger number of correct details overall, compared to participants in the Free Recall condition, $t(38.662) = 2.290$, $p = .028$, $d = 0.65$, CI [0.08, 1.21]. There was no statistically significant difference between conditions for the total errors reported across phases, $t(48) = .427$, $p = .671$, $d = 0.12$, CI [-0.68, 0.44], or for the total accuracy rate across phases, $t(48) = .989$, $p = .328$, $d = 0.28$, CI [-0.28, 0.84].

Discussion

The main aim of the current study was to examine the efficacy of using open-ended questions following a self-administered account with respect to the quantity and the quality of the information reported in response to follow-up questions. The current findings show that a sizeable amount of additional information about the witnessed event was elicited when follow-up questions were asked. This additional information represented 18% of the total information reported in the timeline condition and 22% of the total information reported in the free recall condition. Possibly, the use of follow-up questions led to additional retrieval attempts focused on different components of the event. According to the activation theory of memory, these successive trials could cause the spreading activation of links through the network to retrieve further encoded details (Anderson, 1983). In other words, the use of open-ended prompts to elaborate based on the initially provided report further cued participants' memory for the witnessed event.

However, despite the opportunity to provide more information in response to the follow-up prompts, participants in the free recall condition still reported fewer correct details overall compared to those in the timeline condition. This overall benefit is because, consistent with previous research, more correct details were initially reported with the timeline technique than with the free recall format, without any cost to accuracy (Hope et al., 2013b; Hope et al., 2018). Instead, in the follow-up questioning phase, participants reported a similar amount of new information across conditions. Surprisingly, the two groups also reported a similar amount of attributions of actions to persons, which is inconsistent with Hope et al.'s (2013b) finding that the timeline technique facilitated the correct reporting of such attributions (cf. free recall).

Consistent with existing findings of highly accurate reporting when providing a free narrative, accuracy rates of the initial report were equally high across conditions (Hope et al., 2013b; Memon et al., 2010). In response to follow-up questions, however, the accuracy of the additional information reported was not as high, with an average accuracy rate of approximately 60% in both conditions. Indeed, previous research indicates that with increased reporting there is also a higher risk of erroneous reporting (Kohnken et al., 1999; Memon et al., 2010; Roberts & Higham, 2002). A possible explanation for the lower accuracy of the reported information provided in the follow-up questioning phase 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, by reporting information which 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 to the interviewee that there is an increased expectation for them to be informative (Grice, 1975). It is likely that this expectation resulted in them adopting a more liberal criterion of accuracy in order to still provide informative answers (Goldsmith et al., 2002). Therefore, the findings that the information provided in response to follow-up questions was not as accurate as the information provided in their initial report may have been the result of an accuracy-informativeness trade-off. In other words, when asked follow-up questions, participants were able to report new information, but the new information was not as confidently accurate as the initial report. In order to satisfy an informativeness criterion however, the interviewees likely volunteered more details while risking accuracy.

The current experiment served as a first step to examine the efficacy of follow-up questions based on a free narrative. Given that the new information was not as accurate as the spontaneously reported information and the potential implications for applied contexts, further investigation was required. It could be suggested that accurate reporting was influenced by a shift in format or by introducing a social interaction element. To further investigate interviewees' performance to follow-up questions, a second experiment was conducted to examine whether the follow-up question phase could be refined through instructions that take into account the use of metacognitive processes in reporting. Since there are established benefits for the use of open questions (cf. closed questions), it becomes necessary to examine whether and in what way, reporting of more information can be balanced with high accuracy.

Experiment 2

Introduction

The second experiment focused on improving accurate reporting by encouraging interviewees to use meta-cognitive monitoring strategies when asked follow-up questions. Research on decision-

making mechanisms that are involved in the reporting of information from memory shows that interviewees try to achieve a balance between being informative and accurate (Goldsmith et al., 2002; Koriat & Goldsmith, 1996). To this end, interviewees control how much information they report by volunteering or withholding an answer based on how confident they are about the accuracy of their recollection (Ackerman & Goldsmith, 2008; Koriat & Goldsmith, 1996).

Interviewees can also regulate their answers by adjusting the level of precision or the coarseness of the information they report (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, (Goldsmith et al., 2002), interviewees start by retrieving a fine-grain answer, which they will volunteer if it is rated as likely to be correct, otherwise a coarse-grained answer is provided instead to preserve accuracy. Further to the satisficing model of the minimum-confidence criterion (Goldsmith et al., 2008), the dual-criterion model suggests that the criterion of informativeness also mediates interviewees' reporting (Ackerman & Goldsmith, 2008). Depending on the situational demands, even if coarse-grain responses are more likely to be correct they may not be informative enough compared to fine-grain responses to be reported (Ackerman & Goldsmith, 2008; Yaniv & Foster, 1995). In other words, in the context of a criminal or intelligence investigation, in an effort to provide an accurate memory report an interviewee could report coarse details that are more likely to contain the correct answer (i.e., maximise accuracy). However, if the reported details are very broad, then they might not be reasonably informative for the investigation to progress (Grice, 1975), so they might offer a more specific answer. Therefore, both a confidence and an informativeness criterion are used to decide

on volunteering sufficiently accurate and precise information. Based on the findings of the first experiment, it may be that interviewees initially reported information that they assessed as probably correct but in response to follow-up question they were more willing to risk accuracy to also satisfy a demand for informativeness.

To support the use of follow-up questions, the current experiment examined whether instructions that encourage the exercise of meta-cognitive monitoring can improve accurate reporting to additional prompting. To this effect, half of the participants were instructed that they could withhold from providing an answer and that they could regulate the precision of their answers by providing coarse information (e.g. he wore dark clothes) or fine information (e.g. he wore a grey jumper and black jeans) (Goldsmith et al., 2002; Koriat & Goldsmith, 1996). Previous research that applied 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 even after a delay (Goldsmith et al., 2005) and that they can successfully balance informativeness and accuracy when answering cued-recall questions by regulating precision and controlling reporting output (Weber & Brewer, 2008). Other research examining how interviewees regulate the output and precision of their reporting in various contexts (i.e., reporting in private vs with an audience; receiving a penalty or not for reduced accuracy) showed that often interviewees would rather provide informative (i.e. precise) details, but this tendency is reduced in the presence of an evaluative audience or when they received penalties for inaccurate responses, in which case they report more broad details, which are more likely to be accurate (McCallum et al., 2016). More recently, Brewer, Vagadia, Hope and Gabbert (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 they can be provided in response to cued-recall questions even if they were not spontaneously volunteered in an initial free narrative, thus increasing accuracy. Therefore, based on previous research, interviewees should be able to maintain accuracy in reporting by following the instructions that promote controlling of their memory output and the type of details they report.

Participants were also reminded that they should not guess about any details. Based on results from the meta-analysis on the Cognitive Interview, which showed that along with increased accurate reporting, erroneous reporting increased as well, researchers emphasized that interviewees should be instructed to not guess and to reply "I don't know" or "I don't remember" throughout the interviewing process, so that they effectively monitor reporting (Memon et al., 2010). Similar warnings to not guess are also included in other evidence-based interviewing tools, such as the Self-Administered Interview to encourage interviewees to volunteer information they are certain about (SAI; Gabbert, Hope, & Fisher, 2009). There is evidence that interviewees who used the SAI to report events on two occasions were even more accurate the second time than interviewees who had not used the SAI before, suggesting that among other instructions (e.g., "make sure you provide a complete and accurate account"), warnings to avoid guessing contributed to the interviewees controlling their reporting more carefully over time (Gawrylowicz, Memon, & Scoboria, 2013). Research by Koriat and Goldsmith (1996) also shows that participants are more likely to maintain accurate reporting when they are 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 different types of instructions to control monitoring of memory output can lead to increased accuracy in reporting.

To investigate 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. However, participants witnessed a different stimulus event that initially depicted a meeting of a terrorist group, followed by the perpetrators placing explosives in a target location. A different stimulus was used in order to increase the generalizability and the relevance of our findings for security contexts. 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. In keeping with the procedure of the previous experiment, the same follow-up open-ended questions were used, with the addition of specific instructions to preserve and increase accuracy.

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 accurately reported details (cf. inaccurate) (Fisher, 1995; Roberts & Higham, 2002), while they are more likely to volunteer highly confident responses (Weber & Brewer, 2008) and to withhold lower confidence responses that are more likely to be incorrect (cf. volunteered responses) (Evans & Fisher, 2011). Research on eyewitness memory shows that confidence ratings can be successfully used to discriminate accurate from inaccurate answers to cued recall questions (*resolution* or *relative monitoring accuracy;* e.g. Brewer et al., 2018; Weber & Brewer, 2008). Further research on answering cued recall and general-knowledge questions shows that there is evidence of calibration between confidence and accuracy in that an increase in accuracy is related to an increase in confidence

judgments (*calibration* or *absolute monitoring accuracy;* Luna, & Martin-Luengo, 2012; Luna, Martin-Luengo, & Brewer, 2015). To explore if retrospective confidence judgments correspond to account accuracy, at the end of the session, all participants were asked to indicate how confident they felt about their written and spoken accounts respectively. Confidence ratings were used to explore whether the ratings corresponded to the pattern of the accuracy rates for the overall information provided with the timeline format and for the information provided in response to the follow-up questions. For instance, if the accuracy rate for interviewees' responses to the follow-up questions was lower than the accuracy of the initial written account, as shown in the previous experiment, we were interested to see if interviewees' confidence ratings would follow a similar trajectory. In that case, confidence could serve as an indicator for the interviewees' correct reporting.

It was predicted that, when interviewees received instructions to monitor their answers for accuracy, 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

Participants were randomly allocated to a condition where they either received pre-questioning instructions or not after they all provided an initial account with the timeline technique. 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 enhanced accuracy after receiving instructions to monitor reporting based on previous related

findings (e.g., Goldsmith et al., 2002; Koriat & Goldsmith, 1996; Weber & Brewer, 2008;

Scoboria & Fisico, 2013). A total of 60 participants (50 Females, Age: M = 20.72, SD = 3.73,

Range = 18-33) were recruited through the student participation pool and through advertisements

circulated across campus and were granted course credit or a £5 honorarium for participating.

The dependent variables were the number of correct details, accuracy rates across reporting

phases, and confidence ratings. We also report the number of errors across phases. All

participants 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 enhanced accuracy 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

head out to carry out the plan. At the outset, three of the perpetrators are seen waiting in a room

next to each other. Another individual, who plays the role of the leader of the group, then enters

the room. The film is shot from a first-person perspective to give the impression of the viewer

being present in the room with the group. The leader delivers information to the perpetrators

about the target of the attack and assigns roles to each member; overseeing the operation and

providing the detonator, placing the explosives, acting as a look out while the operation takes

place, and being the getaway driver. There is also a discussion among members about the

explosives to be used and how they are to be detonated and when. The perpetrators arrive at the

selected target, a park, and are seen entering and 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 with a camera 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 perpetrators are seen exiting the park. There is a brief dialogue from inside the car, about the explosives being placed successfully.

Timeline reporting format. The timeline reporting 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 cardboard).

Follow-up open-ended questions. A list of five open-ended questions was composed to probe about gaps in participants' account and about inconsistencies/need to clarify:

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…)

Procedure

Participants were asked to take part in a study which investigates factors that affect people's memory reports for witnessed events. Participants witnessed the stimulus event on a computer screen with headphones on. 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,

participants were then presented with the timeline reporting format to provide their account.

First, participants were given the self-generated cues instruction, as a technique that was used in

Kontogianni et al. (2018). On a piece of paper with six bullet points, they were instructed 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 Enhanced Accuracy Instructions. These instructions were presented in written format

after participants provided their initial account and prior to being asked any follow-up questions.

Participants were instructed to refrain from guessing, to feel free to withhold an answer, and to

consider the level of detail they felt they could accurately report (see Appendix B for verbatim

instructions). With respect to the level of detail in reporting, they were asked to provide all the

information they believed to be accurate from the event, regardless of whether it was precise or

broad. To clarify what precise and broad details could be, they were provided with examples of

fine grained (precise) and coarse grained (general) details, such as describing a car as "small and

dark coloured" (general details), or as "a Volkswagen Golf, British Racing Green, 5-door

hatchback, with tinted windows, and a registration number" (precise details). 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

general and/or specific details about what they remembered. After they had answered the

practice question and 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 interviewer selected between three to five questions from a list to ask, 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 the questioning phase, 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 was again only focused 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 of their responses to the follow-up questions. All participants were debriefed and compensated for their time.

Coding

Coding of the interviews was blind to experimental conditions and was conducted by the researcher and lab research assistants. The same coding scheme as in the previous experiment was used and each detail was coded as a person, action, object and setting detail. Details were coded as correct and awarded with one point if they were present in the stimulus event and

described accurately. Conversely, details were coded as inaccurate and awarded one point if they were not present in the event. Vague and subjective details were not coded for accuracy.

To assess inter-rater reliability, 10 interviews (i.e. 15% of all interviews) were randomly selected and coded by an independent rater who was blind to experimental conditions. Inter-rater reliability was high across coding categories, *ICC* = .98, 95% CI [.965, .984].

Results

Initial reporting phase

A series of independent t-tests were conducted to examine the number of correct details reported with the use of the Timeline Technique and after asking follow-up questions. An analysis of the initial reports showed that there was no statistical difference between conditions for the number of correct details reported, $t(58) = 1.11$, $p = .271$, $d = 0.29$, *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 errors, $t(58) = .870$, $p = .388$, $d = 0.23$, *CI* [-0.73, 0.28], or for the accuracy rate of details reported, $t(58) = 1.33$, $p = .187$, $d = 0.34$, *CI* [-0.17, 0.85]. Table 3.2 shows Means and SDs for correct details, incorrect details, and accuracy rates reported in both conditions, across reporting phases.

Table 3.2

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

		Enhanced accuracy 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

Follow-up questioning phase

We were particularly interested in the results of the follow-up questioning phase as this is the stage where participants either received enhanced accuracy instructions or did not receive instructions to answer the questions. There was no significant difference between the number of follow-up questions that were asked in the enhanced accuracy instructions condition ($M = 4.67$, $SD = 0.55$) and in the no instructions condition ($M = 4.63$, $SD = 0.56$), $t(58) = .234$, $p = .816$.

There was no difference between conditions for the number of correct details, $t(58) = .04$, $p = .970$, $d = 0.01$, CI [-0.50, 0.52], or for the number of errors reported in response to follow-up

questions, $t(58) = .468$, $p = .642$, $d = 0.12$, CI [-0.63, 0.39]. Despite the use of enhanced

instructions about accurate reporting by one of the two groups, there was no significant

difference between the two conditions for the accuracy rate of the additional information, $t(58) =$

.672, $p = .504$, $d = 0.17$, CI [-0.33, 0.68].

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, $F(1,58) = 22.17$, $p < .001$, $\omega^2 = .261$.

Total Reporting across Initial Report and Follow-up Questioning Phase

There was no difference between the two conditions for the number of correct details reported

overall, $t(58) = 1.01$, $p = .317$, $d = 0.26$, CI [-0.25, 0.77]. There was no difference between

conditions for the total number of errors reported, $t(58) = .908$, $p = .368$, $d = 0.24$, CI [-0.74,

0.27]. Nor was there a difference between conditions for the total accuracy rate across reporting

phases, $t(58) = 1.35$, $p = .184$, $d = 0.28$, 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,

$t(57) = 1.42$, $p = .160$, $d = 0.37$, CI [-0.15, 0.88], or in response to follow-up questions, $t(57) =$

.42, $p = .674$, $d = 0.11$, CI [-0.40, 0.62]. A pairwise t-test analysis 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, $t(58) = .142$, $p = .888$, $d = 0.02$, CI [-0.24,

0.27]. Table 3.3 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 3.4. 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.

Table 3.3

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

	Enhanced Accuracy 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 3.4

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

Confidence	Mean Accuracy Initial report	SD	n	Mean Accuracy Follow-up	SD	n
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

Discussion

Contrary to our hypothesis, providing participants with instructions designed to enhance accuracy did not significantly increase the accuracy of the information provided in response to follow-up questions, relative to participants who received no additional instructions. Overall, the use of instructions that aimed to facilitate accurate reporting led neither to an increase in the number of correct details, nor to higher accuracy rates, compared to the condition where participants did not receive any instructions to preserve accuracy.

The current results follow the same pattern as in the first experiment. Participants reported more about the witnessed event when additional prompts were used. Overall, the correct details in participants' responses across conditions represent 22% (enhanced accuracy instructions) and 21% (no instructions) of the overall elicited information, respectively. In terms of accuracy, reporting was again highly accurate in the initial account but the accuracy rate of the additional information in response to the follow-up questions was not as high. However, the accuracy rate of the new information was higher than in Experiment 1, with 75% mean accuracy across conditions in this case. Therefore, participants reported accurate information but to a lesser extent when responding to follow-up questions than when initially providing their own account.

Participants' confidence in the accuracy of their reports remained stable across the initial account and the follow-up questioning phase. On average, participants' confidence was approximately 70% across conditions and did not fluctuate in the direction of the declining accuracy rates for new information. Therefore, confidence estimates in the current study did not offer any diagnostic information to assess the accuracy in reporting across phases. Previous

studies have examined the use of confidence judgments for free recall and found a lack of a strong confidence-accuracy relationship. For instance, in Gwyer and Clifford (1997), confidence ratings were taken for information elicited with the Cognitive Interview for certain recall areas, such as person descriptions and actions. Ibabe and Sporer (2004) found that confidence was lower for accurate responses to open prompts than in response to true or false and alternative-choice prompts. Therefore, the current results fit with existing research which suggests that confidence for elaborate reports may not be as useful in assessing accuracy as it is for cued-recall. However, notably the current study used only two measures of confidence (for initial and follow-up reporting) and thus we cannot make strong conclusions about the relationship between accuracy and confidence. Further research could measure confidence ratings for each response provided to an open prompt, to more closely examine how interviewees consider the accuracy of their reporting.

It should be noted that the two scales used in the current experiment were always administered together and in the same order, with the rating for the initial report presented first followed by the rating for the responses to the follow-up questions. The scales were administered in this manner to match the way that information was reported through the session and to indirectly encourage participants to compare their reports between the different modalities. However, it may be the case that the administration order resulted in an anchoring effect, with the confidence estimates for the information provided to the follow-up questions being biased towards the initial ratings for the information reported in the timeline format (Tversky & Kahneman, 1974). This is plausible since it appears that participants were under-confident in the accuracy of their initial reports but over-confident in the accuracy of their responses to the

follow-up questions. Of course, the latter is also a result of accuracy rates declining from the initial report to the follow-up questions while confidence remained stable.

The current findings suggest that the use of follow-up questions elicits additional information, however, this information might not be as accurate as an initial spontaneous report. Although more research is needed to determine why the instructions to not guess, to withhold an answer if uncertain and to regulate precision, did not improve accurate reporting, the current results suggest that interviewees might, to some extent, already monitor their responses to preserve accuracy and that the instructions did not further contribute to the metacognitive regulation they exercised. Alternatively, if the use of follow-up questions suggests an increased need for informativeness, there will still likely be a cost to accuracy to some degree, as the interviewees might use a less conservative confidence criterion in an informativeness-accuracy trade-off, despite explicit instructions to monitor accuracy. Eventually, the amount of additional details that interviewees can provide is likely to depend on the amount of information they initially reported (Memon et al., 2010; Roberts & Higham, 2002). To some extent the initial report will also include some errors. Although further questioning might lead to interviewees correcting themselves, it is also likely that further questions can lead to the reporting of more errors.

The fact that instructions to preserve accuracy did not contribute to more accurate responses puts further emphasis on the use of open-ended questions, since interviewees are more likely to strategically monitor their memory reports when they have increased control (Evans & Fisher, 2011). Across both experiments, the decrease in accuracy occurred after asking three to five open prompts, but accuracy is likely to vary depending on the number and type of questions asked. For instance, it is already established in the literature that practices such as asking

multiple-choice questions or repeatedly asking a question will increase the amount of erroneous reporting, mainly due to encouraging interviewees to guess if uncertain (Evans & Fisher, 2011; Fisher, 1995; Fisher & Geiselman, 1992; Memon et al., 2010). In sum, open-ended questions are preferable given that they elicit longer and more accurate responses than closed and direct questions and are therefore more efficient (Fisher, Milne, & Bull, 2011; for a review, see Oxburgh et al., 2010). Practitioners should be cautious about the reliability of new information provided in response to follow-up questions, and the number of questions they ask, since regardless of the information gathering techniques they use, there is always a limited pool of accurate details that interviewees can recall but an unlimited pool of inaccurate details to report.

General Discussion

Across two experiments, the results showed that follow-up open-ended questions can be effective for gaining new details. However, the results also indicated that the accuracy rate for the information obtained in response to follow-up questions was not as high as the information reported in an initial account. We should note that there was also a difference between the results across both experiments with respect to the reported accuracy rates. Specifically, the accuracy rates observed for follow-up questions were markedly higher in Experiment 2 (cf. Experiment 1). Determining what factors drove the variance in the magnitude of the drop in accuracy for new information between the two studies is not possible based on the current results. One explanation is that the stimulus events differed between the two studies, with respect to the number of perpetrators (five vs four); the duration (1.15min vs 4.28min) of the events; and the content, as in the second experiment participants witnessed a conversation where the perpetrators plotted and carried out a terrorist attack. More importantly, participants' responses to follow-up questions were found to be less accurate than their initial reports across both witnessed events.

The current findings across experiments are surprising because the accuracy in the follow-up questioning phase was impaired regardless of the use of open-ended questions and instructions that emphasize accurate reporting. Previous research shows that the use of various retrieval attempts, such as techniques included in the Cognitive Interview, can result in improved reporting of correct details with a slight increase in incorrect details as well (cf. standard interviews; Memon et al., 2010). Nevertheless, the use of open-ended prompts is preferable to any other question types and is recommended for applied settings (Oxburgh et al., 2010). Meanwhile, research suggests that specific instructions can assist interviewees in balancing accuracy and informativeness demands when asked follow-up questions (e.g. Evans & Fisher, 2011). However, the similar rates of accurate reporting between conditions suggest that participants potentially already regulated their responses to some extent in order to preserve accuracy. Given that interviewees are more likely to monitor their answers and regulate precision when they have more control over reporting (Ackerman & Goldsmith, 2008; Brewer & Weber, 2008), it is likely that interviewees in the current study were already using metacognitive strategies in responding to open prompts, to the extent that they could not further improve their responses. Examining exactly how participants reported coarse and fine-grain details was beyond the aims of the current study as i) we were primarily interested in whether participants could use this instruction as a useful strategy to preserve and increase the accuracy of their accounts; ii) there is no evidence that coarse and fine-grain details apply to all types of details from the witnessed events (e.g. information about actions) therefore not all errors could be avoided with the use of this strategy alone; iii) it is likely that since there are limitations in coding fine and coarse-grain responses for all types of details, participants may also face difficulties in controlling their responses with this strategy alone. More 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 instructions - potentially because of the increased demand of informativeness that is conveyed by the use of follow-up questions to begin with. Moreover, recent research suggests that interviewees often show a tendency of reporting informative (i.e. precise) details, although this bias can be reduced in the presence of an evaluative audience and via the use of incentives, such as penalties for inaccurate responses (McCallum et al., 2016). Further research could investigate to what extent the interviewees' perceptions of what is required with respect to accuracy, informativeness and precision, interact with their metacognitive decision-making processes over reporting.

The results of both experiments highlight the need to better understand how interviewees' reporting might differ when asked follow-up questions about additional information, compared to when they spontaneously report information. Given that the current findings indicate a trade-off in favour of informativeness than accuracy, despite the use of instructions to enhance accurate reporting, more research is needed on the generation of more errors when additional prompts are used – even if they are open-ended – and the role of monitoring processes when demands for informativeness increase. Although research suggests that there can be accuracy trade-offs at the cost of increased recall when open invitations and varied retrieval attempts are encouraged (Fisher & Geiselman, 2010), the potential costs and benefits of increased recall when follow-up open questions are used have not been systematically examined. For instance, future research could examine the effect that follow-up specific probes which start with "what?", "when?", "where?", "who?", "why?" and "how" (5WH; Milne & Bull, 1999) have on the quality and quantity of additional reports. Another useful direction would be to increase our understanding of how the demands for informativeness and accuracy are communicated to and perceived by

interviewees. Overall, even with the use of evidence-based practices, elicited information will always include both accurate and to some extent, inaccurate details. Therefore, it is crucial that future research helps us increase our understanding of the limitations and boundaries of memory retrieval to further inform practices in applied settings, such as a need for interviewers to seek corroboration for information reported in response to follow-up questions (cf. spontaneously).

Chapter 4: Facilitating recall of repeated events with a multi-method interviewing format

Abstract

Memory reports of repeated experiences tend to include more general information than specific details about the separate incidents. However, interviewers in both forensic and intelligence gathering settings tend to rely more on specific rather than generic reported information. To facilitate recall and particularization of repeated events by adults, we tested the self-generated cues mnemonic, the timeline technique, and follow-up open-ended questions combined in a Multi-Method Interviewing Format (MMIF). Over the course of a week in four separate sessions, 150 participants (121 Females, $Mean_{age} = 21.26$, $SD = 5.21$) watched four scripted videos depicting meetings of a terrorist group who planned and carried out an attack involving an explosive device. Three videos were highly similar while a fourth video was similar (typical content condition) or differed with respect to two critical details to introduce a deviation to the script (changed content condition). A week later, participants returned to provide their account using the MMIF, the timeline technique alone or a free recall format. Consistent with previous research on memory for deviations from the script, more correct details and fewer source confusion errors were expected in the changed content condition compared to the typical content condition. It was predicted that participants in the MMIF condition would provide more information than participants in the Timeline and Free recall conditions. It was also expected that participants in the MMIF condition would report fewer source confusion errors than participants in the other two conditions. The results partly confirmed our hypotheses as more correct details were elicited with the use of the MMIF compared to the timeline technique and the free recall format. However, there was no effect of format or of the presence of deviations from the script on source monitoring. Also, the presence of deviations did not result in increased recall for the

events overall or for the targeted event in the changed content condition. The findings are discussed in relation to eliciting information in applied settings and future research on adults' memory for repeated events.

Introduction

Witnesses, victims or sources may be questioned about events that have occurred repeatedly over a period of time (e.g. incidents of domestic violence, sexual abuse or even meetings of a criminal network). In the context of such investigations, interviewees will likely need to retrieve memories of a single specific incident across these repeated events (e.g. details of a specific assault or a specific meeting). There is a growing body of developmental research which shows that memory for repeated experiences differs from memory for unique experiences, namely with respect to the type of details reported and the focus of recall prompt (e.g. Connolly & Lindsay, 2001; Price & Connolly, 2013). When experiencing a series of similar repeated events, some details recur in the same way across events (e.g., every meeting starts with the leader of the group describing an attack plan), and other details change in each incident (e.g., a different target location is selected to attack every time). Research shows that because of repeated exposure, memory for the recurring details across events is stronger than memory for what occurred during a unique experience; but even a unique experience is better remembered than the details that change from one incident to another in a series of repeated events (Connolly & Lindsay, 2001).

Also, when eliciting details about repeated events, the focus of an investigation may be on what happened during a specific incident or on what usually happened. Memory for the general routine of the events is typically stronger than memory for a specific incident in the series, relative to memory of a unique experience (Connolly & Lindsay, 2001; Price & Connolly, 2013). Given that in court cases of child abuse, evidence is required for specific incidents rather

than for the general routine of events, numerous studies have been conducted to examine effective ways of interviewing children about repeated events and to provide evidence-based recommendations to practitioners (for a review see Brubacher, Powell, & Roberts, 2014; Schneider, Price, Roberts, & Hedrick, 2011). However, there is limited research on adults' memory of repeated events and more particularly on techniques that can be used to elicit information for specific incidents within the series (i.e., particularization) (Cohen & Java, 1995; Leins et al., 2014; MacLean et al., 2018; Means & Loftus, 1991; Theunissen et al., 2017; Willen et al., 2015). Therefore, the present study aims to test the effectiveness of a multi-method interviewing format (MMIF) combining the timeline technique (Hope et al., 2013b) with the self-generated cues mnemonic and follow-up open-ended questions, to facilitate recall for complex repeated events and to improve particularization of specific incidents. To thoroughly test the effectiveness of the MMIF, we used a comparison group where participants only used the timeline in conjunction with the self-generated cues and a baseline group where participants used a free recall format followed by open-ended questions.

Repeated events and schemas

Repeated events are thought to be represented in memory as parts of an overarching schema that is characterized by a common theme. Schemas are higher-order knowledge structures that include general representations of complex concepts - such as sequences of multiple activities – and are thus considered a part of semantic memory that also interact with new episodic information (Ahn et al., 1992; Brewer & Nakamura, 1984). A script is conceptualized as a simple type of schema, including knowledge of what typically occurs in an event (e.g., the typical actions involved when ordering food in a restaurant). Schema theory suggests that one experience can be sufficient to begin building a script. However, further exposure to similar

occurrences informs a more elaborate schematic representation, which shapes our expectations that future occurrences will follow the same pattern (Abelson, 1981; Ahn et al., 1992). Script acquisition occurs as early as after a second similar experience and becomes stronger following more experiences (Farrar & Boyer-Pennington, 1999; Hudson et al., 1992). Over the course of multiple repeated events, some *variations* of the typical actions are likely to occur. Variations represent potential alternatives in the script, such as ordering a different meal in a restaurant at every visit (Abelson, 1981). After a script is established, any new element or *variation* encountered will be integrated to the script resulting in a reconstruction which preserves the schema-consistent information, especially after long intervals (Brewer & Nakamura, 1984). As variations are absorbed by the script they are more likely to be forgotten, whereas general schema-consistent details are more likely to be retrieved over time (Abelson, 1981).

Most studies examining memory for repeated events have participants experience a series of sessions or *instances* (between three to five across studies with varying intervals between instances) which are designed to promote script acquisition. In most studies, participants engage in interactive events such as classroom activities (e.g., Brubacher, Roberts, & Powell, 2012; Roberts et al., 2015) or food tasting sessions (e.g., MacLean et al., 2018; Weinsheimer, Coburn, Chong, MacLean, & Connolly, 2017), where several activities, which always take place in the same order, are performed. Some target activities are manipulated to change from one instance to another, while some remain stable across instances. For instance, participants always listen to music, but a different instrument is used in every instance (e.g., Connolly & Gordon, 2014). Details that change in each instance are termed *variable* details – or variations according to schema theory – whereas the details that remain stable across events are termed *fixed* details – or schema-consistent details in the context of schema theory.

The structure of the four repeated events in the current study follows the same approach that is used in the literature to promote script acquisition, as suggested by schema theory. The four events which depict meetings of a terrorist network planning and carrying out different attacks are overall similar but they also include some variations (e.g., McNichol et al., 1999). During script acquisition, it is argued that events and their sub-components are organized in memory in a hierarchy. For instance, a recurring activity across events consists of separate actions which may differ from one instance to the next (i.e., variations or variable details), and which may be performed by different people, using different objects etc. (Hudson et al., 1992). In this hierarchical representation of the events, the recurring activities in the script are represented on a higher level, while the alternative actions which differ across instances are organized in *slots* that are part of the lower levels of the hierarchy (Hudson et al., 1992; Schank & Abelson, 1977). To encourage script acquisition, the current study aimed to simulate this representation of repeated events, where each event consists of separate activities and actions, each organized on a different level of a hierarchical structure. For an outline of the structure of the recurring activities and their variations in the current study's events, see Figure 4.1.

Repeated event — Planning and carrying out an attack

Typical or fixed activities — Select location | Supply explosives | Plant the explosives

Typical or fixed actions & variations — Spinnaker tower | M48 Mortar rounds | Arrive at location — Spinnaker tower / University building

University building | Composition C4 | Conceal explosives in bag

Exit location
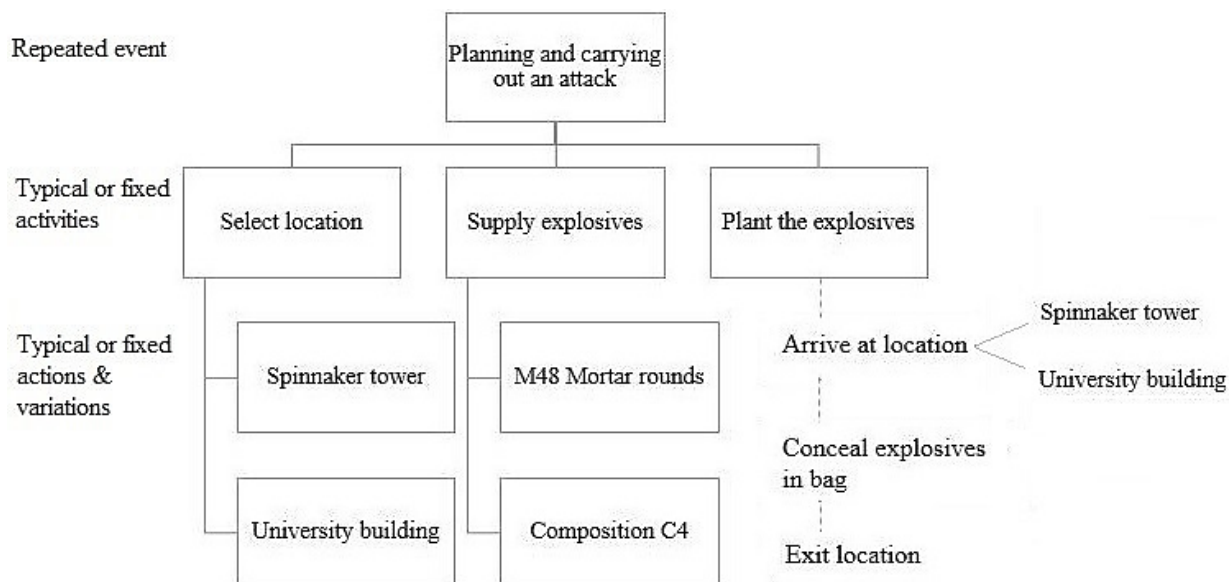
Figure 4.1. Outline of the various event components (i.e., activities, actions and variations) in a series of repeated events represented on (sub)levels within a hierarchical structure. In the top level the repeated event is represented; the main activities of the repeated event are represented directly below; and the actions and variations for each activity are part of the lower levels.

A comparable conceptualization to that of schema theory, and of memories being formed by schema-consistent and variation details, is suggested by Fuzzy-Trace Theory (FTT; Brainerd & Reyna, 1990). According to FTT, experiences are encoded and stored in memory in the form of gist, which represents semantic and relational information, (i.e., patterns or schema-consistent details) and of verbatim traces which represent specific details of an experience including its source (i.e., variations specific to individual incidents). Although both traces relate to the same experience, gist traces refer to a subject's overall understanding of what occurred and are stored and retrieved separately from verbatim traces which refer more to details of the experience itself. Research shows that we tend to extract and process the general pattern of an event before specific details are even stored in memory (see Brainerd & Reyna, 2004, for a review). This variability in how information is processed and stored in memory further dictates that access and retrieval of gist and verbatim traces occurs through the use of different retrieval cues; with cues that prompt for memory of the general pattern of what occurred triggering the retrieval of gist

information, contrary to cues that prompt for specific details of an event triggering the retrieval of verbatim information (*principle of retrieval dissociation*; Brainerd & Reyna, 2004). Similarly to schema theory, FTT suggests a differential loss over time for gist and verbatim details, with the latter appearing to be more sensitive to forgetting than the former. Thus, retrieval of verbatim details depends on the use of cues but also on time, which suggests that even with the use of appropriate cues to retrieve specific details of an event, memory over time inevitably relies more on gist while access to verbatim traces decreases (Brainerd & Reyna, 2002; 2004). In sum, according to both FTT and schema theory, specific details of separate incidents are less likely to be accessed over time compared to the general pattern of events, either because specific details fade out from memory more rapidly than the general routine or because they become absorbed by the memory of the routine itself.

Research on memory of repeated events shows that schema-consistent information is better remembered than the variation details which represent potential alternatives in the script (Price & Connolly, 2013). However, other research shows that recall is stronger for *deviations* from the script*, which are atypical and unpredictable details that as variations, they also change from one instance to another (e.g., a mistake in one's order at the restaurant; e.g., Connolly et al., 2016). Although variations are likely to be absorbed by the script, deviations may be stored in memory separate to the script. According to Bartlett (1932), specific deviations for each instance within a schema are represented separately from the schema itself but at recall, they are both retrieved together. There is evidence that deviations can serve as "tags" for each specific instance (*script pointer plus tag;* Graesser et al., 1979; Abelson, 1981). To the extent that deviations are schema-inconsistent, research suggests that they are highly likely to be recalled because they attract attention and require increased resources to be integrated to the script, therefore

comprising a strong memory trace (Brewer & Nakamura, 1984). According to the associative network model of memory (Anderson, 1983), which proposes that memory traces of events are part of a connected network of nodes that vary in strength, deviation details would receive increased attention at encoding thus forming strong memory traces with strong links in the network. In other words, deviations are more memorable because they violate the script and are therefore distinctive and salient in the memory network (Davidson, 2006; see also Cohen & Java, 1995; Means & Loftus, 1991). Recent studies have manipulated the presence of deviations in a target instance across the series of events, usually by interrupting a session within the series because of some complication (e.g., the researcher administering the experiment was needed elsewhere). These studies show that the presence of deviations in a specific instance leads to increased recall for that instance (targeted effect; MacLean et al., 2018) or even for all the instances in the series (general effect; Connolly et al., 2016; MacLean et al., 2018). Thus, deviations might affect the encoding of events and improve recall both for all events and for the specific instance when they occurred, with potential implications for information elicitation.

In addition to the effectiveness of memory-enhancing techniques to facilitate recall of repeated events, the current study examines how deviations affect recall and particularization of specific instances. To this end, participants witnessed either a similar series of events (typical content) or a similar series where the third instance included one novel and one changed detail, as deviations from the script (changed content). Both a general and a targeted effect on correct recall are expected in the changed content condition where one instance contains deviation details, based on research suggesting that changed details may be used as labels for specific occurrences (Brubacher et al., 2011; Graesser et al., 1979). Conversely, lower levels of correct

recall are expected in the typical content condition, where all instances are similar (Slackman & Nelson, 1984; McNichol et al., 1999).

Witnesses often need to describe instances of repeated experiences with high levels of precision, such as by reporting dates of events, or by identifying perpetrators of specific actions. However, although repeated exposure can strengthen memory for what *usually* occurs in a series of repeated events, it also undermines the particularization of individual instances (Brubacher et al., 2014). Because memory for the general routine of the repeated events is stronger than for individual instances, there is an increased likelihood of interference between events (Farrar & Boyer-Pennington, 1999). In other words, interviewees might not be able to attribute a detail to the specific instance where it occurred. Confusing details between different instances however, increases the likelihood of suggestibility in reporting (Lindsay et al., 2004), and it can have negative consequences for an interviewee's credibility (Brubacher et al., 2014). The process of relating a memory to its source is examined within the Source-Monitoring Framework, which posits that the source of a memory is usually automatically identified based on perceptual (e.g. sound), and contextual (spatial and temporal) characteristics (e.g., where did we have an experience, who gave us a piece of information; Johnson et al., 1993). However, source misattributions are more likely for experiences of similar repeated events (cf. unique experiences) (Johnson et al., 1993; Lindsay & Johnson, 1989). For instance, interviewees might be able to remember what occurred but not when it occurred. Source monitoring can improve when cues are available at retrieval (Johnson, Kahan, & Raye, 1984; Lindsay, 2014; Zaragoza & Lane, 1994) and source confusion errors are reduced when subjects are directed towards making source-monitoring judgments (Lindsay et al., 2004; Oeberst & Blank, 2012). Also, based on the "script pointer plus tag" hypothesis, discrimination between instances that include atypical

details should be easier than for instances that only include schema-consistent details (Graesser et al., 1979; Graesser, Woll, Kowalski, & Smith, 1980), therefore the presence of deviations in an instance might facilitate recall for that specific instance and by extent improve particularization and reduce source monitoring errors.

Although part of the focus of the current study is to examine how deviations can benefit recall for specific instances, the main aim is to investigate how cues at retrieval can promote overall recall and particularization of repeated events. Research suggests that the use of cues plays an important role to access information about both the general routine and specific details of repeated events (e.g., Hudson et al., 1992; Lindsay, 2014). For instance, Means and Loftus (1991) interviewed participants about recurring health care events that they experienced over a prolonged period of time, and found that asking participants to think about specific elements of each medical visit (e.g., type of doctor, length of wait, weather etc.) and then constructing a personal timeline improved particularization. In a recent study, Leins et al. (2014) extended the CI and facilitated recall for multiple meetings by using various mnemonics, including a timeline based on research by Hope et al. (2013b). Developed for use in information elicitation contexts, the timeline technique which uses a physical timeline format and interactive instructions, was found to elicit more accurate details and more correct sequential information about a unique multi-actor event than a free recall format (Hope et al., 2013b). Based on the notion that events are organized temporally, a timeline should similarly facilitate remembering of discrete instances of repeated events by capitalizing on the organization of autobiographical memory for the temporal order of events (Belli, 1998; Conway & Pleydell-Pearce, 2000). Notably, visual timelines have been used before in survey methodology to elicit information about prolonged periods of time (Belli, 1998; Van der Vaart & Glasner, 2007; Yoshihama & Bybee, 2011).

Brubacher et al. (2014) suggest that interviewers in investigations with children should first inquire about the frequency of the witnessed events and then focus on each instance to elicit more episodic information (see also, Connolly & Gordon, 2014). Therefore, the timeline technique is adapted in the current study so that interviewees are first asked to outline the repeated events they witnessed and then to describe each on a separate timeline to prompt particularization. A similarly modified timeline was recently used in a study by Hope et al. (2018), where participants witnessed three different events of conversations between multiple actors.

Based on promising findings in previous research (Kontogianni et al., 2018), the timeline technique is used in conjunction with the self-generated cues mnemonic. Self-generated cues prompt interviewees to list the first six things that come to mind from the event and to consider each one to help them remember more details. The rationale behind the use of the mnemonic is that by generating their own cues, interviewees capitalize on their own subjective experience of the event and so they are likely to retrieve salient details which will prompt recall for other related memories (Anderson & Pitchert, 1978; Nairne, 2002). Use of the mnemonic should facilitate the identification of specific details from each event that can be used as event "labels" for specific instances (McNichol et al., 1999; Nairne, 2002), thereby facilitating discrimination between instances. Similarly, Willen et al. (2015) used context-specific cues derived from the most salient details that participants remembered from a series of dental visits they underwent. When other participants were interviewed about their visits, they recalled more details about specific instances when they used the context-specific cues from a similar perspective to theirs, than when they used more general cues such as "times and dates". It is expected that inherent to the timeline format, temporal mnemonics together with the self-generated cue mnemonics may

also facilitate source monitoring across events (i.e., discrimination between instances).

Therefore, more correct details and fewer internal intrusions should be reported with the use of

the timeline and MMIF than with the use of a free recall format.

Interviewees are likely to omit information when reporting single and repeated

experiences for various reasons: because of forgetting, because more support is needed to access

the information in memory, or because they are unaware of the investigative relevance of

specific details (Fisher & Geiselman, 1992). More relevant to the experience of repeated events,

interviewees might remember that something occurred, but not when (Brubacher et al., 2014). In

addition, due to interference between specific instances in the series of repeated events,

interviewees might make source confusion errors that appear as inconsistent details in reporting

(e.g., confusing a perpetrator's actions in one instance based on what was witnessed in another

instance). Follow-up open-ended questions aimed at elaborating on omitted information and

clarifying inconsistencies for each event, should provide a further attempt to retrieve information

from memory for that specific event. Previous research has shown that use of open questions that

prompt more details for specific instances (in depth), are more likely to elicit episodic rather than

generic information (i.e., information about the general routine; Brubacher et al., 2012).

Therefore, the use of follow-up questions in the MMIF condition should lead to more correct

details and fewer source monitoring errors, compared to the timeline condition.

So far, there are distinct challenges in remembering and reporting repeated events which

have not been fully addressed by the current information elicitation protocols and techniques. In

particular, there is a lack of research on memory-enhancing techniques and tools that can be used

with adults to promote recall of repeated events overall but also of specific instances. Therefore,

the current study tests the effectiveness of different cues, mnemonics and retrieval instructions

combined in what is referred to here as a multi-method interviewing format to facilitate recall of repeated events overall and promote particularization. The motivation behind the use of a multi-method format is to extend the timeline technique and test its effectiveness in combination with other cues and prompts, to contribute to the development of an adaptive toolbox that can be used in applied settings, as suggested by recent research (e.g., Leins et al., 2014; Willen et al., 2015).

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 (Interviewing format: Timeline vs MMIF vs Free recall) x 2 (Content manipulation: Typical vs Changed) between-subjects design. Participants were recruited through the student participation pool and through 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 were excluded from analyses as their data were incomplete. Dependent variables were the number of correct details, correct gist and verbatim details, accuracy rates for all reported details, attributions of action and statements to people and intrusion errors.

Materials

Stimulus events. Five stimulus events were developed. Each event was a short film, between 4 and 5 minutes in duration (see Table 4.1 for details per event), depicting a meeting between four perpetrators (three males, one female) who plot a terrorist attack and then proceed to carry out the plan.

Table 4.1

Duration of all four events presented to participants across content conditions.

	Events categorized according to location			
Typical content	Spinnaker tower	Portland building	Portsmouth & Southsea station	Victoria park
	4.41 min	4.38 min	4.36 min	4.28 min
Changed content	Spinnaker tower	Portland building	Victoria park	Portsmouth & Southsea station
	4.41 min	4.38 min	5.23 min	4.36 min

Each event was shot from a first-person perspective to give the impression of the viewer being present in the room with the group. In each video, the leader delivers information to the perpetrators about the target of the attack and assigns the following roles to each member: overseeing the operation and providing the detonator, placing the explosives, acting as a look out while the operation takes place, and being the getaway driver. There is also a discussion among members about the explosives to be used and how they are to be detonated and when. The perpetrators are seen arriving at the selected location and act according to their assigned roles: one person plants the explosives, one looks out and one waits to hand the detonator (a mobile phone) to the person who planted the explosives. Afterwards they all leave the area in a getaway car. There is a brief dialogue between the group members inside the car, confirming that the explosives have been placed successfully. Four of the stimuli films were highly similar. For example, the meeting place to plan the attack, the people forming the terrorist group, the key topics covered in the plot of each attack, the sequence of what was discussed and how actions were performed within each event, were always fixed across events. Other details varied between

instances, such as the location of each attack (see Table 4.2 for a list of all the variable details across events).

Table 4.2

Overview of variable details and alternative options across events.

	Events			
Type of detail	**1**	**2**	**3**	**4**
Target	The head of the National reform party	Swedish left-wing sociologist	Local activist	Investigative journalist
Attack Location	Spinnaker Tower	Portland building	Victoria park	Portsmouth & Southsea train station
	Base of the tower	Road in front of the building	North-west end of park	(Train arriving) from London
	Back of the café	Space under the stairs	Café lodge	Café inside the station
	Construction equipment	-	In bushes	Next to the sofa
	Seafront hotel	-	Premier hotel	Ibis hotel
	Two blocks away	The Hard	Closest bus stop	Square
	Within 900m	Within 900m	Within 900m	Within 800m
	Hotel lobby	-	Hotel lobby	Parking
Object/ equipment	M48 Mortar rounds	3 kilos	Six packages	9 kilos
	-	C-4	Semtex	106 rounds
	Improvised Explosive Device	-	Improvised Explosive Device	Improvised Explosive Device
	Nokia 3210	Nokia 6210	Nokia 5210	Nokia 3310
	Motorola V3	Motorola V3	Motorola V3	Motorola V3
	Backpack	Backpack	Briefcase	Sports rucksack

Time	Wednesday	Thursday	Saturday	Monday
	10 am	6 pm	12 o'clock	9 am
	10.30 am	6.10 am	Half an hour before	Two hours before
	6.30 am	-	Half an hour upon start (of the event)	10 minutes
	-	-	9 am	6 am
	Until the party starts	Until the reception starts	After the blast	After the blast
Operational roles				
Person planting bomb	Niko (male)	Nina (female)	Niko (male)	Nina (female)
Lookout	Taking pictures	Not taking pictures	Taking pictures	Not taking pictures
Passing the detonator	Nina hands detonator to Niko over discussion[1]	Niko hands detonator to Nina[2]	Nina hands detonator to Niko over discussion	Niko hands detonator to Nina

An alternative version of one of the four events was also developed, and it included two deviations: i) one of the four perpetrators -who always has the role of the lookout when carrying out the operation in every event- was also in charge of the meeting and provided all the information to the group, while the perpetrator -who always had the role of the leader- simply attended the meeting with the other two members; ii) 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 young woman, who was not seen in the other events, briefly entered to hand her an envelope.

---

[1] Nina asks Niko for directions and uses a map to conceal the detonator which she passes over to him
[2] Niko hands the detonator to Nina without any interaction

Participants witnessed four events. Either all four events were highly similar (typical content condition) or three were highly similar and one was the event that included the deviations (changed content condition). The presentation order of the events was counterbalanced across participants to avoid order effects. The event that included the deviations in the changed content condition however, was always presented third in sequence to avoid any primacy or recency effects on recall.

Two-level Timeline Reporting Format. The two-level timeline format for reporting repeated events consists of: (i) A physical cardboard "Scoping" timeline (33 in. x 12 in.) which depicts a horizontal line running at mid-point from one end of the card to the other to provide an overview of all the experienced events; (ii) an A3 cardboard "Specified" timeline (33 in. x 12 in.) which depicts a horizontal line running at mid-point from one end to the other representing the temporal space of each event (iii) Person Description cards (5 in. x 3 in.): blank, white and lined cards; (iv) 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 cardboard); (v) Statement cards (5 in. x 3 in.): blank, pink and lined cards.

Follow-up open-ended questions. A list of five open-ended questions was composed to probe about gaps in participants' account and about inconsistencies/need to clarify:

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…)

Free recall Reporting Format. The free recall format consists of an A4 5-page long lined booklet.

Procedure

Participants were asked to take part in a study which investigated factors that influence memory accounts for a series of witnessed events. Participants visited the lab to witness four events on four separate occasions over the span of seven days. When scheduling participants, the shortest amount of time between visits was one day and the longest was four days. Participants witnessed the stimulus events on a computer screen with headphones on and every time they were instructed to imagine that they are an undercover agent who 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 witnessing the final event, participants were invited to return to the laboratory after a seven-day delay ($M = 7.29$ days, $SD = 0.53$) to provide an account of the events. Therefore, they provided their account two weeks after the first visit and one week after the last visit.

When they returned to provide their account, participants were either provided with instructions for the MMIF, the timeline format alone or the free recall format. All participants were reminded to imagine that they are an undercover agent who infiltrated a terrorist group and that they are in possession of valuable intelligence information. Participants in all conditions were instructed to begin by outlining all the events in the order they witnessed them and to then focus on each one. They were all also instructed to provide as many details as possible about the events and the people involved and to report exactly what was said when possible (verbatim

statements). Also, all participants were asked to not guess about things which they cannot remember. Participants allocated to the MMIF condition used the scoping timeline format to provide an outline of all the events in the order they witnessed them. They were asked to use a white, lined card to label each event and place all the cards on the timeline in the order they witnessed the events. They then used the self-generated cue instructions and a timeline to describe each individual event. On a piece of paper with six bullet points, they were instructed to write down the first six things that they remembered from each 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 from the event. They were also instructed to use the person description, action and statement cards to provide their account and to show "who did/said what and when".

After they finished providing their account in written form, they were asked follow-up open-ended questions. The interviewer selected between three to four questions from a list to ask per event, so that all participants across conditions were asked an equal number of questions. The topics were not pre-selected, instead the questions were asked based on what participants reported (both about the perpetrators discussing and carrying out the attack), using questions such as "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 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". This procedure allowed for interviewers to maintain the same phrasing of questions but avoid using a scripted list of cued recall questions. Participants were not forced 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. All participants were finally asked if

there was anything else they would like to report. During the questioning phase, the participant's written account remained on the table and the interviewer would point to the specific part to which the prompt referred to when asking each question. The follow-up questioning phase was audio and video-recorded, but the camera was only focused on the table and the format with the participant's account.

Participants in the Timeline condition were also asked to complete the scoping timeline for all the events and they then used the self-generated cues instructions and a timeline to describe each event, but they were not asked any follow-up questions. Finally, participants allocated to the free recall condition were provided with the free recall reporting format. After providing their account, participants were asked three to four follow-up open-ended questions per event. The same procedure for the follow-up questioning phase was followed as described in the MMIF condition. At the end, all participants were debriefed and compensated for their time.

Coding

Coding of the interviews was mostly conducted by the lead researcher and partly by two lab research assistants. The same coding scheme as in the third experiment in this thesis was used (Experiment 2, Chapter 3). Each detail was coded as a person, action, object or setting detail. Details were coded as accurate if they were present in the stimulus event and described correctly. Details that were vague or subjective (e.g., "he was young", "he looked satisfied") were not scored for accuracy. Interviews were 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 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 when three verbatim words were reported correctly and as incorrect when two or fewer words corresponded to the script. Additional coding was conducted for the accuracy of attributions of both actions and statements to a person, but only for the initially provided account and not for information reported in response to follow-up questions.

All the accounts were coded for 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 2) and the source of the stimulus events where it was witnessed in. 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 as "1-Target Event-3". Finally, all the accounts were coded for the total number of reported events and the order in which they reported witnessing the events. The same coding scheme was used to code the responses to the follow-up questions, with the exception of coding for the attributions of actions and statements.

Twenty-four interviews (i.e. 15% of all interviews) were randomly selected and independently coded by a rater, who was blind to experimental conditions for Event content (to some extent also for Format; i.e. between the MMIF and Timeline conditions that used the same format). Inter-rater reliability was high across coding categories, $ICC = .97$, 95% CI [.967, .974].

Statistical Analyses

To examine how the independent variables predicted correct reporting, the dependent variables were analysed using general linear mixed models (GLMMs) with fixed effects of reporting

format (categorical: Multi-Method Interviewing Format vs Timeline vs Free recall) and event

content (categorical: typical vs changed), and random intercepts for events nested within

participants[3] (Finch, Bolin, & Kelley, 2014). We compared two models: i) baseline with fixed

predictors only; and ii) fixed predictors with random intercepts for events nested within

participants. We found that the second model was the best 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 event content and reporting format. We used simple contrasts to code

the reporting format and the event content to examine the effects of the independent variables.

Specifically, for format, three contrasts compared reporting with Free recall (reference level) to

reporting with Timeline and MMIF, and reporting with MMIF to Timeline separately; for event

content, the contrasts compared typical vs changed content. The reported estimates for the

categorical predictors show the degree to which the dependent variable changed relative to the

reference level.

The analyses were run in R version 3.5.0 (R Core Team, 2017) using the lme function

from the nlme package (Pinheiro, Bates, DebRoy, Sarkar, & R Core Team, 2017).

Results

We present the results of the analysis on correct recall and recall of gist for total reporting[4] first

and then look more closely to the initial reports (number of details reported and accuracy rates)

---

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

[4] A model was built to examine total reporting instead of reporting in the follow-up phase, in order to compare the effects of all reporting formats across events, which was the main aim of the current study. Meanwhile, the information gain in the follow-up phase is also evident with the use of this model.

and the reporting of internal intrusions (source confusion errors). We also present secondary

analyses on the number of correct person-action attributions, the number of the witnessed events

reported across conditions and the number of probes that were asked in the follow-up

questioning phase. In the interest of parsimony, results on the reporting of verbatim details and

on accuracy rates for both gist and verbatim details compared for initial and total reporting, are

presented in Supplemental materials (see Appendix C), as no significant results emerged across

conditions. For the interested reader, results on the memory for the order of the witnessed events

are also included in Supplemental materials (see Appendix C).

Memory for the number of witnessed events

A chi-square analysis was conducted to examine whether reporting of the number of witnessed

events (i.e., four) was associated with reporting format or event content. Results showed that

numerous participants reported fewer or more than the correct number of events. However, there

was no significant association between the number of reported events and reporting format, $\chi^2(6)$

$= 5.747, p = .452$, or event content condition, $\chi^2(3) = 4.571, p = .206$. Twenty-four percent of the

total number of participants did not report witnessing the correct number of events. Table 4.3

shows the number of participants across format conditions reporting the number of events they

witnessed.

Table 4.3

Number of participants across format conditions reporting correct (i.e., four; cf. incorrect) number of witnessed events.

	Format			
Number of events reported	**MMIF**	**Timeline**	**FR**	**Total**
Two events (-2)	0	1	0	1
Three events (-1)	10	14	9	33
Four events (- 0)	39	34	41	114
Five events (+1)	1	1	0	2
Total	50	50	50	150

Total reporting (after using follow-up questions)

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

Reporting Format significantly predicted the reporting of total correct details, with more correct details being reported in the Timeline condition, $b = 5.48$, CI [-1.02, 11.98], $t(144) = 3.24$, $p = .001$, and in the MMIF condition $b = 7.48$, CI [0.98, 13.98], $t(144) = 4.42$, $p < .001$ (cf. Free Recall). However, reporting of total correct details was equally likely in the Timeline and the MMIF condition, $b = 2.00$, CI [-4.50, 8.50], $t(144) = 1.18$, $p = .240$. Again, neither Event Content, $b = .95$, CI [-6.26, 4.36], $t(144) = .69$, $p = .491$, nor the interaction between Content and Format significantly predicted the reporting of correct details: i) Free recall vs Timeline, $b = .28$, $t(144) = .08$, $p = .936$; ii) Free recall vs MMIF, $b = .46$, $t(144) = .14$, $p = .889$; iii) Timeline vs MMIF, $b = .18$, $t(144) = .05$, $p = .960$. Therefore, the use of different formats only affected the

reporting of correct details (see Figure 4.2, for results at initial and follow-up reporting).
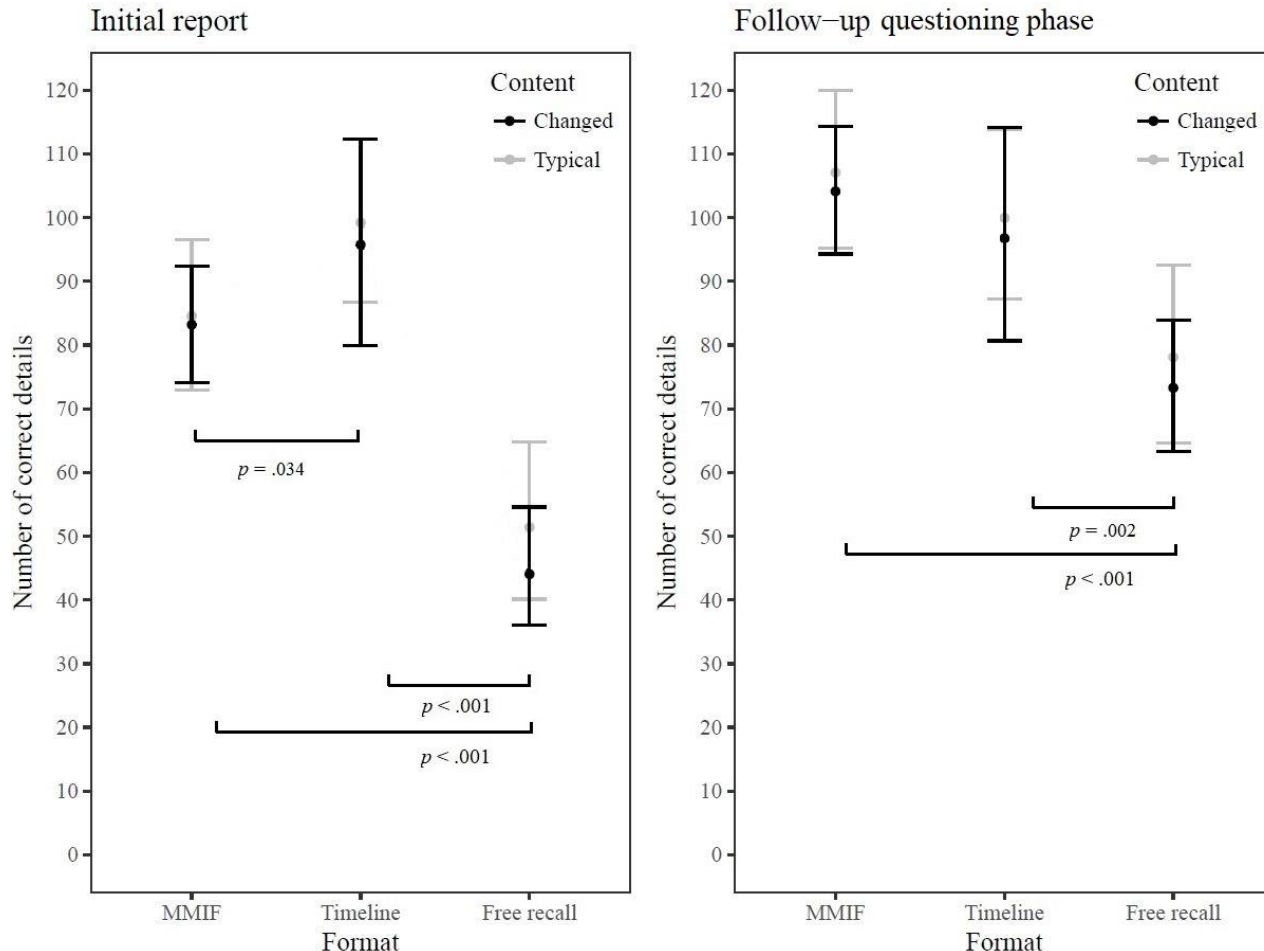

Figure 4.2. Number of correct details as a function of format within Changed and Typical content conditions. Error bars represent 95% Confidence Intervals. Note: reporting in the Timeline condition remained stable (i.e. no follow-up questions were used).

A separate model was built to examine if there is a targeted effect of changed content on the reporting of correct details for the third instance, which contained the deviation details. Content did not significantly predict the reporting of overall correct details $b = 2.53$, $t(144) = 1.38$, $p = .170$. Therefore, the change of content did not affect the reporting of details overall (general effect on recall) or for the specific event (target effect on recall). Figure 4.3 shows the reporting of correct details across events within content conditions (total reporting).
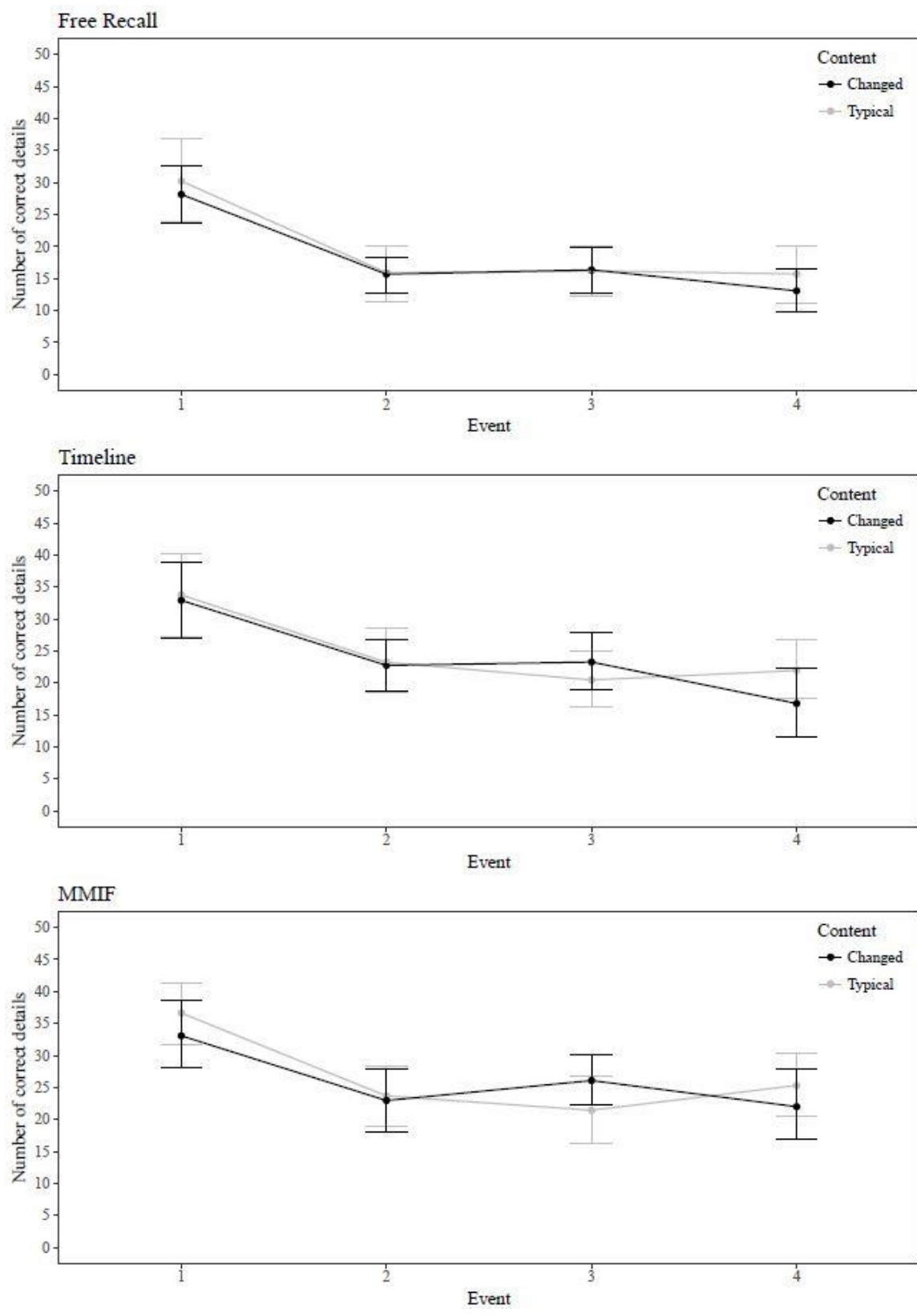
Figure 4.3. Number of correct details reported across events within Changed and Typical content conditions for each format. Error bars represent 95% Confidence Intervals.

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, $F(1, 96) = 32.72$, $p < .001$, $\omega^2 = .245$. However, accuracy in the follow-up questioning phase was not affected by the format participants used to provide their initial report, $F(1, 96) = 1.37$, $p = .245$, $\omega^2 = .003$. Means and standard deviations for accuracy rates can be found in Table 4.4.

Table 4.4

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

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

Reporting Format was a significant predictor of the reporting of the number of correct gist details, with higher likelihood of reporting in the MMIF than in the Free recall condition, $b = .78$, *CI* [0.09, 1.46], $t(144) = 2.24$, $p = .027$, and a higher likelihood of reporting in the Free recall condition than in the Timeline condition, $b = .69$, *CI* [-1.37, -0.01], $t(144) = 2.00$, $p = .047$. The reporting of correct gist details was more likely in the MMIF than in the Timeline condition, $b = 1.47$, *CI* [0.14, 2.79], $t(144) = 4.24$, $p < .001$. Event content did not predict the reporting of correct gist details ($p = .832$). The interaction between Content and Format was not significant: i) Free recall vs Timeline, $b = .32$, $t(144) = .46$, $p = .646$; ii) Free recall vs MMIF, $b = .47$, $t(144) = .68$, $p = .498$; iii) Timeline vs MMIF, $b = .79$, $t(144) = 1.14$, $p = .256$.

Figure 4.4 presents the reporting of correct gist details as a function of format within

content conditions. Figure 4.5 shows the reporting of correct gist details across events within
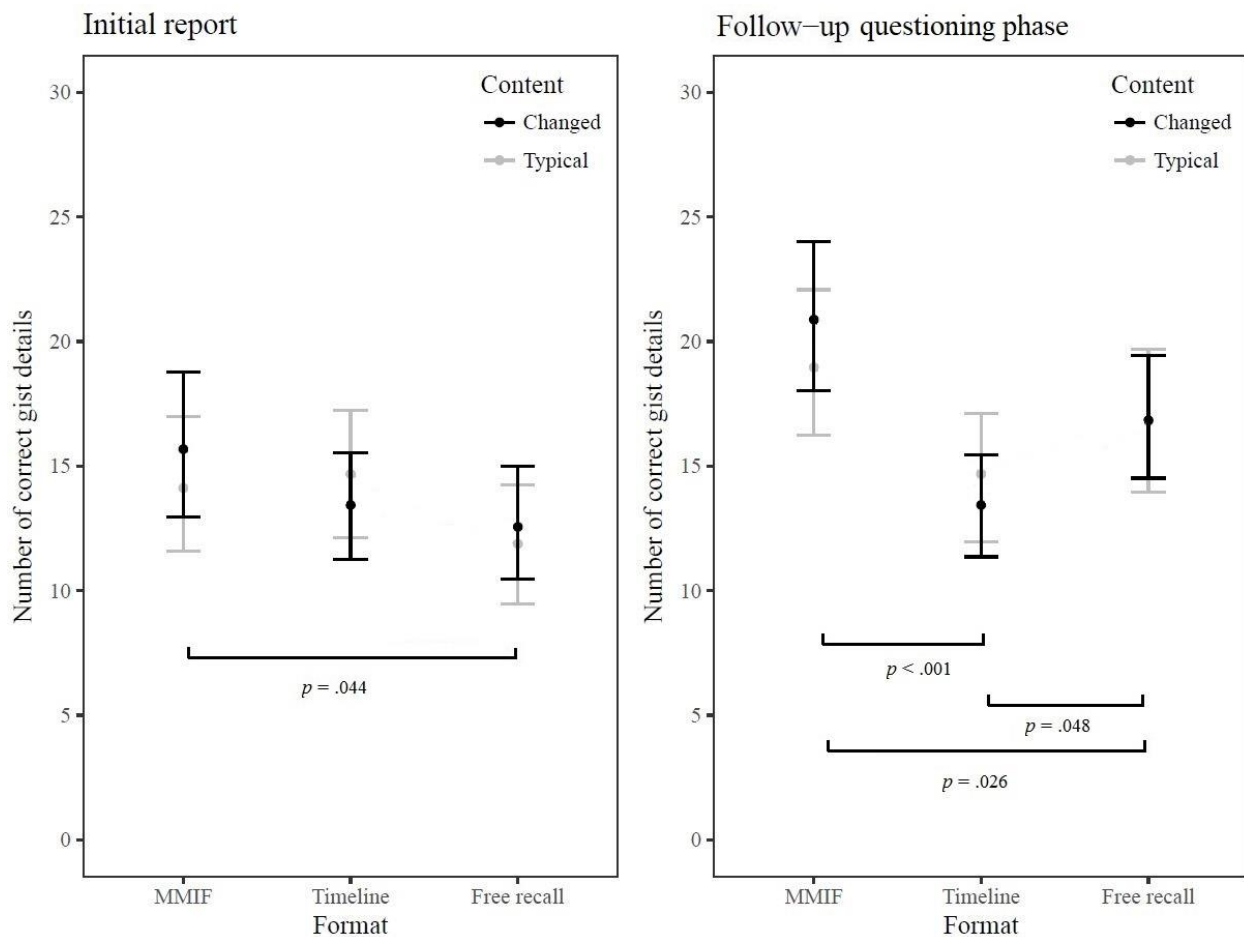
content conditions (total reporting).

Figure 4.4. Number of correct gist details as a function of format within Changed and Typical content conditions. Error bars represent 95% Confidence Intervals. Note: reporting in the Timeline condition remained stable (i.e. no follow-up questions were used).
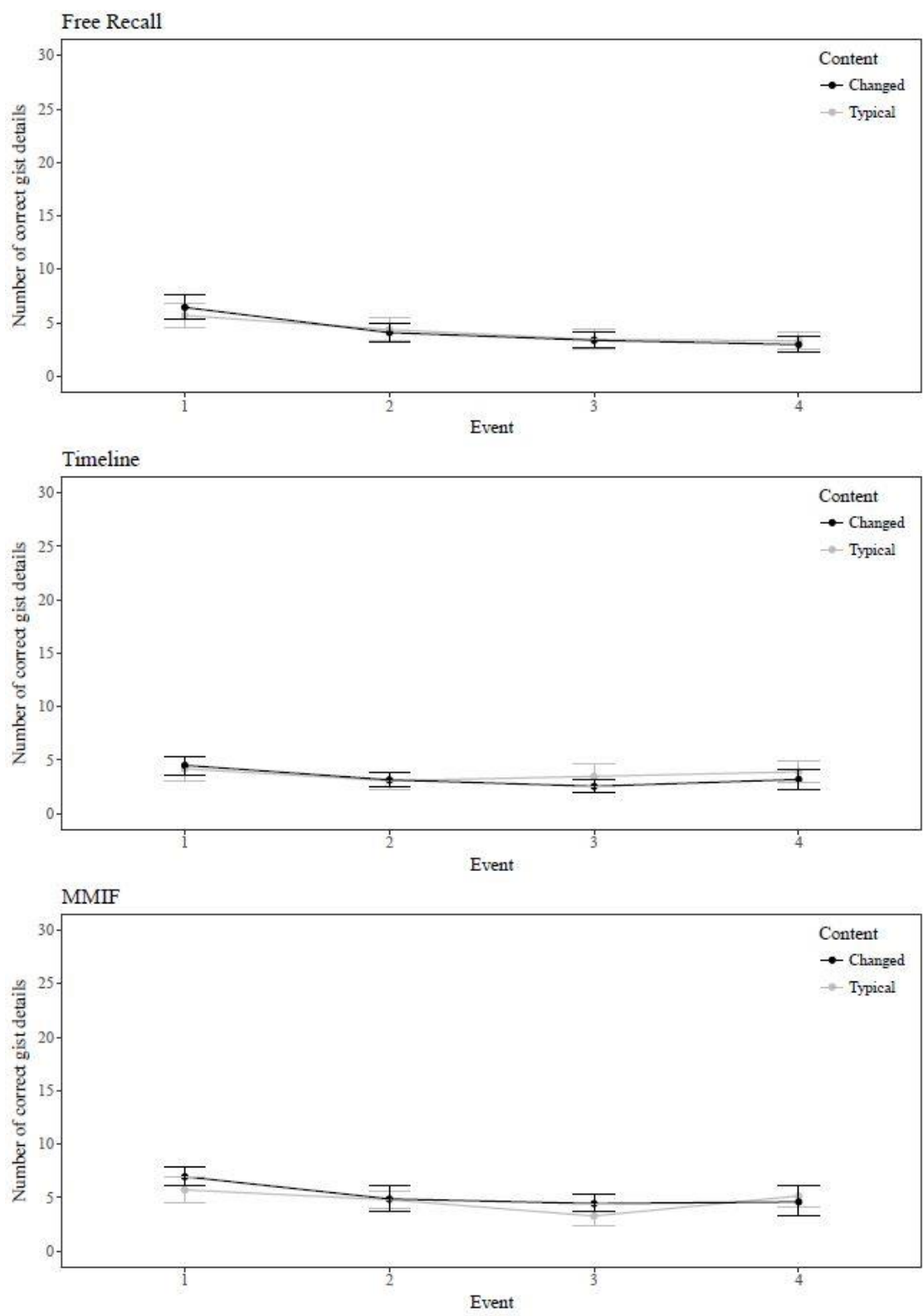
Figure 4.5. Number of correct gist details reported across events within Changed and Typical content conditions for each format. Error bars represent 95% Confidence Intervals.

Initial reporting phase

Reporting Format was a significant predictor of the number of correct details reported, with a higher probability of correct details being reported in the Timeline condition, $b = 12.47$, CI [6.32, 18.62], $t(144) = 7.80$, $p < .001$, and in the MMIF condition, $b = 9.04$, CI [2.89, 15.19], $t(144) = 5.65$, $p < .001$ (cf. Free Recall). Unexpectedly, the Timeline format was associated with the reporting of more correct details than the MMIF format, $b = 3.43$, CI [-2.72, 9.58], $t(144) = 2.14$, $p = .034$. Event content was not a significant predictor, $b = 1.03$, CI [-6.05, 3.98], $t(144) = .79$, $p = .431$ and there was no significant interaction between Content and Format: i) Free recall vs Timeline, $b = .92$, $t(144) = .29$, $p = .772$; ii) Free recall vs MMIF, $b = 1.50$, $t(144) = .47$, $p = .640$; iii) Timeline vs MMIF, $b = .58$, $t(144) = .18$, $p = .857$.

With respect to accuracy rates, we were interested in the overall accuracy reported across events rather than within each event. To this end, a univariate analysis of variance was conducted to examine overall accuracy. The analysis 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 between the MMIF and the Free recall condition ($p = .035$), but not between the Timeline and Free recall conditions ($p = .228$). There was not a significant difference between the MMIF and Timeline conditions ($p = 1$). There was no significant main effect of Event content on the accuracy rate $F(1,144) = .14$, $p = .709$, $\omega^2 = -.006$, and the interaction between Format and Content was not significant $F(2, 144) = 2.47$, $p = .088$, $\omega^2 = .020$.

Reporting Format significantly predicted the reporting of correct gist details, with the MMIF being associated with the reporting of more correct gist details than the Free recall format,

b = .67, *CI* [0.02, 1.32], *t*(144) = 2.03, *p* = .044. However, participants in the Timeline and Free

recall conditions were equally likely to report correct gist details, *b* = .46, *CI* [-0.19, 1.11], *t*(144)

= 1.39, *p* = .167, as were participants in the MMIF and Timeline conditions, *b* = .21, *CI* [-0.86,

0.44], *t*(144) = .64, *p* = .523. Event content did not significantly predict the reporting of correct

gist details (*p* = .758) and the interaction between Format and Content was not significant: i)

Free recall vs Timeline, *b* = .48, *t*(144) = .73, *p* = .467; ii) Free recall vs MMIF, *b* = .22, *t*(144) =

.33, *p* = .742; iii) Timeline vs MMIF, *b* = .70, *t*(144) = 1.06, *p* = .291.

Reporting Format significantly predicted the reporting of correct attributions to persons,

with more correct details reported with the Timeline, *b* = 2.38, *CI* [1.52, 3.24], *t*(144) = 5.49, *p* <

.001 and the MMIF, *b* = 1.58, *CI* [0.72, 2.43], *t*(144) = 3.63, *p* < .001, than with the Free recall

format. Participants in the Timeline and MMIF conditions were equally likely to report correct

attributions, *b* = 0.81, *CI* [-0.05, 1.66], *t*(144) = 1.86, *p* = .065. Event Content did not predict the

correct reporting of attributions, *b* = 0.23, *CI* [-0.93, 0.47], *t*(144) = 0.66, *p* = .510, and the

interaction between Content and Format was not significant, i) Free recall vs Timeline, *b* = .18,

t(144) = .21, *p* = .834; ii) Free recall vs MMIF, *b* = .07, *t*(144) = .08, *p* = .936; iii) Timeline vs

MMIF, *b* = .25, *t*(144) = .29, *p* = 773. Figure 4.6 shows the number of correct attributions made

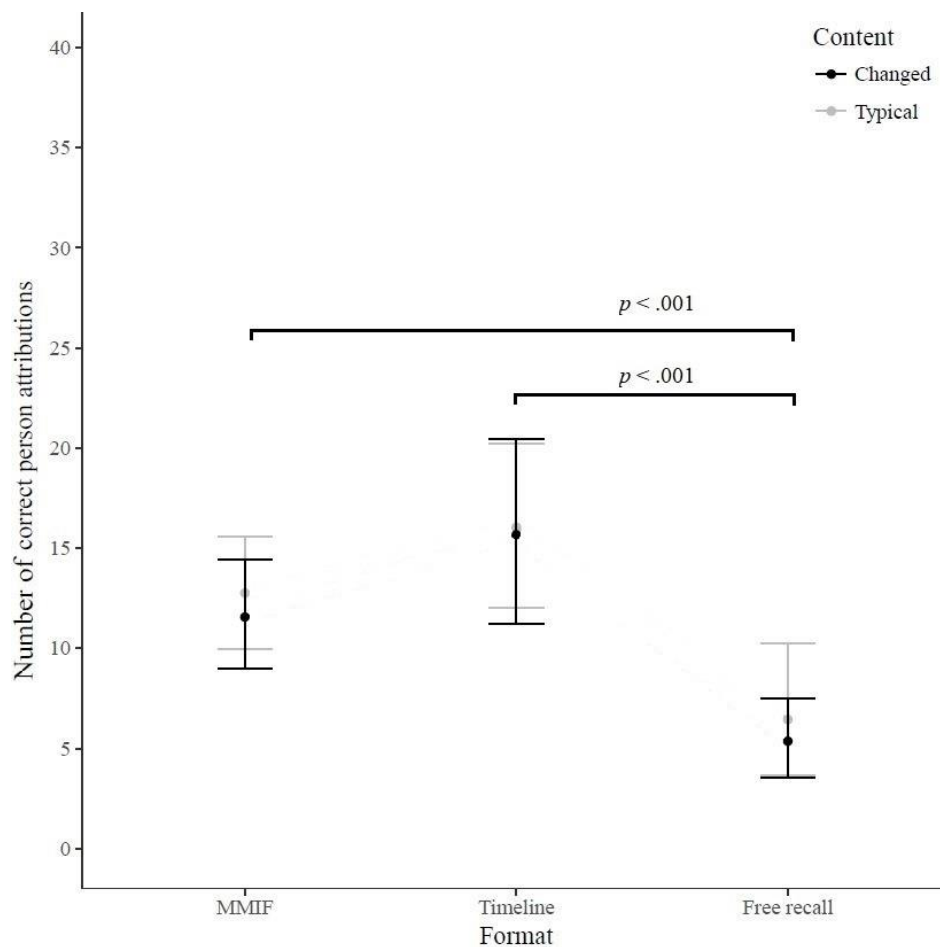as function of format within content conditions.

Figure 4.6. Number of correct attributions made as a function of format within Changed and Typical content conditions. Error bars represent 95% Confidence Intervals.

A univariate analysis of variance showed that Reporting Format had a main effect on the accuracy rate of the reported attributions of actions and statements to persons, $F(2,144) = 4.72$, $p = .010$, $\omega^2 = .048$. Event content did not have a significant effect, $F(1,144) = .18$, $p = .672$, $\omega^2 = -.006$, and the interaction between Format and Content was not significant, $F(2,144) = .11$, $p = .896$, $\omega^2 = -.012$. 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 Timeline ($M = 0.73$, $SD = 0.19$) than in the Free recall condition ($M = 0.61$, $SD = 0.29$), $p = .030$, with no difference between the MMIF and Timeline conditions ($p = 1.00$).

Schema interference: internal intrusions across instances

Table 4.5 shows the distribution of the mean number of intrusions in each event across

conditions. Table 4.6 shows the mean number of intrusions according to type of detail. Intrusions

were coded according to the type of variable details that changed in each instance, namely target,

location, time, object/equipment, and operational role details (as displayed in Table 4.2).

Table 4.5
Mean number and SDs of intrusions reported across format and content conditions.

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

Table 4.6
Means, SDs and total counts of intrusions per type of variable detail.

Type of detail	Mean intrusions	*SD*	Total intrusions
Target	.18	.45	27
Location	.25	.54	37
Object	.99	1.16	149
Time	.28	.75	42
Person	1.29	1.54	194
Gist	.16	.45	24
Verbatim	.21	.48	31
Action	.09	.37	14

Neither Event Content, $b = .01$, *CI* [-0.06, 0.05], $t(144) = .44$, $p = .661$, nor Reporting

Format predicted the reporting of intrusions. The reporting of intrusions was equally likely in the

Free recall condition to the Timeline condition, $b = .00$, *CI* [-.06, 0.06], $t(144) = .00$, $p = 1.00$,

and to the MMIF condition, $b = .02$, *CI* [-0.04, 0.08], $t(144) = 1.23$, $p = .221$, with similar results for the reporting in the Timeline to the MMIF condition, $b = .02$, *CI* [-0.04, 0.08], $t(144) = 1.23$, $p = .221$ (see Table 4.5). The interaction between Content and Format was significant for i) Free recall vs Timeline, $b = .08$, $t(144) = 2.31$, $p = .022$ but not for: ii) Free recall vs MMIF, $b = .02$, $t(144) = .54$, $p = .590$; iii) Timeline vs MMIF, $b = .06$, $t(144) = 1.77$, $p = .079$.

Discussion

The main aim of the current study was to examine the effectiveness of using the timeline combined with self-generated cues and follow-up prompts in a multi-method format, to facilitate recall and particularization of repeated events. The current findings showed that use of the MMIF improved particularization of specific instances and overall recall of the witnessed repeated events, with more correct information reported with the MMIF than with the free recall format, without a cost to accuracy. The results between the MMIF and timeline condition were comparable with respect to the reporting of details about persons, objects, actions and settings but the reporting of gist details (across events and per instance) was higher in the MMIF condition (discussed further below). In line with previous research, the results indicated that the use of follow-up questions about each instance facilitated the reporting of more episodic information (Brubacher et al., 2012). Based on the current results, we cannot reach any conclusions as to whether the use of follow-up questions specifically resolved within-statement inconsistencies in reporting, as the questions were tailored to the participants' reporting and therefore there was no control over the number of questions asked regarding inconsistencies. Most of the questions addressed gaps and omitted information in the participants' reports, therefore the current results address the elicitation of additional information due to using follow-up questions. It should be noted that the information provided in response to follow-up questions

was not as accurate as the spontaneously reported information. However, existing research shows that an increase in the reporting of both errors and correct details is likely when additional information is reported (Memon et al., 2010; Roberts & Higham, 2002). More importantly for the aims of the current study, the overall accuracy rates were similar in all conditions, therefore the MMIF elicited more correct details than a free recall format, without a cost to accuracy.

A closer analysis of the initial reporting phase shows how the use of cues and the timeline format facilitated reporting. Consistent with previous research on the timeline technique, the results showed that initial accounts were more detailed about person, object, action and setting information when participants used the timeline than the free recall reporting format (Hope et al., 2013b). Initial reporting was similar across format conditions for both gist and verbatim details, contrary to the results of a recent study by Hope et al. (2018). However, Hope and colleagues (2018) used a modified timeline tailored to facilitate access to gist and verbatim details by increasing the match between the encoding and the retrieval of conversations (*encoding specificity principle*; Tulving & Thomson, 1973). This was beyond the focus of the current study, as the reporting of gist and verbatim details was only one part of the witnessed events and we were interested in examining participants' recall of a complex series of events overall. With respect to particularization of specific instances, the results indicate that participants were able to retrieve more information for each instance when using the timeline technique than a free recall format. Possibly the temporal markers inherent in the timeline format and the use of separate timelines per event facilitated the retrieval of specific details relative to the structured narrative that one adheres to when using a free recall format. Although, it cannot be certainly concluded from the current results, it is likely that particularization was further facilitated due to the use of the self-generated cue mnemonic, which was used to prompt participants to recall specific details

from each instance prior to describing the event in detail (Brubacher et al., 2011). Consistent with previous research, participants who used the timeline in both conditions also reported more correct attributions of actions and statements to perpetrators than those who used a free recall format (Hope et al., 2013b; Hope et al., 2018). The memory literature suggests that details from witnessed events are associated in memory with the spatial and temporal context in which they were encoded (Tulving, 1983). Following that logic, a perpetrator's actions and statements will be temporally associated to each other and to the temporal and spatial context in which they were encountered. Therefore, the improved reporting of action and statement attributions to persons further suggests that use of the timeline prompted participants to retrieve information and correctly attribute it to each instance or in other words to the associated temporal context. The current results thus provide further evidence of the usefulness of the timeline technique for multi-actor events in both single and repeated experiences.

Importantly, the increased reporting and particularization of instances in both conditions where the timeline format was used (cf. free recall format), occurred without an increased cost in the reporting of intrusions. Certainly, fewer intrusions were expected in the MMIF relative to the timeline and to the free recall condition. However, the mean number of intrusions was overall low across conditions, indicating limited albeit some interference between events, which might still have implications in applied contexts. For instance, although there were few intrusions, the results show that the number of intrusions varied per detail type. Namely, 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 (Roberts, 2003). Given that source confusion errors are unlikely to be completely avoided, the current findings suggest that corroboration is necessary for details that interviewees have

difficulty attributing to an instance. Overall the use of memory-enhancing techniques was beneficial for the recall of specific instances, although there is further room for improvement in reducing source-monitoring errors. For instance, using manipulation checks and inquiring which event interviewees remember the best, would serve to better understand the metacognitive processes involved in discriminating individual instances and to explore useful directions for information elicitation (Danby, Brubacher, Sharman, Powell, & Roberts, 2017). Importantly, confusion of details between events might negatively affect the interviewee's credibility (Brubacher et al., 2014). For instance, limited research indicates that adults' reports of unique experiences are generally perceived as more credible than reports of repeated experiences (Weinsheimer et al., 2017). We should note that to better understand how memory for repeated experiences is perceived and assessed with respect to the reliability and credibility of interviewees' reporting and how such judgments might affect decision-making in applied contexts (e.g., sexual assault allegations, asylum-seeking applications), further research following on the current findings is presently underway.

Contrary to our hypotheses, the use of deviations in the changed content condition affected neither the recall nor the particularization of the witnessed instances. Accurate reporting of all of the events and of specific instances was improved when memory-enhancing techniques were used, relative to a free recall format, yet recall did not further benefit from the presence of deviations, as expected. Participants in both event content conditions reported similar rates of correct information across events and there was no statistically significant increase in the reporting of the third instance. The current results are inconsistent with previous research which shows that the presence of deviations facilitates recall for all the witnessed events (i.e., general effect) and for the targeted instance itself (i.e., target effect) (e.g. Connolly et al., 2016; MacLean

et al., 2018). In addition, the current results showed that witnessing a changed event did not facilitate source monitoring as participants reported a similar number of intrusions across conditions. One explanation for our results may be that the current deviations were not sufficiently memorable[5] to impact recall and particularization. Research suggests that the effect of deviations from the script on recall may depend on whether the deviation has any consequence to the sequence of events (i.e., *continuous deviations*; Connolly et al., 2016; *obstacles, errors and distractions*; Hudson et al., 1992; Schank & Abelson, 1977). Although the comparison between specific types of deviations was beyond the focus of our study, this differentiation might be relevant. In addition, an important limitation is that we did not include a manipulation check to confirm whether participants encoded the deviations (e.g. MacLean et al., 2018).

Alternatively, the current findings might be explained by the deviations becoming less accessible in memory over time than the general routine (Fuzzy-Trace-Theory; Brainerd & Reyna, 2004). In the context of the associative network theory (Anderson, 1983), although deviation details might initially form strong memory traces due to increased attention at encoding, delay causes the strength of the memory and the associated links to other traces to fade. Hence, when searching through the network at retrieval, traces might be less likely to be activated. Conversely, as fixed details across instances also form strong traces and links in the network because of repeated exposure, and schematic processing increases over time (Brewer & Nakamura, 1984), memory traces for fixed details are more easily activated at a delay (cf. deviations). Indeed, memory for deviations is better at immediate recall or after a 24-hour delay, but at longer intervals recall for fixed details increases due to schematic processing (see

---

[5] We can report that at least one of the two deviations we used, where a person that was not present in any other events, was seen in the park, passing by one of the perpetrators to hand her an envelope, was not reported by any participant.

Davidson, 2006), which affects both recall and source-monitoring performance (Graesser et al., 1980; Hudson et al., 1992). Notably, in recent studies which showed increased memory for deviations, participants (both children and adults) 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, where participants provided their account a week after the last witnessed event. Possibly, after the one-week interval, participants could not access the deviation details, since if they did all participants across reporting formats should report more details and fewer intrusions for at least the target instance. However further research on the effects of deviations on memory is needed to determine if this was the cause. Interestingly, deviations are conceptualized differently by FTT and by schema theory. According to schema theory, deviations are schema-inconsistent details which are stored separately to the script and can even serve as tags for specific instances (Graesser et al., 1979). Therefore, they differ from variations which are predictable alternatives to recurring details and gradually become absorbed by the script (e.g., Hudson et al., 1992). Within Fuzzy-Trace-Theory however, deviation details fit with the description of verbatim traces which are specific to an actual experience, as opposed to gist traces which represent the general pattern of an event in memory. As verbatim traces, deviations are more vulnerable to forgetting (cf. gist traces) and might not be retrieved after a delay (Brainerd & Reyna, 2004). Thus, the current findings appear to be more consistent with the conceptualization of deviations within the Fuzzy-Trace-Theory, rather with the proposed view of schema theory.

In conclusion, the current findings suggest that a multi-method interviewing format including the self-generated cue mnemonic, the timeline technique and open-ended follow-up questions can be useful in eliciting detailed and accurate information for repeated events with

multiple actors. Importantly, the null findings with respect to the presence of deviations in the current study indicate the potential limitations in reporting repeated events. Perhaps the finding that 24% of our participants reported witnessing fewer or more than four events is indicative in itself of how repeated similar events can become integrated in memory over time (Brewer & Nakamura, 1984). Further research is needed to investigate how deviations affect (or not) delayed recall and whether this can be used as an advantage for memory-enhancing techniques to improve reporting and reduce source confusion errors.

Chapter 5: Thesis General Discussion

The overarching aim of the current doctoral research was to examine the effectiveness of innovative information elicitation techniques in facilitating recall of multi-actor single and repeated events, to use with cooperative sources. Across four experiments, we tested the use of a self-generated cue in conjunction with the timeline technique for a unique event witnessed under both optimal and sub-optimal encoding conditions; examined the efficacy of follow-up open-ended questions based on a self-report; and tested the use of self-generated cues, the timeline technique and open prompts to facilitate recall of repeated events.

This discussion provides an overview of the main findings, followed by a consideration of the theoretical and practical implications. Finally, some of the limitations of the current research are presented together with suggestions for future research.

Summary of findings

One of the most consistent findings observed in the current programme of research is the effectiveness of the timeline technique, used with additional mnemonics, for improving recall and accurate reporting of complex witnessed events. In Experiment 1, use of self-generated cues in conjunction with the timeline increased the reporting of correct details (cf. other-generated cues and no cues), under full attention conditions at encoding. In Experiment 2, use of the timeline technique elicited more correct details than a free recall reporting format, in line with previous research (Hope et al., 2013b). Moreover, the results of Experiments 2 and 3 showed that the timeline technique can be bolstered by follow-up open-ended questions to elicit information across various multi-actor events. In Experiment 4, use of the timeline, with the self-generated cues and follow-up prompts, increased particularization of specific instances of multi-actor repeated events and elicited more correct information compared to free recall (there was no

additional benefit at retrieval when deviations from the script were present in the repeated events). In addition, use of the timeline technique in conjunction with the self-generated cues, improved the reporting of action and statement attributions to the persons involved in the event. Therefore, across four experiments, self-generated cues and follow-up prompts were successfully added to the timeline technique to facilitate the retrieval of multi-actor events witnessed on both single and repeated occasions. Given the limited research to date on the use of retrieval techniques that can facilitate recall for both single and repeated events witnessed by adults, these experiments make a novel contribution to the literature.

Another interesting finding emerged with respect to the efficacy of additional open prompts based on an initial self-report. In Experiments 2, 3 and 4, we found that asking follow-up open-ended questions led to the reporting of additional information, but that the accuracy of the new information was not as high as the accuracy of the spontaneously reported information. Research shows that with increased reporting, there is also a higher risk of erroneous reporting (Roberts & Higham, 2002). There is evidence which shows that even with the use of the Cognitive Interview, which includes various mnemonics and prompts, an increase in accurate details relative to standard interviewing approaches is also accompanied by a slight increase in erroneous information (Kohnken et al., 1999; Memon et al., 2010). An increase in errors, as part of an increase in overall reported information, has been attributed to the interviewees' (in)effective strategic monitoring of memory output. In other words, the more details interviewees provide the less conservative they tend to be about accurate information (Goldsmith et al., 2002; Memon et al., 2010). However, research shows that interviewees can balance informativeness and accuracy when they have control over reporting and when they are warned to not guess and to reply "I don't know" if uncertain about their recollections (Evans & Fisher,

2011; Koriat & Goldsmith, 1996; Scoboria & Fisico, 2013; Wells & Brewer, 2008). The findings of Experiment 3 are of particular interest because, even when interviewees were provided with explicit instructions to monitor accuracy in reporting, they did not provide more accurate information compared to interviewees who did not receive any instructions when asked follow-up open-ended questions. Therefore, these results indicate that explicit instructions to enhance accuracy are not always successful. It is also suggested that the use of follow-up questions might increase the demand for interviewees to be informative, and so they risk reporting less accurate details.

In summary, our findings suggest that the timeline technique, combined with additional mnemonics and prompts, can facilitate the retrieval and reporting of complex witnessed events, although not under sub-optimal encoding conditions. Below, we discuss the theoretical and practical implications of the research presented in this thesis.

Theoretical implications for research in memory retrieval processes

The current research draws on a rich theoretical background describing how cued-retrieval techniques can improve recall of witnessed events, by capitalizing on the associative nature of memory (e.g., Anderson, 1983; Bower, 1967; Tulving & Thomson, 1973). Our findings are largely consistent with previous research on the use of self-generated cues as distinctive mnemonics (e.g., Mäntylä & Nilsson, 1988). Similarly to Tullis and Benjamin (2005), in Experiment 1 we found that the use of self-generated cues facilitated retrieval more than other-generated context-retrieval cues. The results support the notion that although increasing the match between encoding and retrieval is necessary to facilitate recall, it is the distinctive quality of the cues that moderates how much information is retrieved (Goh & Lu, 2012; Nairne, 2002;

Watkins & Watkins, 1975). Consistent with findings in the literature, results in Experiment 1 showed that divided attention during encoding had a powerful effect on the amount of recalled information (e.g., Craik et al., 1996; Naveh-Benjamin et al., 2000). Although the effects of divided attention are well-documented in the literature, this experiment is among the few to examine whether the use of mnemonics at retrieval can facilitate access to poorly encoded memory traces. Unlike previous research, the results showed that neither self- nor other-generated cues facilitated recall when the event was witnessed under divided attention (e.g., Backman & Nilsson, 1991). The current findings suggest that possibly, if information is poorly encoded to begin with, it might not be available at retrieval despite the use of cues. However, further research is needed to examine what are the underlying mechanisms that contribute to the effectiveness of self-generated cues (cf. other-generated cues) and to delineate the limitations of their effectiveness under various distractions at encoding (e.g., Marsh et al., 2017).

With respect to recall of repeated events, schema theory suggests that memory tends to rely on the general routine of the fixed details across events (Brewer & Nakamura, 1987; Hudson et al., 1992). According to the source-monitoring framework, source misattribution of details is more likely when events are repeated and similar to each other (Johnson et al., 1993). In Experiment 4, however, the use of self-generated cues in conjunction with the timeline facilitated particularization of specific instances compared to a free recall reporting format. Although it cannot be directly concluded from this experiment, it is likely that the use of self-generated cues helped interviewees identify specific details to individual instances which in turn facilitated particularization (Brubacher et al., 2011). Consistent with previous research, using the timeline to report each event facilitated the retrieval of discrete details. The reporting of specific events likely increased due to the inherent temporal markers of the timeline that provided a frame of

reference in recalling the repeated events and which are lacking from the free narrative structure of the free recall format (Hope et al., 2018; Leins et al., 2014; Means & Loftus, 1991). Further, the results of improved reporting for specific details and for attributions of statements and actions to people with the timeline technique (cf. free recall format) support the notion that event-details that are encoded in close spatial and temporal proximity are recalled together (*temporal contiguity effect;* Howard & Kahana, 1999; Tulving, 1983; Unsworth, 2008). Also, in line with previous research which suggests that the use of episodic prompts increases the recall of specific relative to generic details of repeated events, the use of follow-up prompts elicited more information for each instance (Brubacher et al., 2012).

However, in contrast to previous research with both children and adults, the current findings showed that the presence of deviations from the script did not improve memory recall neither for all events nor just for the particular instance in which they occurred (e.g., Connolly et al., 2016; MacLean et al., 2018). The results in Experiment 4 indicated that deviations from the routine did not impact recall in any obvious respect. Given that we did not use a manipulation check to inquire participants if they remembered the deviations, it is possible that the presented deviations were not sufficiently memorable to affect recall. Research shows that deviations are more likely to be memorable when they have a consequence to the sequence of events (i.e., *continuous deviations*; Connolly et al., 2016; *obstacles, errors and distractions*; Hudson et al., 1992; Schank & Abelson, 1977). Although, examining the effect of specific types of deviations was not within the scope of the current research, it might be relevant to the results of Experiment 4 and worth investigating in the future.

Notably, in Experiment 4, there was a longer delay until retrieval compared to previous studies which show improved results for memory due to the presence of deviations (e.g.,

Connolly et al., 2016; MacLean et al., 2018; although see Brubacher et al., 2012). The lack of an effect of deviations on memory recall is interesting because there are competing theoretical interpretations regarding the effects of deviations on memory for repeated events. According to schema theory, deviations are memorable because they violate the schematic representation of the events and are therefore not effortlessly integrated to the script (Brewer & Nakamura, 1984). In fact, they can serve as "tags" for the specific instance in which they occurred, thus facilitating discrimination between instances (Graesser et al., 1979; Hudson et al., 1992). In the context of the Fuzzy-Trace-Theory, deviation details are the same as any other verbatim traces because they refer to the actual experience as opposed to gist traces which represent the general pattern of the event. As verbatim traces they are more vulnerable to forgetting than gist traces and might therefore not be retrieved after a delay (Brainerd & Reyna, 2002; 2004). Our results fit with the conceptualization of deviations as verbatim traces by the Fuzzy-Trace-Theory, rather than with the view of deviations by schema theory. Based on the current findings, future research should consider examining the conditions under which deviations have beneficial effects on recall and whether such benefits are reduced in the case of increased intervals between encoding and retrieval.

Not surprisingly, we found that the use of follow-up prompts can elicit new information that was not provided in an initial report. Consistent with previous research on the use of prompts in interviewing, the results in Experiments 2 and 3 suggest that follow-up questions can act as a cue for a specific part of the event to further encourage a search through memory (Fisher & Geiselman, 2002; 2010; Gabbert et al., 2016; Ibabe & Sporer, 2004). Importantly, the results across both experiments also showed that the accuracy of the additional information was not as high as that of the spontaneously reported information. This result was surprising because

previous research shows that interviewees balance informativeness and accuracy when asked open-ended questions as they have control over what information to volunteer (Evans & Fisher, 2011; Goldsmith et al., 2002; Koriat & Goldsmith, 1996). In contrast to previous research, the use of instructions to improve effective monitoring in Experiment 3, such as to avoid guessing, to feel free to withhold an answer and to consider the precision of their answers was not successful in increasing accurate reporting (Evans & Fisher, 2011; Koriat & Goldsmith, 1996; McCallum et al., 2016; Scoboria & Fisico, 2013; Wells & Brewer, 2008). Overall, the results of Experiments 2 and 3 indicate that there might be limitations in how effectively interviewees can balance informativeness and accuracy when asked follow-up questions. Future research is therefore needed to investigate the potential costs and benefits of additional reporting in response to follow-up questions.

Practical implications for applied information elicitation contexts

The findings presented in this thesis support and build on previous research on the effectiveness of the timeline technique in facilitating recall of multi-actor events (Hope et al., 2013b). In line with Hope et al. (2018), the current results show that additional mnemonics and follow-up questions can be used together with the timeline technique, thus contributing to the development of an adaptive format that can be used flexibly for both single and repeated witnessed events. Eliciting reliable information for multi-actor events can be critical both in forensic and security settings, such as in cases of assault where multiple perpetrators are involved or in gang crimes and terrorist activities performed by groups.

The results of Experiment 1 demonstrated that self-generated cues can be effectively used with the timeline technique to improve the reporting of multi-actor events witnessed under full attention. Consistent with recent research which shows that cues that are derived from a

witness's own perspective can be efficient at retrieval, we found that self-generated cues elicited more information than the use of generic contextual cues provided by the interviewer (Willen et al., 2015). Given that witnessing an event is a subjective experience and that interviewers may be unaware of what information is salient to the interviewee, self-generated cues can be used to support a witness-led largely self-administered report. Further evidence in support of using self-generated cues with the timeline is provided by Experiment 4, where reporting for repeated events was improved relative to the use of a free recall reporting format with no additional mnemonics. When eliciting information about repeated events, adopting a witness-led approach can be particularly important, as research shows that confusing when details occurred is more likely when remembering repeated than single events, especially if they are thematically similar (Johnson et al., 1993; Lindsay et al., 2004). As the likelihood for such misattributions between events increases, so does the interviewee's suggestibility and susceptibility to misinformation (e.g., Zaragoza & Lane; 1994). Although interviewees in Experiment 4 reported an overall low number of internal intrusion errors between events, previous research suggests that intrusions between repeated events are unlikely to be completely avoided (Brubacher et al., 2011; Lindsay et al., 2004; McNichol et al., 1999). Although not directly examined in this experiment, promoting a self-administered report of the events might have facilitated the low reporting of internal intrusions across conditions. Future research could systematically examine if the use of self-generated cues and self-administered reporting limits the potential impact of suggestive questioning for repeated witnessed events.

Cases where interviewees might be questioned about repeated experiences that have occurred over a period of time, are relevant for both police investigations (e.g., domestic abuse, sexual assault, gang crime) and HUMINT contexts (e.g., attending meetings where future

terrorist activities are planned). Further, when multiple perpetrators are involved in such cases, it is crucial to elicit information that helps identify "who did and/or said what and when" to correctly target members in a criminal network and to ensure the prosecution of the guilty party in a trial (Roberts, 2003). The results of Experiment 4 are consistent with previous research on the timeline, which shows that it increases the accurate reporting of attributions of actions and statements to the people involved in the witnessed event (cf. free recall; although see Experiment 2). Therefore, the current findings suggest that the timeline technique bolstered by the use of self-generated cues facilitates retrieval and reporting of multi-actor events.

The results presented in this thesis also show that follow-up open-ended questions can be used to elicit new information for both single and repeated events. Consistent with previous research, the results of Experiment 1 indicated that even with support at retrieval, interviewees may provide accurate and detailed accounts but omit critical details that may be relevant to an investigation (Hope et al., 2013a; Smeets et al., 2004). Also, findings on eyewitness memory research, suggest that when witnessing complex events, not all components are likely to be strongly encoded in memory, which might lead to within-statement inconsistencies (Fisher, Brewer, & Mitchell, 2009). Further, in reporting repeated events inconsistent reporting might also occur due to source misattributions between instances (Johnson et al., 1993). To elicit more information about critical details and to address inconsistencies, interviewers are likely to use follow-up prompts (e.g., Shepherd & Griffiths, 2013).

Across Experiments 2, 3 and 4, follow-up open-ended questions elicited additional information to an initial self-report, suggesting that the questions prompted an additional memory search (Fisher & Geiselman, 1992). However, across experiments, the accuracy of the new information was not as high as that of the spontaneously reported information. Therefore,

practitioners should be aware that although there is a limited pool of accurate details that interviewees can recall, there is an unlimited pool of inaccurate details to report. The current doctoral research suggests that practitioners should be cautious about the use of follow-up questions and consider the need for corroboration of additionally reported information to increase its reliability. Further empirical research could systematically examine the costs and benefits when using other types of probes, such as focused questions which are also likely used to target specific omitted information (5WH; Milne & Bull, 1999). For instance, although there may be limitations in using confidence judgments to assess the accuracy of responses to open prompts (e.g., Ibabe & Sporer, 2004), research shows that they can be indicative of accuracy when used for responses to specific questions (e.g., Goldsmith et al., 2002; Koriat & Goldsmith, 1996). Therefore, future research could investigate whether confidence judgments – and possibly other metacognitive markers – can accompany responses to follow-up focused questions as "signatures" to accurate information (see also Fisher, 1995; Fisher, Schreiber Compo, Rivard, & Hirn, 2014).

Methodological considerations and future directions

One of the limitations in the current experiments is that ecological validity in certain aspects of the experimental design was low or necessarily artificial. In Experiment 1, we instructed participants to perform a concurrent task during the encoding of an event to divide their attention. Although, this measure is commonly used in the literature and has been shown to successfully cause a diversion of attentional resources at encoding (e.g., Naveh-Benjamin et al., 2006), it is not a perfect analogue to a real-world experience of witnessing an event and having to pay attention to multiple aspects on the scene. Further research should investigate the effectiveness of mnemonics with more naturalistic divided attention measures, such as using a

smartphone, conversing (e.g. Marsh et al., 2017), or having an operationally active role (e.g., emergency responder, police or military officer) compared to being a bystander during an event (Hope et al., 2016). Given the limited research in the effectiveness of memory-enhancing techniques for sub-optimal encoding conditions, future research could delineate the limitations of the use of cues and provide guidelines about their deployment by interviewers in the field.

A similar concern with respect to the examination of cues and techniques in these experiments is our use of short-term intervals between the event and the retrieval phase. In most of the experiments reported in this thesis, participants provided a report about the witnessed event after a 10-minute interval. One exception was Experiment 4, where participants witnessed the events on separate occasions and provided their report a week after the last event. Although immediate recall testing is commonly used in laboratory research for an initial examination of the effectiveness of retrieval techniques, interviewers in HUMINT settings often focus on non-recent events that, in contrast to common criminal investigation cases, might not have been previously reported (Evans et al., 2010). Research on eyewitness memory shows that as the interval between witnessing and recalling an event increases, the level of detail and accuracy in reporting decreases (e.g., Penrod, Loftus, & Winkler, 1982). However, meta-analyses on the effectiveness of the Cognitive Interview (CI; Fisher & Geiselman, 1992), which is an established approach for improving eyewitness recall, show that although the beneficial effects of the CI decrease after a delay with respect to both the quantity and quality of the reported information, there is still an advantage on recall compared to standard interviewing protocols (Kohnken et al., 1999; Memon et al., 2010). Further, although research on the timeline technique is so far limited, there is evidence that even after a two-week delay, use of the timeline elicits 32% more correct information relative to a free recall request (cf. 47% more at immediate recall; Hope et al.,

2013b). Taken together, although further research is needed, findings to date suggest that the benefits of using memory-enhancing techniques to improve retrieval are robust to delays between encoding and retrieval (cf. free request of information). However, it is likely that the effects on memory recall would be lessened after prolonged retention intervals.

Another potential limitation of the experiments reported in this thesis, is that the use of the timeline was primarily tested with a sample of literate and motivated university students. Given that the timeline technique relies on a self-administered written format, future research should investigate if it is equally effective for recall with interviewees with different educational backgrounds and whether – or to what extent – its use needs to be facilitated by the interviewer. Similarly, although participation in the current experiments was not restricted to native English speakers, participants had to be fluent in the English language to avoid any effects on performance due to reasons other than not remembering details of the event. Although, the timeline format can be used flexibly in reporting, further research is needed to explore if interviewees who use a language that assumes a different form to English in writing (e.g., Arabic; Chinese) approach the linear direction of the timeline differently to English-speaking users. To date, research on the effectiveness of memory-enhancing techniques across cultures is limited, yet it would be particularly useful in the context of HUMINT settings where interviewers may interact with sources from diverse cultural backgrounds (e.g., Russano, Narchet, Kleinman, & Meissner, 2014).

A potential concern with respect to the procedure in Experiments 2 and 4 could be the choice of free recall as a comparison to the timeline technique, given that any additional support is likely to facilitate retrieval. In keeping with previous research by Hope and colleagues (2013b; 2018), and since this was an initial test of extending the timeline with additional mnemonics and

follow-up prompts for single and repeated events, we examined the effectiveness of the timeline format against a free request of information. Further, using a free recall format in comparison to the timeline allowed for self-administered reporting across conditions, without any additional interactions with the interviewer prior to the stage of the follow-up questions. It is worth noting that the general instructions provided to participants at retrieval remained the same across conditions, and only the instructions pertaining to the reporting formats differed (e.g., using person description and action cards on the timeline). Importantly, the aim of the current doctoral programme was not to compare the timeline combined with additional components to already established effective techniques (e.g., Cognitive Interview, Fisher & Geiselman, 1992; Self-Administered Interview, Gabbert et al., 2009) but to contribute to the development of techniques and tools that practitioners can use in a range of circumstances in applied information elicitation contexts.

With respect to our investigation of the efficacy of follow-up questions, the current results do not demonstrate conclusively why explicit instructions that emphasize accuracy did not improve the quality of reporting, unlike in previous research (e.g., Evans & Fisher, 2011; Koriat & Goldsmith, 1996; McCallum et al., 2016; Scoboria & Fisico, 2013; Wells & Brewer, 2008). Research shows that interviewees generally prefer to report informative (i.e. precise) details, but they strategically regulate their responses based on the context, and whether there are financial (e.g., penalties for inaccuracy) or social consequences (e.g., reporting information in public or in private) (Koriat & Goldsmith, 1996; McCallum et al., 2016). Although in these experiments, there was no manipulation of context, we cannot reject the possibility that social or communicative factors affected reporting, especially since interviewees answered follow-up questions in a direct social interaction with an interviewer after having used a self-administered

format to provide an initial report. Given that even after explicit instructions about accuracy, information in the follow-up phase was not as accurate as in the initial report, it may be that the use of follow-up questions might implicitly communicate an increased demand for informativeness to begin with. Future research is needed to investigate how interviewees' perceptions of what is required of them with respect to accuracy, informativeness and precision, interact with their strategic control over reporting. Given that both cognitive and social factors affect the quantity and quality of reporting (Fisher & Geiselman, 1992), it might be that when asking follow-up questions, using warnings to emphasize accuracy is necessary but not enough, and that interviewers need to implicitly or explicitly communicate what the relevance of the interviewee's contribution is (Grice, 1975). Further research would increase our understanding of interviewees' decision-making strategies at reporting and provide interviewers with guidelines for effectively communicating with sources.

Consistent with previous research, the current results suggest that various mnemonics and techniques can be used together to facilitate access to information in memory (e.g., Fisher & Geiselman, 2010; Hope et al., 2018; Leins et al., 2014). Following recent calls to develop adaptive theory-driven techniques for information elicitation, the current doctoral thesis outlined several benefits in using intuitive and innovative mnemonics and formats for recall of complex witnessed events (Fisher et al., 2014; Granhag et al., 2014). Recently, Fisher and colleagues (2014) suggested that there may be unexplored benefits to the use of multiple response modalities in reporting. For instance, although reporting via writing may allow the interviewee more time to probe and retrieve details from memory, some information might be optimally reported via speech. The current research aimed to be a step towards that direction by introducing the use of follow-up prompts to a self-administered written timeline format. In keeping with the logic of various response

modalities, there is evidence that sketching can facilitate the retrieval and description of physical locations (Dando et al, 2009a; Fisher & Geiselman, 1992), while recent studies show that a family tree mnemonic might be beneficial for reporting information about the hierarchy of networks (Hope et al., 2018; Leins et al., 2014). Therefore, further research could focus on developing non-traditional multi-modal formats informed by the memory literature that can be used flexibly by interviewees to report different types of information.

Conclusion

Across four experiments, the use of theoretically-driven cues and techniques facilitated the retrieval of various multi-actor events witnessed on both single and repeated occasions. These findings are largely consistent with previous research which shows that the use of various cues and retrieval attempts facilitates access to information from memory for witnessed events. The current experiments successfully extended the timeline technique with self-generated cues and follow-up open prompts, thus contributing to the development of flexible evidence-based information elicitation techniques. Therefore, the research presented in this thesis makes a useful contribution to the growing body of evidence-based techniques for use with cooperative sources in security HUMINT contexts.

References

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

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.

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

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

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

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

BÄckman, L., & Nilsson, L. G. (1991). Effects of divided attention on free and cued recall of verbal events and action events. *Bulletin of the Psychonomic Society*, *29*(1), 51-54. doi:10.3758/BF03334767

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University.

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

Belli, R. F., Stafford, F. P., & Alwin, D. F. (Eds.). (2009). *Calendar and time diary methods in life course research.* Thousand Oaks, CA, USA: Sage.

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

Berntsen, D., Staugaard, S. R., & Sørensen, L. M. T. (2013). Why am I remembering this now? Predicting the occurrence of involuntary (spontaneous) episodic memories. *Journal of Experimental Psychology: General*, *142*(2), 426-444. doi:10.1037/a0029128

Borum, R., Gelles, M., & Kleinman, S. (2009). Interview and interrogation: A perspective and update from the USA. In R. Milne, S. Savage, & T. Williamson (Eds.) *International Developments in Investigative Interviewing* (pp. 111-128). Cullompton, UK: Willan Publishing.

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.

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

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory: Dual processes in memory, reasoning, and cognitive neuroscience. *Advances in child development and behaviour*, *28*, 41-100.

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

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

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

Brown, C., Lloyd-Jones, T. J., & Robinson, M. (2008). Eliciting person descriptions from eyewitnesses: A survey of police perceptions of eyewitness performance and reported use

of interview techniques. *European Journal of Cognitive Psychology*, *20*(3), 529-560. doi:10.1080/09541440701728474

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

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

Brubacher, S. P., Powell, M. B., & Roberts, K. P. (2014). Recommendations for interviewing children about repeated experiences. *Psychology, Public Policy, and Law*, *20*(3), 325-335. http://dx.doi.org/10.1037/law0000011

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

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, G., & Java, R. (1995). Memory for medical history: Accuracy of recall. *Applied*

   Cognitive Psychology, *9*(4), 273-288. https://doi.org/10.1002/acp.2350090402

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

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic

   processing. *Psychological review*, *82*(6), 407.

Connolly, D. A., & Gordon, H. M. (2014). Can order of general and specific memory prompts

   help children to recall an instance of a repeated event that was different from the

   others?. *Psychology, Crime & Law*, *20*(9), 852-864.

   https://doi.org/10.1080/1068316X.2014.885969

Connolly, D. A., Gordon, H. M., Woiwod, D. M., & Price, H. L. (2016). What children recall

   about a repeated event when one instance is different from the others. *Developmental*

   Psychology, 52(7), 1038-1051.

Connolly, D. A., & Lindsay, D. S. (2001). The influence of suggestions on children's reports of a

   unique experience versus an instance of a repeated experience. *Applied Cognitive*

   Psychology, *15*(2), 205-223. https://doi.org/10.1002/1099-

   0720(200103/04)15:2<205::AID-ACP698>3.0.CO;2-F

Connolly, D. A., & Price, H. L. (2006). Children's suggestibility for an instance of a repeated event versus a unique event: The effect of degree of association between variable details. *Journal of Experimental Child Psychology, 93*(3), 207-223. https://doi.org/10.1016/j.jecp.2005.06.004

Conway, M. A., & Pleydell-Pearce, C. W. (2000). The construction of autobiographical memories in the self-memory system. *Psychological review*, *107*(2), 261-288.

Craik, F. I., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General*, *125*(2), 159-180. doi:10.1037/0096-3445.125.2.159

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

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

Dando, C., Wilcock, R., & Milne, R. (2009a). The cognitive interview: The efficacy of a modified mental reinstatement of context procedure for frontline police investigators. *Applied Cognitive Psychology*, *23*(1), 138-147. doi:10.1002/acp.1451

Dando, C., Wilcock, R., Milne, R., & Henry, L. (2009b). A modified cognitive interview procedure for frontline police investigators. *Applied Cognitive Psychology: The Official*

Journal of the Society for Applied Research in Memory and Cognition, *23*(5), 698-716.

https://doi.org/10.1002/acp.1501

Davidson, D. (2006). Memory for bizarre and other unusual events: Evidence from script

research. In R. Hunt & J. B. Worthen (Eds.), *Distinctiveness and memory* (pp. 157–179).

New York, NY: Oxford University.

Davis, M. R., McMahon, M., & Greenwood, K. M. (2005). The efficacy of mnemonic

components of the cognitive interview: towards a shortened variant for time-critical

investigations. *Applied Cognitive Psychology*, *19*(1), 75-93.

https://doi.org/10.1002/acp.1048

DeClue, G. (2010). Three things that should be used to guide investigative interviews by military

and intelligence agencies. *The Journal of Psychiatry & Law*, *38*(1-2), 251-262. doi:

https://doi.org/10.1177/009318531003800111

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

Evans, J. R., Meissner, C. A., Brandon, S. E., Russano, M. B., & Kleinman, S. M. (2010).

Criminal versus HUMINT interrogations: The importance of psychological science to

improving interrogative practice. *The Journal of Psychiatry & Law*, *38*(1-2), 215-249.

doi: https://doi.org/10.1177/009318531003800110

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

Farrar, M. J., & Goodman, G. S. (1992). Developmental changes in event memory. *Child Development*, *63*(1), 173-187. https://doi.org/10.1111/j.1467-8624.1992.tb03605.x

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

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

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. P., Brewer, N., & Mitchell, G. (2009). The relation between consistency and accuracy of eyewitness testimony: Legal versus cognitive explanations. *R. Bull, T. Valentine, & T. Williamson (Eds.), Handbook of psychology of investigative interviewing: Current developments and future directions*, 121-136.

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

Fisher, R. P., & Geiselman, R. E. (2010). The cognitive interview method of conducting police interviews: Eliciting extensive information and promoting therapeutic

jurisprudence. *International journal of law and psychiatry*, *33*(5-6), 321-328. https://doi.org/10.1016/j.ijlp.2010.09.004

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

Fisher, R. P., Schreiber Compo, N., Rivard, J., & Hirn, D. (2014). Interviewing witnesses. In T. J. Perfect, & D. S. Lindsay (Eds.), *The SAGE handbook of applied memory*, (559-578). London, UK: SAGE.

Fivush, R., Hudson, J., & Nelson, K. (1984). Children's long-term memory for a novel event: An exploratory study. *Merrill-Palmer Quarterly*, *30*(3), 303-316. Retrieved from: https://www.jstor.org/stable/23086104

Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current issues and advances in misinformation research. *Current Directions in Psychological Science*, *20*(1), 20-23. https://doi.org/10.1177/0963721410396620

Gabbert, F. & Brown, C. (2015). Interviewing for face identification. In T. Valentine, & J. P. Davis, (Eds.), *Forensic Facial Identification: Theory and Practice of Identification from Eyewitnesses, Composites and CCTV.* Chichester, UK: Wiley-Blackwell.

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.

Gabbert, F., MacPherson, I., & Hope, L. (2014, June). Adding a new mnemonic to the Cognitive Interview. Paper presented at the 7th Conference of the International Investigative Interviewing Research Group (IIIRG), Lausanne, Switzerland.

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

Gilbert, J. A., & Fisher, R. P. (2006). The effects of varied retrieval cues on reminiscence in eyewitness memory. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition*, *20*(6), 723-739. https://doi.org/10.1002/acp.1232

Godden, D. R., & Baddeley, A. D. (1975). Context-dependent memory in two natural environments: On land and underwater. *British Journal of psychology*, *66*(3), 325-331. https://doi.org/10.1111/j.2044-8295.1975.tb01468.x

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

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

Graesser, A. C., Gordon, S. E., & Sawyer, J. D. (1979). Recognition memory for typical and

    atypical actions in scripted activities: Tests of a script pointer+ tag hypothesis. *Journal of*

    Memory and Language, *18*(3), 319.

Graesser, A. C., Woll, S. B., Kowalski, D. J., & Smith, D. A. (1980). Memory for typical and

    atypical actions in scripted activities. *Journal of Experimental Psychology: Human*

    Learning and Memory, *6*(5), 503.

Granhag, P. A., Vrij, A., & Meissner, C. A. (2014). Information gathering in law enforcement and

    intelligence settings: Advancing theory and practice. *Applied Cognitive*

    Psychology, *28*(6), 815-816. doi:10.1002/acp.3093

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., Blocksidge, D., Gabbert, F., Sauer, J. D., Lewinski, W., Mirashi, A., & Atuk, E. (2016). Memory and the operational witness: Police officer recall of firearms encounters as a function of active response role. *Law and human behavior*, *40*(1), 23-35.

Hope, L., Gabbert, F., & Fraser, J. (2013a). 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. (2018). Who said what and when? A timeline approach to eliciting information and intelligence about conversations, plots and plans. Paper under review.

Hope, L., Mullis, R., & Gabbert, F. (2013b). 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

Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multi-element events. *Journal of Experimental Psychology: General*, *142*(4), 1370-1383. doi:10.1037/a0033626

Howard, M. W., & Kahana, M. J. (1999). Contextual variability and serial position effects in free recall. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *25*(4), 923–941.

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

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

Ihlebæk, C., Løve, T., Erik Eilertsen, D., & Magnussen, S. (2003). Memory for a staged criminal event witnessed live and on video. *Memory*, *11*(3), 319-327. doi:10.1080/09658210244000018

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

Johnson, M. K., Kahan, T. L., & Raye, C. L. (1984). Dreams and reality monitoring. *Journal of Experimental Psychology: General*, *113*(3), 329-344. http://dx.doi.org/10.1037/0096-3445.113.3.329

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory & Cognition*, *24*(1), 103–109. https://doi.org/10.3758/BF03197276

Kahana, M. J. (2012). *Foundations of human memory.* New York, NY: Oxford University Press.

Kebbell, M., Milne, R., & Wagstaff, G. (1999). The Cognitive Interview: A survey of its forensic effectiveness. *Psychology, Crime, & Law, 5*(1-2)*,* 101–115. doi:10.1080/10683169908414996

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.

Kuebli, J. (1990, March). This way or that way: complexity in children's event knowledge. In *Southeastern Conference on Human Development,* Richmond, VA (March).

Lane, S. M. (2006), Dividing attention during a witnessed event increases eyewitness suggestibility. *Applied Cognitive Psychology, 20*(2)*,* 199–212. doi:10.1002/acp.1177

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

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

Lindsay, D. S., Allen, B. P., Chan, J. C., & Dahl, L. C. (2004). Eyewitness suggestibility and source similarity: Intrusions of details from one event into memory reports of another

event. *Journal of Memory and Language*, *50*(1), 96-111.

https://doi.org/10.1016/j.jml.2003.08.007

Lindsay, D. S., & Johnson, M. K. (1989). The eyewitness suggestibility effect and memory for

source. *Memory & Cognition*, *17*(3), 349-358. https://doi.org/10.3758/BF03198473

Loftus, E. (2003). Our changeable memories: Legal and practical implications. *Nature Reviews*

Neuroscience, *4*(3), 231-234. doi:10.1038/nrn1054

Loftus, E. F. (1979). The malleability of human memory: Information introduced after we view

an incident can transform memory. *American Scientist*, *67*(3), 312-320.

http://www.jstor.org/stable/27849223

London attack: What happened (2018, May). *BBC News.* Retrieved from:

https://www.bbc.co.uk/news/uk-england-london-40147164

Luna, K., & Martín-Luengo, B. (2012). Confidence–accuracy calibration with general

knowledge and eyewitness memory cued recall questions. *Applied Cognitive*

Psychology, *26*(2), 289-295. https://doi.org/10.1002/acp.1822

Luna, K., Martín-Luengo, B., & Brewer, N. (2015). Are regulatory strategies necessary in the

regulation of accuracy? The effect of direct-access answers. *Memory & cognition*, *43*(8),

1180-1192. https://doi.org/10.3758/s13421-015-0534-2

MacLean, C. L., Coburn, P. I., Chong, K., & Connolly, D. L. (2018). Breaking script: Deviations

and postevent information in adult memory for a repeated event. *Applied Cognitive*

Psychology, *32*(4), 474-486. https://doi.org/10.1002/acp.3421

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

Marsh, J. E., Patel, K., Labonté, K., Threadgold, E., Skelton, F. C., Fodarella, C., ... & Vachon, F. (2017). Chatting in the face of the eyewitness: The impact of extraneous cell-phone conversation on memory for a perpetrator. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, *71*(3), 183-190. http://dx.doi.org/10.1037/cep0000101

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

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

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

Memon, A., Meissner, C. A., & Fraser, J. (2010). The Cognitive Interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, public policy, and law*, *16*(4), 340-372.

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

Milne, R. and Bull, R. (1999). *Investigative interviewing: psychology and practice*. Chichester, UK: Wiley.

Mulligan, N. W. (2014). Memory for pictures and actions. In T. J. Perfect, & D. S. Lindsay (Eds.), *The SAGE handbook of applied memory*, (pp. 20-36). London, UK: SAGE.

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

Naveh-Benjamin, M., Craik, F. I., Perretta, J. G., & Tonev, S. T. (2000). The effects of divided attention on encoding and retrieval processes: The resiliency of retrieval processes. *The Quarterly Journal of Experimental Psychology: Section A*, *53*(3), 609-625. https://doi.org/10.1080/713755914

Naveh-Benjamin, M., Kilb, A., & Fisher, T. (2006). Concurrent task effects on memory encoding and retrieval: Further support for an asymmetry. *Memory & Cognition*, *34*(1), 90-101. doi:10.3758/BF03193389

Oberauer, K., & Bialkova, S. (2009). Accessing information in working memory: Can the focus of attention grasp two elements at the same time? *Journal of Experimental Psychology: General*, *138*(1), 64–87. http://dx.doi.org/10.1037/a0014738

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

Ozgul, F. (2016). Analysis of topologies and key players in terrorist networks. *Socio-Economic Planning Sciences*, *56*, 40-54. doi:10.1016/j.seps.2016.07.002

Office for National Statistics (2015). Focus on Violent Crime and Sexual Offences: Year ending March 2015. Retrieved from: https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/compendium/focusonviolentcrimeandsexualoffences/yearendingmarch2015

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.

Pansky, A., Koriat, A., & Goldsmith, M. (2005). Eyewitness recall and testimony. In N. Brewer, & K. D. Williams (Eds.), *Psychology and law: An empirical perspective*, (pp. 93-150). New York, NY, USA: Guilford Press.

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

Penrod, S., Loftus, E. F., & Winkler, J. (1982). The reliability of eyewitness testimony: A psychological perspective. In N. Kerr & R. Bray (Eds.), The psychology of the courtroom (pp. 119-168). New York, NY, USA: Academic Press.

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

Polyn, S. M., Norman, K. A., & Kahana, M. J. (2009). A context maintenance and retrieval model of organizational processes in free recall. *Psychological Review*, *116*(1), 129–156.

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

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

Rivard, J. R., Fisher, R. P., Robertson, B., & Hirn Mueller, D. (2014). Testing the cognitive interview with professional interviewers: Enhancing recall of specific details of recurring events. *Applied Cognitive Psychology*, *28*(6), 917-925. https://doi.org/10.1002/acp.3026

Roberts, K. P. (2002). Children's ability to distinguish between memories from multiple sources: Implications for the quality and accuracy of eyewitness statements. *Developmental Review*, *22*(3), 403-435. https://doi.org/10.1016/S0273-2297(02)00005-9

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

Roberts, K. P., Brubacher, S. P., Drohan-Jennings, D., Glisic, U., Powell, M. B., & Friedman, W. J. (2015). Developmental differences in the ability to provide temporal information about

repeated events. *Applied Cognitive Psychology*, *29*(3), 407-417.

https://doi.org/10.1002/acp.3118

Roberts, W. T., & Higham, P. A. (2002). Selecting accurate statements from the cognitive

interview using confidence ratings. *Journal of Experimental Psychology: Applied*, *8*(1),

33-43. http://dx.doi.org/10.1037/1076-898X.8.1.33

Russano, M. B., Narchet, F. M., Kleinman, S. M., & Meissner, C. A. (2014). Structured

interviews of experienced HUMINT interrogators. *Applied Cognitive Psychology*, *28*(6),

847-859. https://doi.org/10.1002/acp.3069

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

human knowledge structures. Hillsdale, NJ: Erlbaum.

Schneider, L., Price, H. L., Roberts, K. P., & Hedrick, A. M. (2011). Children's episodic and

generic reports of alleged abuse. *Applied Cognitive Psychology*, *25*(6), 862-870.

https://doi.org/10.1002/acp.1759

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.

Slackman, E., & Nelson, K. (1984). Acquisition of an unfamiliar script in story form by young children. *Child Development*, *55*(2), 329-340. doi: 10.2307/1129946

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

Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic bulletin & review*, *8*(2), 203-220. https://doi.org/10.3758/BF03196157

Taylor, P. J., Snook, B., Bennell, C., & Porter, L. (2015). Investigative psychology. In B. L. Cutler, & P. A. Zapf (Eds.), APA handbook of forensic psychology (Vol. 2 Criminal investigation, adjudication, and sentencing outcomes) (pp. 165‑186). Washington, DC, US: American Psychological Association.

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

Thomson, D. M., & Tulving, E. (1970). Associative encoding and retrieval: Weak and strong cues. *Journal of experimental psychology*, *86*(2), 255-262. http://dx.doi.org/10.1037/h0029997

Tullis, J. G., & Benjamin, A. S. (2015). Cue generation: How learners flexibly support future retrieval. *Memory & Cognition*, *43*(6), 922-938. doi:10.3758/s13421-015-0517-3

Tulving, E. (1983). Elements of Episodic Memory. Oxford: Clarendon.

Tulving, E., & Osler, S. (1968). Effectiveness of retrieval cues in memory for words. *Journal of experimental psychology*, *77*(4), 593-601. http://dx.doi.org/10.1037/h0026069

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

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

Unsworth, N. (2008). Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal–contextual search. *Journal of Memory and Language*, *59*(2), 223–236. https://doi.org/10.1016/j.jml.2008.04.002

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

Vrij, A., Hope, L., & Fisher, R. P. (2014). Eliciting reliable information in investigative interviews. *Policy Insights from the Behavioral and Brain Sciences*, *1*(1), 129-136. https://doi.org/10.1177/2372732214548592

Watkins, O. C., & Watkins, M. J. (1975). Buildup of proactive inhibition as a cue-overload effect. *Journal of Experimental Psychology: Human Learning and Memory*, *1*(4), 442-452. doi:10.1037/0278-7393.1.4.442

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

Weinsheimer, C. C., Coburn, P. I., Chong, K., MacLean, C. L., & Connolly, D. A. (2017). Perceptions of credibility for a memory report of a single versus repeated event. *Applied Cognitive Psychology*, *31*(4), 414-423. https://doi.org/10.1002/acp.3340

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

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

Wheeler, R. L., & Gabbert, F. (2017). Using self-generated cues to facilitate recall: a narrative review. *Frontiers in psychology*, *8*, 18-30. https://doi.org/10.3389/fpsyg.2017.01830

Wheeler, R. L., Gabbert, F., Hope, L., & Jones, S. (2017, January). *Evaluating a Self-Generated Cue Mnemonic to Enhance Eyewitness Retrieval.* Paper presented at SARMAC conference, Sydney, Australia.

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.

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

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*(4), 934-945. http://dx.doi.org/10.1037/0278-7393.20.4.934

Appendix A: Supplemental Materials (Chapter 2)

In this Supplemental Materials section, we provide information about coding and analyses for variables which are conventional in this research area (e.g. reporting of incorrect details) but which lie outside our main hypotheses.

Critical Details Coding

Prior to data collection, six legal professionals viewed the stimulus event and independently provided a list of details that they considered critical to pursue an investigation of the assault and relevant legal charges. Details mentioned by at least four of the six legal professionals were included in a final list of 24 critical details. Accounts were then coded for the reporting of these critical details. To calculate a completeness rate for critical details, the total of reported critical details was divided by 24 (i.e. the maximum number of critical details). Higher scores indicated higher levels of completeness.

Supplemental Results (main results reported in chapter 2)

Reporting of Incorrect Details

There was no significant main effect of Cues, $F(2,126) = 1.10$, $p = .337$, $\omega^2 = .001$, or Attention, $F(1,126) = .08$, $p = .777$, $\omega^2 = -.007$, on the total number of incorrect details reported. The interaction between Attention and Cues was not significant, $F(2,126) = .23$, $p = .793$, $\omega^2 = -.012$. Means for incorrect details reported as a function of cue and attention conditions are presented in Table 1.

Table 1

Mean number (SE) of incorrect details by cues (Self-Generated Cues, Other-Generated Cues, No Cues) and attention (Full and Divided).

	Incorrect details					
	SGC		OGC		NC	
Attention	*M (SE)*	95%CI	*M (SE)*	95%CI	*M (SE)*	95%CI
Full	9.9 (0.5)	[7.5, 12.3]	9.3 (0.7)	[6.6, 12.4]	11 (0.6)	[7.6, 13.6]
Divided	10.1 (0.4)	[8.3, 11.9]	8.7 (0.3)	[7.1, 10.4]	11.6 (0.7)	[8.6, 14.7]

Reporting of Incorrect Action Attributions

There was no effect of either Attention, $F(1,126) = .00$, $p = 1.00$, $\omega^2 = -.008$, or Cues, $F(2,126) = .74$, $p = .479$, $\omega^2 = -.004$, on the total number of incorrect person-action details. No significant interaction emerged between Cues and Attention, $F(2,126) = 2.01$, $p = .138$, $\omega^2 = .015$. Means for incorrect person-action details reported as a function of cue and attention conditions are presented in Table 2.

Table 2

Mean number (SE) of incorrect person-action details by cues (Self-Generated Cues, Other-Generated cues, No Cues) and attention (Full and Divided).

	Incorrect Person-Action details					
	SGC		OGC		NC	
Attention	*M (SE)*	95%CI	*M (SE)*	95%CI	*M (SE)*	95%CI
Full	0.86 (0.1)	[0.53, 1.21]	1.5 (0.1)	[0.91, 2.32]	1.05 (0.1)	[0.56, 1.56]
Divided	1.05 (0.1)	[0.58, 1.54]	0.91 (0.1)	[0.54, 1.35]	1.45 (0.1)	[0.95, 2.00]

Reporting of Critical Details

The mean number of reported critical details across conditions was 12 ($SD = 2.9$) out of a total of

24 details. There was a significant main effect of Attention on the total number of reported

crime-related details, $F(1,126) = 28.00$, $p < .001$, $\omega^2 = .174$, but there was no main effect of Cues,

$F(2,126) = .06$, $p = .940$, $\omega^2 = -.014$. No significant Attention by Cue interactions emerged for

reported critical details, $F(2,126) = .51$, $p = .600$, $\omega^2 = -.008$. Finally, there was a significant

main effect of Attention, $F(1,126) = 28.48$, $p < .001$, $\omega^2 = .176$, but not Cues, $F(2,126) = .05$, $p =$

.954, $\omega^2 = 0.014$, on the rate of completeness of participants' accounts. The interaction between

Attention and Cues was not significant for the rate of completeness, $F(2,126) = .44$, $p = .647$, ω^2

$= -.009$. Means for reported critical details as a function of cue and attention conditions are

presented in Table 3.

Table 3

Mean number (SE) of reported critical details by cues (Self-Generated Cues, Other-Generated cues, No Cues) and attention (Full and Divided).

|	**Reported details**					
	SGC		OGC		NC	
Attention	*M (SE)*	95%CI	*M (SE)*	95%CI	*M (SE)*	95%CI
Full	13.1 (0.2)	[12.2, 14.1]	12.7 (0.2)	[11.8, 13.5]	12.9 (0.2)	[11.9, 14]
Divided	10.3 (0.2)	[9.3, 11.6]	10.9 (0.3)	[9.6, 12]	10.2 (0.3)	[8.8, 11.6]

Type of Details Reported

There was a main effect on the total number of person details for Attention, $F(1,126) = 14.55$, p

$< .001$, $\omega^2 = .095$, with more person details reported under full than divided attention. There was

also a main effect of Cues, $F(2,126) = 4.91$, $p = .009$, $\omega^2 = .057$. Post-hoc tests showed that more

person details were reported overall with SGC than with No Cues (p= .011), but not compared to the Other-Generated Cues condition (p = .061). There was also no significant difference in the number of person details reported in the Other-Generated Cues condition in comparison to the No Cues condition (p = 1.00). No significant interaction emerged for the total number of person details, $F(2,126) = 1.40$, $p = .251$, $\omega^2 = .006$. Bonferroni-corrected pairwise comparisons revealed that the use of Self-Generated Cues led to the reporting of more person details comparing to the use of Other-Generated Cues ($p = .039$) and of No Cues ($p = .005$), under the Full attention condition. However, there was no difference between cues under Divided attention conditions ($p > .05$).

There was a main effect of Attention, $F(1,126) = 8.64$, $p = .004$, $\omega^2 = .056$, but not of Cues, $F(2,126) = .24$, $p = .788$, $\omega^2 = -0.011$, on the total number of object details reported. There was no significant interaction between Cues and Attention, $F(2,126) = 1.32$, $p = .272$, $\omega^2 = .005$. Similarly, there was a main effect of Attention, $F(1,126) = 15.57$, $p < .001$, $\omega^2 = .102$, but not of Cues, $F(2,126) = .03$, $p = .966$, $\omega^2 = -0.015$, on the total number of action details reported. The interaction between Attention and Cues was not significant, $F(2,126) = 1.01$, $p = .366$, $\omega^2 = .000$. Levene's test was significant for the analysis of action details ($p = .03$). Finally, there was no effect of Attention, $F(1,126) = .62$, $p = .434$, $\omega^2 = -.003$ or Cue, $F(2,126) = 2.86$, $p = .061$, $\omega^2 = .028$, on the total number of setting details reported. Levene's test was significant ($p = .005$). No significant interaction emerged for the reporting of setting details, $F(2,126) = .70$, $p = .499$, $\omega^2 = -.005$. Boxplots were used to explore the distribution for the total number of both action and setting details. For action details, the distribution was symmetrical however there were seven outliers representing participants who reported a high number of action details. For setting details, the distribution was not symmetrical but positively skewed with three outliers who

reported a high number of setting details. Given the low number particularly regarding setting details ($M = 6.88$, $SD = 3.58$), and the lack of significant results for both type of details, no action was taken due to the Levene's test being significant. The effect of cues on the mean number of person details within Full and Divided attention conditions are presented in Figure 1. Means for action, object and setting details reported within both attention conditions are presented in Tables 4a and 4b.
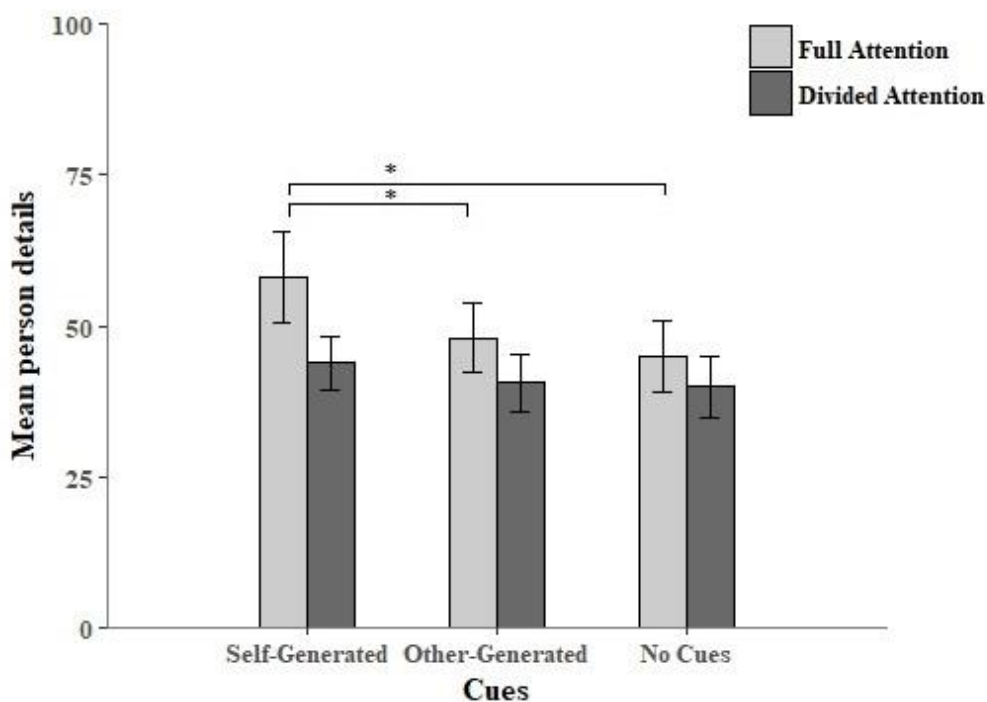
Figure 1. Mean number of person details as a function of cues (Self-Generated Cues vs Other-Generated Cues vs No Cues) within Full and Divided attention conditions. Error bars represent ± 1.96 standard errors (95% confidence intervals). Asterisks indicate significant differences between cue conditions, * $p < .05$.

Table 4a

Mean (SE) number of action, object and setting details by cues (Self-Generated Cues, Other-Generated Cues, No Cues) under Full attention.

| Details type | **Full Attention** | | | | | |
| | SGC | | OGC | | NC | |
	M (SE)	95%CI	*M (SE)*	95%CI	*M (SE)*	95%CI
Action	18.4 (0.6)	[15.5, 21.3]	18.1 (0.8)	[14.6, 21.6]	16.3 (0.6)	[13.8, 19.1]
Object	10 (0.3)	[8.7, 11.3]	9.2 (0.3)	[7.8, 10.7]	9.2 (0.3)	[8.1, 10.4]
Setting	8.4 (0.4)	[6.3, 10.4]	7.2 (0.4)	[5.5, 9.1]	5.8 (0.2)	[4.9, 6.7]

Table 4b

Mean (SE) number of action, object and setting details by cues (Self-Generated Cues, Other-Generated Cues, No Cues) under Divided attention.

| Details type | **Divided Attention** | | | | | |
| | SGC | | OGC | | NC | |
	M (SE)	95%CI	*M (SE)*	95%CI	M (SE)	95%CI
Action	12.3 (0.4)	[10.5, 14.2]	12.9 (0.4)	[10.8, 14.9]	14 (0.5)	[11.6, 16.5]
Object	7.2 (0.2)	[6, 8.4]	8.6 (0.2)	[7.5, 9.8]	7.7 (0.4)	[6, 9.5]
Setting	7.3 (0.3)	[6, 8.5]	6.3 (0.2)	[5.3, 7.3]	6.3 (0.3)	[4.9, 7.9]

Appendix B: Enhanced Accuracy Instructions (Chapter 3)

In this study, you witnessed an event as an undercover agent. You are in possession of valuable intelligence information, so you now need to provide a full report. Keep in mind that this information will be passed on to our analysts for assessment. To find out more about what you know, you will be asked a series of questions based on the account you provided.

Your responses to these questions should be as accurate and detailed as you can provide.

Specifically, your responses should reflect the level or degree of detail you feel you can report accurately.

What do we mean by this?

For example, if you had witnessed someone breaking into a car then in answer to the question "Tell me about the car to which the perpetrator attempted to force entry", you might not remember anything or you might recall some very broad or general details such as remembering that the car was "small and dark coloured".

Alternatively, you might have a very clear memory of the car and be able to report very precise or fine details such as the specific colour, size and exact model of the car (e.g. a Volkswagen Golf, British Racing Green, 5-door hatchback, tinted windows, registration number etc.).

This is what we mean by **level of detail (general or precise).**

Instructions

Often witnesses choose not to report broad or general information (such as "…" in the example above). Here we are specifically interested in **everything** you can actually remember – and the differing level of details (general versus specific) you are able to report.

You should provide **all** the information you believe to be accurate – it doesn't matter whether that information is detailed or not.

Do you understand what we want you to do?

If not, please ask the researcher.

Practice question

Now it's time to put this into practice. Without looking, what can you remember about what footwear the researcher in the room with you is wearing?

If you can remember only general details about the shoes (e.g. they were dark coloured) then write those details down.

If you can remember the specific details such as the <u>exact colour(s)</u> (e.g. black with silver and cherry red trims and white laces) and <u>type of shoe</u> (e.g. NIKE-AIR trainer) then you should write all of these details down instead.

However, <u>you might not be able to recall any details at all</u> – if this is the case and you really don't remember any details at all (either general information or detailed information):

- **Do not guess.**

- **Feel free to refuse to answer the question.**

Remember to provide **all** the information you believe to be **accurate**.

Appendix C: Supplemental materials (Chapter 4)

Description of model stages

A model that included the fixed effects (Reporting Format and Event Content) with random intercepts for events nested within participants (hereby referred to as the random intercept model) was compared to a model that only included the fixed effects.

Total reporting (after asking follow-up questions). The relationship between Format and Content showed significant variance in intercepts across participants with respect to the reporting of mean correct details, $SD = 11.60$, $\chi^2(1) = 23.45$, $p < .001$; the reporting of mean correct gist details, $SD = 2.18$, $\chi^2(1) = 37.43$, $p < .001$; and the reporting of mean correct verbatim details, $SD = .65$, $\chi^2(1) = 51.77$, $p < .001$.

Initial reporting phase. The relationship between Format and Content showed significant variance in intercepts across participants with respect to the reporting of mean correct details, $SD = 10.32$, $\chi^2(1) = 33.36$, $p < .001$; the reporting of mean correct gist details, $SD = 1.83$, $\chi^2(1) = 66.04$, $p < .001$; the reporting of mean correct verbatim details, $SD = .59$, $\chi^2(1) = 54.16$, $p < .001$; and the reporting of mean correct attributions of actions and statements to persons, $SD = 1.75$, $\chi^2(1) = 176.61$, $p < .001$.

Source monitoring errors. The relationship between Format and Content showed significant variance in intercepts across participants with respect to the average reporting of intrusions, $SD = 0.33$, $\chi^2(1) = 231.05$, $p < .001$.

Results

Memory for the order of events

A chi-square analysis was conducted to examine whether reporting of the correct order in which events were witnessed was associated with reporting format or event content. Results showed that there was no significant association between the order reported for events and format conditions, $\chi^2(8) = .656$, $p = .999$, or content conditions, $\chi^2(4) = 3.417$, $p = .491$, in the reporting of the witnessed events in the correct order (i.e., as witnessed). A total of 73 out of 150 participants reported all four events in the correct order. Tables 1a and 1b show the total number of participants reporting events in correct order across format and content conditions respectively.

Table 1a

Number of events reported in the order they were witnessed across format conditions

Events in correct order	Format			
	MMIF	Timeline	FR	Total
No events (-4)	2	2	2	6
One event (-3)	7	7	6	20
Two events (-2)	17	14	10	41
Three events (-1)	2	4	4	10
Four events (- 0)	22	23	28	73
Total	50	50	50	150

Table 1b

Number of events reported in the order they were witnessed across content conditions.

| | Content | | |
Events in correct order	Typical	Changed	Total
No events (-4)	2	4	6
One event (-3)	10	10	20
Two events (-2)	25	16	41
Three events (-1)	4	6	10
Four events (- 0)	34	39	73
Total	75	75	150

Reporting of accuracy rates for gist reporting across reporting phases

A separate univariate analysis of variance was conducted on the mean gist accuracy rates for information reported after asking follow-up questions. Neither Format, $F(2,144) = .10$, $p = .905$, $\omega^2 = -.01$, nor Content, $F(1,144) = 1.27$, $p = .262$, $\omega^2 = .002$, had a main effect on the mean gist accuracy rate for reported information. The interaction between Format and Content was not significant, $F(2,144) = 2.81$, $p = .064$, $\omega^2 = .024$.

A repeated-measures analysis of variance showed that there was no difference in the accuracy rate of gist details between the follow-up questioning phase and the initial reporting phase, $F(1,96) = .51$, $p = .477$, $\omega^2 = -.01$. Accuracy in the follow-up questioning phase was not affected by reporting format, $F(1,96) = 2.14$, $p = .147$, $\omega^2 = .011$. A univariate analysis of variance showed that there were no significant main effects of Reporting Format, $F(2,144) = 2.40$, $p = .094$, $\omega^2 = .019$, or Event Content, $F(1,144) = .12$, $p = .730$, $\omega^2 = -.006$, on accuracy rate of the initially reported gist details. The interaction between Format and Content was not significant either, $F(2,144) = 1.67$, $p = .192$, $\omega^2 = .009$.

Reporting of verbatim details and accuracy rates for total reporting

Neither Event Content, b = .16, CI [-0.02, 0.34], $t(144)$ = 1.79, p = .076, nor Reporting Format

significantly predicted the reporting of correct verbatim details. Likelihood of reporting of

correct verbatim details was the same in the Free recall and the Timeline conditions, b = .03, CI

[-0.25, 0.19], $t(144)$ = .27, p = .788, and in the MMIF condition, b = .09, CI [-0.13, 0.30], $t(144)$

= .78, p = .437. Participants in the Timeline and MMIF conditions were also equally likely to

report correct verbatim details, b = .11, CI [-0.33, 0.10], $t(144)$ = 1.05, p = .295. The interaction

between Content and Format was not significant: i) Free recall vs Timeline, b = .14, $t(144)$ = .64,

p = .523; ii) Free recall vs MMIF, b = .19, $t(144)$ = .89, p = .375; iii) Timeline vs MMIF, b = .05,

$t(144)$ = .23, p = 820.

A univariate analysis of variance examining the accuracy rate of verbatim details reported

after asking follow-up questions showed that neither Reporting Format, $F(2,144)$ = .36, p = .698,

ω^2 = -.01, nor Event Content, $F(1,144)$ = .002, p = .964, ω^2 = -.01, had a main effect on the

overall accuracy of verbatim details. The interaction between Format and Content was not

significant, $F(2,144)$ = .90, p = .409, ω^2 = -.001.

A repeated-measures analysis of variance showed that there was no difference in the

accuracy rate of verbatim details between the follow-up questioning phase and the initial

reporting phase, $F(1,94)$ = .82, p = .368, ω^2 = -.002. Mean accuracy rates and SDs for gist and

verbatim details across format conditions for initial and follow-up reporting can be found in

Table 2.

Reporting of verbatim details for initial reporting

Neither Event Content, $b = .16$, CI [-0.16, 0.47], $t(144) = 1.91$, $p = .058$, nor Reporting Format

significantly predicted the reporting of correct verbatim details. Participants in the Free recall

condition were equally likely to report correct verbatim details to participants in the Timeline

condition, $b = .08$, CI [-0.31, 0.47], $t(144) = .80$, $p = .425$, and in the MMIF condition, $b = .06$,

CI [-0.33, 0.45], $t(144) = .60$, $p = .549$. The likelihood of reporting was the same for participants

in the Timeline and MMIF conditions, $b = .02$, CI [-.18, .22], $t(144) = .20$, $p = .842$. The

interaction between Content and Format was not significant: i) Free recall vs Timeline, $b = .00$,

$t(144) = .00$, $p = 1.00$; ii) Free recall vs MMIF, $b = .10$, $t(144) = .50$, $p = .618$; iii) Timeline vs

MMIF, $b = .00$, $t(144) = .00$, $p = 1.00$.

A univariate analysis of variance showed that neither Format, $F(2,144) = .35$, $p = .705$, ω^2

$= -.005$, nor Content, $F(1,144) = 2.97$, $p = .087$, $\omega^2 = 0.01$, had a main effect on the accuracy rate

of verbatim details reported in the initial phase. The interaction between Format and Content was

not significant, $F(2,144) = .85$, $p = .430$, $\omega^2 = -.002$.

Table 2

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

| | Initial report | | | | Follow-up | | | |
| | Gist | | Verbatim | | Gist | | Verbatim | |
Format	Mean Accuracy rate	*SD*	Mean Accuracy rate	*SD*	Mean Accuracy rate	*SD*	Mean Accuracy rate	*SD*
MMIF	.88	.12	.09	.09	.83	.18	.06	.09
Timeline	.85	.18	.09	.09	-	-	-	-
FR	.81	.18	.08	.09	.84	.21	.08	.12

Appendix D: Ethical Approval (Chapters 2, 3 and 4)

Each of the four experiments presented in this thesis received ethical approval from the University of Portsmouth's Science Faculty Ethics Committee (SFEC). Below, the letter of favourable opinion is attached for each of the experiments reported in this thesis.

Faculty of Science

University of Portsmouth

Science Faculty Ethics Committee
Science Faculty Office
University of Portsmouth
St Michael\s Building
White Swan Road
PORTSMOUTH
PO1 2DT

Feni Kontogianni
Department of Psychology
University of Portsmouth

feni.kontogianni@port.ac.uk

T: 023 9284 3379
ethics-sci@port.ac.uk

25 November 2015

FAVOURABLE ETHICAL OPINION

Study Title: Introducing a new mnemonic to the timeline technique: retrieval support for suboptimally encoded events.

Reference Number: SFEC 2015-087

Date Resubmitted: 19 November 2015

Thank you for resubmitting your application to the Science Faculty Ethics Committee (SEFC) for ethical review in accordance with current procedures[1], and for making the requested changes following the first SFEC review, and for the clarifications provided.

I am pleased to inform you that SFEC was content to grant a favourable ethical opinion of the above research on the basis described in the submitted documents listed at Annex A, and subject to standard general conditions (*See Annex B*).

Please note that the favourable opinion of SFEC does not grant permission or approval to undertake the research. Management permission or approval must be obtained from any host organisation, including the University of Portsmouth or supervisor, prior to the start of the study.

Wishing you every success in your research

Dr Jim House
Vice-Chair Science Faculty Ethics Committee

Annexes

A - Documents reviewed
B - After ethical review - Guidance for researchers

---

[1] Procedures for Ethical Review, Science Faculty Ethics Committee, University of Portsmouth, October 2012 (to be updated).

Faculty of Science

University of **Portsmouth**

Science Faculty Ethics Committee
Science Faculty Office
University of Portsmouth
St Michael's Building
White Swan Road
PORTSMOUTH
PO1 2DT

Ms Feni Kontogianni
Department of Psychology
University of Portsmouth

T: 023 9284 3379
ethics-sci@port.ac.uk

feni.kontogianni@port.ac.uk

25th Aptil 2016

FAVOURABLE ETHICAL OPINION WITH MINOR CONDITIONS

Study Title: **Using a multi-modal interviewing format to facilitate memory recall for repeated events**
Reference Number: **SFEC 2016-036** (Please quote this in any correspondence)

Thank you for submitting your application to the Science Faculty Ethics Committee (SFEC) dated 5th April in accordance with current procedures[1].

I am pleased to inform you that SFEC was content to grant a favourable ethical opinion of the above research on the basis described in the submitted documents listed at Annex A, and subject to standard general conditions[2] and the following minor conditions/recommendations:

1. The Universities data retention policy states that data should be kept for 20 years, or if the study is being conducted through external funding, researchers must abide by the funding bodies retention policy.

There is no requirement for you to confirm these conditions have been met in writing to the committee.

Please note that the favourable opinion of SFEC does not grant permission or approval to undertake the research. Management permission or approval must be obtained from any host organisation, including the University of Portsmouth or supervisor, prior to the start of the study.

Wishing you every success in your research

Yours sincerely,

Dr Chris Markham
Chair, Science Faculty Ethics Committee

---

[1] Procedures for Ethical Review, Science Faculty Ethics Committee, University of Portsmouth, October 2012 (to be updated).
[2] After ethical review – Guidance for researchers (Please read).

Science Faculty Ethics Committee
Science Faculty Office
University of Portsmouth
St Michael's Building
White Swan Road
PORTSMOUTH
PO1 2DT

Feni Kontogianni
Department of Psychology
University of Portsmouth

T: 023 9284 3379
ethics-sci@port.ac.uk

Feni.Kontogianni@port.ac.uk

20 July 2017

FAVOURABLE ETHICAL OPINION

Study Title: Using a multi-modal interviewing format to facilitate memory recall for repeated events

Reference Number: SFEC 2017-074

Date Submitted: 30 June 2017

Thank you for submitting your proposal to the Science Faculty Ethics Committee (SEFC) for ethical review in accordance with current procedures.

I am pleased to inform you that SFEC was content to grant a favourable ethical opinion of the above research on the basis described in the submitted documents listed at Annex A, and subject to standard general conditions (*See Annex B*).

Please note that the favourable opinion of SFEC does not grant permission or approval to undertake the research. Management permission or approval must be obtained from any host organisation, including the University of Portsmouth or supervisor, prior to the start of the study.

Wishing you every success in your research.

Dr Paul Morris
Vice Chair, Science Faculty Ethics Committee

Annexes

A - Documents reviewed
B - After ethical review - Guidance for researchers

Information:

Eva Rubinova - Co-Investigator
Professor Lorraine Hope - PhD Supervisor

Appendix E: UPR16 Form

FORM UPR16
Research Ethics Review Checklist

Please include this completed form as an appendix to your thesis (see the
Research Degrees Operational Handbook for more information

**UNIVERSITYof
PORTSMOUTH**

Postgraduate Research Student (PGRS) Information						Student ID:	782775	
PGRS Name:	Foteini Kontogianni							
Department:	Psychology		**First Supervisor:**		Professor Lorraine Hope			
Start Date: (or progression date for Prof Doc students)			1-10-2015					
Study Mode and Route:		Part-time	☐	MPhil	☐	MD		☐
		Full-time	☒	PhD	☒	Professional Doctorate		☐

Title of Thesis:	Eliciting information from cooperative sources about single & repeated multi-actor events
Thesis Word Count: (excluding ancillary data)	37.672

If you are unsure about any of the following, please contact the local representative on your Faculty Ethics Committee for advice. Please note that it is your responsibility to follow the University's Ethics Policy and any relevant University, academic or professional guidelines in the conduct of your study

Although the Ethics Committee may have given your study a favourable opinion, the final responsibility for the ethical conduct of this work lies with the researcher(s).

UKRIO Finished Research Checklist:
(If you would like to know more about the checklist, please see your Faculty or Departmental Ethics Committee rep or see the online version of the full checklist at: http://www.ukrio.org/what-we-do/code-of-practice-for-research/)

a)	Have all of your research and findings been reported accurately, honestly and within a reasonable time frame?	YES NO	☒ ☐
b)	Have all contributions to knowledge been acknowledged?	YES NO	☒ ☐
c)	Have you complied with all agreements relating to intellectual property, publication and authorship?	YES NO	☒ ☐
d)	Has your research data been retained in a secure and accessible form and will it remain so for the required duration?	YES NO	☒ ☐
e)	Does your research comply with all legal, ethical, and contractual requirements?	YES NO	☒ ☐

Candidate Statement:

I have considered the ethical dimensions of the above named research project, and have successfully obtained the necessary ethical approval(s)

Ethical review number(s) from Faculty Ethics Committee (or from NRES/SCREC):	SFEC 2015-087, 2016-036, 2017-074
If you have *not* submitted your work for ethical review, and/or you have answered 'No' to one or more of questions a) to e), please explain below why this is so:	
N/A	
Signed (PGRS):	Date: 28-09-2018

UPR16 – April 2018