Eyewitness confidence : the relation between accuracy and confidence in episodic memory
Odinot, G.

2

Repeated recall, retention interval and the accuracy - confidence relation in eyewitness memory*

* This chapter is published as: Odinot, G., & Wolters, G. (2006). Repeated recall, retention interval and the accuracy-confidence relation in eyewitness memory. *Applied Cognitive Psychology*, 20, 973-985. http://dx.doi.org/10.1002/acp.1263

Summary

People can evaluate the quality of their memories by giving a confidence judgment concerning the perceived accuracy of what is recalled or recognized. Even when people strive for accuracy and claim great confidence they may, however, not remember what actually happened. Both accuracy and confidence can be affected by various factors. In this study we investigated the effects of retention interval (either 1, 3 or 5 weeks delay before first testing) and of repeated questioning (initial recall after 1 week, repeated after 3 and 5 weeks) on accuracy and confidence of recall of a naturalistic videotaped event. Longer retention intervals before initial testing resulted in lower accuracy and lower confidence scores. Repeated recall, however, had little effect on accuracy and confidence. Relatively high accuracy-confidence correlations were found in all delay and repetition conditions. Practical implications of these findings for questioning eyewitnesses are discussed.

Introduction

In reconstructing the exact nature of events, like crimes and accidents, witness reports are often essential because other records are lacking. Witness reports are a major source for fact or truth finding in police investigations, and the testimony of actual witnesses carries considerable weight in the outcome of criminal and civil trials. A substantial body of research on memory for everyday events has made it abundantly clear that these memories are fallible and prone to errors. Many variables affect the accuracy of memory, supplementing or altering it, or even more dramatically, creating conditions where people can be made to believe they remember events that never happened (e.g., Deffenbacher, 1991; Wells & Loftus, 2003).

One of the factors that may contribute to inaccuracies in memory-based reports is repeated recall. This may create problems in real-life situations, such as crime investigations. In large criminal investigations repeated interviews are almost inevitable. A standard scenario is that the police initially questions witnesses for a first-hand account. If the witness has important information, he or she is likely to be questioned again by the police, and by prosecutor or defense lawyers, over subsequent weeks or months. Finally, the witness may be called to the stand to present their recollection of the event when a case is brought to trial.

One of the reasons to question witnesses several times is that witnesses may provide new information during follow-up questioning. Information that could not be remembered initially may be remembered later. However, repeated recall may also introduce distortions of memory. From a review of the literature, Roediger, McDermott, and Goff (1997) concluded that repeated recall can have both facilitating and detrimental effects on later retention. To understand the effects of repeated recall it is important to note that retrieval is not a neutral process, leaving memory unaffected. Rather, probing memory and (re)activating memory

traces is itself a learning experience. It is an active process that selectively strengthens or alters the contents of memory thus irrevocably affecting future retention (Bjork, 1975).
Prolonged retrieval periods and repeated retrieval attempts may lead to the recall of previously inaccessible memories. This phenomenon has been shown in laboratory studies under the headings of reminiscence, spontaneous recovery, or hypermnesia (e.g., Roediger et al., 1997; Scrivner & Safer, 1988; Turtle & Yuille, 1994). After studying lists of unrelated items, multiple subsequent retrieval attempts cause cumulatively more items to be remembered, although the absolute number of remembered items in each following attempt is likely to decrease. Gains have also been reported for more naturalistic stimuli, like videotaped events (Scrivner & Safer, 1988) and remembering names of former classmates (Williams & Hollan, 1981). These gains are possibly due to a tendency to recall items that became inaccessible during a former retrieval session (Raaijmakers & Shiffrin, 1981), or to the dissipation of inhibitory effects of retrieval practice over sessions (Levy & Anderson, 2002).
Recall of information generally increases the likelihood that it is recalled again later. So, retrieval consolidates memory, either by strengthening a memory trace, or by linking it to additional retrieval cues. However, not only correct, but also incorrect information that has been recalled before is more likely to be remembered in subsequent retrieval attempts. For instance, Roediger, Jacoby, and McDermott (1996) showed that incorrect recall of misinformation given after watching a series of slides, increases the likelihood that it is recalled again in subsequent tests. Subjects also became more certain that the incorrectly recalled information was correct, as was shown by an increase of the probability that it was judged as 'remembered' (instead of 'known'). Apparently, recall of incorrect information also makes it more easily accessible, causing it to be remembered with increasing confidence. Other studies have shown that if participants are forced to guess on a first test, they tend to accept these guesses as true memories on later test (e.g., Roediger, Wheeler, & Rajaram, 1993). Repeated attempts to recall imagined or suggested information have even been shown to be a powerful force for the creation of false memories (e.g., Ceci, Huffman, Smith, & Loftus, 1994; Hyman, Husband, & Billings, 1995).
Incorrect information may come from several sources, both internally (e.g., by guessing or imagination) or externally (e.g., information provided or suggested by others). As was shown by Loftus (1979), externally provided misinformation is easily integrated into the memory of an original event, and it becomes impossible for subjects to distinguish between original information and later presented misinformation. Obviously, repeated retrieval and longer retention periods increase the chances that new (and possibly erroneous) information is received from other sources and is integrated in a memory, causing source monitoring errors (Johnson, Hashtroudi, & Lindsay, 1993).
In evaluating the reports of eyewitnesses, the major concern is to determine their accuracy. However, outside the laboratory 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 memory. There is a widely held intuitive belief that confidence can be used to infer accuracy, both in the general public as well as by legal professionals (Cutler, Penrod, & Stuve, 1988; Leippe, 1980; Lindsay, Wells, & O'Connor, 1989; Luus & Wells, 1994; Penrod & Cutler, 1995; Wise & Safer, 2004). A large body of research, however, has shown that the relationship between confidence and accuracy is far from perfect. Meta-analyses of studies on eyewitness identifications found that the average correlation between confidence and accuracy tends to be relatively small, on the order of 0.25 (e.g., Bothwell, Deffenbacher, & Brigham, 1987; Sporer, Penrod, Read, & Cutler, 1995).

Various reasons have been suggested as to why experimental studies on identification may underestimate the relation between accuracy and confidence, such as impoverished viewing conditions resulting in a homogeneous data set, determining correlations between instead of within subjects, and the use of forced-choice paradigms (e.g., Busey, Tunnicliff, Loftus, & Loftus, 2000; Lindsay, Read, & Sharma, 1998; Olsson, 2000; Smith, Kassin, & Ellsworth, 1989). It may be noted that studies addressing the accuracy-confidence relation in recall or general knowledge recognition tasks have shown somewhat higher correlations (e.g., Bornstein & Zickafoose, 1999; Perfect, Watson, & Wagstaff, 1993; Robinson & Johnson, 1996).

Although various aspects of repeated recall have been studied quite extensively, surprisingly few studies have addressed repeated recall of complex naturalistic events and the accompanying confidence judgments over the course of a relatively long time interval. We only found two studies investigating this situation with recall tasks. Turtle and Yuille (1994) repeatedly tested participants with a recall task concerning memories for a videotaped mock crime. Their data showed that participants in the immediate recall group were more accurate and more confident than the 3-week delay group. Repeated testing resulted in the recall of some additional information across attempts, but the net amount of recall was highest in immediate testing and dropped over a 3-week interval. The authors reported that repeated recall of the same information did not enhance confidence, but quantitative relations between accuracy and confidence were not reported.

Slightly different results were reported by Ebbesen and Rienick (1998). The participants in their study listened to a story read aloud by an unfamiliar person. Memory was tested after 1, 7, and 28 days by asking for recall of story details. They did not find recall of additional information over recall attempts. The mean number of correct story facts remained the same over recall attempts. Mean confidence did not change over repeated testing, either.

Other studies used naturalistic stimulus material, but measured retention with forced-choice recognition tasks. In a study by Shaw and McClure (1996; see also Shaw, 1996) participants witnessed a staged interruption of a classroom meeting. Recognition tests were given after different intervals, followed by a final test after four weeks. Again, repeated testing did not lead to increased accuracy, but it did increase confidence both for correct and incorrect answers, probably due to enhanced retrieval fluency. The correlation between accuracy and confidence was low and not significant in the first test, and dropped even further when the test was repeated. Calibration curves showed that participants were generally overconfident and became more so with repeated tests. From these results Shaw and McClure (1996) concluded that 'unnecessarily repetitive witness questioning that characterizes many criminal investigations must be minimized'.

Some investigators have used immediate memory tests and focused on delayed confidence judgments. Allwood, Ask and Granhag (2005) found a high level of accuracy in an immediate interview recall procedure and high levels of confidence after a delay of 2 weeks. Data showed good calibration and very little overconfidence. This finding stands in contrast with the overconfidence found in a study with the same design but using a recognition task (Granhag, Jonsson, & Allwood, 2004). Roberts and Higham (2002) measured immediate recall of a videotaped staged crime. One week later, the previously recalled information was presented again in small units that were to be rated for confidence. Accurately recalled units were given higher confidence scores. This effect was stronger for units judged to be relevant for a criminal investigation than for irrelevant units. The accuracy-confidence correlations for relevant and irrelevant information units were 0.63 and 0.36, respectively.

In sum, relatively few studies have simultaneously addressed the effect of repeated recall of naturalistic events and systematically related accuracy and confidence measures. Moreover, the results reported are not completely consistent. The present study aims at examining the effect of repeated recall under conditions that resemble the situation of eyewitnesses in real life. Participants were shown a videotape of an extended natural event. Subsequently they were asked to recall as much as possible in a cued recall task and to rate their confidence in the accuracy of the answer. The cues consisted of open-ended questions that did not need to be answered if the participant did not remember. The initial test was given after 1 week and was repeated after 3 and 5 weeks. To gain more insight into the relationship of delay and repetition effects, control groups received the test twice (after 3 and 5 weeks) or only once (after 5 weeks). In all groups accuracy-confidence relationships were determined.

Method

Participants

A group of 67 undergraduate students (50 female and 17 male) were recruited through publication board announcements and by a computerized sign-up system. Participants were randomly assigned to one of three conditions. All received either course credits or were paid between 10 and 20 Euros, depending on the experimental condition in which they participated.

Design

The participants were randomly assigned to the three conditions. Condition 1 (n=23) consisted of three recall sessions 1, 3 and 5 weeks after the video presentation, Condition 2 (n= 24) consisted of two recall sessions 3 and 5 weeks after the video presentation. Condition 3 (n=20) had only one recall session, 5 weeks after the video presentation.

Materials

Videotape. A 21 minutes long videotape, previously broadcasted on the Dutch television, was shown individually to the participants on a high quality 17-inch computer screen. The video depicts two storylines; one of a man who is helping a neighbor to get some things from a shop, and the other of a young man who recently received a motor-bike for his birthday. The two storylines converge in an accident between the car and the motorbike at the end of the video.

Questionnaires. For the recall sessions, a questionnaire was constructed consisting of 23 open-ended questions. The questionnaire started with a very general question wherein the participants were asked to describe the two story lines in general terms. This question was asked in order to reinstate and refresh the memory of the video before proceeding with the more specific questions. The other 22 questions were all open-ended recall questions concerning several aspects of the video. Some questions were cued more specifically (e.g., "describe the car of the man") than others (e.g., "give a full description of the accident and try to be as complete as possible"). The questionnaire for all recall sessions in all conditions was identical.

Procedure

During the first session, participants watched the video individually on a computer monitor. They were told to pay attention because they would have to recall the event later. In the initial and subsequent recall sessions participants were instructed to try to recall information from the original video. They were told to imagine that they where the only witnesses, and that it was important therefore to answer as accurate as possible about details they remembered from the original video-presentation. It was also stressed that if they could not remember the answer from the video, they should refrain from answering by indicating "do not know".

To allow a fine-grained analysis of the recall data, participants were instructed to write the answers to the questions in small units of information. A unit was described as a single element or aspect of information. In practice this was realized by providing participants a series of lines on the answering sheet. The following example was given; *question;* 'What did the dog do when it came out of the water?', *answer;* 'it climbed on the bank'; 'it shook off the water'; 'it ran to his boss'. To encourage the subjects to give single elements of information, the lines on the answering sheet were restricted in length. Participants could answer with as many units of information as they needed. Finally, participants were asked to indicate their confidence regarding the accuracy of each unit of information given on a 7-point scale (1= very uncertain, 7 = absolutely certain).

Completed questionnaires were inspected for incomprehensible or ambiguous answers, and when necessary, participants were asked if they could be more specific about the answer. The experimenter also judged if the information was given in small units and if confidence indications were made to every unit of information. If not, participants were asked to do so afterwards. After completing the final session, participants were debriefed and paid or given credits.

All units of information provided by the participants were scored as correct or incorrect. Information was scored *correct* when it corresponded with information from the video. *Incorrect* information consists of units of information not present in the video, which were either incorrectly remembered or fantasized by the participant. Two experimenters did the scoring, and in case of a disagreement, a third experimenter settled the dispute. Of all units generated 1.6% could not be classified as correct or incorrect; these units of information were discarded from further analysis. The mean number of information-units to answer a question was 2.72 (with a minimum of 1 and a maximum of 25).

Results

We were interested both in the effects of retention interval and repeated questioning on the quantity of recall (the number of "do not know" responses) and the quality of recall (the proportion of correct responses, and mean confidence).

First, we will analyze the effect of retention interval. To that end, the first recall attempts in the 1-, 3- and 5-week retention interval conditions were compared and tested for differences in accuracy and confidence. Second, the effects of repeated recall were tested within conditions 1 (comparing initial recall after 1 week with repeated recall after 3 and 5 weeks) and 2 (comparing initial recall after 3 weeks with repeated recall after 5 weeks). Third, the relation between accuracy and confidence was determined by analyzing the relationship between confidence levels and proportions accurate recall in all interval and repetition conditions, and by calculating gamma correlations.

Retention interval

The effect of retention interval was analyzed by comparing the initial recall sessions only for all three conditions. We first examined the effect of retention interval on the number of questions that could not be recalled (i.e., the number of "do not know" answers per subject). An ANOVA showed a significant effect of retention interval ($F(2, 64) = 5.07$, $p < 0.01$). Post-hoc Bonferroni tests showed a significant difference between 1- and 5-week intervals only

(M = 0.70 and M = 2.35, $p < .05$). The differences between the 1- and 3-week intervals, and between the 3- and 5-week intervals, were not significant. These results indicate that with an increasing retention interval a smaller number of questions can be answered.

A similar analysis on the total units of information recalled by the participants also showed a significant effect of delay ($F(2, 64) = 4.93$, $p < .01$). Post-hoc Bonferroni tests showed significant differences in the number of information units that could be recalled after 1- and 5-week intervals (M = 61.5 and M = 49.0, $p < .05$), and between 3- and 5-week intervals (M = 60.3 and M = 49.0, $p < .05$).

For an analysis of accuracy of recall, we determined the number of correct and incorrect units of information given by the subjects. Proportions accurate and inaccurate answers and corresponding mean confidence judgments are shown in Table 1. Analysis of correctly recalled units showed a significant decrease with retention interval, $F(2, 64) = 14.82$, $p < 0.01$. Bonferroni post-hoc tests showed significant differences between 1- and 5-week intervals (proportions correct 0.85 and 0.71, respectively, $p <.01$), and between 3- and 5-week intervals (proportions correct 0.81 and 0.71, respectively, $p <.05$).

Table 1 Proportions correct and incorrect units of information, and corresponding average confidence ratings (sd in parentheses), as a function of retention interval and repeated recall.

		Retention interval					
		1 week		3 weeks		5 weeks	
		Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Condition 1	Accuracy	.86	.14	.87	.13	.87	.13
N=23	Confidence	6.33 (.39)	5.06 (.89)	6.40 (.36)	5.29 (.87)	6.47 (.29)	5.29 (.98)
Condition 2	Accuracy			.80	.20	.80	.20
N=24	Confidence			6.31 (.44)	4.79 (1.22)	6.27 (.84)	5.24 (.90)
Condition 3	Accuracy					.71	.29
N=20	Confidence					5.89 (.56)	4.58 (.89)

The *confidence* in accurately recalled information also showed a retention interval effect. An analysis of accurate recall confidence in the first recall sessions of the 1-, 3- and 5-week delay conditions, showed a significant effect of interval ($F(2, 64) = 6.04$, $p < .01$). Bonferroni post-hoc tests indicated that the mean level of confidence for correctly recalled information was higher after 1- and 3-weeks intervals (M = 6.33 and 6.31, respectively) than after 5 weeks (M = 5.89, both $p < 0.05$).

Mean confidence levels for incorrectly recalled information also seems to decrease with longer delays, but the difference after one, three and five weeks (5.06, 4.79 and 4.58, respectively) was not significant ($F(2, 64) = 1.9$, *NS*). During initial recall, participants were always significantly more confident about correct information than about incorrect information ($t(22) = 10.06$, $p < 0.01$; $t(23) = 6.59$, $p < 0.01$; $t(19) = 8.41$, $p < 0.01$, for conditions 1, 2, and 3, respectively).

Repeated recall

The effect of repeated recall on the number of unanswered questions was analyzed for condition 1 (3 recall sessions) and condition 2 (2 recall sessions), separately. The mean number of 'do not know' answers in condition 1 decreased over recall sessions after one, three and five weeks (.70, .57 and .52, respectively), but this decrease was not significant (F (2, 21) = .32, *NS*). Similarly, in condition 2 the number of 'do not know' answers did not change significantly as a result of repetition (1.13 and 1.17 in the recall sessions after three and five weeks, respectively; F (1, 23) = .02, *NS*).

The mean proportion of correctly recalled units of information remained almost the same across the subsequent recall sessions, both in condition 1 (0.85, 0.86 and 0.86, respectively) and in condition 2 (0.81 and 0.80, respectively). Also, the mean levels of confidence for correctly recalled units of information were not significantly influenced by repeated recall (condition 1: 6.33, 6.40 and 6.47, respectively; F (2, 21) = 2.30, *NS*; condition 2: 6.31 and 6.27, respectively, F (2, 21) = 0.57, *NS*).

When a participant has recalled incorrect information, it is of particular interest to determine whether repeated recall of incorrect information has an influence on confidence. To test this, we selected all 81 incorrect units of information that were recalled incorrectly during all three sessions of condition 1. Although confidence on these repeated errors increased slightly with repetition (5.16, 5.23 and 5.33 in the first, second and third sessions, respectively), this increase was not significant (F (2, 79) = 0.67, *NS*). In a similar manner, 138 repeated errors made by the participants in condition 2 were selected and tested. Here, a paired-sample t-test showed that the mean confidence given at first recall was significantly lower than the mean confidence given on the identical errors during the later recall (4.91 and 5.30, respectively, t (137) = 3.01, $p < 0.01$).

To determine the effect of repeated recall on confidence, we also analyzed confidence after the same retention interval of 5 weeks, but with different numbers of preceding recall attempts in conditions 1, 2 and 3. This analysis showed that confidence was significantly higher when final recall was preceded by more previous recall attempts. Average confidence of both correct and incorrect answers was 6.30 after two recall attempts, 6.08 after one recall attempt and 5.50 after zero recall attempts (F (2, 64) = 15.2, $p < 0.01$). However, in interpreting this result it should be realized that this difference between the conditions was already present in the first recall attempt after different delays.

Accuracy-confidence relations

To analyze accuracy-confidence relations, we determined the number of correct and incorrect units of information recalled for each confidence level. Goodman-Kruskal gamma correlation coefficients were calculated overall and per subject for each retention interval and repeated recall condition. These correlations are presented in Table 2.

Table 2 Gamma-correlations between confidence and accuracy. Normal script: mean of individual correlations (sd in parentheses), bold script: overall correlations.

		Retention interval		
		1 week	**2 weeks**	**5 weeks**
Condition 1 N=23	Overall	**0.63***	**0.61***	**0.61***
	Individual	0.69 (.14)	0.64 (.22)	0.70 (.24)
Condition 2 N=24	Overall		**0.58***	**0.57***
	Individual		0.65 (.22)	0.63 (.21)
Condition 3 N=20	Overall			**0.49***
	Individual			0.60 (.19)

* p < .01

As can be inferred from this table, repetition had no effect on these correlations. Apparently, once information is retrieved it 'survives' and the content (accuracy) and confidence ratings remain stable. Although longer intervals for initial testing seem to result in lower confidence-accuracy correlations, the differences are not significant (all $z < 1.0$).

The confidence-accuracy relations found in this study seem to indicate that confidence is a reasonable predictor of accuracy. This is also illustrated in Table 3, showing the distribution of the total number of recalled units of information, and the proportions of incorrect units, as a function of the level of confidence expressed by the participants. This table shows an obvious relation between accuracy and confidence, with larger proportions of incorrect units at lower levels of confidence.

Table 3 Total units of information for each confidence level per condition; the proportions of incorrect units are presented in parentheses.

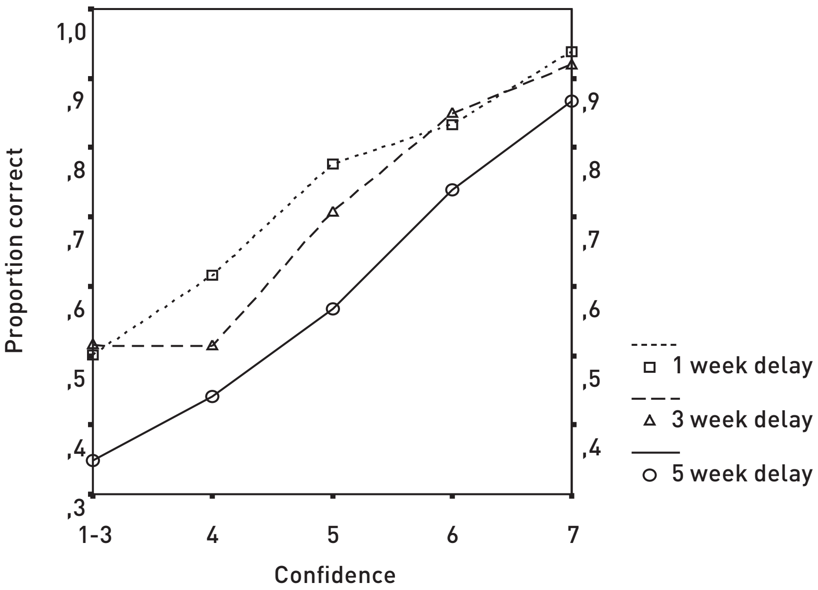
Condition	Retention interval	Confidence Scale 1	2	3	4	5	6	7	Total units of information
Condition 1 N=23	1 week	5 (.60)	34 (.44)	45 (.51)	78 (.40)	143 (.24)	222 (.16)	887 (.06)	1414 (.14)
	3 weeks	2 (1.0)	23 (.48)	49 (.30)	64 (.39)	132 (.29)	250 (.16)	889 (.06)	1409 (.13)
	5 weeks	9 (1.0)	11 (.55)	37 (.51)	76 (.37)	139 (.26)	232 (.12)	915 (.07)	1419 (.13)
Condition 2 N=24	3 weeks	7 (.71)	31 (.68)	52 (.50)	105 (.49)	171 (.31)	209 (.29)	880 (.11)	1455 (.20)
	5 weeks	6 (.66)	28 (.50)	66 (.53)	76 (.58)	148 (.26)	197 (.25)	862 (.11)	1383 (.20)
Condition 3 N=20	5 weeks	16 (.81)	48 (.48)	66 (.55)	109 (.50)	154 (.36)	201 (.23)	386 (.15)	980 (.29)

Table 3 seems to suggest a systematic relationship between delay of first recall and confidence level, both in the number of information units recalled and the proportion accurate recall. To analyze these relationships, we first performed an ANOVA on the number of information units recalled, with delay (1-week in condition 1, 3-weeks in condition 2, and 5-weeks in condition 3) and confidence level as a between and a within subjects factor, respectively. Given the small number of recalled units of information with low levels of confidence, confidence levels 1-3 were combined for this analysis. Both main effects of delay (F (2, 64) = 5.58, $p < 0.01$) and confidence level (F (4, 61) = 41.06, $p < 0.01$) were significant. There was no significant interaction effect (F (8, 124) = 0.82, *NS*).

A similar analysis on the proportions correct (after an arcsin transformation) showed significant effects of confidence level (F (4, 61) = 53.83, $p < 0.01$) and a significant delay X confidence interaction (F (8, 124) = 2.42, $p < 0.05$). The main effect of delay just failed to reach significance (F (2, 64) = 2.74, $p = 0.07$). An analysis of the proportions correct as a function of confidence showed a highly significant linear trend (F (1, 64) = 169.8, $p < 0.01$). The effect of delay and level of confidence on proportions correct recall is shown in Figure 1.

Table 3 shows that in all conditions the proportion of incorrect units associated with higher confidence levels drops steadily. Especially with short recall intervals, the proportion of errors with high confidence ratings drops to a level where confidence seems to become useful as a predictor of accuracy. Five weeks after seeing the video, 15 % of all units of information given the highest confidence score were incorrect. After an interval of only 1 week, just 6% in the highest confidence category was incorrect. So, after a longer retention interval witnesses not only provide less information but they also provide more inaccurate information with the highest confidence rating.

Figure 1 Proportion correct recall as a function of delay of first recall (1-, 3- or 5-weeks) and level of confidence (1 = very uncertain, 7 = absolutely certain; levels 1-3 have been combined due to the small number of data).



Discussion

The purpose of this study was to investigate the influence of retention interval and repeated recall on the accuracy and confidence of episodic eyewitness memory in an ecologically more valid situation than has been used in previous studies. The main findings were that longer intervals before first questioning resulted in more 'do not know' answers, fewer correct units of information recalled, and lower confidence ratings. In contrast, repeated questioning did not affect any of these measures. Repetition of recall did not influence the number of 'do not know' answers, the proportion of correct units of information recalled, the confidence ratings, or the confidence-accuracy correlation.

The findings in this study have some important practical implications. First, longer retention intervals resulted in reduced memory performance. This is not a new finding, of course, but it emphasizes the importance that witnesses should be questioned as soon as possible after an event. Any delay reduces the amount recalled and the confidence in recall.

Second, the results do not show clear indications of memory enhancement with repeated recall attempts. Repeated questioning seems to do no more than to consolidate the information that was retrieved in previous attempts, and it does not seriously affect the subjective confidence in the accuracy of what is recalled. This implies that repeated questioning is not effective in remembering additional information. On the other hand it does not harm the eyewitness report either, because we found no evidence of more incorrect recall, or of inflated confidence. It must be noted, however, that this conclusion is only valid for the conditions as used in our study, i.e., asking recall of details of an original experience, but using the same questions in subsequent retrieval attempts. It cannot be ruled out that changes in subsequent retrieval attempts, e.g., by asking different questions or by using a cognitive interview to follow-up on a free recall attempt, would produce additional information.

Interestingly, we found no evidence of confidence inflation by repeated retrieval as was reported for instance by Shaw (1996) and Shaw and McClure (1996). In these studies, errors that were repeated in subsequent recognition sessions were accompanied by increases in confidence. This increase in confidence was explained by the authors with a retrieval-fluency hypothesis. According to this hypothesis, the ease with which an item can be retrieved from memory may be used for confidence judgments, with greater ease of retrieval yielding higher confidence judgments. This hypothesis is also used to explain 'imagination inflation' findings showing that repeated imagination causes an increased tendency to judge an imagined event as an event that actually happened (Thomas & Loftus, 2002). Our results suggest that in the conditions used here (repeated recall instead of repeated recognition or imagination) retrieval fluency was not associated with inflated confidence judgments.

The findings in this study are important for the way in which researchers and experts think about the reliability and trustworthiness of eyewitness memory. Our results indicate that in ecologically valid conditions, recall of eyewitnesses is reasonably accurate, and that neither accuracy nor subjective confidence is strongly affected by repeated recall attempts.

Ebbesen and Rienick (1998) came to a similar conclusion. This conclusion seems to contradict several experimental studies reporting various detrimental effects of repeated testing (Roediger et al., 1997; Shaw, 1996; Shaw & McClure, 1996). We believe this discrepancy may be caused by the fact that in these latter studies often procedures are used that probably favor the occurrence of repeated testing effects (e.g., using recognition instead of recall, and presenting artificial, unrelated or ambiguous material).

Our results are also at odds with the general belief among memory experts that we should be reluctant in using confidence as an indication of accuracy. This may be true for eyewitness identification (e.g., Bothwell et al., 1987; Deffenbacher, 1991; Sporer et al., 1995), but probably not for recall of events. We found that confidence-accuracy correlations were relatively high, especially with the shortest recall interval, and that they did not change much with repeated recall attempts. An interesting observation is the high level of accuracy for recalled items given the highest confidence ratings, especially after a brief delay. This suggests that in the conditions used in this study (i.e., stressing the importance of being accurate and allowing a witness to withhold an answer if uncertain) participants were quite able to set and apply an internal standard to evaluate the correctness of their recall (Koriat & Goldsmith, 1996).

The distribution of correct recall as a function of confidence suggests that confidence may be useful to distinguish between responses that are likely to be accurate and responses that are less likely to be accurate. Simply by selecting only responses that receive the highest confidence rating, a large proportion of inaccurate information would be filtered out. Unfortunately, however, there always remain incorrect items that are given the maximum confidence score. Therefore, no single witness statement can be accepted as certainly correct on the basis of confidence alone. Although the proportion of highly confident but incorrect recall may be small, it is a significant factor because it is potentially dangerous during a police investigation and can be disastrous in a courtroom.

References

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.

Bjork, R. A. (1975). Retrieval as a modifier. In R. L. Soslo (Ed.), *Information processing and cognition: The Loyola symposium* (pp. 123-144). Hillsdale, NJ: Erlbaum.

Bornstein, B. H., & Zickafoose, D. J. (1999). "I Know I Know It, I Know I Saw It": The stability of the confidence-accuracy relationship across domains. *Journal of Experimental Psychology: Applied, 5*, 76-88.

Bothwell, R. K., Deffenbacher, K. A., & Brigham, J. C. (1987). Correlation of eyewitness accuracy and confidence: Optimality hypothesis revisited. *Journal of Applied Psychology, 72*, 691-695.

Busey, T. A., Tunnicliff, J., Loftus, G. R., & Loftus, E. F. (2000). Accounts for the confidence-accuracy relation in recognition memory. *Psychonomic Bulletin & Review, 7*, 26-48.

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: An International Journal, 3*(3-4), 388-407.

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision making in eyewitness identification cases. *Law and Human Behavior, 12*, 41-55.

Deffenbacher, K. A. (1991). A maturing of research on the behavior of eyewitnesses. *Applied Cognitive Psychology, 5*, 377-402.

Ebbesen, E. B., & Rienick, C. B. (1998). Retention interval and eyewitness memory for events and personal identifying attributes. *Journal of Applied Psychology, 83*, 745-762.

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

Granhag, P. A., Jonsson, A.-C., & Allwood, C. M. (2004). The cognitive interview and its effect on witnesses' confidence. *Psychology, Crime & Law, 10*, 37-52.

Hyman, I. E., Husband, T. H., & Billings, J. F. (1995). False memories of childhood experiences. *Applied Cognitive Psychology, 9*, 181-197.

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

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

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.

Levy, B. J., & Anderson, M. C. (2002). Inhibitory processes and the control of memory retrieval. *Trends in Cognitive Sciences, 6*, 299-305.

Lindsay, D. S., Read, J. D., & Sharma, K. (1998). Accuracy and confidence in person identification: The relationship is strong when witnessing conditions vary widely. *Psychological Science, 9*, 215-218.

Lindsay, R. C. L., Wells, G. L., & O'Connor, F. J. (1989). Mock-juror belief of accurate and inaccurate eyewitnesses. *Law and Human Behaviour, 13*, 333-339.

Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.

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.

Olsson, N. (2000). A comparison of correlation, calibration, and diagnosticity as measures of the confidence-accuracy relationship in witness identification. *Journal of Applied Psychology, 85*, 504-511.

Penrod, S. D., & Cutler, B. L. (1995). Witness confidence and witness accuracy: Assessing their forensic relation. *Psychology, Public Policy, and Law, 1*, 817-845.

Perfect, T. J., Watson, E. L., & Wagstaff, G. F. (1993). Accuracy of confidence ratings associated with general knowledge and eyewitness memory. *Journal of Applied Psychology, 78*, 144-147.

Raaijmakers, J. G. W., & Shiffrin, R. M. (1981). Search of associative memory. *Psychological Review, 88*, 93-134.

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.

Robinson, M. D., & Johnson, J. T. (1996). Recall memory, recognition memory, and the eyewitness confidence-accuracy correlation. *Journal of Applied Psychology, 81*, 587-594.

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.

Roediger, H. L., McDermott, K. B., & Goff, L. M. (1997). Recovery of true and false memories: Paradoxical effects of repeated testing. In M. A. Conway (Ed.), *Recovered memories and false memories* (pp. 118-149). Oxford: Oxford University Press.

Roediger, H. L., Wheeler, M. A., & Rajaram, S. (1993). Remembering, knowing and reconstructing the past. In D. L. Medin (Ed.), *The psychology of learning and motivation* (Vol. 30, pp. 97-134). New York: Academic Press.

Scrivner, E., & Safer, M. A. (1988). Eyewitnesses show hypermnesia for details about a violent event. *Journal of Applied Psychology, 73*, 371-377.

Shaw, J. S. (1996). Increases in eyewitness confidence resulting from postevent questioning. *Journal of Experimental Psychology: Applied, 2*, 126-146.

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.

Smith, V. L., Kassin, S. M., & Ellsworth, P. C. (1989). Eyewitness accuracy and confidence: Within- versus between- subjects correlations. *Journal of Applied Psychology, 74*, 356-359.

Sporer, S. L., Penrod, S. D., Read, J. D., & Cutler, B. L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118*, 315-327.

Thomas, A. K., & Loftus, E. F. (2002). Creating bizarre false memories through imagination. *Memory & Cognition, 30*, 423-431.

Turtle, J. W., & Yuille, J. C. (1994). Lost but not forgotten details: Repeated eyewitness recall leads to reminiscence but not hypermnesia. *Journal of Applied Psychology, 79*, 260-271.

Wells, G. L., & Loftus, E. F. (2003). Eyewitness memory for people and events. In A. M. Goldstein (Ed.), *Handbook of psychology: Forensic psychology* (Vol. 11, pp. 149-160). New York: John Wiley and Sons, Inc.

Williams, M. D., & Hollan, J. D. (1981). The process of retrieval from very long-term memory. *Cognitive Science, 5*, 87-119.

Wise, R. A., & Safer, M. A. (2004). What US judges know and believe about eyewitness testimony. *Applied Cognitive Psychology, 18*, 427-443.