

Odinot, G., Wolters, G., & Van Koppen, P.J. (2009).
Eyewitness memory of a supermarket robbery: A case
study of accuracy and confidence after 3 months. *Law
and Human Behavior, 33*, 506-514.

Eyewitness Memory of a Supermarket Robbery: A Case Study of Accuracy and Confidence After 3 Months

Geralda Odinot · Gezinus Wolters ·
Peter J. van Koppen

Published online: 22 August 2008

© American Psychology-Law Society/Division 41 of the American Psychological Association 2008

Abstract In this case study, 14 witnesses of an armed robbery were interviewed after 3 months. Security camera recordings were used to assess memory accuracy. Of all information that could be remembered about 84% was correct. Although accurately recalled information had a higher confidence level on average than inaccurately recalled information, the mean accuracy–confidence correlation was rather modest (0.38). These findings indicate that confidence is not a reliable predictor of accuracy. A higher level of self-reported, post-event thinking about the incident was associated with higher confidence levels, while a higher level of self-reported emotional impact was associated with greater accuracy. A potential source of (mis)information, a reconstruction of the robbery broadcasted on TV, did not alter the original memories of the witnesses.

Keywords Eyewitness memory · Eyewitness accuracy · Eyewitness confidence · Case study

Eyewitnesses are very important in the criminal justice system, first during police investigations, and later as sources of evidence when a case is brought to trial. In evaluating the reports of eyewitnesses, the major concern is to determine their accuracy. Outside the laboratory, however, it is generally not possible to verify the content of witness reports objectively. In that case, the level of confidence expressed by a witness becomes a potentially useful diagnostic to discriminate between accurate and inaccurate memories. There is a widely held intuitive belief that confidence expressed about a memory can be used to infer its accuracy, both among the general public and by legal professionals (Cutler, Penrod, & Stuve, 1988; Leippe, 1980; Lindsay, Wells, & O'Connor, 1989; Luus & Wells, 1994; Penrod & Cutler, 1995). The confidence expressed by an eyewitness in his or her testimony appears to be a strong determinant of the perceived credibility of the eyewitness (Leippe, Manion, & Romanczyk, 1992; Lindsay et al., 1989). Studies examining the relationship between accuracy and confidence, however, have found low correlations in person identification tasks (e.g., Deffenbacher, 1991; Penrod & Cutler, 1995), and modest correlations in event recall (Roberts & Higham, 2002; Robinson & Johnson, 1996; Odinet & Wolters, 2006).

Although there are some exceptions (for example, Christianson & Hubinette, 1993; Fisher, Geiselman, & Amador, 1989; Read, Tollesstrup, Hammersley, McFadden, & Christensen, 1990; Woolnough & MacLeod, 2001; Yuille & Cutshall, 1986), most research on eyewitness memory is laboratory based. Because experimental designs are attractive, under laboratory conditions the accuracy of memory reports can be measured and various conditions can be manipulated. A major problem with such studies, however, is how well the results can be generalized to real-life situations. Participating in an eyewitness experiment to

G. Odinet (✉) · P. J. van Koppen
Netherlands Institute for the Study of Crime and Law
Enforcement (NSCR) Leiden, P.O. Box 792, 2300 AT Leiden,
The Netherlands
e-mail: Odinet@nscr.nl

G. Wolters
Department of Psychology, Leiden University, Leiden,
The Netherlands

P. J. van Koppen
Maastricht University, Maastricht, The Netherlands

P. J. van Koppen
Free University Amsterdam, Amsterdam, The Netherlands

gain credits or money, for instance, is a relative neutral event in the life of a student. Obviously, this stands in stark contrast with the level of stress real-life witnesses may experience. Even when conditions are created to induce some ecological validity (e.g., using complex emotional stimulus material, asking open-ended questions), important characteristics of real events (e.g., unexpectedness, emotional stress, personal involvement, and aftermath events) are lacking (Wells, Memon, & Penrod, 2006).

Only a few studies have investigated accuracy in the memories of persons who witness a real crime. Those studies can be divided into two categories based on their research method: archival and field studies (Woolnough & MacLeod, 2001). Archival studies look for patterns in the amount and the type of information that is filed in police reports. Field studies mostly focus on the consistency of memory reports in subsequent interviews with the witnesses.

Yuille and Cutshall (1986) conducted a field study in which participants who had witnessed a shooting incident were interviewed. Both police interviews and research interviews were analyzed and the witnesses appeared to be highly consistent and accurate in their accounts. Furthermore, the witnesses' perceived stress level at the time of the event appeared to have no negative effects on subsequent memory. Similarly, Christianson and Hubinette (1993) found no significant relationship between self-rated degree of emotional stress and the number of details in the memories of robbery witnesses after an extended time interval. Woolnough and MacLeod (2001) compared videotapes of an incident with the statements given by victims and bystanders to the police immediately following the incident. They find very high levels of recall accuracy in the memories of witnesses (96%). Moreover, they reported that with higher ratings of emotional impact of the incident upon bystanders, more action details were reported. Archival studies on offender descriptions were conducted by Wagstaff et al. (2003) and Van Koppen and Lochun (1997); both studies reported that witness descriptions were very likely to be more correct than incorrect. Wagstaff et al. (2003) also tested for a 'weapon focus' effect and found no significant results.

The studies described above seem to indicate that witnesses from real-life events provide consistent and accurate information in their accounts. Both archival and field studies, however, have their limitations. In archival studies, for instance, verifying perpetrator descriptions given by witnesses is sometimes impossible, because the perpetrator is not always known. Moreover, errors of omission in police reports are unknown and are therefore not included in the accuracy assessment (van Koppen & Lochun, 1997). As Macleod and Shepherd (1989) pointed out, a limitation of field studies is that witnesses willing to participate in

such studies are most likely to be more confident about what they remember, thereby inflating accuracy estimates. Moreover, most field studies to date have focused on descriptive aspects of the event (e.g., the appearance of the offender) rather than on memory of the event itself (e.g., what the offenders or other bystanders were doing). The accuracy of action details (e.g., who did what and to whom), however, is central to the judicial process (Woolnough & MacLeod, 2001). In this respect, the method employed by Yuille and Cutshall (1986) represents an important milestone in eyewitness research as they were one of the first to investigate descriptive aspects of the event, together with memory for action and person details. Finally, it has to be noted that none of the actual case studies have related accuracy with confidence.

The study presented here allowed us to overcome a number of the criticisms raised against prior field studies. It posed a rare opportunity to determine the accuracy and confidence in the memories about details of an armed robbery that was witnessed 3 months prior to testing. This was possible by comparing what was remembered by the victims with actual video-recordings of the crime. In addition, we asked the witnesses to provide confidence ratings allowing us to determine the accuracy–confidence relation. In this case study we interviewed 14 real eyewitnesses 3 months after the event took place. Accuracy–confidence correlations were calculated to see whether confidence is an indicator for memory accuracy in a real-life situation 3 months after the event. During this period the memories of the witnesses were exposed to influences as they appear in real life, like repeated recall, and exposure to co-witness information and misinformation. This case study does not test specific hypotheses, but it may provide important insights into the reliability of the memory of witnesses 3 months after witnessing a crime.

THE WITNESSED EVENT

On a Friday night, February 9, 2007, a supermarket was robbed in Gorinchem, the Netherlands. It was just after closing time, 9.04 p.m., and all customers had left the store, when a car with two men inside parked at the back of the supermarket. The back entrance of the supermarket is used for the delivery of supplies. New supplies had just arrived and approximately 28 employees were at work inside the supermarket. The two men came out of the car and walked to the back entrance. They were both armed with a gun, and wearing a balaclava. Two employees were beaten and held at gunpoint while they were told to get inside the supermarket. A group of 10 employees noticed the robbers when they came in and were able to escape. The smaller of the

two robbers ran straight to the office where the cash drawers are brought after closing time. The robber forced a young cashier to come with him and ordered her to open the safe. Meanwhile, the other robber, a very big and tall man, walked around holding his gun in his outstretched arm threatening the remaining employees.

In the office, cash drawers were placed in a large shopping bag by the robber while the cashier had to wait in a corner just outside the office. She was in great distress. The tall robber walked to the front of the office and asked if his companion was ready to go. When he was, the two ran off, carrying the large bag filled with cash drawers, jumped in their car, and drove away.

The perpetrators needed less than 3 min to take what they wanted and leave the employees behind in shock and confusion. The police, called by the employees who had been able to escape, arrived a few minutes after the robbers had left. The first statements of the witnesses were taken that evening. After talking with the police, the witnesses spoke a lot with each other about what happened, both during that night and in the days and weeks that followed. Due to the stress experienced, some witnesses underwent psychotherapy to cope with the traumatic event.

Because the police investigation made little progress, a Dutch television program about unsolved crimes, *Opspor-ing Verzocht*, featured this robbery. In the program, broadcasted on March 13, 2007, 5 weeks after the robbery, the police asked for information from the general public. During the program, descriptions and pictures were provided of how the robbers were dressed, the appearance of the bag they were carrying to collect the money trays, and what type of car the robbers had used. Moreover, a reconstruction of the robbery was shown; however, the details of this reconstruction were not completely accurate. We tried to identify effects of this television program on the memory of the witnesses.

METHOD

Participants

This study is based upon interviews with fourteen witnesses (7 males and 7 females), all employees of the supermarket. The age of the witnesses ranged from 15 to 63 years, with a mean age of 27.5. In total, 28 employees were present at the time of the robbery, however, almost half of them were able to escape or hide when the robbers entered. The 14 witnesses who agreed to an interview including the main victim, were all inside the supermarket during the robbery and were able to provide information that was recorded by the security cameras.

Videos

In the supermarket there were 16 digital security cameras installed, of which 9 recorded relevant images of the robbery. The videos were digital recordings in both black-and-white and full color, all without sound recording. The different camera positions make it possible to follow all the actions of the perpetrators and employees, who were recorded from different angles and positions, inside and at the back entrance of the supermarket. The video recordings were made available to us by the police and the store management.

Interview Procedure

The interviews were conducted by two researchers 3 months after the robbery had taken place. Both interviewers had followed interview training classes. The witnesses chose the time and location of the interviews. Every interview followed the same procedure and was recorded on audiotape.

Interviewers started with explaining the goal of the interview and attempting to make the witnesses feel comfortable. It was explained to the witnesses that everything they said would be anonymous and used for research purposes only. The goal and especially the importance of the confidence judgments were explained. The witnesses were asked to indicate the perceived accuracy of their memories on a seven-point scale, where 1 indicates very uncertain and 7 absolutely certain. The scale was visually displayed in front of the witnesses during the interview.

The witnesses were asked to think back to the night of the robbery. First, they were asked to tell in their own words what they had seen and a floor plan of the supermarket was used to illustrate the exact location and movements of the robbers and other persons. The floor plan appeared to be very helpful for remembering and describing the event. The witnesses were not interrupted during free recall, and the interviewers made notes about information that needed more clarification in a later stage of the interview. Sometimes, witnesses provided spontaneous confidence judgments, but systematic confidence judgments were only requested with follow-up questions. After a witness had finished free recall, the interviewer asked more specific questions that followed-up on the general information provided during free recall, i.e., no additional information was introduced by the interviewer. These questions focused on forensically relevant details, like a full description of the robbers, the guns, the bag used, the position and acts of the robbers, and the position and acts of the witness and his/her colleagues. These questions were open-ended like ‘Can you tell us more about....?’ or ‘You mentioned a robber, can you describe how this person

looked?’ Again, witnesses were not interrupted while answering, but they were asked to provide corresponding confidence judgments after answering the question.

After the interview, the witnesses were asked the following three questions: ‘Have you talked with other people about the robbery?’, ‘Did you ever think back about the robbery?’, and ‘How much did the incident affect you emotionally?’ Answers to these questions again had to be indicated on a 7-point scale, where 1 indicates ‘never’ (questions 1 and 2) or ‘not at all’ (question 3), and 7 ‘very often’ (questions 1 and 2) or ‘very much’ (question 3). Witnesses were also asked when and how often had they been interviewed by the police, and if they had watched the television program about the robbery. The duration of the interviews ranged from 12 to 45 min, with a mean of 28 min. The main factor determining duration was the amount of information that the witnesses could provide, which depended on the position and the role of the witness during the robbery.

Scoring Procedure

We used a scoring procedure set out by Yuille and Cutshall (1986) in which statements are parsed into separate units of information. First, after transcribing the audio recordings, repetitions and hesitations were removed. Second, statements about speech, noises, or sounds were removed because the security videos were without sound recording, and it was, therefore, not possible to score the accuracy of this information. Then, the remaining statements were parsed into single units of information. For example, the statement ‘one robber had a black gun’ contains two separate units of information: ‘one robber had a gun’ and ‘the gun was black.’ In some cases, witnesses provided multiple details in one sentence with only one confidence judgment. In these cases, the units of information given in these sentences all received the same confidence score. Next, each unit of information was classified in one of the three types of information: (a) person descriptions (i.e., details concerning the appearance or location of people), (b) object descriptions (i.e., details concerning the appearance or location of objects), and (c) action details (i.e., details related to all actions). Any discrepancies in coding were agreed upon by the two researchers after further discussion.

Accuracy was scored by two independent judges who compared each unit with the information on the security videos. Information given by the witnesses that could not be verified as correct or incorrect from the security video was kept out of the analyses. A Cohen’s Kappa coefficients showed that the inter-rater reliability was high, $\kappa = .91$. The few units on which the judges disagreed, even after conferring, were removed from the analyses.

RESULTS

We were interested in measuring the accuracy of the witness statements using objective records, 3 months after the witnessed event. Therefore, the accuracy level of the statements is analyzed first, followed by the analysis of the confidence judgments. Then, the memory mistakes and the content of the television program in combination with the memory statements are described on a qualitative level. Finally, the effect of post-event thinking, speaking, and the emotional impact on the level of accuracy and confidence is analyzed (Table 1).

Number of Details

The witnesses reported a total of 1,485 units of information, of which 84% were accurate. Of these units, 726 were given during the original free recall and 759 during the subsequent more specific questioning. Units provided during free recall were significantly more often accurate (90%) than units provided with specific questions (78%), ($\chi(1) = 39.1, p < .01$).

The number of units provided by individual witnesses ranged from 22 to 204, and the accuracy rates ranged from 0.75 to 0.97. For a further analysis, we separated the witnesses in two groups: witnesses who were directly involved in the events and who were interviewed by the police (the ‘central’ witnesses, $N = 9$), and ‘peripheral’ witnesses who were less involved and were not questioned by the police ($N = 5$). As expected, central and peripheral witnesses differ in the mean number of units of information they provided. The central witnesses recalled significantly more units of information ($M = 129.3$) than the peripheral witnesses ($M = 64.2$), ($t(12) = 3.2, p < .05$, effect size $r = .55$). Accuracy of recall in both groups, however, did not differ. The proportion of correct information recalled by the central witnesses ($M = .84$) was the same as for the peripheral witnesses ($M = .84$).

Table 1 Total units of information provided by the central and peripheral witnesses per category and the proportions correct

		Central witnesses	Peripheral witnesses
Person descriptions	Units	490	133
	Proportion correct	.84	.81
Object descriptions	Units	261	63
	Proportion correct	.82	.85
Action details	Units	413	125
	Proportion correct	.85	.85
All details $N = 14$	All units	1164	321
	Proportion correct	.84	.84

An ANOVA on the number of recalled details over the categories showed a significant effect ($F(2, 39) = 3.5, p < .05$). Post-hoc test (Bonferroni, $p > .05$) showed that all witnesses recalled significantly more people details ($M = 44.5$) than object details ($M = 23.14$; effect size $r = .86$). The number of action details ($M = 38.49$) did not differ significantly from either people or object details. No differences were found in the accuracy levels of the categories ($F(2, 39) = .28, NS$).

Confidence

Because no confidence judgments were asked in free recall, these judgments are only available for answers to the subsequent specific questions. In total, confidence judgments were available for 759 units of information, with a range of 12–92 for the individual witnesses, and a mean of 63.7 units for the central and 37.2 units for the peripheral witnesses.

A paired t -test, calculated over the mean confidence scores for each witness showed that witnesses were significantly more confident about correctly recalled units of information ($M = 6.11$) than incorrectly recalled units ($M = 5.63$), ($t(13) = 3.17, p < .01$, effect size $r = .68$).

Accuracy–Confidence Relations

To be able to analyze accuracy–confidence relations, we determined the number of correct and incorrect units of information recalled as a function of confidence level for each witness. Goodman–Kruskal gamma correlations between accuracy and confidence were calculated for each witness, and over all data. Because one witness had only provided confidence judgments on correct information, it

was not possible to calculate an individual accuracy confidence correlation for this subject.

The gamma correlations per subject ranged from 0.09 to 0.96 with an average of 0.38. A gamma correlation was also determined over the pooled data of all witnesses. This correlation (0.29) was slightly lower than the average over individual subjects.

Table 2 shows the distribution of the proportion of correct information as a function of confidence expressed by the witnesses. From this table, it can be inferred that the proportion of correct units of information increases with higher levels of confidence. Most answers (about 60%) are given with the maximum level of confidence. Overall, 78% of the units of information provided were correct, and of the information that was recalled with maximum confidence, 84% was correct. Still, a substantial proportion of the answers, 16%, given with the highest level of confidence are incorrect.

Another way of interpreting the information in Table 2 is to note that the accuracy rates vary from 0.63 to 0.84 across the whole confidence scale. This reflects underconfidence at the low end of the scale and overconfidence at the high end of the scale, a finding often reported in studies using calibration to express the accuracy–confidence relationship (e.g., Brewer & Wells, 2006; Juslin, Olsson, & Winman, 1996).

Memory Errors

The memory errors the witnesses made were diverse. Some mistakes may have their origin in making assumptions. For instance, when a pay desk is not in use in the supermarket, it is closed with a small gate. For some witnesses this knowledge may have been enough to presume that one of

Table 2 Total units of information and proportion correct for each confidence level per category

		Confidence scale							Total units of information
		Low						High	
		1	2	3	4	5	6	7	
Central witnesses $N = 9$	Person descriptions	2	2	8	9	15	19	81	136
	Prop. correct	0.5	0.5	0.37	0.78	0.47	0.74	0.79	0.71
	Object descriptions	3	7	13	21	35	38	158	275
	Prop. correct	0.66	0.71	0.77	0.76	0.69	0.79	0.89	0.83
	Action details	0	3	6	9	22	35	87	162
	Prop. correct	0	0.67	0.5	0.67	0.82	0.63	0.82	0.76
Peripheral $N = 5$	All details	0	3	3	17	24	9	129	185
	Prop. correct	0	0	0.34	0.59	0.86	0	0.81	0.81
All witnesses $N = 14$	All details	5	14	32	56	96	101	455	759
	Prop. correct	0.67	0.72	0.63	0.7	0.83	0.74	0.84	0.78

the robbers jumped over the gate when they left. The gate of the specific pay desk, however, was open.

Other mistakes concerned mixing of details that had been witnessed. Both robbers were described as big, but one of them was clearly taller than the other, and all witnesses talked about the tall one and the shorter one. Both carried a weapon: the tall man had silver colored gun and the short man carried a black colored gun. Although the guns were clearly visible to most witnesses, some witnesses mixed up the color of the gun with the wrong robber.

While all witnesses remembered the correct day on which the robbery took place (a Friday), the exact date appeared to be difficult to remember. Just 4 witnesses answered this question correctly. This is in line with the findings of Wagenaar (1986) showing that the exact date of events is quickly forgotten, and with findings showing that timing of events is often better on the basis of local temporal schemata (like days of the week) than on the basis of specific dates (Friedman, 2004).

Effects of Post-event Information in a Reconstruction

Because the investigation made no progress, a television program about unsolved crimes featured this specific robbery. We asked all witnesses if they had seen this program. All, except three witnesses, had seen the program. During the program, a reconstruction of the robbery was shown and descriptions were given (and pictures shown) of the robbers, how they were dressed, how the specific bag they were carrying to collect the money trays looked like, and which car they had used.

To determine the effect of seeing the television program on later recall, we compared the data provided by the witnesses who did and did not see the program. We compared the average number of details ($M = 111.6$ for the non-viewers and $M = 104.5$ for the viewers), the proportion of accurately recalled ($M = 0.85$ for the non-viewers and $M = 0.84$ for the viewers), the confidence in the details inaccurately recalled ($M = 5.19$ for the non-viewers and $M = 5.70$ for the viewers), and finally the confidence for the accurate details ($M = 6.12$ for the non-viewers and $M = 6.10$ for the viewers). None of these averages differed significantly.

Further evidence supporting the claim that seeing the program did not greatly affect later recall was derived from an analysis of the fate of a few details that were specifically mentioned. One of these details was the presence of white stripes on the jacket of one of the robbers. Although this was explicitly shown in a picture, none of the witnesses recalled this detail. Another highlighted detail was the shopping bag the robbers used to carry the money drawers. This bag was shown in a picture from the security videos, and a look-alike bag was standing in front of the desk of the

presenters of the program. However, the witnesses who had seen the program still made mistakes about the colors, shape, and print on the bag.

Other interesting observations come from the fact that the reconstruction was incorrect on a few details. For security reasons, the exact location of the safe was changed. This alternation, however, was so evident to all employees that no one made a mistake about this issue. The reconstruction was also incorrect concerning the truth about the location and position of the cashier. In the reconstruction, the cashier was shown sitting on her heels inside the office, while the security cameras show that the exact location of the cashier was just outside the office, standing with her back against the wall. The two witnesses, who mentioned that they saw her during the robbery, were not influenced by the reconstruction. Both explicitly mentioned that the cashier was standing instead of sitting and that she was outside the office instead of inside.

In sum, we have found no source monitoring errors related to seeing a reconstruction of the robbery. Incorrect information was not recalled, and correct information (some of which was mentioned several times by the presenters or explicitly shown in a picture) did not clearly affect the accuracy or confidence of the memories of the witnesses who saw the program, when contrasted against witnesses who did not see it. Apparently, an original memory record of a significant event is not easily altered by seeing a staged reconstruction of the event.

Post-event Speaking and Thinking, and Emotional Impact of the Event

After finishing the interview the witnesses were asked how many times they had been interviewed by the police. They were also asked to rate on a 7-point scale how much the incident had affected them emotionally, and how often they had thought back and spoken with others about the robbery.

Four of the fourteen witnesses we have interviewed gave a statement to the police on two occasions, the evening of the robbery and the next day at the police office. These four witnesses were the owner of the supermarket, the store manager, and two employees of which one had been hit by a robber. Five witnesses were interviewed once by the police, including the cashier. She gave an extensive interview later on the evening of the robbery. The police had separated her from the other witnesses to avoid exchanging information. Five witnesses had not spoken to the police at all. They were not as closely involved in the event (e.g., standing at a greater distance) as the interviewed witnesses. These five peripheral witnesses also recalled less units of information on average ($M = 64.2$) than the witnesses who were interviewed once ($M = 128$) or twice ($M = 132$).

All witnesses indicated that they had spoken very often about the robbery. This made it impossible to determine any differential effect on accuracy and confidence of recall. Witnesses differed, however, in their answers to the questions about how often they thought back about the robbery and to what extent the robbery had affected them emotionally. It is possible that a high emotional impact and post-event thinking are closely related, but there was no significant correlation between the answers to both questions ($\tau = 0.20$).

To analyze the effect of post-event thinking, the witnesses were divided into a group with high scores (5 and higher) and a group with low scores (4 and lower). An independent *t*-test showed that the group that indicating that they “think back often” about the robbery was significantly more confident than the other group, both on correctly recalled units ($M = 6.31$ and $M = 5.81$, respectively, $t(12) = 2.18$, $p < .05$, $r = .53$), and on incorrectly recalled units ($M = 6.13$ and $M = 4.88$, respectively, $t(12) = 2.55$, $p < .05$, $r = .59$). The groups did not differ, however, in the number of details recalled and the proportion of correctly recalled details.

For an analysis of the effect of emotional impact, the witnesses were again divided into a group with high scores (5 and higher) and a group with low scores (4 and lower). On average, women reported more emotional stress ($M = 5.8$) than men ($M = 3.4$). We do not know, however, to what extent this difference may be due to a gender bias in reporting emotion.

Although the group indicating lower emotional impact recalled less ($M = 87.1$) than the high emotional group ($M = 125.0$), the difference was not significant ($t(12) = 1.24$, *NS*). An independent *t*-test showed, however, that the level of accuracy differed significantly between the low-emotional impact group ($M = 0.81$) and the high emotional group ($M = 0.88$, $t(12) = 2.83$, $p < .05$, $r = .63$). In other words, the group who indicated that the robbery had a high emotional impact appeared to be more accurate than the group indicated less emotional impact. No significant effects were found for emotional impact on confidence.

One could argue that the high-emotion witnesses may have been more closely involved or may have had a better view on the ongoing event than the low emotional witnesses. The central and the peripheral groups of witnesses, however, indicated similar levels of emotional impact ($M = 4.44$ and 4.60 , respectively, $t(12) = -.147$, *NS*).

DISCUSSION

The availability of video footage and the cooperation of all people involved allowed us to investigate the accuracy and confidence in the recall of details of an actual crime by a

group of eyewitnesses after 3 months. The main findings are that: (a) details provided in initial free recall are more accurate than details recalled in subsequent questioning, (b) about 84% of all remembered information was correct, and (c) correctly recalled details on average have a higher confidence than incorrectly recalled details. The distribution of correct and incorrect recalled units as a function of confidence shows an increase in accuracy with increasing confidence, but the accuracy–confidence relationship is rather modest as indicated by an average within subject correlation of 0.38.

Although these findings are significant, it has to be noted that their forensic usefulness is limited because all effects are a matter of degree, and they do not allow strong inferences. Free recall is more accurate than subsequent cued recall, but still about 10% of the details provided are incorrect. Most details remembered are correct, but even closely involved witnesses sometimes provide inaccurate details. Details remembered with high confidence are more often correct than details remembered with less confidence. However, even the maximum level of confidence does not guarantee accuracy, and the accuracy–confidence correlation is modest.

Interestingly, the accuracy and confidence findings in this study rather closely follow the pattern of results found in a laboratory study (Odinot & Wolters, 2006). In this study, participants watched a video of a complex event and were tested with cued recall questions about details 1, 3, or 5 weeks later. Also in this study, accuracy rates after 5 weeks were about 80%. Confidence was higher for correct than incorrect details, and a modest (although somewhat higher than in the present study) accuracy–confidence correlation was found. Yuille and Cutshall (1986), who interviewed their witnesses with a 4–5 months delay, reported an overall accuracy of 84.5% for central witnesses and 79.3% for peripheral witnesses. The striking similarity between the results of the present field study, the findings of Yuille and Cutshall (1986), and our previous laboratory study indicates a consistent pattern of results that may be generalizable to other situations where people have to recall details of a complex event after weeks or months.

One particular feature of the present study was that the witnesses are all colleagues who interact on an almost daily basis. It is likely that the witnesses have extensively discussed the event under study, and indeed all witnesses indicated having talked about the event very often. When eyewitnesses discuss an event, they may influence each other, and, in subsequent recall, report what they heard from others. This phenomenon has been described as collaborative storytelling (Crombag, 1999) or memory conformity (Gabbert, Memon, & Allen, 2003; Gabbert, Memon, Allen, & Wright, 2004). Nevertheless, the statements of the witnesses in this study still showed a large

variation in the amount of recalled information and in the number and variety of memory mistakes. This gives the impression that the memories of the witnesses are not heavily affected by the effect of memory conformity. Unfortunately, we were unable to determine the effects of collaborative storytelling more thoroughly. Moreover, watching a reconstruction of the robbery on television did not enhance or alter the original memories of the witnesses.

Post-event thinking did not affect accuracy, but it did enhance confidence (both for correct and incorrect answers). Such confidence inflation has been reported earlier by Shaw (1996) and Wells and Bradfield (1999). In both studies, participants who engaged in reflective thought about their previously given answers showed confidence inflation. This process may be similar to what occurs when repeatedly thinking about an imaginary event leads to false, but confident, memories (Ceci, Huffman, Smith, & Loftus, 1994; Roediger, Jacoby, & McDermott, 1996; Ryan & Geiselman, 1991). Confidence inflation has also been found as a result of repeated recall attempts (Shaw, 1996; Shaw & McClure, 1996), although this could not be corroborated in other studies (Ebbesen & Rienick, 1998; Odinet & Wolters, 2006; Turtle & Yuille, 1994).

Although the question about emotions was meant to ask for emotion at the time of the crime, we cannot rule out that some witnesses have interpreted the question as referring to post-event emotion. Concerning the effect of emotional stress, we found that high levels of self-reported emotional impact had a significant effect on the accuracy of recalled details. Higher levels of emotion also resulted in a larger number of recalled details, but this effect was not significant. Woolnough and MacLeod (2001), however, reported a significant effect of emotional impact on the number of (action) details reported, but they did not find an effect of emotion on accuracy. These findings, and a review of the literature, clearly indicate a complex relationship between emotion and memory. Emotion can have both positive and negative effects on memory, and this may lead to contradictory findings. For example, Christianson and Hubinette (1993) and Yuille and Cutshall (1986) concluded that emotional stress had no negative effect on the recall of the details of a crime. A meta-analytical review by Deffenbacher, Bornstein, Penrod, and McGorty (2004), however, found considerable support for the hypothesis that high levels of emotional stress have a negative effect on the recall of details of a crime.

In an attempt to account for the data on memory and emotion, Reisenberg and Heuer (2007) concluded that emotion promotes memory of the central parts of an event, but it also makes people less likely to notice, and less likely to recall, information that is more peripheral in an event. Indeed, in our study, most details recalled were related to what might be called central aspects of the situation (e.g.,

descriptions of the guns, person, and action details pertaining to the robbers). Observing and monitoring such details probably is most relevant for surviving a threatening situation (Woolnough & MacLeod, 2001). One could argue that the high-emotion witnesses may have been more closely involved or may have had a better view on the ongoing events than the low-emotion witnesses. However, the central and peripheral groups of witnesses in this study did not indicate different levels of emotional impact.

The robbery used in this study is an ordinary case, and the witnesses represent ordinary people. Therefore, our study is a good example of how memory of a crime fares over time. It is also an example of the potential fruitfulness of collaborations among memory researchers and law enforcement professionals (see, e.g., Cutler & Bull Kovera, 2008). As is clear from the results, most of the information remembered by the witnesses was correct. Still, a substantial proportion was incorrect. Moreover, it is clear that confidence cannot be used to distinguish clearly between accurate and inaccurate memories. Confidence may be used as a cautious indicator for accuracy during police investigations (e.g., Odinet & Wolters, 2006), but it should never be allowed as evidence for memory accuracy in the courtroom.

Acknowledgments The authors would like to thank the Dutch police and the management of the Plus Supermarket in Gorinchem, The Netherlands, for their cooperation. We would also like to thank Marleen Pronker who assisted with collecting and scoring the data. This study would not have been possible without the cooperation of the witnesses. Many thanks to them for telling their story!

REFERENCES

- Brewer, N., & Wells, G. L. (2006). The confidence–accuracy relationship in eyewitness identification: Effects of lineup instructions, foil similarity, and target-absent base rates. *Journal of Experimental Psychology Applied*, 12, 11–30. doi:10.1037/1076-898X.12.1.11.
- Ceci, S. J., Huffman, M. L. C., Smith, E., & Loftus, E. F. (1994). Repeatedly thinking about a non-event: Source misattribution among preschoolers. *Consciousness and Cognition*, 3, 388–407. doi:10.1006/ccog.1994.1022.
- Christianson, S.-Å., & Hubinette, B. (1993). Hands up! A study of witnesses' emotional reactions and memories associated with bank robberies. *Applied Cognitive Psychology*, 7, 365–379. doi:10.1002/acp.2350070502.
- Crombag, H. F. M. (1999). Collaborative story-telling: A hypothesis in need of experimental testing. *Psychology, Crime & Law*, 5, 279–289. doi:10.1080/10683169908401770.
- Cutler, B. L., & Bull Kovera, M. (2008). Introduction to commentaries on the Illinois pilot study of lineup reforms. *Law and Human Behavior*, 32, 1–2. doi:10.1007/s10979-007-9120-x.
- Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision making in eyewitness identification cases. *Law and Human Behavior*, 12, 41–55. doi:10.1007/BF01064273.
- Deffenbacher, K. A. (1991). A maturing of research on the behavior of eyewitnesses. *Applied Cognitive Psychology*, 5, 377–402. doi:10.1002/acp.2350050502.

- Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, K. E. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior*, 28, 687–706. doi:10.1007/s10979-004-0565-x.
- Ebbesen, E. B., & Rienick, C. B. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *The Journal of Applied Psychology*, 83, 745–762. doi:10.1037/0021-9010.83.5.745.
- Fisher, R. P., Geiselman, E. R., & Amador, M. (1989). Field test of the cognitive interview: Enhancing the recollection of actual victims and witnesses of crime. *The Journal of Applied Psychology*, 74, 722–727. doi:10.1037/0021-9010.74.5.722.
- Friedman, W. J. (2004). Time in autobiographical memory. *Social Cognition*, 22, 591–605. doi:10.1521/soco.22.5.591.50766.
- Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology*, 17, 533–543. doi:10.1002/acp.885.
- Gabbert, F., Memon, A., Allan, K., & Wright, D. B. (2004). Say it to my face: Examining the effects of socially encountered misinformation. *Legal and Criminological Psychology*, 9, 215–227. doi:10.1348/1355325041719428.
- 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.
- Leippe, M. R. (1980). Effects of integrative memorial and cognitive processes on the correspondence of eyewitness accuracy and confidence. *Law and Human Behavior*, 4, 261–274. doi:10.1007/BF01040618.
- Leippe, M. R., Manion, A. P., & Romanczyk, A. (1992). Eyewitness persuasion: How and how well do fact finders judge the accuracy of adults' and children's memory reports? *Journal of Personality and Social Psychology*, 63, 181–197. doi:10.1037/0022-3514.63.2.181.
- Lindsay, R. C. L., Wells, G. L., & O'Connor, F. J. (1989). Mock-juror belief of accurate and inaccurate eyewitnesses. *Law and Human Behavior*, 13, 333–339. doi:10.1007/BF01067033.
- Luus, E. C. A., & Wells, G. L. (1994). Eyewitness identification confidence. In D. F. Ross, J. D. Read & M. P. Toglia (Eds.), *Adult eyewitness testimony* (pp. 348–361). Cambridge: Cambridge University Press.
- MacLeod, M. D., & Shepherd, J. W. (1989). Interviewing victims of crime. In E. Viano (Ed.), *Crime and its victims: International research and public policy*. Washington, DC: Hemisphere Publishing Corporation.
- Odinot, G., & Wolters, G. (2006). Repeated recall, retention interval and the accuracy–confidence relation in eyewitness memory. *Applied Cognitive Psychology*, 20, 973–985. doi:10.1002/acp.1263.
- Penrod, S. D., & Cutler, B. L. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law*, 1, 817–845. doi:10.1037/1076-8971.1.4.817.
- Read, J. D., Tollsstrup, P. A., Hammersley, R., McFadzen, E., & Christensen, A. (1990). The unconscious transference effect: Are innocent bystanders ever misidentified? *Applied Cognitive Psychology*, 4, 3–31. doi:10.1002/acp.2350040103.
- Reisenberg, D., & Heuer, F. (2007). The influence of emotion on memory in forensic settings. In R. C. L. Lindsay, D. F. Ross, J. D. Read & M. P. Toglia (Eds.), *Handbook of eyewitness psychology* (Vol. 1, pp. 81–116). Mahwah, NJ: Lawrence Erlbaum Associates, publishers.
- 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. (1996). Recall memory, recognition memory, and the eyewitness confidence–accuracy correlation. *The Journal of Applied Psychology*, 81, 587–594. doi:10.1037/0021-9010.81.5.587.
- Roediger, H. L., Jacoby, J. D., & McDermott, K. B. (1996). Misinformation effects in recall: creating false memories through repeated retrieval. *Journal of Memory and Language*, 35, 300–318. doi:10.1006/jmla.1996.0017.
- Ryan, R. H., & Geiselman, E. R. (1991). Effects of biased information on the relationship between eyewitness confidence and accuracy. *Bulletin of the Psychonomic Society*, 29, 7–9.
- Shaw, J. S. (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., & 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.
- Turtle, J. W., & Yuille, J. C. (1994). Lost but not forgotten details: Repeated eyewitness recall leads to reminiscence but not hypernesia. *The Journal of Applied Psychology*, 79, 260–271. doi:10.1037/0021-9010.79.2.260.
- van Koppen, P. J., & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior*, 21, 661–685. doi:10.1023/A:1024812831576.
- Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over 6 years. *Cognitive Psychology*, 18, 225–252. doi:10.1016/0010-0285(86)90013-7.
- Wagstaff, G. F., MacVeigh, J., Boston, R., Scott, L., Brunas-Wagstaff, J., & Cole, J. (2003). Can Laboratory findings on eyewitness testimony be generalized to the real world? An archival analysis of the influence on violence, weapon presence, and age on eyewitness accuracy. *The Journal of Psychology*, 137, 17–28.
- Wells, G. L., & Bradfield, A. L. (1999). Distortions in eyewitnesses' recollections: Can the postidentification-feedback effect be moderated? *Psychological Science*, 10, 138–144. doi:10.1111/1467-9280.00121.
- Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, 7, 45–75. doi:10.1111/j.1529-1006.2006.00027.x.
- Woolnough, P. S., & MacLeod, M. D. (2001). Watching the birdie watching you: Eyewitness memory for actions using CCTV recordings of actual crimes. *Applied Cognitive Psychology*, 15, 395–411. doi:10.1002/acp.717.
- Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *The Journal of Applied Psychology*, 71, 291–301. doi:10.1037/0021-9010.71.2.291.