Who said what and when? Using a timeline technique to elicit information and

intelligence about conversations, plots and plans

Lorraine Hope

University of Portsmouth

Fiona Gabbert

Goldsmiths, University of London

Max Kinninger

University of Portsmouth

Feni Kontogianni

University of Portsmouth

Alexandra Bracey

University of Portsmouth

Adam Hanger

University of Portsmouth

Correspondence concerning this article should be addressed to Lorraine Hope, Department of Psychology, University of Portsmouth, King Henry I Street, Portsmouth PO1 2DY, UK.
Email: Lorraine.Hope@port.ac.uk

Abstract

The verbal content of interactions (*what was said* and *who said what)* can be important as evidence and intelligence information. Across three empirical studies, we examined memory for details of an overheard (Experiment 1) or witnessed (Experiments 2 and 3) conversation using a timeline technique adapted for the reporting of conversations between multiple speakers. Although participants in all conditions received the same general instructions, participants assigned to timeline reporting format reported more verbatim information and made fewer sequencing errors than those using a free recall format. In Experiments 2 and 3, using an extended version of the technique, participants using the timeline reporting format also reported more correct speaker attributions and provided more information about the individuals involved, without compromising accuracy rates. With a large effect size across experiments, these findings suggest that timeline reporting formats facilitate the reporting of episodic memories and benefit the reporting of conversations.

Public Significance Statement

In criminal investigations and human intelligence gathering contexts obtaining reliable information about the verbal content of criminal interactions (*what was said* and *who said what*) may be critical. We found that mock witnesses who reported using a self-administered timeline format reported more information overall, more verbatim information, made fewer sequencing errors, and were more likely to correctly identify 'who said what' than those participants who used a free recall reporting format. These findings contribute to the work of practitioners in investigative and intelligence gathering contexts by facilitating the elicitation of detailed accounts, including conversational details.

Who said what and when? Using a timeline technique to elicit information and intelligence about conversations, plots and plans

In both criminal investigations and intelligence gathering contexts, obtaining reliable information from witnesses and sources is critical. To date, the research literature has largely focused on eliciting information about *what happened.* However, the verbal content of interactions (*what was said* and *who said what)* can be equally important as evidence and intelligence. First, there are a number of criminal and civil case types that can revolve around the content of interpersonal verbal interactions, for example, verbal and sexual harassment, bullying, stalking and bribery. Second, crimes may take place under conditions where only overheard information is available because the witness was blindfolded, hooded, hiding or otherwise unable to see the perpetrators (e.g. hostage-taking). Third, intelligence gleaned from clandestine groups or terrorist organizations by undercover sources may take the form of recalled accounts of meetings and conversations during which the details of plots or other plans were discussed. However, despite repeated calls for focus on this topic, memory for conversations remains the "orphan child of witness memory researchers" (Davis & Friedman, 2007, p.3). Furthermore, there have been few attempts to develop or adapt interviewing techniques to enhance the recall and reporting of *who said what* and *when did they say it*. Across three experiments, the current research aims to address this shortcoming, using the timeline technique (Hope, Mullis, & Gabbert, 2013) to elicit information about overheard conversations and the people involved in them.

Memory for Conversations

Failures of memory for conversations have been well documented, from Neisser's forensic analysis of John Dean's account of conversations with President Nixon

during the Watergate scandal (Neisser, 1981) to Brian Williams' entirely false memory for conversations with soldiers in Iraq (see Rechdan, Sauerland, Hope, & Ost, 2016) and, more recently, Neil Degrasse Tyson's misrecollections of George Bush's speech to Congress after the 9/11 attacks (see Chabris & Simons, 2014). Beyond these examples of quite dramatic error, research on memory for everyday conversations demonstrates a tendency to report the gist of such conversations in free recall tests (Campos & Alonso-Quecuty, 2006; Stafford & Daly, 1984; Stafford, Burggraf, & Sharkey, 1987). Gist memory for conversations reflects the meaning of content, usually in an inferred or synopsized form while verbatim memory for surface details of the conversation, such as the exact words or phrases used, is very rarely reported in free recall (Miller, deWinstanley, & Carey, 1996; Stafford et al., 1987; Stafford & Daly, 1984; Stafford, Waldron, & Infield, 1989; see also Corbin, Reyna, Weldon, & Brainerd, 2015). Because of our tendency to recall gist, errors can occur when people infer that something was actually said when it was only implied (e.g. Harris, 1978). Furthermore, the quantity of information recalled about conversations tends to be low (Stafford et al., 1987; see also Hjelmquist & Gidlund, 1985) although performance in recognition tests suggests that verbatim memory for conversations can be accessed (e.g. MacWhinney, Keenan, & Reinke, 1980).

In the forensic context, the findings of a small literature on memory for *criminal conversation* illustrates both the transformation of such conversations into gist-based free recall reports and the superior recall of the gist of criminal conversations (cf. verbatim recall). For example, motivated by a high profile sexual harassment case, Pezdek and Prull (1993) examined memory for utterances in an audiotaped dialogue which included sexually explicit target sentences. At short delays, participants were able to discriminate between the original and new lure sentences, showing better

performance for target versus control sentences. Verbatim memory was better for sexual than non-sexual sentences at short delays but memory performance declined over longer delays. Similarly, in a test of earwitness memory for a criminal conversation concerning a planned theft, Campos and Alonso-Quecuty (2006) noted that when participants were asked to provide a free recall they tended to report the gist of the conversation and only few verbatim details. They also confirmed the importance of modality noting that participants who had been exposed to the conversation in audio-visual mode (vs. auditory only) showed superior gist recall, particularly after a delay (cf. auditory only).

While there are different approaches to accounting for memory performance when recalling conversations, Campos and Alonso-Quecuty (2006, 2008) have drawn on Fuzzy-Trace Theory (FTT; Brainerd & Reyna, 2004, 2002; Reyna & Brainerd, 1995) to account for their findings. In short, according to FTT, correct recall is underpinned by both gist and verbatim memory traces, but gist traces are likely to be preferentially accessed and reported (Reyna & Brainerd, 1995). While this account provides a useful explanatory framework for accounting for deficits in free recall performance, it also offers a tantalizing suggestion as to how the reporting of conversational details might be enhanced. Specifically, the FTT proposes that the test format can bias the recollection of gist or verbatim information and the extent to which verbatim or gist traces are accessed is dependent on the nature of cues present at retrieval (Brainerd & Reyna, 2004). Cues that reinstate a surface form of the original experience facilitate access to verbatim information (e.g. cued recall, recognition) while cues that reinstate meaning are more likely to facilitate access to gist information (e.g. unstructured free recall; e.g. Brainerd, Wright, Reyna, & Payne, 2002; Henry & Gudjonsson, 2004; Holliday, 2003).

To date, research has failed to capitalize on this conceptualization of memory in order to elicit more informative details of a conversation. Instead, research on the elicitation of information about conversations has focused on determining the efficacy of existing investigative interviewing approaches. For instance, Campos and Alonso-Quecuty (2008) examined whether the Cognitive Interview (CI; Fisher & Geiselman, 1992) enhances the reporting of conversational details. Using a slightly modified version of the CI, Campos and Alonso-Quecuty (2008) found that the CI elicited more correct information about an audio-visually presented conversation than a standard free recall interview. Verbatim recall was also slightly improved relative to the standard interview although, consistent with previous research, gist recall of the conversation predominated. Prescott, Milne, and Clarke (2011) also found that participants interviewed using the CI provided significantly more information about a witnessed event including an incidental conversation than mock witnesses interviewed using a modified structured interview. However, there was no benefit of the CI for the recall of verbatim information about the conversation (see also Öhman, Eriksson, & Granhag, 2013). As such, research to date broadly confirms that a high quality interviewing approach can elicit more information about conversations. However, research has not explored whether an alternative retrieval format might (a) provide more effective cues to enable participants to access verbatim details, and (b) capitalize on the structure of a conversation during retrieval. Furthermore, research has not systematically examined the ability of witnesses to correctly attribute statements to specific individuals in multi-speaker (> 2) contexts.

One information elicitation approach that dispenses with the standard interview format involving an interviewer requesting a verbal linear narrative from an interviewee, which usually occurs in response to a request to "Tell me everything that

happened", is the Timeline Technique (Hope et al., 2013; see Hope & Gabbert, 2018).

This approach involves a self-administered recall and reporting technique designed to

enhance an interviewee's ability to recall information from a particular time period in

sequence, and to identify the people involved and link those people with their specific

actions. The timeline technique attempts to capitalise on the notion that episodic

memory is temporally ordered and temporal context plays an important role in the

retrieval process. Tulving (1983) argued that information in episodic memory is

associated with the temporal-spatial context in which it was encoded. The basic

memory literature also demonstrates that temporal context plays an important role in

the retrieval process during free recall (Howard & Kahana, 1999; Kahana, 1996;

Unsworth, 2008). Indeed, the fact that items encoded in close temporal proximity tend

to be recalled in close proximity has led researchers to conclude that temporal

clustering of items is a "ubiquitous property" of sequence recall (Polyn, Norman, &

Kahana, 2009, p.130; see also Kahana, Howard, & Polyn, 2008). Although only a

small number of articles have reported the use of timeline formats to facilitate

episodic retrieval for adult witnesses or informants (Hope et al., 2013; Leins et al.,

2014; Kontogianni, Hope, Vrij, Taylor, & Gabbert, 2018), timeline-type methods

have been used previously as a survey methodology to elicit information about

autobiographical events (Belli, 1998; Belli, Bilgen, & Al Baghal, 2013; Belli,

Stafford, & Alwin, 2009; Van der Vaart, 2004), including experiences of violence

(Yoshihama, Gillespie, Hammock, Belli, & Tolman, 2005).

    In the first test of timeline interviewing for an episodic event reported by Hope

et al. (2013), participants provided their account of a witnessed event on a 'timeline'

of the relevant time period for the target event. Additional retrieval support was

provided through the use of instructions and interactive reporting materials. Testing

also included a comparison of component elements of the timeline technique (i.e. instructions, reporting cards, visual timeline; Experiment 2) but optimal performance was observed when the complete timeline format was used. In sum, mock witnesses who provided their accounts about a multi-perpetrator event using a timeline technique provided more (i) person-description details (ii) person-action details and (iii) sequence details than when using a free report format, at no cost to accuracy, with an overall large effect size for correct information reported. In the current paper, we examine whether these benefits of reporting using a timeline approach might accrue for the reporting of a conversation. Like any other episodic event, conversations occur within a particular time-frame (Lu, Harter, & Graesser, 2009) and, as such, witnesses can identify a start point, a finish point and a sequence of information, including who said what, when, and to whom.

There is also another reason to predict that the timeline reporting format may benefit the recall of conversational details, including attribution and sequencing of information. Given the prevalence of text and in-app messaging (e.g., SMS, Instant Messaging, WhatsApp) and availability of online formats for synchronous and asynchronous communication, communicating via real-time text chat is commonplace (Flanagin, 2005; Gergle, Millen, Kraut, & Fussell, 2004, Quan-Haase, 2008). Over 15 million texts are sent every minute of every day worldwide (Domo, 2017). Thus, it is possible that visualising a conversation along a timeline is now intuitive for many people. For instance, text messages often represent 'conversations' in a vertical time-stamped format with each side of the screen showing the messages from each party to the conversation. One rationale for representing interactions in this way is the preference for linguistic co-presence and visual co-presence in an interaction (i.e. what is said and who said it; e.g. Kraut, Gergle, & Fussell, 2002) which is also

associated with conversational efficiency (Gergle, Millen, Kraut, & Fussell, 2004). Given this preferred, and likely familiar structure for representing conversations, the current study examined whether a timeline reporting format reflecting this structure might facilitate recall of conversations.

Current Experiments

Across three empirical studies, we examined memory for details of an overheard (Experiment 1) or witnessed (Experiments 2 and 3) conversation using a version of the Timeline Technique developed by Hope et al. (2013) adapted for the reporting of an overheard conversation [for a visual of the adapted timeline format, see Supplemental Materials; for a visual of the original timeline technique see Kontogianni et al., 2018]. To achieve this, we revised the original timeline format in two ways. First, we oriented the timeline vertically. A vertical conversation flow best matches the structure of the written conversation formats common in text, mobile and online settings and also reflects traditional formats representing conversations, such as scripts or dialogues. Pilot testing of both formats for the reporting of a conversation suggested that mock witness participants found the vertical format more intuitive and easy to understand. Second, we replaced the person and action cards of the original timeline technique with speech record cards on which each recalled element of the conversation could be reported. In Experiment 1, we examined the information reported about an overheard conversation between two perpetrators where no additional visual information was available. Witness-participants either provided their account using the adapted timeline or provided a free recall account. In Experiment 2, witness-participants in the experimental condition viewed a film of conversation involving several gang members and provided an account using an extended version of the timeline that also enabled them to provide descriptions of the target individuals.

Finally, in Experiment 3, witness-participants, in the role of undercover officers acting as members of a gang, witnessed three different filmed conversations between gang members on three separate occasions over a seven-day period and, in the experimental condition, provided an account of these conversations using the timeline to provide their report.

General Methodology

Participants. Participants aged between 18-45 years who were native English speakers qualified to participate in this research. Our participants were predominantly white British citizens. Across the three experiments, conducted over a period of approximately 24 months, participants were recruited from student and community-based samples via online platforms and local advertising. The experiments were approved by the Department of Psychology Ethics Committee at the University of Portsmouth.

Instructions. General instructions about the mock witness task and the need to provide a detailed and accurate account were held constant between reporting conditions in all experiments; i.e. only the instructions for the reporting formats differed [see Supplemental Materials for timeline reporting instructions]. In each experiment, the control comparison group responded to a free recall request for information about the conversation they had heard/seen. Of course, an interview format incorporating any form of retrieval support is likely to elicit more information than a free recall. However, given that (a) previous research suggests that no other format tested to date strongly outperforms a free recall for details (especially verbatim details) of a conversation, and; (b) both the timeline and free recall formats can be administered without adding the potential confounding factor of interviewer-behaviour, we retained the simplest comparison in the experimental design.

Coding. The data were coded by three coders trained by the first author. The coding approach drew on coding practices described elsewhere in the literature for the quantification of details reported from memory (e.g. Hope, Mullis & Gabbert, 2013; Kontogianni, Hope, Vrij, Taylor & Gabbert, 2018). Detailed Coding Protocols were prepared for each stimulus event and these protocols specified the details for coding in each script (Experiments 1, 2 and 3) and other details (e.g. person descriptions; Experiments 2 and 3). Across experiments, in the first phase of coding, data were coded for the number of correct and incorrect details reported. Each detail reported was given one point and scored as either correct or incorrect with reference to the stimulus event. Number sequences (e.g. phone number, safe code) correctly recalled were given one point per number correct and an extra point was awarded if the numbers were in the right sequence. In the second phase of coding, a verbatim statement was coded as one point if three or more words were exactly the same as the sentence heard in the conversation (there were two exceptions to this rule in the case of two word sentences or lone words in quotations). For the coding of gist statements, a point was award for the correct extraction a 'gist' of a conversation element (i.e. gist level correct information that was not reported in verbatim detail). Sequence errors were coded by assessing whether information was reported in the wrong order. For example, if the correct sequence for what occurred was ABCD and a participant reported this information in ACBD order then C would count as a sequence error as it is the first detail reported out of sequence. Note we would not code both C and B as sequence errors in this example, only C. In Experiments 2 and 3, speech attributions were coded as correct when a statement was correctly attributed to a specific speaker (e.g. "The male wearing the green sweatshirt said 'Let's meet at 4pm'"). Each person description detail provided was coded as correct or incorrect against detailed

description coding template. Additionally, each detail provided concerning the leadership and relationships between the members of the group was allocated one point. For Experiment 3, the coding protocol also incorporated the occurrence of internal intrusions across events (i.e. reporting details from one event as part of another event) although as no intrusions were observed this issue will not be discussed further. Vague or subjective information was not coded (e.g. "he sounded angry"; "he looked ugly"). [See Supplemental Materials for sample coding templates].

Accounts were randomly selected for inter-reliability coding. A fourth coder independently coded around 15 per cent of accounts for each experiment. As coders were inevitably aware of the reporting condition (although not necessarily the specific hypotheses) due to the different reporting formats, the data were not tallied or otherwise aggregated until all data had been coded. Accuracy rates were calculated once all coded data had been collated in a spreadsheet by dividing the total correct number of items by total items reported (correct and incorrect).

Hypotheses. Consistent with previous research, we predicted that participants in timeline reporting conditions would report (i) overall more correct details about the target conversations; (ii) more verbatim details of the conversations; and (iii) more correct information about who said what during the conversation. We also predicted that there would be fewer sequential errors in the accounts using the timeline format (cf. free recall accounts). All statistical tests were performed with a preset alpha of .05. Where homoscedasticity was an assumption of a statistical test, Levene's (1960) test for equality of variance was assessed and corrected values reported as necessary.

Experiment 1

13

In Experiment 1, mock witness participants listened to an audio-only recoding of a conversation between two male bank robbers. On the grounds that a timeline format capitalizing on both the structure of a conversation and memorial benefits accruing from a temporal format, we predicted that participants in the timeline reporting condition would report more verbatim information than control participants and, consistent with previous timeline research (e.g. Hope et al., 2013) make fewer sequencing errors.

<div align="center">Method</div>

Design and Participants

In a between subjects design, 44 participants (30 female), aged 18-39 years old ($M =$ 21.68, $SD = 4.39$) were randomly allocated to either the Timeline (Timeline; n = 22) or Free Recall (FR; n = 22) reporting format conditions and took part either for course credit (student participants) or on a voluntary basis (non-student participants). The dependent measures were the number of overall correct and incorrect details reported, verbatim and gist details, sequence errors and accuracy rate.

Materials

      Stimulus conversation. A criminal conversation taking place during a bank robbery was scripted and recorded. Two male actors took the roles of the perpetrators committing the robbery to create a conversation lasting 272 secs. The scripted conversation included numerous details of potential forensic value such as safe codes, names of contacts, locations, vehicles etc. At the end of the conversation, the perpetrators discussed the getaway plan and escaped through a fire exit [see Supplemental Materials for conversation script].

      Conversation Timeline Technique. The conversation timeline interview was adapted from the timeline technique reported in Hope et al. (2013). The timeline task

had two elements: (i) A physical 'timeline' comprising a length of cardboard measuring 33 inches x 12 inches with a line running across the mid-point to each end and (ii) lined speech record cards comprising small cards measuring 1 inch x 3.5 inches.

Procedure

Participants were randomly assigned to condition and tested individually. Prior to hearing the criminal conversation, all participants were given instructions that asked them, when listening to the conversation, to put themselves in the position of someone being held hostage in the bank where the robbery was taking place. They were asked to imagine that they were in a location (e.g. under a desk) where they could not see the perpetrators but could clearly hear their conversation. After clarifying that they understood the task and their instructions, participants listened to the pre-recorded conversation through a headset. After listening to the conversation, participants completed unrelated filler tasks for 15 mins. Participants were then asked to provide a detailed report about what they had heard during the 'robbery'. In the timeline condition participants were given the timeline format to structure their report of the overhead conversation. Although the timeline was initially placed vertically, participants were informed they were free to orient the timeline as they wished (all participants placed the timeline in a vertical orientation). They were asked to use the speech cards to record what they remembered about the conversation using one card per statement, and place the cards on the timeline in the order in which the conversation had occurred. They were further instructed to place the speech cards for the different perpetrators on the left and right sides of the conversation timeline respectively and advised that where the attribution of the statement to a specific perpetrator was not possible they should just place these cards in the centre of the

timeline. In the free recall condition, participants were provided with an A4 booklet with blank pages in which to write their account of the overhead conversation in the order in which they heard it, noting which perpetrator had said what. Importantly, participants in both conditions received the same general recall instructions requiring them to report as much detailed information as they could about the overhead conversation, and instructing them to report verbatim statements if they could remember exactly what was said. These instructions emphasised the importance of reporting the conversation in the right order and attributing statements to individuals if possible. Participants were also instructed to avoid guessing. No time restrictions were imposed in either reporting condition. Inter-rater agreement for all data across the first phase of coding was Kappa = .79 and verbatim-gist coding was Kappa = .68.

Results

There was no significant difference between conditions for the overall number of correct, $t(42) = 0.48$, $p = .63$, $d = 0.15$, 95% CI [-.45, 0.74] or incorrect, $t(42) = 0.57$, $p = .57$, $d = 0.17$, 95% CI [-0.42, 0.76] details reported about the overheard conversation between the perpetrators. There was no difference between conditions in the overall accuracy rate of information provided, $t(42) = 0.14$, $p = .89$, $d = 0.04$, 95% CI [-0.55, 0.63] and nor was there any difference between conditions for the number of gist statements reported, $t(42) = 1.08$, $p = .29$, $d = 0.32$, 95% CI [-0.27, 0.92]. However, participants in the timeline condition reported significantly more verbatim statements correctly than participants in the free recall condition, $t(42) = -3.37$, $p = .002$, $d = 1.01$, 95% CI [0.38, 1.64]. Participants in the timeline condition also made fewer sequencing errors when reporting details of the overheard conversation, $t(33) = -2.53$, $p = .016$, $d = -.76$, 95% CI [-1.37, -0.15], see Table 1.

Table 1 here

Discussion

The main aim of Experiment 1 was to conduct an initial test of the timeline technique as a method for eliciting information about conversation exchanges. Although the timeline method did not elicit significantly more information overall about the content of the conversation, participants in the timeline reporting condition reported over twice the amount of correct verbatim details about the conversation at no cost to accuracy. These participants also made fewer sequencing errors in their account of the conversation.

Previous attempts have often failed to improve verbatim recall of conversations (e.g. Prescott et al., 2007). However, by adapting a timeline format in the current study, we made a number of important changes to the retrieval context that may underpin the gains observed – even though participants in both conditions received the same instruction to report as much verbatim detail as possible. First, consistent with predictions of FTT (Brainerd & Reyna, 2004), the timeline format may promote a more 'surface' form of the original experience than a more unstructured retrieval attempt and facilitate the reporting of verbatim information. Second, the 'surface' match between the original overheard conversation and the reporting format may have also been enhanced by common representational formats for conversations e.g. SMS texts, Instant Messaging, and other online formats.

However, one aspect of conversational remembering we were unable to address in Experiment 1 was whether the timeline format improves the attribution of certain statements to certain speakers ('Who said what?'). In the absence of visual cues or information about the two male protagonists, participants appeared to find the task of discriminating between the speakers and attributing statements difficult and,

although some statements were attributed, others were simply placed on in the middle of the timeline. Therefore, in Experiment 2, we extended the timeline technique to facilitate the differentiation of each speaker in the conversation.

Experiment 2

The 'Who said what?' question is critical in many intelligence gathering and evidentiary contexts, for example, to identify which member of a terrorist cell gave certain orders or to confirm which gang member made direct threats. Yet research to date has not focused on the attribution of particular statements to particular speakers in recall. Further, research has not addressed whether the recall of this speaker-statement link might be improved through the use of alternative information elicitation approaches. The main aim of Experiment 2 was to examine whether the timeline format also facilitated the attribution of speech to specific speakers. We were particularly interested in whether, as in Hope et al. (2103), examination of person-action links, the correct attribution of speech to specific individuals was possible when multiple speakers contributed to a conversation.

These aims necessitated a number of methodological changes to the timeline interview procedure to facilitate the reporting of information about the speakers. In the original version of the timeline technique, participants were provided with person description cards on which they were instructed to provide detailed descriptions of the target individuals. Thus, in Experiment 2, we introduced a speaker description component to the conversation timeline format used in Experiment 1. We hypothesized that this revised format would (i) produce more detailed descriptions of target speakers, and (ii) facilitate the attribution of speech to speaker. Further, we hypothesized that participants reporting details of a conversation using this format

would report more information overall, more verbatim information, and also make fewer sequencing errors than participants in the free recall condition. We located the context of the research in an intelligence gathering setting; all participants were instructed that they were in the role of an undercover officer with the task of reporting on the plans discussed at the meeting of a terrorist group.

Method

Design and participants

In a between subjects design, 40 participants (30 female) aged 18 – 28 years (M = 20.29, SD = 2.76) were randomly allocated to either the enhanced Timeline or Free Recall reporting format conditions and took part either for course credit (student participants) or on a voluntary basis (non-student participants). The dependent measures were the number of correct and incorrect details reported (overall, verbatim and gist details), speaker attributions, sequence errors and accuracy rate.

Materials

Briefing. A short 'pre-deployment briefing' informed all participants that they were working as an undercover officer tasked with infiltrating a group of activists. The briefing advised that the activists had invited the undercover officer, who they believed to be a member of their group, to an important meeting. Participants were informed that they would now attend that meeting and their task was to report back everything they could remember about what was discussed.

Stimulus Conversation. A stimulus event involving a detailed planning conversation between five members of a violent activist group was scripted, recorded and edited into a short film (412 secs). The event was filmed from a first person perspective in order to immerse the viewer into the role of an undercover officer. The

'viewer' entered a room in which there were five other people. One member of the group initiated the meeting and the group discussed a future attack they planned to carry out. The conversation contained detailed planning information including how the attack would take place, timings, names of relevant buildings and structures, vehicle details, which explosives will be used, how the explosives will be detonated etc. The roles and responsibilities of each group member with respect to planning and conduct of the attack were also discussed [see Supplemental Materials for conversation script].

Timeline Technique. The timeline technique used in Experiment 1 was extended to include instructions for the reporting of person descriptions. Drawing on previous research (e.g. Gabbert et al., 2009), a person description reporting form was prepared which, using neutral prompts and cues (e.g. age? hair? eyes? build? clothing? any unique features?) instructed participants to provide as much information as they could about each group member.

Procedure

Participants were randomly assigned to condition and tested individually. At the outset, participants were given the pre-deployment briefing informing them that, acting in the role of an undercover officer, they would be attending a meeting of an activist group and that their role was to remember as much as they could about this meeting. After clarifying that they understood the task and their instructions, participants watch the meeting video and listened to the associated audio via a headset. After watching the video, all participants completed unrelated filler tasks for 15 mins. Participants were then asked to provide a detailed report about what they had heard during the meeting.

In both the timeline and free recall condition, the first part of the task asked

participants to provide descriptions of the five individuals they had encountered in the meeting. The general instructions for this task were exactly the same in both conditions and asked participants to report as many details as they could about each of the individuals attending the meeting, including description cues (i.e. what clothes they wore, colour hair, how they spoke, personality, did anything stand out about them? etc.). Instructions requested participants to provide as much details as possible but to avoid guessing. Participants in the timeline condition provided their descriptions using person description forms while participants in the free recall condition completed a free recall description for each individual following the same general instructions. Before commencing the main recall task, participants in both conditions were given different coloured adhesive markers and instructed that they should allocate a colour to each group member they had described and then use a coloured marker to reference this individual during the main recall task of reporting what was said during the meeting. Then all participants were instructed to report everything they could remember about content of the meeting. Again, the general instructions were the same in both conditions and, as in Experiment 1, asked participants report as much detailed information as they could about what was said during the meeting, instructing them to report verbatim statements if they could remember exactly what was said. For both conditions, the instructions also emphasised the importance of reporting the conversation in the right order and attributing statements to individuals where possible. Participants were instructed to avoid guessing. Participants in the timeline condition then reported their account using the timeline, attributing different statements to the people described using the coloured markers. Participants in the free recall condition reported their accounts in the free recall booklet using the coloured markers in the same way to link the

individuals described to their account. No time restrictions were imposed in either reporting condition. Inter-rater agreement for all data across the first phase of coding was Kappa = .82 and for verbatim-gist coding was Kappa = .65.

Results

Information provided in the person description task and in the timeline report was combined to establish the overall performance of the interview formats in terms of the total amount of information elicited. There was a significant difference between conditions for the total number of correct details reported with participants in the timeline condition reporting more correct details than participants in the free recall condition, $t(38) = 3.85$, $p < .001$, $d = 1.22$, 95% CI [0.53, 1.89]. There was no difference between conditions for the number of incorrect details reported, $t(38) = 1.73$, $p = .09$, $d = 0.55$, 95% CI [-0.09, 1.17] or the overall accuracy rate, $t(38) = 1.11$, $p = .27$, $d = 0.35$, 95% CI [-0.27, 0.97]; see Table 2.

Table 2 here

Conversation Details: Verbatim, Gist and Attribution[i]

Participants in the timeline condition reported significantly more correct verbatim details than free recall participants, $t(36) = 3.11$, $p = .004$, $d = 1.01$, 95% CI [0.32, 1.68]. There was no difference in the number of correct gist details reported between conditions, $t(38) = 0.68$, $p = .50$, $d = 0.22$, 95% CI [-0.41, 0.83]. In line with predictions, participants in the timeline condition correctly attributed more statements to the correct speaker than participants in the free recall condition, $t(21) = 2.62$, $p = .01$, $d = .83$, 95% CI [0.17, 1.47]. Consistent with previous results, participants in the timeline condition made fewer sequence errors, $t(30) = -2.23$, $p = .03$, $d = -0.71$, 95%

22

CI [-1.34, -0.06].

Person Description Details

The significant information gain in the timeline condition derived largely from the detailed person descriptions, with participants in the timeline condition providing almost double the number of correct person details, $t(38) = 4.56, p < .001, d = 1.44$, 95% CI [0.73, 2.13]). Although there were more description errors in the timeline group, ($t(38) = 2.94, p = .006, d = 0.93$, 95% CI [0.27, 1.58]), this did not result in a cost to overall accuracy rate, $t(38) = 0.44, p = .66, d = 0.14$, 95% CI [-0.48, 0.76].

Discussion

In Experiment 2, participants who provided their account of the conversation that took place between five people using the timeline format reported significantly more correct information, and more correct verbatim information, than those who provided free recall reports. Experiment 2 extended the proof-of-concept work conducted in Experiment 1 by permitting an examination of the attribution of statements to specific speakers. As predicted, timeline participants made more correct speech-to-speaker attributions and fewer sequence errors than free recall participants

In terms of overall information gain, the person description format completed by participants in the timeline condition bolstered the amount of information reported, even though participants in the free recall condition received the same general person description instruction (although it should also be noted that there was an increase in the number of person description errors reported in the timeline condition). However, even when the person description information is discounted, the information gain remained in the timeline reporting condition for verbatim and speaker attribution details. Given that we did not see an overall information gain for the timeline

condition in Experiment 1, but this was evident in Experiment 2, it may be that additional retrieval support in the person description form used in the experimental condition worked to promote the recall of additional conversation information. It is also likely that having visual access to the scene (cf. Experiment 1) meant that witnesses accessed significantly more information about the interaction that prompted recall of the conversation.

Although participants in the timeline condition reported more verbatim information than free recall participants, the actual number of details reported in both conditions was small (and much fewer than reported in Experiment 1). Again, this may be due to the nature of the stimuli used. In Experiment 1 the only information available to the mock witnesses was the audio of the conversation whereas in Experiment 2 participants also had visual input from the scene. The number of speakers contributing to the conversation in Experiment 2 was also greater than in Experiment 1 (five versus two). Thus, perceptual load in terms of visual input and cognitive load in terms of a larger number of speakers to monitor likely placed additional demands on processing resources, resulting in an increased reliance on gist (see, for example, Raveh & Lavie, 2015). Nonetheless, the timeline reporting format facilitated the reporting of more information about the conversations, more correct speaker attributions and fewer sequencing errors cf. the control condition.

Although the results for the enhanced conversation timeline technique are promising, the immediacy of testing in Experiment 2 (only 15 minute delay between exposure to the target conversation and reporting recall of that conversation) is a shortcoming, particularly with respect to the application of the technique in applied contexts where delay is likely to be significantly longer. Therefore, one aim of Experiment 3 was to examine recall for conversations after a period of delay. Further,

and relevant to the applied context, we extended the timeline technique to facilitate the reporting of several conversations involving multiple speakers.

Experiment 3

One challenging feature of eliciting information in policing and security contexts is that an interviewee may have information about multiple similar events. In police investigation contexts, such information may concern repeated instances of domestic violence, stalking or harassment – or even information about repeated criminal activity. In intelligence contexts, it may concern repeated meetings with other people or groups (e.g. see Rivard, Fisher, Robertson, & Hirn Mueller, 2014) or more extensive reports about longer time periods involving similar routines or activities e.g. training at a terrorist camp. In an intelligence-gathering context, eliciting accurate information in debriefings or interviews about what was said in such interactions is likely to be particularly important, particularly if the interactions involve planning future activities or provide additional information about existing networks or how individuals are organised in terms of hierarchy or relationships in cells or gangs.

However, although there is a sizeable body of developmental research examining differences in the reporting of memory for unique and repeated experiences (e.g. Connolly & Lindsay, 2001; Price & Connolly, 2013) and how best to interview children in such contexts (e.g. Brubacher, Powell, & Roberts, 2014), little research has examined adult memories for multiple or repeated similar events or techniques to maximise recall and reporting in this context. Indeed, Willen, Granhag, Stromwall, and Fisher (2015) identified only five previous studies that had investigated how adult recall of multiple similar events can be particularized (Cohen & Java, 1995; Leins, Fisher, Pludwinski, Rivard, & Robertson, 2014; Means, Nigama,

Zarrow, Loftus, & Donaldson, 1989; Means & Loftus, 1991; Philips & Fisher, 1998; Rivard, et al., 2014).

Even if the interviewee is entirely cooperative, it may be that repeated events are represented in memory by a schema (e.g. I always met X at the café on Harbour Street at 11am on Fridays) and unique event details may be difficult to recall (e.g., Bartlett, 1932; Farrar & Goodman, 1990; Hudson, Fivush, & Kuebli, 1992). In their study examining recall of dental visits over an extended period, Willen et al. (2015) noted the benefits of context-specific cueing to promote particularization in the reporting of specific visits. This finding resonates with the suggestion by Leins et al. (2014) that a timeline that includes specific cues, such as relevant temporal markers, might well be beneficial in intelligence gathering contexts for capturing discrete events. Therefore, in the current study we adapted the timeline technique to facilitate the reporting of three conversations observed at separate meetings of the same gang. The three 'meetings' took place over the course of one week and recall for all three events was tested after a one week delay. We predicted that participants in the timeline condition would, as in Experiment 2, provide significantly more information overall about the 'meetings' attended than those in the free recall condition, including more correct speaker attributions and verbatim details.

Method

Design and participants

Fifty-one participants were recruited but one participant failed to attend the final test session. Thus, in a between subjects design, 50 participants (35 female) aged 18 – 32 years ((M = 21.62, SD = 3.18) were randomly allocated to either the enhanced timeline (n= 25) or free recall (n = 25) reporting format conditions. An additional

recruitment criterion in Experiment 3 was that participants were unfamiliar with the TV show 'Sons of Anarchy'. All participants were paid £10 on completion of all four test sessions. The dependent measures were the number of overall correct and incorrect details reported, verbatim and gist details, speaker attribution details, sequence errors and accuracy rate.

Materials

Stimulus Events. Three short film clips ($M = 115.7s$, $SD = 15.3s$) of the American TV-show 'Sons of Anarchy' were selected as suitable stimuli for testing recall of similar repeated events. Each of these clips displayed a different group meeting of the same biker gang. The clips were selected against a number of criteria, including viewpoint (i.e. to replicate the sense of participant involvement as a infiltrator), accessible detail (i.e. the key information presented had to be self-contained and not cross-reference other unseen events) and similarity of nature and length of discussion. Clips were pilot tested to confirm they met these criteria. In the first clip (136 secs), five members of the gang discussed an upcoming police investigation and potential for distracting the police with an artificial crime scene. In the second clip (112 secs), three members discussed a potential traitor in their gang who had been arrested by the police. In the third clip (99 secs), the gang discussed how they can increase their income and which industries might offer suitable opportunities. The order of presentation of the film clips was randomized across participants [see Supplemental Materials for conversation scripts].

Timeline Technique. Participants reported descriptions of the individuals involved and then reported their recall for details of the meeting on the timeline using the same basic reporting materials as used in Experiments 1 and 2 with one minor exception. Instead of using coloured markers as in Experiment 2 participants were

given different coloured cards to record their recall of the statements for each person (which they could then link with the relevant descriptions).

Procedure

Participants were randomly assigned to condition and tested individually. At the outset, participants were given a pre-deployment briefing informing them that, acting in the role of an undercover officer, they would be attending three meetings of a criminal gang they had infiltrated and that their role was to remember as much as they could about these three meetings so that they could report who said what and when, as well as any other relevant information in detail after attending all three meetings. They were instructed to report as much detail as possible about the group conversations and the individuals involved. Participants were informed that they would attend three sessions during which they would 'attend a meeting of the gang' (i.e. watch one film clip in each session) and they would be asked to report on all of these events in detail at an interview one week later. All participants watched the video clips on three separate days during one week and arranged a return visit to the laboratory for the following week (interval between last stimulus event and recall session: $M = 118.78$ hours, $SD = 31.71$).

After the delay all participants were interviewed for their recall of the three events. Participants in both reporting conditions were given the same general recall instructions and asked to provide detailed descriptions of each person involved in the meetings. Participants in the timeline condition provided these descriptions using the person description form while control participants followed free recall instructions to provide person descriptions, as in Experiment 2. As an extension to the person description task designed to elicit information about the hierarchy within the group of target individuals, participants in both conditions were asked if they could identify

who the leader was and what the relationships were between the different gang members. Participants in both conditions were given the same general instruction that they should think about the people they had just described and asked to consider what the relationships were between them in order to describe their role or position in the group (e.g. leader, organiser). Participants in the timeline condition were additionally instructed to complete this task by forming a spider diagram or mind-map on a blank sheet of paper and describe the group hierarchy and the links between its members (see 'family tree technique' described by Leins et al., 2014).

After completing the person description task, all participants were asked to identify the meetings in the order in which they experienced them and then, provide a report of what had occurred in each meeting. As before, participants in both conditions received the same overall recall instructions encouraging them to report as much and as detailed information as possible about each meeting while refraining from guessing. Participants in the timeline condition then reported their account using the timeline technique as in Experiment 2. Participants in the free recall condition reported their accounts in the free recall booklet. A separate account (using the timeline or free recall format) was made for each event. No time restrictions were imposed in either reporting condition. Inter-rater agreement for all data across all data across the first phase of coding was Kappa = .83 and for verbatim-gist coding was Kappa = .69.

Results

There was a significant difference between conditions for the total number of correct details reported across the three events with participants in the timeline condition reporting more correct details than participants in the free recall condition, $t(39) =$

5.45, $p < .001$, $d = 1.57$, 95% CI [0.89, 2.18]. There was no difference between conditions for the number of incorrect details reported, $t(46) = 1.59$, $p = .12$, $d = 0.46$, 95% CI [-0.12, 1.03] or for the overall accuracy rate, $t(29) = 1.21$, $p = .24$, $d = 0.35$, 95% CI [-0.22, 0.93]; see Table 3.

Table 3 here

Conversation Details: Verbatim, Gist and Attribution

Participants in the timeline conditions reported significantly more correct verbatim details than free recall participants, $t(30) = 2.28$, $p = .03$, $d = 0.64$, 95% CI [0.63, 1.12]. Participants in the timeline condition also reported significantly more gist details, $t(48) = 2.16$, $p = .04$, $d = 0.61$, 95% CI [0.04, 1.17]. Consistent with predictions, participants in the timeline condition correctly attributed more statements to the correct speaker than participants in the free recall condition, $t(37) = 3.79$, $p = .001$, $d = 1.07$, 95% CI [0.47, 1.66]. The number of sequence errors overall was low with no difference between conditions, $t(48) = -0.54$, $p = .59$, $d = 0.15$, 95% CI [-0.71, 0.40]. Timeline participants also provided significantly more correct information about the network hierarchy than free recall participants, $t(47) = 2.35$, $p = .02$, $d = 0.67$, 95% CI [0.09, 1.25].

Person Description Details

Timeline participants reported significantly more correct descriptive details about the gang members, $t(41) = 7.12$, $p < .001$, $d = 2.01$, 95% CI [1.32, 2.69]. There was a significant difference between conditions in terms of person description errors with timeline participants reporting more erroneous details, $t(48) = 3.47$, $p < .001$, $d = 0.98$, 95% CI [0.39, 1.56]. However, there was no difference between the conditions in terms of accuracy rates for person descriptions, $t(32) = 0.12$, $p = 0.90$, $d = 0.03$, 95% CI [-0.52, 0.59].

Discussion

Participants assigned to using the timeline technique reported significantly more information than participants who provided free recall reports about the target events after a one-week delay. This information gain using the timeline technique was not associated with a cost to accuracy – although we note that participants in the timeline condition did make more person description errors (discussed below). Further, as in Experiment 2, participants using the timeline reported more verbatim details and made more correct speaker attributions than participants in the comparison group.

It may be that for repeated events over a period of time, the timeline was particularly beneficial due to a close match between the structure of these memories and the reporting structure. Anderson and Matessa (1997) proposed that for recall of autobiographical events, several distinct information units at a basic level are assembled in chronological groups on a higher level. These groups are again collated under one general, extended event (see also Conway, 1996; Conway & Bekerian, 1987; Conway & Rubin, 1993). In the current study, the reporting procedure for participants in the timeline condition mapped this structure in that participants first had to recall the general event that occurred the previous week ('undercover mission') and then the sub-events within that general event ('different meetings') and then finally report the conversation details specific to each meeting ('statements made in conversation'). Although memories will have been organised similarly for participants in the free recall condition, the 'narrative' reporting procedure inherent to free recall likely did not support retrieval to the same extent as the timeline. The reporting structure of the timeline may have also allowed witnesses to capitalise on the temporal clustering of items and accrue benefits for information associated with temporal clusters (Howard & Kahana, 1999; Kahana, 1996; Polyn, Norman, &

Kahana, 2009). As it was not the aim of the current research to determine the structure of memory for multiple events and how it relates to reporting format, this interpretation is speculative. However, an interesting direction for future research would be to further explore the notion of capitalizing on memory structure when eliciting information – particularly for complex or multiple events or events that took place sometime in the past.

We did not observe particularization problems pertaining to internal intrusions or evidence of any tendency to confuse content of the three different conversations which may be due to the nature of the meetings. Although the context of the witnessed meetings (i.e. meeting location, key gang members present) was consistent for all three meetings, the meetings were relatively short and the nature of topics discussed in each meeting was different. Future research should examine recall performance when there is greater overlap between the conversation topics across meetings, as more source monitoring errors would be expected (see Lindsay, 1990; Lindsay, Allen, Chan, & Dahl, 2004).

General Discussion

The main aim of the research reported here was to develop and test a novel reporting method to assist the recall of overheard conversations. In three experiments, we examined the recall of conversation details, incrementally extending the initial technique to elicit further information about speakers including descriptions and relationships within groups. In Experiment 1, participants using the conversation timeline reported over twice as much verbatim detail about an overheard conversation than participants asked to provide a free recall account. In Experiment 2, we extended the technique to include descriptions of the speakers and facilitate the attribution of

statements to the relevant speakers in a group conversation setting. Here participants in the timeline condition reported more information overall about the speakers and gist and verbatim details of the conversation. Participants using the timeline also made more correct speaker attributions and fewer sequencing errors. These findings were replicated in Experiment 3, which tested recall of multiple conversations between group members after a delay of one week.

Meta-Analyses BETTER HEADING?

To quantify effects across the three experiments for our primary variables, we conducted mini meta-analyses using a fixed effects approach (Goh, Hall & Rosenthal, 2016; see https://osf.io/6tfh5/). The meta-analytic effect sizes for total correct details and verbatim statements were large and significant. For gist statements, the meta-analytic effect size was small and significant. For total incorrect details, the meta-analytic effect size was small and significant. This result is notable as total incorrect details did not differ significantly between conditions in any experiment. A limitation of the current set of experiments is likely low statistical power relating to limited sample sizes. However, we note that for overall accuracy rates, the meta-analytic effect was small but was not significant suggesting no overall difference in accuracy rates between the reporting conditions (see Table 4).

Table 4 here

Recall of Conversations

Participants using the timeline technique to report their memories provided more information about the conversations they heard/observed than participants who reported via free recall. As such, these data may broadly support the notion that test format can bias the recollection of gist or verbatim information depending on the

nature of cues present at retrieval (Brainerd & Reyna, 2004). One possible interpretation is that the timeline format reinstates a surface form of the original experience and, in doing so, facilitates greater access to verbatim information. The timeline reporting method could be considered to reflect a surface form of conversations in that (i) the temporal sequence of the conversation is clearly represented and (ii) the timeline format replicates other 'conversation' formats (vertical, speaker attributions, temporally ordered) such as those commonly in use in digital and social media. Having said that, there was a sizeable discrepancy between the proportion of verbatim detail reported as a function of the total conversation details reported (i.e. verbatim and gist) across experiments. In Experiment 1, verbatim details accounted for 26% of information reported about the target conversation in the timeline condition but only 11% and 5% in Experiments 2 and 3, respectively. This finding is also consistent with results reported by Campos & Alonso-Quecuty (2006) It is likely this seeming discrepancy reflects the nature of the stimuli and resultant differences in attention and associated cognitive load as the conversations comprised audio only in Experiment 1 but audio-visual in Experiments 2 and 3. This explanation would also be appear to align with the increase in gist reporting seen in Experiments 2 and 3 (cf. Experiment 1) suggesting that as more visual information became available at encoding, participants resorted to a normative gist style of reporting for conversations. Our rationale for making this methodological change was to advance from a relatively simplistic target conversation to a more complex and ecologically valid group conversation setting. As such, manipulating modality lay outside our research aims. However, in light of these results, future work should explore differences of recall for conversations encountered by different or multiple

modalities (e.g. via telephone, Skype) and written 'conversations' via social media e.g. online chatrooms or forums, instant messages.

Another aspect of conversational remembering deliberately not targeted in the current work is recall of conversation in which the rememberer has contributed to the conversation directly and interacted with the other speakers. Holding participant contribution as speaker constant, in this case making no contribution, allowed us full control of the conversation details encoded without idiosyncratic variations that might occur in a naturalistic interaction. However, future work should examine the memorial effects of contribution on both ability to remember a conversation one has contributed to and how performance might interact with different recall elicitation formats. There is evidence to suggest that conversation partners have better memory for their own contributions particularly when they perceive the topic to be important (Miller, deWinstanley, & Carey, 1996). Thus, active involvement appears to increase memory, possibly as a result of generation effects (e.g. Slamecka & Graf, 1978), production effects (e.g. Ozubko, Gopie, & MacLeod, 2012) or enactment effects (e.g. Engelkamp, 1998).

Active involvement also has implications for source monitoring with research on conversation recall typically showing impaired external source monitoring to be associated with active involvement (see e.g. Jurica & Shimamura, 1999; Koriat, Ben-Zur, & Druch, 1991; Mulligan, 2004, 2011). Koriat et al. (1991) suggest that active involvement impairs source monitoring as a result of reduced context integration due to a focus on one's own information output. However, when people are receivers of information (input), context integration is enhanced as contextual features accompanying the information are also absorbed (e.g. Jurica & Shimamura, 1999; see also Gopie & MacLeod, 2009). Using ecologically valid conversation dyads, Fischer,

Schult, and Steffens (2015) observed this trade-off in memory for conversations such that context is remembered better for input events, while information is remembered better for output invents. Given this interesting role of context to prompt additional recall for conversation, it may be the case that participants in timeline conditions who consistently reported more correct speaker attributions benefited particularly from the nature of the reporting context. Future research needs to determine whether the benefits of the reporting format are retained when the rememberer has been actively involved in the conversation.

These results replicate the beneficial effect of timeline reporting observed by Hope et al. (2013) and Kontogianni et al. (2018) and, indeed, other research using different timeline formats to elicit information (e.g. Leins et al., 2014; van der Vart & Glasner, 2007). While the current research is not designed to extricate an explanatory account of why this benefit accrues for reports made in a non-traditional format, it is likely that the format facilitates the retrieval of closely associated details, including details of conversations, potentially as a result of capitalising on associative retrieval processes such as temporal contiguity (Estes, 1955; Howard & Kahana, 2002). These effects emerge when "retrieved contextual states overlap with the encoding context of nearby items" (Howard, Kahana, & Polyn, 2008, p.25). The timeline reporting format, by virtue of letting the interviewee clearly see what related information has been reported, may make this overlap more explicit and, hence, beneficial for cuing recall. Of course, this suggestion remains speculative in the absence of direct testing.

Beyond any benefits accrued from the timeline format, it is likely also the case that the additional retrieval tasks in Experiments 2 and 3 (describing the group members) prior to reporting the conversation details may have also cued further recall. These additional tasks were incrementally included as part of the reporting task

across experiments as our goal was to develop a holistic reporting method that would also elicit relevant contextual information (e.g. details about speakers). However, drawing on the retrieval practice literature (e.g. Roediger & Butler, 2011), future work might consider both the nature and timing of such tasks to maximise recall in the substantive phase of the interview.

It is also worth noting that the provision of additional retrieval support can be associated with a pattern of increased error reporting. Across experiments, the meta-analytic effect size was mid-way between the small and medium conventions and significant, $d = .39$, $Z = 2.23$, $p = .01$, 95% CI [0.05, 0.73] suggesting an overall pattern of more incorrect details being reported in the timeline condition. However, as noted above, this is offset by results indicating no overall difference in accuracy rates between the reporting conditions. This pattern of increased errors is particularly marked for Person Description details in Experiments 2 and 3 although, again, the magnitude of correct information gained (approximately twice as many details) offsets any absolute difference in overall accuracy rates. Similar patterns of results for errors and accuracy rates have been documented for other information gathering techniques, such as the Cognitive Interview and may be associated with loosened threshold or criteria for reporting (see meta-analysis by Memon, Meissner, & Fraser, 2010). Practitioners should be aware of the risk of increased errors and, where possible, use warning instructions to emphasize the need to maximise both the accuracy and completeness of reports.

One possible criticism of the current experiments concerns the nature of the comparison group where we compared timeline recall performance with that of free recall performance. For this initial development and extension of the technique, we wanted to assess the performance of the timeline reporting instructions against a basic

request for information where, in both the experimental and control groups, additional questioning by an interviewer was eliminated. Of course, as acknowledged at the outset, providing any form of retrieval support is likely to elicit more information than a free recall. However, it is worth noting that the general instructions about the tasks and the requirement for detailed and accurate information were held constant between conditions. Second, once the instructions were administered, the reporting task in both conditions was entirely self-administered with no further questions or prompts provided by the researchers. Finally, the purpose of developing new techniques for eliciting different information in different contexts is not necessarily to pitch them against effective existing techniques (e.g. Cognitive Interview) but rather to provide professionals charged with eliciting information in different investigative and intelligence gathering contexts with additional tools and techniques to maximise the information obtained.

Another potential criticism is that as we deliberately did not set any time restrictions on reporting, the amount of time spent on reporting likely varied between participants and conditions. Although we did not measure reporting time, it is plausible that the longer someone spends reporting information, the more information they are likely to report. However, eliciting more information from an interviewee is not a question of equating the reporting time but rather providing additional retrieval support to maximize reporting. Across the current experiments, both reporting formats were entirely self-administered. As such, it is informative that the timeline technique format sufficiently engaged compliant interviewees to, overall, provide significantly more information than those interviewees asked to provide a free recall.

Application of timeline reporting methods in investigative and intelligence gathering contexts

The current results are promising and, in conjunction with previous findings, suggest that timeline reporting formats facilitate the reporting of episodic memories and extend to enhancing recall of conversations. Furthermore, and in keeping with Kontogianni et al. (2018) who tested the efficacy of self-generated cues as a mnemonic within the timeline technique, this set of experiments illustrates that other additional mnemonics can be effectively introduced to the basic timeline technique.

Of course, we are not proposing that this conversation timeline technique, or other related timeline formats, constitute the means by which to conduct a comprehensive interview with a witness or an intelligence debriefing with a cooperative source. Like other self-administered formats, this technique may be useful for eliciting detailed initial accounts or accounts about specific target events, uncontaminated by interviewer interjections or agendas, prior to more detailed interrogation of those accounts through the use of additional questioning and follow-up interviewing. A phased approach may be particularly important where the interviewer does not have incident outcome information or is unaware of the full scope of intelligence information the interviewee holds (see Gabbert, Hope, Carter, Boon & Fisher, 2016 for a discussion of the importance of obtaining detailed initial accounts in investigative contexts). To this end, further research is underway to determine the outcomes of follow-up questioning and collation of information across multiple accounts. Echoing the call by Fisher, Schreiber Compo, Rivard, and Hirn (2013), this research contributes to a wider goal of developing flexible and adaptive reporting techniques informed by the memory literature to support the elicitation of detailed and accurate investigative and intelligence information.

References

Anderson, J. R., & Matessa, M. (1997). A production system theory of serial

    memory. *Psychological Review, 104*, 728-748. http://dx.doi.org/10.1037/0033-

    295X.104.4.728

Bartlett, F. C. (1932). *Remembering: An experimental and social study.* New York,

    NY, USA: Cambridge University Press.

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*, 383–406. https://doi.org/10.1080/741942610

Belli, R. F., Bilgen, I., & Al Baghal, T. (2013). Memory, communication, and data

    quality in calendar interviews. *Public opinion quarterly*, *77*, 194-219.

    https://doi.org/10.1093/poq/nfs099

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.

Brainerd, C. J., & Reyna, V. F. (2002). Fuzzy-trace theory and false memory. *Current*

    Directions in Psychological Science, *11*, 164-169. https://doi.org/10.1111/1467-

    8721.00192

Brainerd, C. J., & Reyna, V. F. (2004). Fuzzy-trace theory and memory

    development. *Developmental Review*, *24*, 396-439.

    https://doi.org/10.1016/j.dr.2004.08.005

Brainerd, C. J., Wright, R., Reyna, V. F., & Payne, D. G. (2002). Dual-retrieval

    processes in free and associative recall. *Journal of Memory and Language*, *46*,

    120-152. https://doi.org/10.1006/jmla.2001.2796

Brubacher, S. P., Powell, M. B., & Roberts, K. P. (2014). Recommendations for

    interviewing children about repeated experiences. *Psychology, Public Policy,*

    and Law, *20*, 325-335. http://dx.doi.org/10.1037/law0000011

Campos, L., & Alonso‑Quecuty, M. L. (2008). Language crimes and the cognitive

interview: Testing its efficacy in retrieving a conversational event. *Applied*

Cognitive Psychology, *22*, 1211-1227. doi: 10.1002/acp.1430

Campos, L., & Alonso-Quecuty, M. (2006). Remembering a criminal conversation:

Beyond eyewitness testimony. *Memory*, *14*, 27-36.

https://doi.org/10.1080/09658210444000476

Chabris C. F., & Simons D. J., (2014, December 1). Why our memory fails us. *The*

New York Times. Retrieved from:

http://www.nytimes.com/2014/12/02/opinion/why-our-memory-fails-

us.html?_r=0

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

recall. *Applied Cognitive Psychology*, *9*, 273-288. doi: 10.1002/acp.2350090402

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*, 205-223. doi: 10.1002/1099-

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

Conway, M.A. (1996). Autobiographical knowledge and autobiographical memories.

In D.C. Rubin (Ed.), *Remembering our past: Studies in autobiographical*

memory (pp. 67- 93). New York, NY, USA: Cambridge University Press.

Conway, M.A., & Bekerian, D.A. (1987). Organization in autobiographical memory.

Memory & Cognition, *15*(2), 119-132. https://doi.org/10.3758/BF03197023

Conway, M.A., & Rubin, D.C. (1993). The structure of autobiographical memory. In

A.E. Collins, S.E. Gathercole, M.A. Conway, & P.E.M. Morris (Eds.), *Theories*

of memory (pp.103-137). Hove, UK: Lawrence Erlbaum Associates Ltd.

Corbin, J. C., Reyna, V. F., Weldon, R. B., & Brainerd, C. J. (2015). How reasoning, judgment, and decision making are colored by gist-based intuition: A fuzzy-trace theory approach. *Journal of Applied Research in Memory and Cognition*, *4*, 344-355. https://doi.org/10.1016/j.jarmac.2015.09.001

Davis D., Friedman R. D. (2007). Memory for conversation: the orphan child of witness memory researchers. In M. P. Toglia, J. D. Read, D. F. Ross &, R. C. L. Lindsay (Eds), *Handbook of Eyewitness Psychology: Memory for Events Vol. 1* (pp. 3–52). New York, NY, USA: Lawrence Erlbaum Associates.

Domo (2017). Data never sleeps 5.0. Retrieved from https://www.domo.com/learn/data-never-sleeps-5.

Engelkamp, J. (1998). *Memory for actions*. Hove, England: Psychology Press.

Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological review*, *62*, 369-377.

Farrar, M. J., & Goodman, G. S. (1990). Developmental differences in the relation between scripts and episodic memory: Do they exist? In R. Fivush & J. A. Hudson (Eds.), *Emory symposia in cognition, Vol. 3. Knowing and remembering in young children* (pp. 30-64). New York, NY, USA: Cambridge University Press.

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

Fischer, N. M., Schult, J. C., & Steffens, M. C. (2015). Source and destination memory in face-to-face interaction: A multinomial modeling approach. *Journal of Experimental Psychology: Applied, 21*, 195-204. http://dx.doi.org/10.1037/xap0000046

Flanagin, A. J. (2005). IM online: Instant messaging use among college students. *Communication Research Reports, 22,* 175–187.

Gabbert, F., Hope, L., Carter, E., Boon, R., & Fisher, R. P. (2015). The role of initial witness accounts within the investigative process. In G. Oxburgh, T. Myklebust, T. Grant, & R. Milne (Eds.), *Communication in investigative and legal contexts: integrated approaches from forensic psychology, linguistics and law enforcement.* (pp. 107-132). Wiley-Blackwell.

Gergle, D., Millen, D. R., Kraut, R. E., & Fussell, S. R. (2004, April). Persistence matters: Making the most of chat in tightly-coupled work. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 431-438). New York, NY, USA: ACM. doi: 10.1145/985692.985747

Goh, J. X., Hall, J. A., & Rosenthal, R. (2016). Mini Meta-analysis of your own studies: Some arguments on why and a primer on how. *Social and Personality Psychology Compass, 10,* 535-549.

Gopie, N., & MacLeod, C. M. (2009). Destination memory: Stop me if I've told you this before. *Psychological Science*, *20,* 1492-1499. https://doi.org/10.1111/j.1467-9280.2009.02472.x

Henry, L. A., & Gudjonsson, G. H. (2004). The effects of memory trace strength on eyewitness recall in children with and without intellectual disabilities. *Journal of Experimental Child Psychology*, *89,* 53-71. https://doi.org/10.1016/j.jecp.2004.05.002

Hjelmquist, E., & Gidlund, Å. (1985). Free recall of conversations. *Text-Interdisciplinary Journal for the Study of Discourse*, *5,* 169-186. https://doi.org/10.1515/text.1.1985.5.3.169

Holliday, R. E. (2003). Reducing misinformation effects in children with cognitive

    interviews: Dissociating recollection and familiarity. *Child development*, *74*,

    728-751. doi: 10.1111/1467-8624.00565

Hope, L., Mullis, R., & Gabbert, F. (2013). Who? What? When? Using a timeline

    technique to facilitate recall of a complex event. *Journal of Applied Research in*

    Memory and Cognition, *2*, 20-24. https://doi.org/10.1016/j.jarmac.2013.01.002

Hope, L. & Gabbert, F. (2018). Interviewing victims and witnesses. In N. Brewer &

    A. Douglass (Editors). *Psychology and Law.* Guilford Publications, New York.

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*, 923–941. http://psycnet.apa.org/doi/10.1037/0278-

    7393.25.4.923

Hudson, J. A., Fivush, R., & Kuebli, J. (1992). Scripts and episodes: The development

    of event memory. *Applied Cognitive Psychology*, *6*, 483-505.

    doi: 10.1002/acp.2350060604

Jurica, P. J., & Shimamura, A. P. (1999). Monitoring item and source information:

    Evidence for a negative generation effect in source memory. *Memory &*

    Cognition, *27*(, 648-656. https://doi.org/10.3758/BF03211558

Kahana, M. J. (1996). Associative retrieval processes in free recall. *Memory &*

    Cognition, 24, 103–109. https://doi.org/10.3758/BF03197276

Kahana, M. J., Howard, M. W., & Polyn, S. M. (2008). Associative retrieval

    processes in episodic memory. In H. L. Roediger III (Ed.), *Cognitive*

    psychology of memory. Vol. 2 of Learning and memory: A comprehensive

    reference. Oxford, UK: Elsevier.

Kontogianni, F., Hope, L., Vrij, A., Taylor, P. J., & Gabbert, F. (2018). The Benefits
of a Self-Generated Cue Mnemonic for Timeline Interviewing. *Journal of
Applied Research in Memory and Cognition.*
https://doi.org/10.1016/j.jarmac.2018.03.006

Koriat, A., Ben-Zur, H., & Druch, A. (1991). The contextualization of input and
output events in memory. *Psychological Research*, *53*, 260-270.
https://doi.org/10.1007/BF00941396

Kraut, R. E., Gergle, D., & Fussell, S. R. (2002, November). The use of visual
information in shared visual spaces: Informing the development of virtual co-
presence. In *Proceedings of the 2002 ACM conference on Computer supported
cooperative work* (pp. 31-40). New York, NY, USA: ACM.

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*, 926-935. doi: 10.1002/acp.3041

Levene, H. 1960. "Robust Tests for Equality of Variances". In I. Olkin (Ed.),
Contributions to Probability and Statistics, (pp. 278–92). Palo Alto, CA,
USA: Stanford University Press.

Lindsay, D. S. (1990). Misleading suggestions can impair eyewitnesses' ability to
remember event details. *Journal of Experimental Psychology: Learning,
Memory, and Cognition*, *16*, 1077-1083. http://dx.doi.org/10.1037/0278-
7393.16.6.1077

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*, 96-
111.

Lu, S., Harter, D., & Graesser, A. C. (2009). An empirical and computational

    investigation of perceiving and remembering event temporal

    relations. *Cognitive Science*, *33*, 345-373. 10.1111/j.1551-6709.2009.01016.x

MacWhinney, B., Keenan, J. M., & Reinke, P. (1982). The role of arousal in memory

    for conversation. *Memory & Cognition*, *10*, 308-317.

    https://doi.org/10.3758/BF03202422

Means, B., & Loftus, E. F. (1991). When personal history repeats itself: Decomposing

    memories for recurring events. *Applied cognitive psychology*, *5*, 297-318.

    doi: 10.1002/acp.2350050402

Means, B., Nigam, A., Zarrow, M., Loftus, E. F. & Donaldson, M. S. (1989).

    Autobiographical memory for health-related events. *National Center for Health*

    Statistics. Vital and Health Statistics, 6, 1–37. Retrieved from:

    https://stacks.cdc.gov/view/cdc/6145/cdc_DS1_6145.pdf

Memon, A., Meissner, C. A., Fraser, J. (2010). The cognitive interview: A meta-

    analysis and study space analysis of the past 25 years. *Psychology, Public*

    Policy, & Law 16, 350–372 doi: 10.1037/a0020518.

Miller, J. B., deWinstanley, P., & Carey, P. (1996). Memory for conversation.

    Memory, *4*, 615-631. doi: 10.1080/096582196388816

Mulligan, N. W. (2004). Generation and Memory for Contextual Detail. *Journal of*

    Experimental Psychology: Learning, Memory, and Cognition, 30, 838-855.

    http://dx.doi.org/10.1037/0278-7393.30.4.838

Mulligan, N. W. (2011). Generation disrupts memory for intrinsic context but not

    extrinsic context. *The Quarterly Journal of Experimental Psychology*, *64*, 1543-

    1562. https://doi.org/10.1080/17470218.2011.562980

Neisser, U. (1981). John Dean's memory: A case study. *Cognition*, *9*, 1-22.

https://doi.org/10.1016/0010-0277(81)90011-1

Öhman, L., Eriksson, A., & Granhag, P. A. (2013). Angry Voices from the Past and

Present: Effects on Adults' and Children's Earwitness Memory. *Journal of*

Investigative Psychology and Offender Profiling, *10*, 57-70.

doi: 10.1002/jip.1381

Ozubko, J. D., Gopie, N., & MacLeod, C. M. (2012). Production benefits both

recollection and familiarity. *Memory & Cognition*, *40*, 326-338.

https://doi.org/10.3758/s13421-011-0165-1

Pezdek, K., & Prull, M. (1993). Fallacies in memory for conversations: Reflections on

Clarence Thomas, Anita Hill, and the Like. *Applied Cognitive Psychology*, *7*,

299-310. doi: 10.1002/acp.2350070404

Philips, M. R. & Fisher, R. P. (1998). *Enhancing cooperative suspects' memories of*

crime: A cued recall approach. Poster presented at the 106th annual convention

of the American Psychological Association, San Francisco, CA.

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, 129–156.

Popolov, D., Callaghan, M., & Luker, P. (2000, May). Conversation space:

visualising multi-threaded conversation. In *Proceedings of the working*

conference on Advanced visual interfaces (pp. 246-249). New York, NY, USA:

ACM.

Prescott, K., Milne, R., & Clarke, J. (2011). How effective is the enhanced cognitive

interview when aiding recall retrieval of older adults including memory for

conversation? *Journal of Investigative Psychology and Offender Profiling*, *8*, 257-270. doi: 10.1002/jip.142

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*, 89-94. https://doi.org/10.1016/j.jarmac.2013.03.001

Quan-Haase, A. (2008). Instant messaging on campus: Use and integration in university students' everyday communication. *The Information Society*, 24, 105–115.

Rechdan, J., Sauerland, M., Hope, L., & Ost, J. (2016). Was that how it happened?. *In-Mind Magazine (English Edition)*, *7*(31). Retrieved from: http://www.in-mind.org/article/was-that-how-it-happened-shaping-our-memory-

Raveh, D., & Lavie, N. (2015). Load-induced inattentional deafness. Attention, Perception, & Psychophysics, 77, 483-492. doi:10.3758/s13414-014-0776-2for-personal-experiences-in-conversation-with

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and individual Differences*, *7*, 1-75. https://doi.org/10.1016/1041-6080(95)90031-4

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*, 917-925. doi: 10.1002/acp.3026

Roediger, H. L. & Butler, A. C. (2011). The critical role of retrieval practice in long-term retention. Trends in Cognitive Science, 15, 20-27.

Slamecka, N. J., & Graf, P. (1978). The generation effect: Delineation of a

phenomenon. *Journal of experimental Psychology: Human learning and

Memory*, *4*, 592-604. http://dx.doi.org/10.1037/0278-7393.4.6.592

Stafford, L., & Daly, J. A. (1984). Conversational memory. *Human Communication

Research*, *10*, 379-402. doi: 10.1111/j.1468-2958.1984.tb00024.x

Stafford, L., Burggraf, C. S., & Sharkey, W. F. (1987). Conversational memory the

effects of time, recall, mode, and memory expectancies on remembrances of

natural conversations. *Human Communication Research*, *14*, 203-229.

doi: 10.1111/j.1468-2958.1987.tb00127.x

Stafford, L., Waldron, V. R., & Infield, L. L. (1989). Actor-observer differences in

conversational memory. *Human Communication Research*, *15*, 590-611. doi:

10.1111/j.1468-2958.1989.tb00200.x

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

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*, 223–236. https://doi.org/10.1016/j.jml.2008.04.002

Van der Vaart, W. (2004). The time-line as a device to enhance recall in standardized

research interviews: A split ballot study. *Journal of Official Statistics*, *20*, 301-

317. Retrieved from: http://www.jos.nu/Articles/abstract.asp?article=202301

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*,

227-238. doi: 10.1002/acp.1338

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*, 28-37. doi/10.1111/sjop.12180

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

Table 1.

Experiment 1. Means, standard deviations and 95% confidence intervals for

conversation details by condition.

	Timeline Reporting		Free Recall Reporting	
	Mean (SD)	95% CI	Mean (SD)	95% CI
Correct details	39.23 (14.34)	[32.87, 45.59]	36.91 (17.48)	[29.16, 44.66]
Incorrect details	6.27 (4.09)	[4.46, 8.09]	5.59 (3.84)	[3.89, 7.29]
Accuracy	0.87 (.06)	[0.84, 0.90]	0.87 (.09)	[0.83, 0.90]
Verbatim details**	10.36 (6.16)	[7.63, 13.09]	4.73 (4.87)	[2.57, 6.89]
Gist details	1.91 (2.11)	[0.97, 2.85]	1.32 (1.46)	[0.67, 1.96]
Sequence errors*	5.00 (2.83)	[3.75, 6.25]	8.09 (4.99)	[5.88, 10.30]

*p < 0.05; **p < 0.01

Table 2.

Experiment 2. Means, standard deviations and 95% confidence intervals for reported

details by condition. [See Supplemental Materials for additional descriptives and

associated analyses]

	Timeline Reporting		Free Recall Reporting	
	Mean (SD)	95% CI	Mean (SD)	95% CI
Account Detail				
Correct***	29.00 (9.57)	[24.52, 33.48]	17.98 (8.49)	[14.02, 21.96]
Incorrect	7.33 (3.27)	[5.80, 8.86]	5.63 (2.96)	[4.24, 7.02]
Accuracy rate	0.80 (0.07)	[0.76, 0.83]	0.76 (0.11)	[0.71, 0.81]
Conversation Detail				
Verbatim correct**	1.15 (1.06)	[0.64, 1.66]	0.26 (0.65)	[-0.05, 0.58]
Gist correct	9.35 (6.61)	[6.25, 12.44]	8.10 (4.88)	[5.82, 10.38]
Attributions correct**	4.10 (5.27)	[1.63, 6.57]	0.95 (1.10)	[0.44, 1.46]
Sequence error*	1.15 (1.35)	[0.52, 1.78]	2.55 (2.46)	[1.39, 3.70]
Person Detail				
Correct***	23.70 (8.87)	[19.55, 27.85]	12.08 (1.17)	[8.73, 15.44]
Incorrect**	5.89 (2.99)	[4.49, 7.29]	3.42 (2.24)	[2.38, 4.47]
Accuracy rate	0.80 (0.08)	[0.76, 0.83]	0.78 (0.15)	[0.71, 0.85]

*p < 0.05; **p < 0.01; ***p < 0.001

Table 3.

Experiment 3. Means, standard deviations and 95% confidence intervals for reported details by condition. [See Supplemental Materials for additional descriptives and associated analyses]

	Timeline Reporting		Free Recall Reporting	
	Mean (SD)	95% CI	Mean (SD)	95% CI
Account Detail				
Correct***	62.56 (24.29)	[52.53, 72.59]	31.74 (13.74)	[25.75, 37.72]
Incorrect	6.71 (3.25)	[5.33, 8.08]	5.00 (4.15)	[3.25, 6.75]
Accuracy rate	0.90 (0.05)	[0.87, 0.92]	0.86 (0.13)	[0.81, 0.92]
Conversation Detail				
Verbatim correct*	0.84 (1.21)	[0.34, 1.34]	0.25 (0.44)	[0.06, 0.44]
Gist correct*	15.00 (10.28)	[10.75, 19.24]	9.72 (6.64)	[6.98, 12.46]
Attribution correct***	8.20 (5.74)	[5.83, 10.57]	3.24 (3.14)	[1.94, 4.54]
Sequence error	0.64 (0.91)	[0.26, 1.01]	0.80 (1.15)	[0.32, 1.28]
Hierarchy correct*	6.56 (2.84)	[5.39, 7.73]	4.87 (2.09)	[2.99, 5.76]
Person Detail				
Correct***	31.96 (10.11)	[27.79, 36.13]	14.80 (6.57)	[12.09, 17.51]
Incorrect***	3.84 (2.19)	[2.93,4.74]	1.76 (2.05)	[0.91, 2.60]
Accuracy rate	.089 (0.06)	[0.86, 0.91]	0.88 (0.15)	[0.82, 0.95]

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

Table 4.

Statistics for mini meta-analyses of primary variables reported in Experiments 1, 2 and 3

	d	*Z*	*p*	95% CI
Correct details	0.93	5.01	<.001	[0.57, 1.29]
Incorrect details	0.39	2.23	0.01	[0.05, 0.73]
Accuracy rate	0.25	1.42	0.08	[-0.09, 0.59]
Verbatim correct	0.87	4.78	<.001	[0.51, 1.22]
Gist correct	0.39	2.26	0.01	[0.05, 0.74]

i In this section in Experiment 2 and Experiment 3, we report analyses for variables pertaining directly to our hypotheses. As is usual in this area of research, we also coded the number of incorrect details for all measures in all experiments. Given that (i) we had no hypotheses pertaining to incorrect details, and; (ii) there were typically no differences between conditions in terms of the number of incorrect details reported, these analyses are reported in Supplemental Materials in order to streamline the Results sections. All main analyses (and significant differences) are reported in the manuscript.