Effects of different types of forensic information on eyewitness' memory and confidence accuracy

Farhan Sarwar[a]*, Carl Martin Allwood[b], and Åse Innes-Ker[a]

[a]*Department of Psychology, Lund University, Sweden.*
[b]*Department of Psychology, University of Gothenburg, Sweden.*

ABSTRACT

This study investigated eyewitnesses' memory and confidence accuracy for action information (what happened at the crime scene), and detail information (descriptions of persons, objects, time and place). In Experiment 1, 89 participants watched a film and participated in one of four conditions: Laboratory discussion, Family discussion, Retell and Control, the first three with five meetings each. Three weeks later all participants open free recalled the events, and confidence judged their answers. The participants showed better free recall and confidence accuracy for action than for detail information. Participants in the two discussion conditions and in the Retell condition recalled more items and those in the Lab-discussion and Retell conditions more correct items for action information, than those in Control group. However, the four conditions did not differ for proportion correct of all action items recalled and confidence accuracy for action items. In brief, Experiment 1 showed that witness discussions and retellings of the experienced event with others improved recall for action information but had had no, or small, effects on confidence accuracy. Experiment 2 investigated recall and confidence accuracy performance for action and detail information using focused questions. Seventy-seven participants watched a film, answered and confidence judged 63 questions about action and detail information about the events. Again, participants showed better memory and confidence accuracy for action information. Overall, the results indicate that, for both free recall and focused questions, witnesses' recall and confidence accuracy is better for action information than for detail information, thus extra precaution is needed in the forensic system when detail information from witnesses is considered.

© 2014 Colegio Oficial de Psicólogos de Madrid. All rights reserved.

Efectos de diferentes tipos de información forense en la exactitud de la confianza y memoria del testigo visual

RESUMEN

Palabras clave:
Testigos
Precisión de la confianza
Información de acción
Recuerdo libre
Preguntas específicas

Este estudio analizó la precisión de la memoria y confianza de los testigos para la información de acción (lo que sucedió en la escena del crimen), y la información de detalles (descripción de personas, objetos, tiempo y lugar). En el Experimento 1, 89 participantes vieron una película y participaron en una de las cuatro posibles condiciones: Discusión en el laboratorio, Discusión familiar, Volver a contar la historia, y Control, las tres primeras condiciones con cinco encuentros cada una. Tres semanas después, todos los participantes relataron los hechos mediante recuerdo libre, y juzgaron la confianza que tenían en sus respuestas. Los participantes mostraron mejor recuerdo y precisión para la información de acción que para la de detalle. Los participantes de las dos condiciones de discusión y de la condición de volver a contar la historia recordaron más ítems, y aquellos que participaron en la discusión de laboratorio y en volver a contar la historia mostraron más ítems correctos para información de acción, que el grupo control. Sin embargo, las cuatro condiciones no mostraron diferencias en la proporción de ítems de acción correctos recordados y la confianza en su juicio. Sucintamente, el Experimento 1 mostró que las discusiones de testigos y volver a relatar los hechos vividos a otros mejora el recuerdo de la información de acción, pero tenía poco, o ningún, efecto en la precisión de la confianza de los juicios. El Experimento 2 analizó la precisión del recuerdo y de los juicios de la información de acción y detalle usando preguntas específicas. Setenta y siete participantes

*Correspondence concerning this article should be sent to Farhan Sarwar, Lund University, Box 213, 22100 Lund, Sweden. E-mail: farhan.sarwar@psy.lu.se

vieron una película, respondieron y juzgaron sesenta y tres preguntas sobre información de acción y de detalle en relación a los hechos. De nuevo, los participantes mostraron mejor memoria y confianza para la información de acción. En general, los resultados indican que, tanto para recuerdo libre como para preguntas específicas, la precisión en recuerdo y confianza es mayor para la información de acción que para la información de detalle, por tanto, es necesario tomar precauciones en el sistema forense cuando tiene en consideración información de detalles proporcionada por testigos.

Eyewitness testimony plays a central role in police investigative work and in later phases of the forensic process. However, eyewitnesses cannot always provide the required information, and many variables have been found to influence and bias performance (Blackwell-Young, 2008). Examples of such variables are the extent to which the witness has retold the event or discussed it with others and what type of memory recall instructions are used by the forensic personnel. In addition, eyewitness error has been reported to be the main cause of faulty convictions (Wise, Pawlenko, Safer, & Mayer, 2009). Moreover, eyewitness' confidence has been found to be the most influential factor when personnel in the forensic process evaluate the correctness of eyewitness memory reports (Boyce, Beaudry, & Lindsay, 2007), but witnesses' confidence accuracy has commonly been found to be lacking (Allwood, Ask, & Granhag, 2005; Allwood, Innes-Ker, Homgren, & Fredin, 2008; Leippe & Eisenstadt, 2007). A possibility is that different parts of witness' memory statements, such as information about actions and details, may differ with respect to their recall properties and their metacognitive accuracy. If this is the case, having knowledge about such differences could help forensic personnel to make better use of witnesses' testimony.

This study investigated whether action (what happened at the crime scene) or detail information (descriptions of persons, objects, time and place) is associated with better memory and confidence accuracy. Confidence accuracy refers to the match between a person's confidence judgments and the correctness of their answers (also called metacognitive realism). The study also investigated whether performance is influenced by repeated recall, by discussing the experienced event repeatedly, and finally, by the format of the questions asked. The two types of question formats used were: Open free recall, henceforth called *free recall* (Experiment 1) and focused questions (Experiment 1 and 2). It is of forensic relevance to understand memory performance for action and detail information in free recall, as eyewitnesses' early free recall can have an important influence on the subsequent course of the forensic investigation. Focused questions ask for specific information and are usually used as follow-up questions to cover relevant aspects of the event that the witness did not mention in their free recall report.

Effects of Multiple Retelling and Discussions on Memory Performance

Witness' memory performance is influenced by the fact that they retell their story multiple times, for example to the police and other personnel in the justice system and to their family and friends (Paterson & Kemp, 2006; Skagerberg & Wright, 2008). Not only how many times a memory is shared (repetition) but also *how* an event is shared affect the memory of the event (Hyman, 1994; Marsh, 2007; Pasupathi, 2001), and may affect memory and confidence accuracy differently for different types of forensic information.

Merely retelling an event can be considered the same as *actively repeating* the learned materials. Research shows that active repetition tends to improve memory performance in later recall, compared to the same amount of rehearsal based on reading the original material (Karpicke & Roediger, 2007, 2008; Roediger & Karpicke, 2006a).

Roediger and Karpicke called this *the testing effect* (Cull, 2000; Roediger & Karpicke, 2006a, 2006b).

In some cases, multiple retellings can hamper the retrieval of information because of *retrieval induced forgetting* (Coman, Manier, & Hirst, 2009; MacLeod, 2002). The retrieval induced forgetting effect is most likely to hold when information in the learned material is not well integrated. Thus, when the encoded content is well integrated there may be less or no retrieval induced forgetting (Anderson, 2003; Anderson & McCulloch, 1999).

In general, action information has been found to be recalled better than detail information (e.g., Migueles & Garcia-Bajos, 1999). However, it can be speculated that, given the described properties of retrieval induced forgetting, the testing effect and the retrieval induced forgetting effect may depend on the type of information being repeated. For action information, the testing effect may be more relevant. The reason is that actions are often well integrated in the sense that retrieval of the first action is likely to act as a cue to the next action and so on. Detail information (i.e., descriptive details of individuals and things) may often be more loosely integrated. For example, the colour of hair may have only a very weak association with the colour of shirt. Consequently, retrieval induced forgetting may affect the memory of detail information (Coman et al., 2009) more than action information. The effects of multiple retellings were investigated in Experiment 1.

Discussing an event is different from merely retelling it since eyewitness' memory is often influenced by information supplied by their discussion partners. For example, co-witnesses discussing the event often engage in social interactions and cognitive work that result in a more common version of the event (e.g., Principe & Schindewolf, 2012). This may decrease eyewitnesses' memory performance (Nourkova, Bernstein, & Loftus, 2004; Wright, Self, & Justice, 2000). Consequently, how a witnessed event is retold (simply repeated or discussed), and if discussed, how, can affect the memory of the event in subsequent recalls (Hyman, 1994; Marsh, 2007; Skowronski & Walker, 2004). The effects of multiple discussions were also investigated in Experiment 1.

Confidence Accuracy: Action vs. Detail Information

The results of previous research on confidence accuracy for action and detail information has not looked at free recall, but only at different types of focused questions (Ibabe & Sporer, 2004; Migueles & Garcia-Bajos, 1999). Confidence ratings were given on a scale from 1 to 5 in both studies and neither study reported the relation between level of confidence and the proportion correct memory reports for different types of information. However, it can be deduced from Table 2 and 3 in Ibabe and Sporer's (2004) reported results that their participants were somewhat better at separating correct from incorrect answers for detail information (.75) than for action information (.55). This is the slope-measure discussed below.

Free recall and focused questions differ in cognitive demand because focused questions test recognition memory and free recall test participants' own-regulated recall (compare the memory report model presented by Koriat & Goldsmith, 1996). Consequently, par-

ticipants' confidence accuracy may differ for action and detail information depending on question type.

Roberts and Higham (2002) investigated students' memory performance and confidence for central and peripheral information using the Cognitive interview (Geiselman, Fisher, MacKinnon, & Holland, 1986). Initially, four police officers and one crown council were asked to identify *events* (what happened), *people* and *weapon(s)* in a criminal event (called relevant information for solving the case and here called central information). Other information was seen as peripheral information. After context reinstatement, the participants were interviewed in three phases: report of the event information in natural order, reverse order, and with changed perspective. Finally, in phase four, they answered focused questions. When analyzing confidence accuracy, Roberts and Higham (2002) collapsed the data for the four phases in the interview. Therefore it is not clear which types of recall cause the effects with respect to confidence accuracy.

Repeated post-event questioning may cause increased confidence without a corresponding change in memory performance. One reason for this could be that the ease of information retrieval, also called retrieval *fluency*, may increase confidence (Kelley & Lindsay, 1993). Fluency has been reported to increase with repeated retrievals of the same information (Shaw, 1996; Shaw & McClure, 1996; Shaw, McClure, & Dykstra, 2007). This increase in retrieval fluency may affect detail information more than action information because detail information may be less well integrated than action information. In brief, when less well-integrated information is recalled multiple times it will be easier to retrieve and will therefore be associated with higher retrieval fluency, which in turn may increase confidence.

Also, the content and nature of discussions are relevant for confidence accuracy because they may influence memory performance. For example, discussing an event with friends and partners may result in better memory compared to discussing the same event with unknown people (Andersson & Rönnberg, 1995).

Calibration methodology was used in the present research since it gives a more differentiated and informative understanding of confidence accuracy compared to simply computing the correlation between confidence and correctness of the memory report, which is the traditional measure in lineup research (Allwood, 2010; Brewer, 2006; Juslin, Olsson, & Winman, 1996; Weingardt, Leonesio, & Loftus, 1994; Wells, Olson, & Charman, 2002).

The present study consists of two experiments. Experiment 1 was a part of a larger project and the overall findings of participants' performance have been reported elsewhere (Sarwar, Allwood, & Innes-Ker, 2011).

Experiment 1

Experiment 1 focused on participants' memory and confidence accuracy for action and detail information using free recall and focused questions. Participants first watched a short film about a kidnapping then were separated into four conditions. After three weeks, their memory for events in the film was tested, and 3 to 4 days later participants gave confidence judgments of their recall. The four conditions were: *Lab-discussion, Family discussion, Retell* and *Control.* In the first three of these conditions the participants retold and discussed the witnessed event 5 times with an unfamiliar person; retold and discussed the witnessed event with their family and friends; only retold the witnessed event to a person, respectively. In the Control condition the participants neither retold nor discussed the event.

The Lab-discussion condition can be seen as the critical condition. However, due to the very controlled nature of this condition, the Family discussion condition was created to increase the ecological validity and reduce the impact of strict lab conditions. To tap the effects of discussion, the Retell condition was introduced to control for the effects of merely retelling the event.

Hypotheses

Hypothesis 1. Proportion correct of all recalled information, and confidence accuracy were predicted to be better for action information than for detail information. The reasoning was that people in general are better at describing actions than describing details (Migueles & Garcia-Bajos, 1999). Consequently, not only the memory of action information will be better but the better memory is also likely to improve the confidence accuracy.

Hypothesis 2. For action information, the participants in the Lab- and Family discussion conditions were expected to report a lower proportion of correct pieces of information and have poorer calibration than participants in the Retell condition. The reason is that the exchange of information in the two discussion conditions can be expected to introduce new information that may be confused by the participants for information from the original event (Nourkova et al., 2004; Wright et al., 2000). In contrast, participants in the Retell condition were expected to benefit because of the testing effect (Roediger & Karpicke, 2006a). The confidence of the participants in the two discussion conditions would be inflated due to increase in retrieval fluency (Shaw, 1996; Shaw & McClure, 1996) caused by the post-event discussions in addition to the increase in fluency that may have taken place as an effect of the mere retelling in all three experimental conditions. Consequently, participants in the two discussion conditions were expected to show poorer confidence accuracy than those in the Retell condition on the confidence accuracy measures that focus on the relation between proportion correct and confidence (i.e., calibration and over-/underconfidence). Thus, the participants in the Retell condition were expected to show a higher proportion correct and to show better results on the calibration measures for action information than the participants in the discussion conditions.

Hypothesis 3. For detail information, participants in the Retell condition was expected to show better results than the Control condition on the two confidence accuracy measures that involve difference between the confidence level and proportion correct (i.e., calibration and over-/underconfidence). All four conditions were expected to show a somewhat low proportion correct, since the participants' memory for detail information has been consistently found to be weak in previous research (e.g. Migueles & Garcia-Bajos, 1999; Yuille & Cutshall, 1986). This would leave space for a beneficial repetition effect to be pertinent for the detail information in the three experimental conditions. At the same time, the level of proportion correct was expected to be somewhat dampened for the two discussion conditions due to that the participants in these conditions might mix up the discussion partner's contributions with information from the witnessed event. An increase for confidence due to repetition (especially in the discussion conditions) was also expected, but this increase was not expected to be as large as for proportion correct. Put together, this lead to an expectation of a higher proportion correct and somewhat inflated confidence for the Retell condition and as an effect of this better calibration and less over-/underconfidence than for the Control condition, with the participants in the two discussion conditions performing in between the other two conditions. In the Results section, both number of items and number of correct items are analyzed, as those results add to the understanding of memory performance, but no predictions about those results were made.

Method

Participants

The participants were 89 students (62 women) from Lund University, Sweden, with the mean age of 25 (18-47 years). At first, 92 participants were recruited with the aim to allocate 23 participants to

each condition. The dropouts from the Lab-discussion condition, Family discussion condition, the Retell and the Control condition were 7, 6, 4, and 1 respectively. Dropouts during the study were all replaced except for the three who dropped out at the end. Participants who successfully completed the experiment received a movie ticket worth 90 SEK (approximately 9€).

Design

The experiment conformed to a 4 X 3 mixed design. The between-subjects factor was the condition type: Lab-discussion (*n* = 21; one participant only attended 4 out of 5 sessions), Family discussion (*n* = 23), Retell (*n* = 23) and Control condition (*n* = 22). The within-subjects factor was the information type: *action* information and *detail* information. In addition, as described below, information about surroundings, etc. that can be seen as *non-forensic* was separated out.

Materials

Videotape. The witnessed event was a film about two men kidnapping a woman from a bus stop. The film was 3 min and 50 s long and was shown on a 28-inch color television. This film has been used in previous research (Allwood, Ask, & Granhag, 2005; Granhag, 1997).

Focused questions about the film. The questionnaire consisted of 44 questions about the film. The questionnaire was found productive and valid for the purpose of the study in previous research (Allwood, Innes-Ker, Homgren, & Fredin, 2008; Allwood, Jonsson, & Granhag, 2005). Each question had one correct and one incorrect answer alternative. The questions were about different details of the people's appearance, clothes, ages, as well as the surrounding environment such as letter boxes, cars, busses, and, specifically, the offenders' car. Some example questions were: His/her hair color was ---- (A. red; B. blond). How old was he/she ---- (A. 21 – 30 years; B. 31 – 40 years). What was the colour of the car ---- (A. blue; B. white).

Confidence judgment scales. An 11-point scale was used for the participants' judgments of the detailed parts of the free recalls. This scale started at 0% ("Completely sure that I remember wrong") and increased in steps by 10 (10, 20, 30 …) to 100% ("Completely sure that I remember correct"). This scale was inserted below each recalled item to be confidence judged. A 6-point scale was also used for participants' judgments of their answers to the 44 focused questions about the film. As the probability to randomly choose the right answer was 50%, the confidence scale ranged from 50% ("Guessing"), in steps by ten, to 100% ("Completely sure").

Procedure. The participants were received in the lab in small groups of between 4 to 8 individuals. They were told: "We are investigating human perception in different forensic situations". First, participants watched the video film, and then they were randomized to one of the four conditions. Participants in the three experimental conditions received a time schedule for the five occasions over a 20-day period when they would retell the film.

In each of the five sessions, the participants in the Lab-discussion and Retell conditions visited the laboratory. Participants in the Lab-discussion condition discussed the events in the film with a confederate. The confederate asked spontaneous questions about the film after listening to the story of the film. One-hundred and five confederates were recruited solely for the discussion purpose, and each confederate participated only in one discussion. The participants in the Retell condition simply told the story about the film to the experimenter. Participants were instructed to tell whatever they remembered about the film. The experimenter

posed no questions to the participants. All discussions and retellings were recorded on a MP3 recorder. Participants in the Family discussion condition did five discussions at home with their family or friends, each time with a new individual. The participants carried out their discussions on the scheduled dates and times. Further, participants confirmed their discussions to the experimenter by 7:00 pm by sending an SMS (Short Message Service). If any participant failed to send the SMS by 7:00 pm on the discussion day, the experimenter called the participant and reminded him/her of the task. In the Control condition, participants were instructed to not tell the contents of the film to anybody.

All participants returned to the lab on the 21st day for a memory test. The participants first typed whatever they remembered about the events they witnessed in the film (free recall test) and then answered 44 focused questions.

Finally, all participants returned to the laboratory on the 24th or 25th day to give their confidence judgments. First, participants gave their confidence judgments about their free recall items (which had been prepared as described below). Second, the participants gave their confidence judgments for the 44 focused questions.

Preparation of Material for Participants' Confidence Judgments

In order to record participants' confidence judgments for each piece of information, each participant's free recall was broken down into single pieces of information using the method described by Allwood, Ask and Granhag (2005). The principles followed were: (1) items about actors and actions carried out were considered a single piece of information. For example "the woman looked at the timetable" was considered as a single piece. (2) An object with one associated characteristic was considered as one piece. For example "the yellow letterbox" was considered as a single piece. (3) If an object was associated with more than one characteristic, the additional characteristics were regarded as separate pieces of information. For example "the bald old guy" was considered as two pieces. To help participants recall the context in which they mentioned each piece of information, one or two sentences adjacent to that piece were attached. These reference items were enclosed in brackets while the items to be confidence judged were underlined. Finally, an 11-point confidence scale was placed under each piece of information.

Division of Memory Reports into Forensic and Non-forensic Subcategories

In preparation for the ensuing analyses, participants' memory reports were classified into different information categories in two steps. In the first step, to separate out the irrelevant information, participants' statements were divided into forensic and non-forensic items. Information such as houses, roads, surroundings, etc., was seen as non-forensic because it is either not needed to solve the crime or can be collected from the crime scene without the help of an eyewitness.

In the second step, the forensic items were divided into action and detail categories by following the plot relevancy model by Heuer and Reisberg (1990). In this model, information or facts related to the story (actions involved in the event) that cannot be changed without changing the story are regarded as central (e.g., the suspect put a gun to the victim's head) and other information (detail information) as peripheral.

Inter-Judge Reliability. One coder divided all the participants' memory reports into action, detail, and non-forensic categories. In order to check the reliability of coding done by the first coder a second coder coded the 10% of the cases into the three categories. Inter-rater reliability was calculated from the 10% of the cases coded by the two coders using Cohen's Kappa. The resulting Cohen's Kappa value was .76, which shows a good inter-rater reliability.

Measurements

The five dependent measures used were proportion correct, confidence, calibration, over-/underconfidence and slope. *Proportion correct* is simply the number of correct items divided by the total number of reported items of the kind considered (e.g., all items, or all action items). Calibration, over-/underconfidence and slope are used to measure the confidence accuracy and are calculated from the scores for proportion correct and confidence. *Calibration* and *over-/underconfidence* measure the relationship between a person's level of confidence and proportion correct. Calibration compares a person's proportion correct and the corresponding confidence for each confidence level and represents the person's squared deviance from perfect calibration at all confidence levels (depending on the type of confidence scale used, e.g., 50, 60, 70, …). Over-/underconfidence is the difference between the person's average confidence and his/her percentage of correct items. The value zero represents perfect confidence accuracy for both calibration and over-/underconfidence. For the over-/underconfidence measure, positive and negative values show overconfidence and underconfidence respectively. *Slope* is a measure of the person's ability to separate his/her correct and incorrect answers by means of the level of his/her confidence judgments. It is computed by subtracting the person's mean confidence level for incorrect items from the mean confidence level of the correct items. Yates (1994) provided the specific formula for calibration and a more thorough discussion of these measures.

Results

Free recall

First the analyses for the number of recalled items of the different types (i.e., action, detail, and non-forensic items), and the total number of correct recalled items for each information type and for each condition are presented. Next follows the analyses pertaining to the three hypotheses, and finally the results for the focused questions are presented.

Number of Recalled Items by Category and Condition

The total amount of items reported (both correct *and* incorrect) indicates whether the different conditions affected the *quantity* of information volunteered. All items reported were submitted to a 4 X 3 mixed ANOVA in line with previous research (e.g., Migueles & García-Bajos, 1999; Wessel & Merckelbach, 1997) and since no other alternative was feasible. The between-subjects factor was condition (Lab-discussion, Family discussion, Retell and Control condition) and the within-subject factor was information type (action, detail, and non-forensic). The means and standard deviations are shown in Table 1.

The results showed that there was a significant main effect of information type, Wilks' Lambda = .19, $F(2, 84)$ = 178.3, $p < .001$; η^2 = .62. Contrasts with Bonferroni corrected significance levels revealed that action items were reported significantly more often than detail items $F(1, 85)$ = 185.3, $p < .001$: η^2 = .65, and non-forensic items, $F(1, 85)$ = 328.2, $p < .001$; η^2 = .77.

There was also a significant main effect of the four conditions, $F(3, 85)$ = 6.34, $p < .001$; η^2 = .18. Contrasts for the four conditions and the type of information with Bonferroni corrected significance levels revealed for action information that the Lab-discussion condition, $t(41)$ = 1.90, $p < .001$; Cohen's d = 1.5, the Retell condition, $t(43)$ = 3.84, $p < .001$; Cohen's d = 1.2, and the Family discussion condition, $t(43)$ = 2.75, $p < .01$; Cohen's d = .83, reported significantly more items as compared to the Control condition. No differences between the conditions were found for detail and non-forensic information.

The results also showed that there was a significant interaction effect for condition and information type, Wilks' Lambda = .83, $F(2, 86)$ = 2.74, $p < .01$; η^2 = .04. Inspecting the means in Table 1, it is unclear what is driving this interaction, as the discussion type seems to have differential impact on what type of information is recalled. It does appear that the variation between conditions for action items recalled is larger than for the two other types of information. To further probe the data for simple effects is unlikely to yield reliable insights.

Number of Correct Items for Different Kinds of Information

The number of correct items is an indicator of the quality of the recalled information. A 4 X 3 mixed ANOVA with a similar construction as above was conducted. The number of correct items is shown in Table 1.

The results showed that there was a significant main effect of information type, Wilks' Lambda = .19, $F(2, 80)$ = 170.9, $p < .001$; η^2 = .69. Contrasts with Bonferroni corrected significance levels revealed that significantly more correct action items were recalled as compared to the detail items $F(1, 81)$ = 293.04, $p < .001$; η^2 = .75, and the non-forensic items, $F(1, 81)$ = 260.5, $p < .001$; η^2 = .74. The contrasts also revealed that the number of correct non-forensic items was also significantly higher than the number of correct detail items, $F(1, 81)$ = 11.91, $p < .001$; η^2 = .12.

There was also a significant main effect of condition, $F(3, 81)$ = 6.45, $p < .001$; η^2 = .19. Contrasts with Bonferroni corrected significance levels revealed that the participants in the Lab-discussion condition, $t(41)$ = 4.52, $p < .001$, Cohen's d = 1.4, and Retell condition, $t(43)$ = 3.62, $p < .001$, Cohen's d = 1.1, recalled a significantly higher number of correct items than the participants in the Control condition. Further contrasts for each type of information revealed that the participants in the Lab-discussion recalled significantly more correct items than the Control condition for action information, $t(41)$ = 5.16, $p < .001$, Cohen's d = 1.6, and non-forensic information, $t(39)$ = 2.33, $p < .03$, Cohen's d = 0.75. The participants in the Retell

Table 1
Experiment 1: Results for the Free Recall. Mean Number of Items and Number of Correct Recalled Items for Different Types of Information for the Four Conditions

	Lab-discussion *M(SD)*	Family discussion *M(SD)*	Retell *M(SD)*	Control *M(SD)*
Mean number of items				
Action information	31.24(8.3)	25.83(9.0)	28.48(8.9)	18.86(7.9)
Detail information	13.81(8.5)	10.56(9.6)	7.13(6.7)	9.23(7.3)
Non-forensic	12.71(7.5)	10.69(7.6)	10.34(8.2)	7.40(5.2)
Correct items				
Action information	27.43(7.9)	22.00(8.2)	26.95(8.8)	16.75(6.3)
Detail information	8.52(5.5)	6.52(6.4)	4.71(4.9)	5.50(5.2)
Non-forensic	10.81(6.9)	8.91(6.1)	10.30(6.7)	6.75(3.6)

condition also recalled significantly more correct items than the participants in the Control condition for action information, $t(43) = 4.63$, $p < .001$, Cohen's $d = 1.4$, and the non-forensic information, $t(39) = 2.10$, $p < .05$, Cohen's $d = 0.67$.

The results showed that there was a significant interaction effect for condition and information type, Wilks' Lambda = .82, $F(2, 80) = 2.88$, $p < .01$; $\eta^2 = .04$. As with the analysis of the total number of items recalled, the interaction appears to have complex reasons. Inspection of the means again suggest that the variation between discussion conditions for action items recalled was larger than for the other two types.

Confidence Accuracy: Calibration Curves

Figure 1 shows the calibration curves for the free recall for action, detail, and non-forensic information for all participants. The x-axis shows the confidence levels (from 0 to 100%) and the y-axis shows the percent of correct answers. The numbers inside the graph give the percent of answers for each confidence level in each condition. The diagonal shows perfect calibration. As can be seen in the Figure, the calibration curves for the action, detail, and non-forensic information show very low percentages for the confidence scores between 0 and 50. Consequently, these levels will not be further discussed. For confidence scores 50 and above, the scores for action and non-forensic information are clustered close to the diagonal, indicating that confidence is fairly accurate for these two types of information. For detail information, the calibration curve shows overconfidence from 60% to 100% confidence. In addition, for both action and non-forensic information, about half of the confidence judgments are at the 100% level. For detail information, the confidence judgments are fairly evenly distributed between 50% and 100%. For all information types the participants show some overconfidence at the 100% confidence level.

Proportion Correct, Confidence, and Confidence Accuracy for Action, Detail and Non-forensic Information and for the Four Conditions

The next set of analyses addressed the three main hypotheses. To compare the participants' performance in the four conditions for the three information types, the data were subjected to 4x3 mixed ANOVAs, separately for proportion correct, confidence, calibration, over-/underconfidence, and slope. The means and standard deviations are shown in Table 2.

Proportion correct of all recalled items. The results showed a significant main effect of information type, Wilks' Lambda = .43, $F(2, 66) = 43.56$, $p < .001$; $\eta^2 = .50$. Contrasts with Bonferroni corrected significance levels revealed that the participants reported a significantly higher proportion correct of action information than detail information $F(1, 67) = 73.52$, $p < .001$; $\eta^2 = .25$. The participants also a reported higher proportion correct for non-forensic information than detail information, $F(1, 67) = 85.68$, $p < .001$; $\eta^2 = .30$. There was no main effect of condition, and no interaction effect was found.

Confidence. The results showed a significant main effect for confidence for information type, Wilks' Lambda = .43, $F(2, 66) = 44.64$, $p < .001$; $h^2 = .46$. Contrasts with Bonferroni corrected significance levels revealed that the participants were significantly more confident for action information than detail information $F(1, 67) = 79.60$, $p < .001$; $h^2 = .25$. The participants were also more confident for the non-forensic information as compared to the detail information, $F(1, 67) = 75.64$, $p < .001$; $\eta^2 = .28$. There was no main effect of condition, and no interaction effect was found.

Calibration. The results showed a significant main effect for calibration for information type, Wilks' Lambda = .65, $F(2, 83) = 22.02$, $p < .001$; $\eta^2 = .27$. Contrasts with Bonferroni corrected significance levels revealed that the participants were significantly better calibrated for action information than detail information $F(1, 84) = 44.15$, $p < .001$; $\eta^2 = .16$. The participants were also better calibrated for non-forensic information as compared to detail information, $F(1, 84) = 29.04$, $p < .001$; $\eta^2 = .14$. There was no main effect of condition, and no interaction effect was found.

Over-/underconfidence. The results showed a significant main effect for over-/underconfidence for information type, Wilks' Lambda = .75, $F(2, 83) = 14$, $p < .001$; $\eta^2 = .19$. Contrasts with Bonferroni corrected significance levels revealed that the participants showed
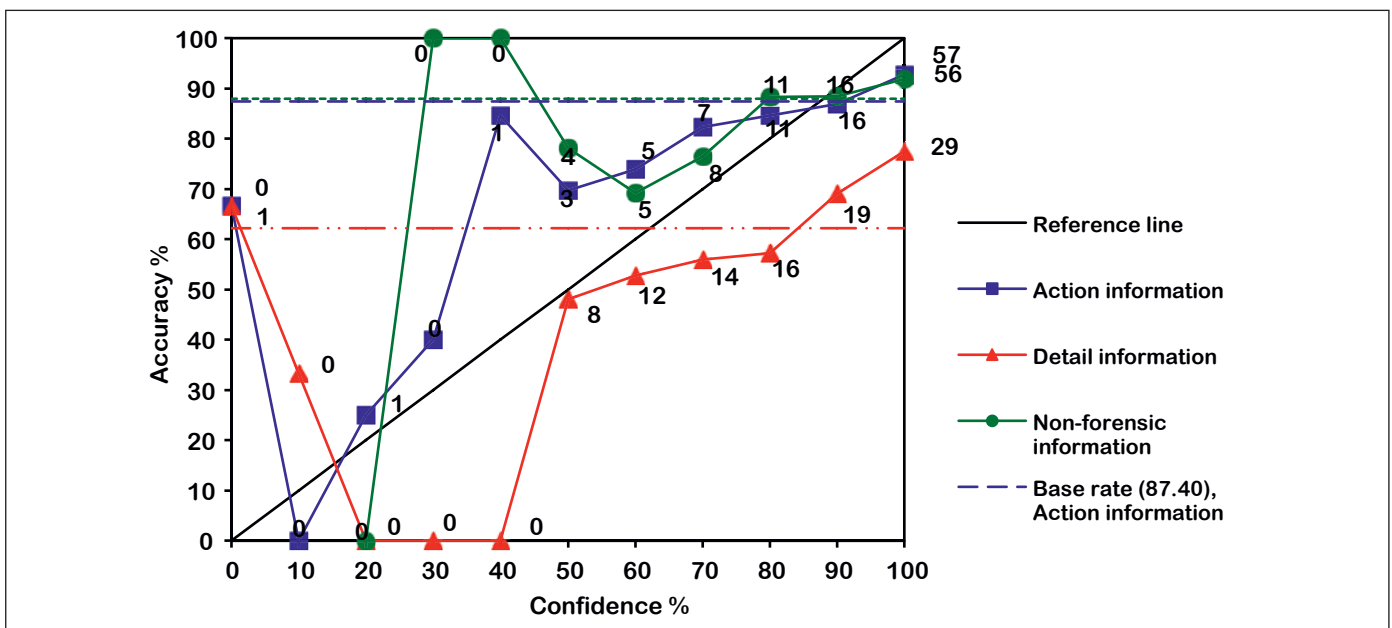

Figure 1. Experiment 1: Calibration curves for the free recall for the different information types (action, detail and non-forensic) collapsed over the four conditions. Digits at each point show the % of confidence judgments that used this confidence level at that particular level of proportion correct. 0 % on the confidence axis (x) means "Completely sure that I remember wrong" and 100 % means "Completely sure that I remember correct".

Table 2
Experiment 1: Results for the Free Recall. Action, Detail, and Non-forensic Information: Means (and SDs) for Proportion Correct, Confidence, Calibration, Over-/underconfidence and Slope

	Lab-discussion M(SD)	Family discussion M(SD)	Retell M(SD)	Control M(SD)
Proportion correct				
Action information	87.60(10.1)	84.20(10.0)	92.56(5.2)	85.10(12.9)
Detail information	61.10(21.2)	63.91(22.0)	68.15(25.9)	57.38(27.9)
Non-forensic	86.15(13.8)	91.32(11.3)	93.84(6.8)	86.18(13.1)
Confidence				
Action information	87.96(6.7)	90.45(5.9)	90.17(4.9)	86.86(7.7)
Detail information	75.71(10.4)	80.60(12.1)	80.44(7.2)	73.73(12.8)
Non-forensic	89.76(6.8)	91.43(8.3)	90.53(7.1)	86.57(11.9)
Calibration				
Action information	0.04(.03)	0.05(.03)	0.03(.02)	0.05(.03)
Detail information	0.11(.09)	0.12(.11)	0.11(.12)	0.12(.10)
Non-forensic	0.05(.05)	0.05(.07)	0.02(.02)	0.06(.06)
O/U confidence				
Action information	0.00(0.10)	0.05(0.11)	-0.02(0.07)	0.01(0.13)
Detail information	0.14(0.19)	0.16(0.24)	0.12(0.26)	0.16(0.22)
Non-forensic	0.03(0.14)	0.01(0.13)	-0.03(0.08)	0.02(0.16)
Slope				
Action information	8.11(9.9)	16.63(14.9)	0.04(8.2)	7.10(7.8)
Detail information	7.15(9.2)	7.27(13.1)	11.25(6.1)	13.88(14.7)
Non-forensic	10.79(16.0)	10.63(18.3)	21.98(11.9)	11.64(7.9)

Note. O/U confidence = Over-/underconfidence

significantly less over-/underconfidence for action information than for detail information $F(1, 84) = 26.22$, $p < .001$; $\eta^2 = .12$. The participants also showed less over-/underconfidence for non-forensic information as compared to detail information, $F(1, 84) = 24.18$, $p < .001$; $\eta^2 = .11$. There was no main effect of condition, and no interaction effect was found.

Slope. There was no significant difference for slope between the three types of information.

The Focused Questions

Separate between-subjects one-way ANOVAS were conducted for each of the dependent measures for the focused questions (see Table 3). There were no significant effects for any of the measures.

Discussion

The results for the free recall showed that participants performed significantly better for action information as compared to detail and non-forensic information on the mean number of items and correct items reported in each information category. The participants' better performance for action information is in line with previous research (Roebers, Moga, & Schneider, 2001). The lower number of detail items could, at least partly, be due to repetition. Hershkowitz and Terner (2007) found that in the second recall attempt, participants recalled less detail information than in the first attempt. It is difficult to say whether the low number of detail items recalled in Experiment 1 was due to the multiple recall attempts or because of the other factors that might have influenced the memory of detail information during the time between watching the event and final recall. A partial explanation for the lower number of non-forensic information units may be that participants exert control over what to share on the basis of their understanding of what is expected of them (Grice, 1975; Russell & Schober, 1999). Consequently, the participants may only have shared such information that they considered relevant to the forensic event.

There were also differences between the conditions for the different types of information. Participants in the Retell and the Lab-discussion conditions benefited from the retellings and discussions

Table 3
Experiment 1: Results for Focused Questions. Means (and SDs) for Proportion Correct, Confidence, Calibration, Over-/underconfidence and Slope

	Lab-discussion M(SD)	Family discussion M(SD)	Rehearsal M(SD)	Control	F	η^2
Proportion correct	53.68(6.68)	55.34(7.14)	56.13(5.83)	54.03(6.54)	0.67	.02
Confidence	62.97(4.45)	65.55(6.76)	64.33(6.97)	62.28(5.57)	1.29	.04
Calibration	0.05(0.02)	0.05(.03)	0.05(0.02)	.05(.03)	0.32	.01
O/U confidence	0.09(0.08)	0.10(.08)	0.08(0.07)	.08(.09)	0.33	.01
Slope	4.02(4.37)	3.50(4.94)	4.22(5.73)	3.64(5.46)	0.09	.00

Note. O/U confidence = Over-/underconfidence

respectively; they reported significantly more action items as compared to the Control condition. However, there was no similar benefit for either the number of detail items or correct detail items.

The first hypothesis expected that the participants' would have a higher proportion correct and better confidence accuracy for action information than detail information. The results indeed showed that participants had higher proportion correct, higher confidence, better calibration, and over-/underconfidence for action information as compared to detail information. However, the slope measure did not differ between the information types.

Participants' superior performance for action information over detail information is in line with the previous findings where focused questions were used for analyzing confidence accuracy (Ibabe & Sporer, 2004; Migueles & Garcia-Bajos, 1999). Note the large effect sizes indicating a strong effect of information type. The present findings for confidence accuracy extend the previous findings to eyewitness' free recall. The results for the measures of confidence accuracy further add to our knowledge in that the participants showed very good confidence accuracy for the measures that relate the participants' level of confidence to their proportion correct (i.e., the calibration and over-/underconfidence measures) for action information but not equally impressive results for these measures for detail information. Interestingly, the results for the slope measure showed that participants' ability to discriminate between correct and incorrect responses did not differ for action and detail information.

According to the second hypothesis, participants in the two discussion conditions were expected to report a lower proportion correct and to show poorer confidence accuracy for action information than the participants in Retell condition. The results did not support this hypothesis in that no difference were found between the conditions for proportion correct, confidence, calibration, over-/underconfidence, and slope. This result should be taken with caution because the mixed ANOVA is a bit stringent, and it excluded those participants from the analysis who did not report either type of information. Consequently, the ANOVA results presented were based on 71 participants (total N = 89). This lowered the power of the analysis. This issue should be explored further in future research with a different design and with more participants.

The results did not support the third hypothesis that for detail information, participants in the Retell condition would show better confidence accuracy than the Control condition. The reason may be that the standard deviations were large for the proportion correct for detail information as evidenced when comparing with action information.

The performance on the 44 focused questions was overall poor, and also, there were no differences between conditions. One possible reason is that the focused questions were only about detail information. Since there were no differences between the conditions for the detail information in the free recall, findings for focused questions further strengthen the free recall results that detail information is difficult to remember accurately.

Experiment 2

The testing part of Experiment 1 was designed to follow the standard police procedure of interviewing the witnesses. The Participants first did a free recall and then answered the focused questions. These questions were only about detail information; hence it was not possible to compare the participants' performance for action and detail information for the focused questions. To further explore this issue a follow-up experiment was conducted. Given that the results for the focused questions did not differ between the four conditions, the different conditions of Experiment 1 were not included in Experiment 2. Just as in Experiment 1it is predicted that the proportion correct of all recalled information, and confidence accuracy would be better for action information than for detail information.

Method

Participants

Seventy-seven students from Lund University participated. The mean age of the participants was 26.5 years, ranging from 20 to 65 years. Each participant was given a lottery ticket worth 25 SEK (approximately € 2.5).

Design

A within-subject design was used. The within-subject factor was the two types of focused questions used: questions on action and detail information.

Materials

Videotape. The same video was used as in Experiment 1.

Focused questions about the film. The questionnaire consisted of 63 questions about the film. Eighteen questions were about action information of the events shown in the film. Forty-five questions were about details of the events. The different number of questions in the two categories is due to that there was more detail than action information in the film. Questions were arranged in the order events happened in the film. Each question had three alternative answers where one was always correct. The greater number of options was chosen to lower the guessing level as compared to the two options used in the 44 focused questions in Experiment 1. The position of the correct answer was randomized for the questions.

Confidence judgment scale. An 11-point confidence scale from 0% ("Completely sure that I remember wrong") to 100% ("Completely sure that I remember correct") confidence was used. This scale was placed under each question.

Procedure

The participants first watched the film. In order to avoid covert rehearsal, participants then read a one and half page long article in English as a filler task for 10 minutes. Next, participants answered the questionnaire and made a confidence judgment immediately after answering each question.

Results

Calibration Curves

Figure 2 shows the calibration curves for action and detail information. Action information shows low percentages of items for the respective confidence levels between 0 and 60. The calibration curve for detail information shows that the number of items at each confidence level is fairly equally spread from the 0% level to the 90% level. The graph suggests that participants were relatively well calibrated for action information compared to detail information. The calibration curve for detail information is fairly flat. This shows that participants had difficulty in separating correct from incorrect answers by means of their confidence judgments.

Proportion Correct, Confidence, Over-/underconfidence, and Slope

Paired sample t-tests were conducted to compare the participants' scores for proportion correct, confidence, calibration, over-/underconfidence, and slope for action and detail information (Table 4). Participants showed significantly higher proportions correct,
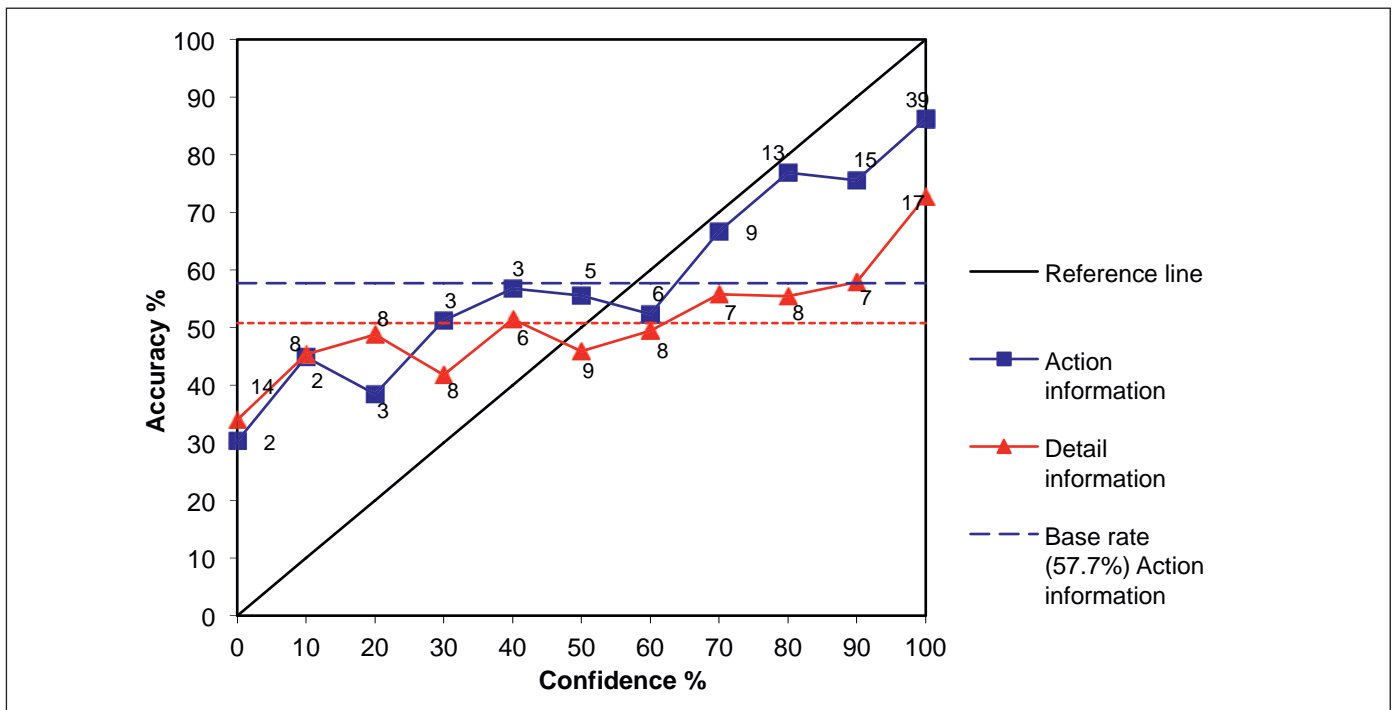
Figure 2. Experiment 2: Calibration curves for the focused questions for action and detail information. Digits at each point show the % of confidence judgments that used this confidence level at that particular level of proportion correct. 0 % on the confidence axis (x) means "Completely sure that I remember wrong" and 100 % means "Completely sure that I remember correct".

$t(76) = 16.46$, $p < .001$, Cohen's $d = 2.19$, confidence, $t(76) = 25.13$, $p < .001$, Cohen's $d = 2.36$, better calibration, $t(76) = -3.06$, $p < .001$, Cohen's $d = -0.55$, and steeper positive slope, $t(76) = 2.73$, $p < .008$, Cohen's $d = .39$, for action information than for detail information. In contrast, participants showed less overconfidence, $t(76) = 3.96$, $p < .001$, Cohen's $d = .40$, for detail information as compared to action information.

Discussion

Experiment 2 was conducted to investigate if participants would show better memory performance and confidence accuracy for action information as compared to detail information for focused questions. The results showed that the participants reported higher proportion correct, higher confidence, had better calibration, and larger slope for action information. In contrast, participants showed less over-/underconfidence for detail information as compared with action information. The results for proportion correct and confidence were in line with previous studies, using open-ended questions (Yuille & Cutshall, 1986), focused questions (Parker & Carranza, 1989), and both focused (2 and 4 answer alternatives) and open-ended questions (Ibabe & Sporer, 2004), that show people are better at remembering central information as compared to peripheral information.

As noted, the results showed that the participants showed less overconfidence for detail information than for action information. This result was unexpected because the participants could be expected to show less overconfidence for action information due to better memory for this type of information. However, this result can be explained by looking at the calibration curves in Figure 2. Here it can be seen that the action items are, on average, closer to the diagonal (indicating perfect confidence accuracy) than the detail items. This conclusion remains also when the percentages of the items at each confidence level are weighted in. The good result for the over-/underconfidence measure for the detail items may be due to the underconfidence found for confidence classes below 50% being "compensated" by the overconfidence found for the confidence classes above 50%, giving an aggregate result of 0.00 for the peripheral items.

Parenthetically, it should also be noted that the participants used the confidence class 0 (ostensibly meaning "absolutely certain I'm wrong") for 14% of the detail items whereas their performance for these items was at chance level. The explanation for this is not clear, except that the detail items were in general fairly difficult and may have appeared so for the participants. The participants appear not to have considered that they could expect to have about one-third of the items correct just by chance. In total, in the over-/underconfi-

Table 4
Experiment 2: Means (and SDs) for Proportion Correct, Confidence, Calibration, Over-/underconfidence and Slope

	Action information $M(SD)$	Detail information $M(SD)$	t	Cohen's d
Proportion correct	72.01(11.05)	51.60(7.25)	16.46*	2.18
Confidence	76.97(10.27)	51.41(11.49)	25.13*	2.34
Calibration	0.09(0.05)	0.12(0.06)	-3.06*	-0.54
O/U confidence	0.05(0.12)	0.00(0.13)	3.96*	0.40
Slope	20.79(15.60)	15.50(11.61)	2.73*	0.38

Note. O/U confidence = Over-/underconfidence; *$p < .01$.

dence measure, these two tendencies (under- and overconfidence at the different halves of the scale) balance out. In general, the results quite clearly indicate poorer confidence accuracy for detail questions over action questions when the calibration and slope measures are also considered. In line with this, the participants showed greater overconfidence for detail items than for action items for all confidence levels above 50%.

Responses to all focused questions, aggregated across both action and detail questions, showed lower levels of overconfidence than has been observed in previous research (e.g Allwood et al., 2008; Allwood, Knutsson, & Granhag, 2006). A possible reason might be that the three answer alternatives may have induced more uncertainty (Robinson, Johnson, & Robertson, 2000), especially for the more difficult questions (i.e., the questions on details).

General Discussion

This research investigated eyewitness' capacity to correctly report and give confidence judgments with good confidence accuracy of their recall of action and detail information. The issue was investigated using eyewitness' free recall reports and their answers to focused questions. For the free recall, non-forensic information was also analyzed.

The results for the free recall showed that the number of items reported and the number of correct items reported was higher for action information than for detail and non-forensic information, whereas participants had a higher proportion correct both for action and non-forensic information than for detail information. These results are in line with the previous findings that the participants reported more items for action information (Jelicic et al., 2006; Yuille & Cutshall, 1986) and had a higher proportion correct (Ibabe & Sporer, 2004; Migueles & Garcia-Bajos, 1999) than for detail information. This suggests that people can describe the time-course of an event in a forensic situation quite well, whereas descriptions of individuals and objects suffer in accuracy comparatively. One possible reason for these results can be that it is not only the motion *per se* that draws attention. Rather, the onset of motion would draw attention if it causes regrouping and new interpretation of the scene (Abrams & Christ, 2003; Hillstrom & Yantis, 1994). In case of a forensic event, actions or movements of different people and things would cause a continuous regrouping and reinterpretation of the whole scene over the period of time the crime happens. Thus, the witness' attention is likely to switch between the different actions initiated by the actors to interpret the situation in the light of the regrouping of the stimuli. Although a blue shirt that moves would display a moving color it would not usually signal a motion onset with a regrouping effect. Considering that attention is a limited resource, it is consequently likely that attention used to interpret the regrouping of the situation makes the eyewitnesses unintentionally blind to detail information. Hence, the police need to be more careful when taking into account the description of the offender provided by an eyewitness compared with the witness' description of the action information of the event.

This issue needs further research. Such research could use a methodology that study eyewitness gaze behavior during a real time event similar to a forensic event. Use of eye tracking device is a possible option to study such behavior.

The results for the confidence accuracy measures for the free recall show that confidence accuracy for action information (in contrast to detail information) can be fairly good. These results indicate that confidence may be seen as a fairly reliable indicator of proportion correct for action information in witnesses' free recall.

This research also addressed the impact of retellings, discussions, and the type of memory question on the memory and confidence accuracy of different types of information. The results showed that the participants in the two discussion conditions reported significantly more action information than the participants in the Control

condition. However, the four conditions did not differ in terms of proportion correct, confidence, calibration, and over-/underconfidence and slope. These results are relevant for the courts' decision making since courts place more trust on eyewitnesses who provide *more* information than eyewitnesses who provide less information (Bell & Loftus, 1988; Heath, Grannemann, Sawa, & Hodge, 1997). This suggests that courts need to be careful when using the amount of information reported as a criterion for the credibility of an eyewitness.

The results showed no difference between the participants' performance in the four conditions for detail information. However, a limitation of the results concerning the free recall in this study was the frequency of detail information recalled was quite low. For this reason, differential effects of discussion or repetition on the proportion correct and confidence accuracy for detail information should be investigated in future research using bigger samples so as to get more stable results.

The results for focused questions from the two experiments are in line with other empirical findings using the Plot relevancy model showing that people, in general, are better at remembering action information (Hershkowitz & Terner, 2007; Roebers & Schneider, 2000; Yuille & Cutshall, 1986) than detail information.

The results for the focused questions from Experiment 2 indicate that people also show better confidence accuracy for action information over detail information. An exception to this was that the participants showed no overconfidence for detail information but as discussed above this result appears to be because the underconfidence for items with confidence under 50% was "compensated" by the overconfidence for the items with confidence over 50%. The specific reason for this pattern is not clear and should be investigated in future research.

In brief, the results indicate that although repeated retellings and discussions risk introducing specific errors into the testimony, the overall effect of these activities seen in this study for free recall was positive for memory but not very noticeable for confidence accuracy performance for action information. In contrast, there was, overall, less effect for detail information. When, as often may be the case, focused questions concern detail information, the results from Experiment 1 suggest that the effect of repeated retellings and discussion over a short time period of about three weeks will not be very noticeable.

Conflict of interest

The authors of this article declare no conflict of interest.

References

Abrams, R. A., & Christ, S. E. (2003). Motion onset captures attention. *Psychological Science, 14*, 427-432.

Allwood, C. M. (2010). Eyewitness confidence. In P. A. Granhag (Ed.), *Forensic psychology in context* (pp. 281-303). Uffculme, Devon, UK: Willan Publishing.

Allwood, C. M., Ask, K., & Granhag, P. A. (2005). The Cognitive Interview: Effects on the realism in witnesses' confidence in their free recall. *Psychology, Crime & Law, 11*, 183-198. doi: 10.1080/10683160512331329943

Allwood, C. M., Innes-Ker, Å., Homgren, J., & Fredin, G. (2008). Accuracy and realism in confidence in children's and adults' event recall in response to free recall and focused questions. *Psychology, Crime & Law, 14*, 529-547. doi: 10.1080/10683160801961231

Allwood, C. M., Jonsson, A.-C., & Granhag, P. A. (2005). The effects of source and type of feedback on child witnesses' metamemory accuracy. *Applied Cognitive Psychology, 19*, 331-344. doi: 10.1002/acp.1071

Allwood, C. M., Knutsson, J., & Granhag, P. A. (2006). Eyewitnesses under influence: How feedback affects the realism in confidence judgements. *Psychology, Crime & Law, 12*, 25-38. doi: 10.1080/10683160512331316316

Anderson, M. C. (2003). Rethinking interference theory: Executive control and the mechanisms of forgetting. *Journal of Memory and Language, 49*, 415-445. doi: 10.1016/j.jml.2003.08.006

Anderson, M. C., & McCulloch, K. C. (1999). Integration as a general boundary condition on retrieval-induced forgetting. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 25*, 608-629. doi: 10.1037//0278-7393.25.3.608

Andersson, J., & Rönnberg, J. (1995). Recall suffers from collaboration: Joint recall effects of friendship and task complexity. *Applied Cognitive Psychology, 9*, 199-211.

Bell, B. E., & Loftus, E. F. (1988). Degree of detail of eyewitness testimony and mock juror judgments. *Journal of Applied Social Psychology, 18*, 1171-1192.

Blackwell-Young, J. (2008). Witness evidence. In G. Davies, C. Hollin, & R. Bull (Eds.), *Forensic psychology*. West Sussex, UK: John Wiley & Sons.

Brewer, N. (2006). Uses and abuses of eyewitness identification confidence. *Legal and Criminological Psychology, 11*, 3-23. doi: 10.1348/135532505x79672

Boyce, M., Beaudry, J. L., & Lindsay, R. C. L. (2007). Belief of eyewitness identification evidence. In R. C. Lindsay, D. F. Ross, J. Don Read & M. P. Toglia (Eds.), *Handbook of eyewitness psychology* (Vol. 2, Memory for people, pp. 501–525). Mahwah, NJ: Lawrence Erlbaum Associates.

Coman, A., Manier, D., & Hirst, W. (2009). Forgetting the unforgettable through conversation: Socially shared retrieval-induced forgetting of September 11 memories. *Psychological Science, 20*, 627-633. doi: 10.1111/j.1467-9280.2009.02343.x

Cull, W. L. (2000). Untangling the benefits of multiple study opportunities and repeated testing for cued recall. *Applied Cognitive Psychology, 14*, 215-235. doi: 10.1002/(SICI)1099-0720(200005/06)14:3<215::AID-ACP640>3.0.CO;2-1

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1986). Enhancement of eyewitness memory with the cognitive interview. *The American Journal of Psychology, 99*, 385-401. doi: 10.2307/1422492

Granhag, P. A. (1997). Realism in eyewitness confidence as a function of type of event witnessed and repeated recall. *Journal of Applied Psychology, 82*, 599-613. doi: 10.1037/0021-9010.82.4.599

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics*: Speech acts (Vol. 3, pp. 41-58). New York, NY: Academic Press.

Heath, W. P., Grannemann, B. D., Sawa, S. E., & Hodge, K. M. (1997). Effects of detail in eyewitness testimony on decisions by mock jurors. *Journal of Offender Rehabilitation, 25*(3), 51-71. doi: 10.1300/J076v25n03_04

Hershkowitz, I., & Terner, A. (2007). The effects of repeated interviewing on children's forensic statements of sexual abuse. *Applied Cognitive Psychology, 21*, 1131-1143. doi: 10.1002/acp.1319

Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & Cognition, 18*, 496-506. doi: 10.3758/BF03198482

Hillstrom, A. P., & Yantis, S. (1994). Visual motion and attentional capture. *Perception & Psychophysics, 55*, 399-411. doi: 10.3758/bf03205298

Hyman, I. E., Jr. (1994). Conversational remembering: Story recall with a peer versus for an experimenter. *Applied Cognitive Psychology, 8*, 49-66. doi: 10.1002/acp.2350080106

Ibabe, I., & Sporer, S. L. (2004). How you ask is what you get: on the influence of question form on accuracy and confidence. *Applied Cognitive Psychology, 18*, 711-726. doi: 10.1002/acp.1025

Jelicic, M., Smeets, T., Peters, M. J. V., Candel, I., Horselenberg, R., & Merckelbach, H. (2006). Assassination of a controversial politician: remembering details from another non-existent film. *Applied Cognitive Psychology, 20*, 591-596. doi: 10.1002/acp.1210

Juslin, P., Olsson, N., & Winman, A. (1996). Calibration and diagnosticity of confidence in eyewitness identification: Comments on what can be inferred from the low confidence-accuracy correlation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*, 1304-1316. doi: 10.1037/0278-7393.22.5.1304

Karpicke, J. D., & Roediger, H. L., III. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language, 5*, 151-162. doi: 10.1016/j.jml.2006.09.004

Karpicke, J. D., & Roediger, H. L., III. (2008). The critical importance of retrieval for learning. *Science, 319*(5865), 966-968. doi: 10.1126/science.1152408

Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basis for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32*, 1-24.

Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review, 103,* 490-517. doi: 10.1037/0033-295X.103.3.490

Leippe, M. R., & Eisenstadt, D. (2007). Eyewitness confidence and the confidence accuracy relationship in memory for people. In R. C. Lindsay, D. F. Ross, J. D. Read & M. P. Toglia (Eds.), *Handbook of eyewitness psychology* (Vol. 2, pp. 377–425). Mahwah, NJ: Erlbaum.

MacLeod, M. (2002). Retrieval-induced forgetting in eyewitness memory: forgetting as a consequence of remembering. *Applied Cognitive Psychology, 16*, 135-149. doi: 10.1002/acp.782

Marsh, E. J. (2007). Retelling is not the same as recalling: Implications for memory. *Current Directions in Psychological Science, 16*, 16-20.

Migueles, M., & Garcia-Bajos, E. (1999). Recall, recognition, and confidence patterns in eyewitness testimony. *Applied Cognitive Psychology, 13*, 257-268. doi: 10.1002/(SICI)1099-0720(199906)13:3<257::AID-ACP566>3.0.CO;2-7

Nourkova, V., Bernstein, D., & Loftus, E. (2004). Altering traumatic memory. *Cognition and Emotion, 18*, 575-585.

Parker, J. F., & Carranza, L. E. (1989). Eyewitness testimony of children in target-present and target-absent lineups. *Law and Human Behavior, 13*, 133-149. doi: 10.1007/BF01055920

Pasupathi, M. (2001). The social construction of the personal past and its implications for adult development. *Psychological Bulletin, 127*, 651-672.

Paterson, H. M., & Kemp, R. I. (2006). Co-witnesses talk: A survey of eyewitness discussion. *Psychology, Crime & Law, 12*, 181-192. doi: 10.1080/10683160512331316334

Principe, G. F., & Schindewolf, E. (2012). Natural conversations as a source of false memories in children: Implications for the testimony of young witnesses. *Developmental Review, 32,* 205-223. doi: http://dx.doi.org/10.1016/j.dr.2012.06.003

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

Robinson, M. D., Johnson, J. T., & Robertson, D. A. (2000). Process versus content in eyewitness metamemory monitoring. *Journal of Experimental Psychology: Applied, 6,* 207-221. doi: 10.1037/1076-898X.6.3.207

Roebers, C. M., Moga, N., & Schneider, W. (2001). The role of accuracy motivation on children's and adults' event recall. *Journal of Experimental Child Psychology, 78*(4), 313-329. doi: 10.1006/jecp.2000.2577

Roebers, C. M., & Schneider, W. (2000). The impact of misleading questions on eyewitness memory in children and adults. *Applied Cognitive Psychology, 14,* 509-526. doi: 10.1002/1099-0720(200011/12)14:6<509::AID-ACP668>3.0.CO;2-W

Roediger, H. L., III, & Karpicke, J. D. (2006a). The power of testing memory: Basic research and implications for educational practice. *Perspectives on Psychological Science, 1,* 181-210. doi: 10.1111/j.1745-6916.2006.00012.x

Roediger, H. L., III, & Karpicke, J. D. (2006b). Test-enhanced learning. *Psychological Science, 17,* 249-256. doi: 10.1111/j.1467-9280.2006.01693.x

Russell, A. W., & Schober, M. F. (1999). How beliefs about a partner's goals affect referring in goal-discrepant conversation. *Discourse Processes, 27*, 530-541. doi: 10.1080/01638539909545048

Sarwar, F., Allwood, C. M. Innes-Ker, Å. (2011). Effects of communication with a non-witness on eyewitnesses' recall correctness and meta-cognitive realism. *Applied Cognitive Psychology 25,* 782-791.

Shaw, J. S., III. (1996). Increases in eyewitness confidence resulting from postevent questioning. *Journal of Experimental Psychology: Applied, 2*, 126-146. doi: 10.1037/1076-898X.2.2.126

Shaw, J. S., III, & McClure, K. A. (1996). Repeated postevent questioning can lead to elevated levels of eyewitness confidence. *Law and Human Behavior, 20*, 629-653. doi: 10.1007/BF01499235

Shaw, J. S., III, McClure, K. A., & Dykstra, J. A. (2007). Eyewitness confidence from the witnessed event through trial. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. Lindsay (Eds.), *Handbook of eyewitness psychology,* (Vol. 1, pp. 555-590). Mahwah, NJ: Erlbaum.

Skagerberg, E. M., & Wright, D. B. (2008). The prevalence of co-witnesses and co-witness discussions in real eyewitnesses. *Psychology, Crime & Law, 14,* 513–521.

Skowronski, J. J., & Walker, W. R. (2004). How describing autobiographical events can affect autobiographical memories. *Social Cognition, 22*, 555-590. doi: 10.1521/soco.22.5.555.50764

Weingardt, K. R., Leonesio, R. J., & Loftus, E. F. (1994). Viewing eyewitness research from a metacognitive perspective. In J. Metcalfe & A. P. Shimamura (Eds.), *Metacognition: Knowing about knowing* (pp. 155–184). Cambridge, MA: MIT Press.

Wells, G. L., Olson, E. A., & Charman, S. D. (2002). The confidence of eyewitnesses in their identifications from lineups. *Current Directions in Psychological Science, 11*, 151-154. doi: 10.1111/1467-8721.00189

Wessel, I., & Merckelbach, H. (1997). The impact of anxiety on memory for details in spider phobics. *Applied Cognitive Psychology, 11*(3), 223-231.

Wise, S., Pawlenko, N. B., Safer, M., & Meyer, D. (2009). What US prosecutors and defense attorneys know and believe about eyewitness testimony. *Applied Cognitive Psychology, 23,* 1266–1281.

Wright, D. B., Self, G., & Justice, C. (2000). Memory conformity: Exploring misinformation effects when presented by another person. *British Journal of Psychology, 91*, 189-203. doi: 10.1348/000712600161781

Yates, J. F. (1994). Subjective probability accuracy analysis. In G. Wright & P. Ayton (Eds.), *Subjective probability* (pp. 381-410). New York, NY: John Wiley & Sons.

Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *Journal of Applied Psychology, 71*, 291-301. doi: 10.1037/0021-9010.71.2.291