Universiteit
Leiden
The Netherlands

# Eyewitness confidence : the relation between accuracy and confidence in episodic memory

Odinot, G.

**Citation**
Odinot, G. (2008, December 16). *Eyewitness confidence : the relation between accuracy and confidence in episodic memory*. Retrieved from https://hdl.handle.net/1887/13360

| | |
|---|---|
| Version: | Not Applicable (or Unknown) |
| License: | Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden |
| Downloaded from: | https://hdl.handle.net/1887/13360 |

**Note:** To cite this publication please use the final published version (if applicable).

# Eyewitness confidence

The relation between accuracy and confidence in episodic memory

Geralda Odinot

Odinot_4.indd 1 10-11-2008 13:19:00

Eyewitness confidence: The relation between accuracy and confidence in episodic memory

Thesis, University of Leiden, The Netherlands

ISBN 978 90 79969 01 2
Author Geralda Odinot
Cover Textcetera, Den Haag, The Netherlands
Lay-out Textcetera, Den Haag, The Netherlands
Print Haveka, Alblasserdam, The Netherlands

© G. Odinot, Leiden, 2008
All rights reserved. No part of this thesis may be reproduced or transmitted in any form by any means, without permission of the copyright owner.

Odinot_4.indd 2 10-11-2008 13:19:00

# Eyewitness confidence

The relation between accuracy and confidence
in episodic memory

**Proefschrift**

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus prof.mr. P.F. van der Heijden,
volgens besluit van het College voor Promoties
te verdedigen op dinsdag, 16 december 2008
klokke 11.15 uur

door

**Geralda Odinot**

Geboren te Utrecht op 2 januari 1970

Odinot_4.indd 3 10-11-2008 13:19:00

**Promotiecommissie**

| | |
|---|---|
| Promotoren: | Prof. dr. W.A. Wagenaar<br>(Universiteit Leiden en Universiteit Utrecht)<br>Prof. dr. P.J. van Koppen<br>(formerly at NSCR, Universiteit Maastricht en Vrije Universiteit Amsterdam) |
| Co-promotor: | Dr. G. Wolters<br>(Universiteit Leiden) |
| Overige leden: | Mr. J.A. Coster van Voorhout<br>Prof. dr. P.A. Granhag<br>Prof. mr. dr. J.F. Nijboer |

Odinot_4.indd 4 10-11-2008 13:19:00

# Table of contents



Odinot_4.indd 5 10-11-2008 13:19:00

# 1

# Introduction

Odinot_4.indd 7 10-11-2008 13:19:00

"I was certain, but I was wrong...."

*Jennifer Thompson*, New York Times, June 18, 2000

In daily life, confidence is often used to express a degree of certainty about the accuracy of information retrieved from memory. It seems a matter of common sense that confidence in a memory is strongly related to the actual accuracy of the memory. "Are you sure about that?" is a legitimate question after hearing someone tells about a prior experience. Subjective confidence about some information provides possible directions for future actions, decisions and beliefs, when objective records to check the correctness of this information are lacking. And clearly, it must work most of the time, otherwise we would have dismissed this rule of thumb a long time ago.

Many decisions in the legal system are based on eyewitness evidence. Witnesses testify what they remember and because objective records to determine the accuracy of these memories are lacking most of the time, indicators to infer the accuracy in witness statements become important. It seems to be a matter of common sense that the level of confidence that is expressed by a witness can be used as a diagnostic tool to discriminate between accurate and inaccurate memories. Research has shown that there is indeed 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, 1994a; Penrod & Cutler, 1995). Contrary to this general belief, the bulk of empirical evidence collected in laboratory and field experiments over the past 25 years indicates that the relationship between confidence and accuracy is far from perfect. In meta-analyses of studies on eyewitness identifications it was found that the average correlation between confidence and accuracy tends to be relatively small, i.e., in the order of 0.25 – 0.30 (Bothwell, Deffenbacher, & Brigham, 1987; Sporer, Penrod, Read, & Cutler, 1995).

Most research, however, has focused on the identification of persons and relatively little is known about the relationship between accuracy and confidence in remembering events. Obviously, the relationship between accuracy and confidence in event memory is very important in the legal system. Therefore, the aim of the present dissertation is to explore the question whether confidence in the memories of a witnessed event can be used as an indicator for accuracy.

In this introductory chapter, first the concept of episodic eyewitness memory is discussed, followed by an overview of prior research on the confidence-accuracy relationship in this area. Next, I will introduce the source monitoring theory as a framework to understand the determinants of confidence judgments. Subsequently, different statistical methods are discussed to measure and to express the relationship between confidence and accuracy. Finally, I will introduce the empirical studies that will be presented in the following chapters.



Odinot_4.indd 8 10-11-2008 13:19:00

## Episodic eyewitness memory

Episodic memory refers to the memory for particular events (episodes) that we experienced in our own life, for instance, when and where something happened, what happened and who were involved. So, episodic memory holds information about our personal past, ranging from everyday experiences (like remembering having had pasta for dinner the other day), to the most significant events in a lifetime (like recalling a graduation or wedding day). Depending on the interval between the event and the time of retrieval it may be called short-term memory (for retention periods of a few tens of seconds), or long-term memory (for retention periods ranging from tens of seconds to a lifetime). Episodic memory is to be distinguished from semantic memory, a form of memory that contains general knowledge of the world (Tulving, 1972). Semantic memory consists of knowledge about concepts, objects, relationships and rules, which are either learned as facts or derived from many experiences.

Together, episodic and semantic memory are known as declarative or explicit forms of memory. Characteristic of declarative memory is the fact that its contents are accessible for conscious inspection and can be reported. In contrast, non-declarative (procedural or implicit) forms of memory refer to information that is stored and that can be expressed in behavioural performance, but to which we have no conscious access. Examples of non-declarative memory are conditioning, learning of skills and repetition priming. It has been suggested that declarative and non-declarative memories are based on different types of learning processes and involve different neural systems and pathways (e.g., Eichenbaum & Cohen, 2001; Squire, 1992).

Since the pioneering work of Ebbinghaus, most human memory research has focused on episodic memory. Experimental studies on episodic memory require participants to encode material (e.g., a set of stimuli, like words or pictures), and to try to remember it after a varying interval. This work has shown that what is remembered strongly depends on active processes during encoding, retention and retrieval (e.g., Wolters, 1983).

Encoding refers to the process by which a trace of an experience becomes registered in memory. However, there are limitations of the cognitive system, not all information experienced is encoded. We selectively pay attention to certain aspects of an event and ignore others (Baddeley, 1997). Moreover, during encoding we actively search for meaning and perform various kinds of mental operations that determine what is stored in memory (Craik & Lockhart, 1972). During retention, the encoded and stored information can be lost or transformed by the encoding of related information (Loftus, 1979). The final step in remembering involves the retrieval of information. Memories are not randomly retrieved but triggered by retrieval cues. Retrieval cues can be general, as in free recall tasks, or specific, as in recognition tasks. Retention performance generally depends on the correspondence of retrieval cues with what was encoded and stored (Tulving & Thomson, 1973). Retrieval cues are often incomplete, however, necessitating a more or less extensive search process



Odinot_4.indd 9 10-11-2008 13:19:00

(e.g., Raaijmakers & Shiffrin, 1981). Moreover, the memory records are often incomplete and require active processes to reconstruct the original experience (Bartlett, 1932).
The testimony of eyewitnesses is based on their episodic memories of the witnessed event. As was indicated above, episodic memories are not passive records of witnessed events. Instead, we selectively and actively interpret our experiences during encoding, we integrate novel information into existing memories during retention, and we reconstruct an original event on the basis of incomplete memory records. Because of this (re)constructive nature of memory, reports may not only be incomplete but even incorrect, caused by factors that intrude at the point of encoding of the event, during storage of the event, or at time of retrieval of the event (Bartlett, 1932; Ceci & Bruck, 1995; Loftus, 1979, 2003). Although our episodic memories are not veridical records of the past, it can be assumed that generally they will be more or less correct. This is insufficient in a legal context however, where eyewitness memory should be accepted only if it is accurate.
The fallibility of the memory of eyewitness was noticed already in the beginning of the 20th century (e.g., Münsterberg, 1908; Stern, 1902; see also Van Strien, 2000). Apart from the work of Bartlett (1932), the malleability and fallibility of episodic memory did not receive much attention until interest was renewed by the pioneering work of Loftus on the effects of misleading post-event information. Since then, many studies have investigated the extent to which event memory is open to distortion, and the results are not comforting. It has been shown, for instance, that suggested or fantasized events may be 'remembered' as actual experiences (Loftus, 1997; Wade, Garry, & Lindsay, 2002). Roediger and McDermott (1995) showed that non-presented words that are strongly associated to a set of presented words, can be 'recognized' to the same degree and with the same level of confidence than the actually presented words. The fallibility of episodic memory leads to questions about the validity of eyewitness testimony, and more specifically to questions how accurate and inaccurate memories can be distinguished.



## Research on accuracy and confidence

Inaccuracy of memories of prior experiences would not be a problem if people were able to assess correctly the level of accuracy, for instance in the form of confidence judgments. Trial simulation studies have shown that jurors indeed give weight to eyewitness confidence when evaluating the accuracy of eyewitness testimony (e.g., Cutler, Penrod, & Dexter, 1990; Cutler et al., 1988; Lindsay et al., 1989). In addition, jurors report that they find it difficult to consider any other alternatives to a confident statement by an eyewitness, a situation which is called the "tyranny of the eyewitness" (Haber & Haber, 2000).
These findings imply that jurors rely on a factor that has no strong value on the evaluation of the accuracy of eyewitness testimony. Available experimental evidence indicates that the

relationship between accuracy and confidence is quite modest. Meta-analyses of studies in recognizing and identifying persons have found average correlations between accuracy and confidence of 0.25 (Bothwell et al., 1987) and 0.29 (Sporer et al., 1995). These modest correlations have led several researchers to look for determining factors. On the one hand, explanations have been sought in methodological factors, such as a limited variability of performance due to the difficulty of the task, absence of supporting contextual information, and the use of between-subject designs (Gruneberg & Sykes, 1993; Lindsay, Read, & Sharma, 1998). Others looked for mediating causal factors influencing the accuracy-confidence relationship (see, e.g., Read, Lindsay, & Nicholls, 1998).

Most research on the accuracy-confidence relationship has looked at performance in person identification tasks. Only relatively recently, studies have begun exploring this relationship in other memory tasks. It has been shown, for example, that the accuracy-confidence relationship for general knowledge questions and for episodic event memory is considerably higher (in the order of 0.40 to 0.60), although still far from perfect (e.g., Koriat & Goldsmith, 1996; Odinot & Wolters, 2006; Perfect, Watson, & Wagstaff, 1993; Robinson & Johnson, 1996). In contrast, other studies have found considerable discrepancies between accuracy of remembered details and confidence (Neisser, 1982). Even in cases where the event was of great personal significance or international importance, and were witnessing resulted in a so called 'flashbulb memory' (e.g., Neisser & Harsch, 1992; Talarico & Rubin, 2003; Wolters & Goudsmit, 2005). Therefore, the general conclusion is that the weight of evidence indicates that eyewitness confidence is not by itself a reliable predictor of eyewitness accuracy (see for a review; Shaw, McClure, & Dykstra, 2007). It should be noted, however, that the research on which this conclusion is drawn, is mainly based on testing episodic memory with wordlists or recognition-type questionnaires.

When someone has to point out a suspect in a line-up, obviously, recognition is the memory process that is involved. However, during an interview with a witness, information is actively retrieved from memory and recall processes are at stake. Therefore, to provide new information about the relation between accuracy and confidence in episodic eyewitness memory it is necessary to make a distinction between recall and recognition memory processes.

In recognition tasks participants have to discriminate between studied and non-studied items. Dual process models of recognition postulate that two qualitative different processes, i.e., recollection and familiarity, are involved in recognition judgments (e.g., Kelley & Jacoby, 1998; Mandler, 1980). Recollection is assumed to be based on the retrieval of specific details of the original presentation. Familiarity is assumed to be a fast process reflecting the global familiarity or strength of an item. It has been suggested that familiarity judgments could be based on the perceptual and conceptual fluency with which an item is processed. Although the dual process model of recognition is not unchallenged, and the debate about dual or single process models still continues (see e.g., Hirshman & Master, 1997; Yonelinas, 2002), the weight of the evidence seems to favour the dual process account.



Odinot_4.indd 11 10-11-2008 13:19:01

Several studies, in which the contributions of recollection and familiarity were separated, have shown that these processes are affected differently by many variables (e.g., response speed, forgetting rates, and levels of processing).
In recall tasks information has to be retrieved on the basis of less specific cues often requiring an active search process. Recall performance strongly depends on the organizational structure of a memory trace because the generation of any part of the trace is used subsequently as a cue to retrieve other parts. In brief, recall memory tends to be characterized by an intentional and effortful retrieval stage, whereas recognition memory tends to be based on the use of a less intentional and less effortful familiarity heuristic (Raaijmakers & Shiffrin, 1992).
It is clear that recall and recognition are two different memory processes that may result in different accuracy-confidence relations. Because recall memory processes are typically used when witnesses are interviewed and information is actively retrieved form memory, the testing methods used in this dissertation will be recall based.

## Source monitoring

The main idea tested in this dissertation is derived from the source monitoring framework of Johnson and Raye (1981, see also; Johnson, 2006; Johnson, Hashtroudi, & Lindsay, 1993). Source monitoring refers to the cognitive processes by which mental experiences (e.g, thoughts or memories) are attributed to particular origins or sources in our past (Johnson et al., 1993). According to this framework, source monitoring is based on characteristics of memories in combination with flexible decision processes. Errors in source monitoring can lead us to report true memories but erroneously situate them in time and place, or to report as actual memories events that we only heard about, saw on television, or imagined (Lindsay, Allen, Chan, & Dahl, 2004).
According to Johnson et al. (1993) the characteristics that are used to monitor the source of a remembered event are perceptual, conceptual, affective and contextual details, as well as information about cognitive operations performed when the memory trace was created. The more such details become available during retrieval of an event, the more likely it is that the event was actually experienced. In addition, strategic deliberations may be taken into account, such as the plausibility of an event given other knowledge.
Source monitoring errors may occur in many contexts. A well-known example from daily experience is the gnawing doubt after having left the house for a holiday trip whether one locked the doors (or turned off the lights), or just though about doing so. A widely publicized source monitoring failure probably happened to Hillary Clinton when she was campaigning for the democratic presidential candidacy in the Spring of 2008. In an interview she told in detail about a memory of becoming under sniper fire during a visit to the former Yugoslavia.



Odinot_4.indd 12 10-11-2008 13:19:01

Video recording of the visit, however, proved her memory to be wrong. Source monitoring issues are also central, for example, to concerns about the accuracy of recovered memories and children's reports of sexual abuse (Johnson, 2006).

The source monitoring framework is somewhat similar to suggestions made by Koriat and Goldsmith (1996) and Brainerd, Wright, Reyna, and Payne (2002). According to these researchers, memory accuracy is under strategic control and people regulate their memory reports in the service of achieving a particular, situation dependent, accuracy level. They proposed a two process model for the regulation of memory accuracy: monitoring effectiveness and the response criterion. Monitoring effectiveness is the subjective assessment of the accuracy of a retrieved answer, and the response criterion is a threshold value influenced by situational demands, which determines whether or not to output the answer. This model, however, does not address the issue why incorrect answers may be retrieved in the first place, and how accuracy of memories is assessed and expressed as a confidence judgment.

In this dissertation, we will follow the source-monitoring framework for distinguishing between true and false memories. We assume that not only the source of a memory, but also confidence judgments about its accuracy, is largely based on the ability to retrieve details of the original experience. Confidence about a memory is likely based on the quality or the strength of the memory trace (see, e.g., Burke, MacKay, Worthley, & Wade, 1991; Clark, 1997; Hintzman, 1988). The more elaborate or stronger the memory trace, the greater will be the number of perceptual, conceptual and contextual details. Robinson and Johnson (1996) suggested that in recall, additional diagnostic information may be provided by retrieval efforts, and that this may also offer valid insight into both the accuracy and the confidence in the accuracy of a memory.

We suggest, therefore, that generally it is to be expected that both accuracy and confidence will increase when more detailed information can be retrieved. However, not being able to retrieve details does not necessarily imply that what is remembered is incorrect. Absence of memory for detail, therefore, may result in a low confidence for an accurate memory. Conversely, if details are remembered incorrectly (e.g., due to source confusions or reconstruction errors) this may result in an inaccurate memory with high confidence judgments. So, although in principle a perfect relation between accuracy and confidence is possible, in reality the relation will suffer either from the loss of details in original memory traces (as may occur for instance with longer retention intervals), or the presence of incorrect details (as may occur with suggested or fantasized events). Another distortion of the accuracy-confidence relation may occur when the same memory trace (or thought) is repeatedly retrieved.



Odinot_4.indd 13 10-11-2008 13:19:01

## Delayed recall and repeated retrieval

In this dissertation we will explore the effects of the length of the retention interval and of repeated retrieval on the accuracy-confidence relation. In criminal investigations, it is not uncommon that it takes a while before a witness is interviewed. As yet, however, few studies have investigated the effect of retention interval on the relation between accuracy and confidence in event memory. Therefore, an important question in the empirical chapters of this dissertation is what the effect of longer retention intervals has on the accuracy-confidence relation.

In addition to being interviewed after a delay, witnesses are also often interviewed more than once. One of the reasons to question witnesses several times is the idea that witnesses may provide new information during follow-up questioning. Information that could not be remembered initially may be remembered at a later moment. However, also the investigation procedure itself induces that 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 defence 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.

Repeated recall may also introduce distortions of memory. Distortions of accuracy and confidence may occur simply by repeated questioning (or repeated reflective thought). For instance, repeated attempts to recall once imagined or suggested information has been shown to be a powerful force in the creation of false memories (e.g., Ceci, Huffman, Smith, & Loftus, 1994; Hyman, Husband, & Billings, 1995). Roediger, McDermott, and Goff (1997) concluded that repeated recall can have both facilitating and detrimental effects on later retention. To understand these effects of repeated recall it is important to note that retrieval is not a neutral process, which leaves 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). Several authors have suggested that repeated recall may cause confidence inflation because it enhances ease of retrieval or response fluency (e.g., Robinson, Johnson, & Robertson, 2000; Shaw, McClure, & Wilkens, 2001). In a recent review Shaw et al. (2007) concluded that repeated questioning generally leads to increases in the confidence ratings.

Moreover, repeated post-event questioning offers an opportunity for retrieval practice. Practicing retrieval of a subset of memories may even suppress access to related memories, a phenomenon known as retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994; Anderson & McCulloch, 1999; Anderson & Spellman, 1995; Barnier, Hung, & Conway, 2004; MacLeod, 2002; Shaw, Bjork, & Handal, 1995).



Odinot_4.indd 14 10-11-2008 13:19:01

As noted before, the conclusions about the relation between accuracy and confidence in event memory are mainly based on testing memory with wordlists or recognition-type questionnaires. Although various aspects of repeated recall have been studied quite extensively, surprisingly few studies have tested repeated retrieval of complex naturalistic stimuli with a recall task, and accompanying confidence judgments over the course of a relatively long retention interval.

## Measuring confidence and the accuracy-confidence relation

The methods for measuring confidence in memory reports are quite similar among researchers. Generally participants are simply asked to rate their confidence on 5, 7, 9 or 10 point scales. The anchoring poles of these scales vary for instance, from "not at all confident" to "extremely confident" (Memon, Hope, & Bull, 2003), "completely uncertain" to "certain enough to testify in court" (Fleet, Brigham, & Bothwell, 1987), and "not at all confident" to "very confident" (Luus & Wells, 1994b). An alternative for the anchored scales is the measurement of confidence by asking the participants for percentages or probability indications to indicate their level of certainty. They may be asked for instance to indicate their confidence on a scale ranging from 0% to 100% that is anchored by "not at all certain" and "totally certain" (e.g., Bradfield, Wells, & Olson, 2002; Juslin, Olsson, & Winman, 1996; Weber & Brewer, 2003). In daily live, however, it is very unusual to speak of 80% or 20% certainty, when we remember something. In this dissertation, confidence will be measured with a 7-point Likert scale, labeled with the anchors "very uncertain" to "absolutely certain".

In the majority of studies, the relation between accuracy and confidence is expressed with the point-biserial correlation. The point-biserial correlation is a measure of the linear relation between a dichotomous and a continuous or categorical variable. One of the problems of the point-biserial correlation is that confidence scores are often not uniformly distributed over the scale values, which may cause an underestimation of the actual accuracy-confidence relation.

Therefore, some researchers (Juslin et al., 1996; Olsson, 2000; Weber & Brewer, 2003) have argued that calibration may be a more informative measure of the relation between accuracy and confidence. In memory research, calibration involves plotting the *subjective* probability of being correct (confidence) against the *objective* probability of being correct (accuracy). By plotting the mean accuracy for each defined confidence interval against the mean confidence for the same interval, a calibration curve is created (see, e.g., Brewer, Keast, & Rishworth, 2002; Granhag, Stromwall, & Allwood, 2000; Wagenaar, 1988). Perfect calibration would be indicated by a linear function, with 100% accuracy for witnesses who were 100% confident, 80 % accuracy for witnesses who were 80% confident, etc. Studies using calibration suggest that the relation between accuracy and confidence is more clearly



Odinot_4.indd 15 10-11-2008 13:19:01

visible when expressed in terms of calibration than when expressed in terms of a correlation. Generally, however, eyewitnesses also do not show good calibration. Most participants in experiments on memory tend to be overconfident. This is especially true in the higher confidence range.

Although a calibration curve is an excellent way to visualize the relationship between accuracy and confidence, it has a few drawbacks. One of these drawbacks is that it tends to obscure somewhat the presence of errors (i.e., inaccurate answers given with a relatively high confidence). For instance in the situation where 80% and 20% confidence correspond with 20% and 80% inaccurate memories, respectively, one has a perfect calibration. Of course, in such a case the calibration logic implies that errors are made and that the proportion of errors at each confidence level can be derived. However, it also does suggest good performance, whereas it should suggest great concern because a substantial proportion of inaccurate memories are recalled with high confidence. This aspect of the data is probably better captured in a correlation coefficient. Also, calibration curves are not easily interpretable as correlations when multiple conditions are compared.

Secondly, to be reliable, a calibration curve needs a large amount of data points that are preferably evenly distributed among the confidence scale. Data gathered with the type of experiment as described in this dissertation (free or cued recall with the option to withhold an answer) shows a very skewed distribution on the confidence scores. When participants make confidence judgments about the perceived accuracy of their memories they do not often use the lower part of the scale. Moreover, when accuracy-confidence relations are calculated within the participants, relative small sample sizes are available. The data gathered in the experiments in this dissertation also violate the assumptions for the point biserial correlations. To overcome these problems, frequency tables and the non-parametric Goodman and Kruskal gamma-correlations are used to present the data.



## Outline of this dissertation

The following four chapters of this dissertation all investigate the effects of a number of variables on the relation between accuracy and confidence in episodic eyewitness memory. The first three chapters are experiments from the laboratory and the fourth chapter is a case study in which real live witnesses of a robbery were interviewed.

Chapter 2 presents an experimental study on the effect of repeated recall and retention interval on the accuracy-confidence relation.

Chapter 3 also presents an experimental study on the effect of repeated recall and retention interval on the accuracy-confidence relation. This study also investigates the effect of suggestive questioning. In addition, confidence and consistency are compared as potential indicators for accuracy.

Odinot_4.indd 16 10-11-2008 13:19:01

Chapter 4 describes an experimental study on the question if repeated retrieval of a subset of memory leads to the suppression of related memories (i.e., retrieval induced forgetting), and whether this affects confidence judgments.
Chapter 5 describes a case study on the memories of real life witnesses, three months after witnessing a robbery on a supermarket.
Finally, chapter 6 summarizes and discusses the results of the four studies in relation to the central research question of this dissertation.

## References


Anderson, M. C., Bjork, R. A., & Bjork, E. A. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1063-1087.

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.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102*, 68-100.

Baddeley, A. (1997). *Human memory: Theory and practice*. Hove: Psychology Press.

Barnier, A. J., Hung, L., & Conway, M. A. (2004). Retrieval- induced forgetting of emotional and unemotional autobiographical memories. *Cognition and emotion, 18*, 457-477.

Bartlett, F. C. (1932). *Remembering*. Cambridge: University press.

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.

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.

Bradfield, A. L., Wells, G. L., & Olson, E. A. (2002). The damaging effect of confirming feedback on the relationship between eyewitness certainty and identification accuracy. *Journal of Applied Psychology, 87*, 112-120.

Brainerd, C. J., Wright, R., Reyna, V. F., & Payne, D. G. (2002). Dual-retrieval processes in recall. *Journal of Memory and Language, 46*, 120-152.

Brewer, N., Keast, A., & Rishworth, A. (2002). The confidence-accuracy relationship in eyewitness identification: The effects of reflection and disconfirmation on correlation and calibration. *Journal of Experimental Psychology: Applied, 8*, 44-56.

Burke, D. M., MacKay, D. G., Worthley, J. S., & Wade, E. (1991). On the tip of the tongue: What causes word finding failures in young and old adults? *Journal of Verbal Learning and Verbal Behavior, 6*, 325-337.

Ceci, S. J., & Bruck, M. (1995). *Jeopardy in the courtroom: A scientific analysis of children's testimony*. Washington, D.C.: American Psychological Association Press.

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.

Clark, S. E. (1997). A familiarity-based account of confidence-accuracy inversions in recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 232-238.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of verbal learning and verbal behavior, 11*, 671-684.





Odinot_4.indd 17 10-11-2008 13:19:01

Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior, 14*, 185-191.
Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision making in eyewitness identification cases. *Law and Human Behavior, 12*, 41-55.
Eichenbaum, H., & Cohen, N. J. (2001). *From conditioning to conscious recollection*. New York: Oxford University Press.
Fleet, M. L., Brigham, J. C., & Bothwell, R. K. (1987). The confidence-accuracy relationship: The effects of confidence assessment and choosing. *Journal of Applied Psychology, 17*, 171-187.
Granhag, P. A., Stromwall, L. A., & Allwood, C. M. (2000). Effects of reiteration, hindsight bias, and memory on realism in eyewitness confidence. *Applied Cognitive Psychology, 14*, 391-420.
Gruneberg, M. M., & Sykes, R. N. (1993). The generalisability of confidence-accuracy studies in eyewitnessing. *Memory, 1*, 185-189.
Haber, R. N., & Haber, L. (2000). Experiencing, remembering and reporting events. *Psychology, Public Policy and Law, 6*, 1057-1097.
Hintzman, D. L. (1988). Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychological Review, 95*, 528-551.
Hirshman, E., & Master, S. (1997). Modelling the conscious correlates of recognition memory: Reflections of the remember-know paradigm. *Memory & Cognition, 25*, 345-352.
Hyman, I. E., Husband, T. H., & Billings, J. F. (1995). False memories of Childhood experiences. *Applied Cognitive Psychology, 9*, 181-197.
Johnson, M. K. (2006). Memory and reality. *American Psychologist, 61*, 760-771.
Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin, 114*, 3-28.
Johnson, M. K., & Raye, C. L. (1981). Reality monitoring. *Psychological Review, 88*, 67-85.
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.
Kelley, C. M., & Jacoby, L. L. (1998). Subjective reports and process dissociation: Fluency, knowing and feeling. *Acta Psychologica, 2-3*, 127-140.
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.
Lindsay, D. S., Allen, B. P., Chan, J. C. K., & Dahl, L. C. (2004). Eyewitness suggestibility and source similarity: Intrusions of details from one event into memory reports of another event. *Journal of memory and language, 50*, 96-111.
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 Behavior, 13*, 333-339.
Loftus, E. F. (1979). *Eyewitness testimony*. Cambridge, MA: Harvard University Press.
Loftus, E. F. (1997). Creating false memories. *Scientific American, 277*, 70-75.
Loftus, E. F. (2003). Our changeable memories: Legal and practical implications. *Nature Reviews: Neuroscience, 4*, 231- 234.
Luus, E. C. A., & Wells, G. L. (1994a). Eyewitness identification confidence. In D. F. Ross & J. D. Read & M. P. Toglia (Eds.), *Adult eyewitness testimony* (pp. 348-361). Cambridge: Cambridge University Press.
Luus, E. C. A., & Wells, G. L. (1994b). The malleability of eyewitness confidence: Co-witness and perseverance effects. *Journal of Applied Psychology, 79*, 714-723.



Odinot_4.indd 18 10-11-2008 13:19:01

MacLeod, M. D. (2002). Retrieval-induced forgetting in eyewitness memory: Forgetting as a consequence of remembering. *Applied Cognitive Psychology, 16*, 135-149.
Mandler, G. (1980). Recognising: The judgement of a previous occurrence. *Psychological Review, 27*, 252-271.
Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology, 94*, 339-354.
Münsterberg, H. (1908). *On the witness stand: Essays on psychology and crime*. New York: Doubleday.
Neisser, U. (1982). *Memory observed: Remembering in natural contexts*. San Francisco: Freeman.
Neisser, U., & Harsch, N. (1992). Phantom flashbulbs: False recollections of hearing the news about Challenger. In E. Winograd & U. Neisser (Eds.), *Affect and accuracy in recall: Studies of "flashbulb" memories* (Vol. 4, pp. 9-31). New York: Cambridge University Press.
Odinot, G., & Wolters, G. (2006). Repeated recall, retention interval and the accuracy-confidence relation in eyewitness memory. *Applied Cognitive Psychology, 20*, 973-985.
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.
Raaijmakers, J. G. W., & Shiffrin, R. M. (1992). Models for recall and recognition. *Annual Review of Psychology, 43*, 205-234.
Read, J. D., Lindsay, D. S., & Nicholls, T. (1998). The relation between confidence and accuracy in eyewitness identification studies: Is the conclusion changing? In C. P. Thompson & D. J. Herrman & J. D. Read & D. Bruce & D. G. Payne & M. P. Toglia (Eds.), *Eyewitness memory: Theoretical and applied perspectives* (pp. 107-130). Mahwah, New Jersey: Erlbaum.
Robinson, M. D., & Johnson, J. T. (1996). Recall memory, recognition memory, and the eyewitness confidence-accuracy correlation. *Journal of Applied Psychology, 81*, 587-594.
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.
Roediger, H. L., & McDermott, K. B. (1995). Creating false memories: Remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803-814.
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.
Shaw, J. S., Bjork, R. A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review, 2*, 249-253.
Shaw, J. S., 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. L. Lindsay (Eds.), *Handbook of eyewitness psychology* (Vol. 1, pp. 371-397). Mahwah, New Jersey: Erlbaum.
Shaw, J. S., McClure, K. A., & Wilkens, C. E. (2001). Recognition instructions and recognition practice can alter the confidence-response time relationship. *Journal of Applied Psychology, 86*, 93-103.
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.
Squire, L. R. (1992). Declarative and non-declarative memory: Multiple brain systems supporting learning and memory. *Journal of Cognitive Neuroscience, 99*, 195-231.
Stern, W. (1902). Zur Psychologie der Aussage: experimentelle Untersuchungen zur Erinnerungstreue. *Zeitschrift für die gesamte Rechtswissenschaft, 22, (1901/2)*.

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science, 14*, 455-461.
Tulving, E. (1972). Episodic and semantic memory. In E. Tulving & W. Donaldson (Eds.), *Organization of memory*. New York: Academic Press.
Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80*, 352-373.
Van Strien, P. J. (2000). Wild-west in de collegezaal. *De Psycholoog*, 35, 464-466.
Wade, K. A., Garry, M., & Lindsay, S. A. (2002). A picture is word a thousand lies. *Psychonomic Bulletin & Review, 9*, 597-603.
Wagenaar, W. A. (1988). Calibration and the effects of knowledge and reconstruction in retrieval from memory. *Cognition, 28*, 277-296.
Weber, N., & Brewer, N. (2003). The effect of judgment type and confidence scale on confidence-accuracy calibration in face recognition. *Journal of Applied Psychology, 88*, 490-499.
Wolters, G. (1983). *Episodic memory*. (Doctoral dissertation). Leiden: Leiden University.
Wolters, G., & Goudsmit, J. J. (2005). Flashbulb and event memory of September 11, 2001: Consistency, confidence and age effects. *Psychological Reports, 96*, 605-619.
Yonelinas, A. P. (2002). The nature of recollection and familiarity: A review of 30 years of research. *Journal of Memory and Language, 46*, 441-517.



Odinot_4.indd 20 10-11-2008 13:19:01

# 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

Odinot_4.indd 21 10-11-2008 13:19:01

## 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



Odinot_4.indd 22 10-11-2008 13:19:01

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




Odinot_4.indd 23 10-11-2008 13:19:01

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



Odinot_4.indd 24 10-11-2008 13:19:01

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 N=23 | Accuracy | .86 | .14 | .87 | .13 | .87 | .13 |
| | Confidence | 6.33 (.39) | 5.06 (.89) | 6.40 (.36) | 5.29 (.87) | 6.47 (.29) | 5.29 (.98) |
| Condition 2 N=24 | Accuracy | | | .80 | .20 | .80 | .20 |
| | Confidence | | | 6.31 (.44) | 4.79 (1.22) | 6.27 (.84) | 5.24 (.90) |
| Condition 3 N=20 | Accuracy | | | | | .71 | .29 |
| | 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).



Odinot_4.indd 28 10-11-2008 13:19:01

## 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 | Overall | **0.63*** | **0.61*** | **0.61*** |
| N=23 | Individual | 0.69 (.14) | 0.64 (.22) | 0.70 (.24) |
| Condition 2 | Overall | | **0.58*** | **0.57*** |
| N=24 | Individual | | 0.65 (.22) | 0.63 (.21) |
| Condition 3 | Overall | | | **0.49*** |
| N=20 | 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** | | | | | | | **Total units of information** |
|---|---|---|---|---|---|---|---|---|---|
| | | **1** | **2** | **3** | **4** | **5** | **6** | **7** | |
| 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) |

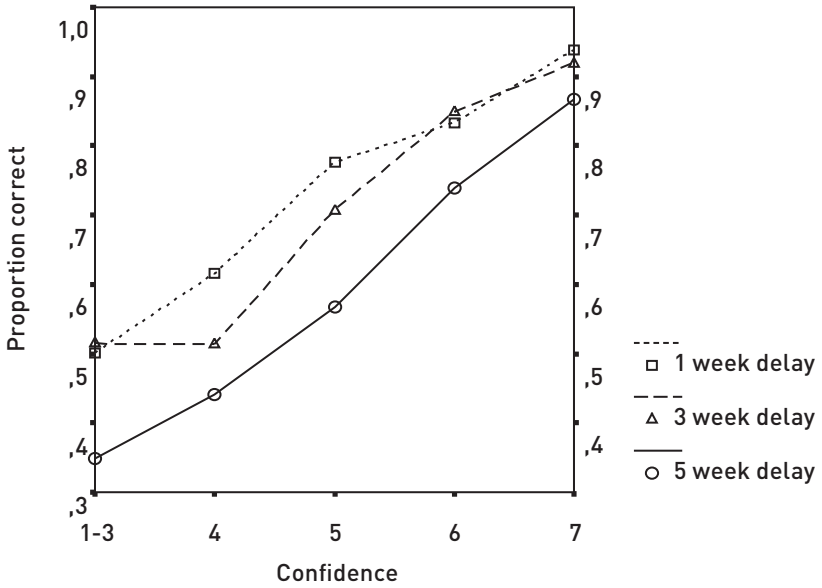
Table 3 seems to suggest a systematic relationship between delay of first recall and confidence level, both in the number of information units recalled and the proportion accu-

rate 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.



Odinot_4.indd 32 10-11-2008 13:19:02

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.

Odinot_4.indd 33 10-11-2008 13:19:02

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.



Odinot_4.indd 34 10-11-2008 13:19:02

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.





Odinot_4.indd 35 10-11-2008 13:19:02

Odinot_4.indd 36 10-11-2008 13:19:02

# 3

# Repeated suggestive questioning, accuracy, confidence and consistency in eyewitness event memory*

* This chapter is submitted as: Odinot, G., Wolters, G., & Giezen van, A. E. (submitted). Repeated suggestive questioning, accuracy, confidence and consistency in eyewitness event memory.

Odinot_4.indd 37 10-11-2008 13:19:02

## Summary

*In legal practice, both confidence and consistency of the testimony of eyewitnesses are used as indicators for accuracy. The present study was designed to assess the effects of repeatedly asking correct and suggestive questions on accuracy, confidence and consistency in recall of an episodic memory. Witnesses viewed a video and answered a questionnaire containing questions that were correct and suggestive in nature. The initial cued recall test was given after 1 week and was repeated after 3 and 5 weeks. Of the questions containing suggestive information, almost half were answered with incorrect details. For the questions correct in nature, correlations between accuracy, confidence and consistency were determined for final recall. Confidence appeared to be a better predictor for accuracy than consistency. Inconsistencies consisted mainly of omitted or committed units of information that were almost as accurate as consistently recalled units. We did not find any evidence for confidence inflation with repeated questioning.*

## Introduction

A substantial body of research on memory for everyday events has made it abundantly clear that memory is not only fallible, i.e., prone to forgetting, but also malleable, i.e., prone to errors (e.g., Deffenbacher, 1991; Wells & Loftus, 2003). Misleading post-event information or suggestive questioning can lead to inaccuracies in eyewitness memory (e.g., Loftus, 1980, 2005). People can be led to develop even completely false memories, i.e., brought to believe they remember events that never happened (e.g., Ceci, Huffman, Smith, & Loftus, 1994; Hyman, Husband, & Billings, 1995).

To distinguish between correct and incorrect memories people often rely on expressions of confidence, or on the consistency of multiple reports of the same memory. In the study reported here, we will examine the relationship between confidence, consistency and accuracy of repeatedly probed memories of a complex event.

It is generally believed that confidence about a memory is strongly related to the accuracy of a memory. This belief is not only held by laypeople but also by members of the legal profession (Cutler, Penrod, & Stuve, 1988; Leippe, 1980; Lindsay, Wells, & O'Connor, 1989; Luus & Wells, 1994; Penrod & Cutler, 1995). The confidence expressed by an eyewitness regarding his or her testimony is a strong determinant of the perceived credibility of the eyewitness (Leippe, Manion, & Romanczyk, 1992; Lindsay et al., 1989). Studies on 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 (Robinson & Johnson, 1996). Among experts, therefore, it is generally concluded that confidence is not a reliable predictor of memory accuracy (e.g., Leippe & Eisenstadt, 2007; Shaw, McClure, & Dykstra, 2007).



Odinot_4.indd 38 10-11-2008 13:19:02

The relationship between accuracy and confidence may further deteriorate when misleading information is presented, either by giving misinformation after witnessing an event, or by suggestive questioning. Shaw, Garcia, & Robles (1997) showed that misinformation indeed can interfere with the accuracy-confidence relation. In three studies, participants first viewed a videotape of a simulated robbery. Afterwards, they read a post-event narrative that contained consistent, inconsistent or neutral information with respect to details in the video. In all studies participants showed overconfidence during the final memory test in all conditions. The overconfidence was, however, significantly higher in the consistent and inconsistent conditions than in the neutral condition. Similar results were reported by Ryan and Geiselman (1991). Their findings indicated that the participants were more confident about biased knowledge (i.e., answers to both leading and misleading post-event information), than about their memories of unbiased information. The effects of misleading information can be enhanced by multiple exposures. Zaragoza and Mitchell (1996) showed that participants in a three-exposure condition were significantly more likely to falsely remember the misinformation with high confidence than participants in a single-exposure condition.

The effect of suggestive questioning on accuracy and confidence of memory was studied by Shaw, Garven and Wood (1997). They reported that although answering misleading questions sometimes resulted in incorrect answers, these answers were given with less confidence. Also Roebers (2002) has reported that confidence in the responses on misleading questions was lower than in the responses for correct questions. Gerrie, Belcher and Garry (2006) asked participants about missing action details in a video. In a recognition memory test, participants confidently, but falsely, remembered some of the suggested but missing action details. In a second experiment the missing details were either crucial or not crucial for the event. In this situation participants were more likely to falsely recognize, and to be more confident, with missing noncrucial than missing crucial information. Repeating suggestive questions seems to augment its effect on confidence. Blagrove and Akehurst (2000) questioned their participants twice. The repetition led to significant increases of confidence in suggested responses to the misleading questions. Recently, also Pezdek, Sperry and Shana (2007) reported that answering the same suggestive questions twice led to an increase in confidence.

Distortions of accuracy, confidence and consistency may also occur simply by repeated questioning (or repeated reflective thought). For instance, several authors have suggested that repeated recall may cause confidence inflation because it enhances ease of retrieval or response fluency (e.g., Robinson, Johnson, & Robertson, 2000; Shaw, McClure, & Wilkens, 2001). In a recent review Shaw et al. (2007) concluded that repeated questioning generally leads to increases in the confidence ratings. These conclusions are, however, mainly based on testing memory with recognition tasks. In contrast, two other studies using cued recall for testing memory of details of complex naturalistic events failed to find evidence of confidence inflation with repeated recall of the same information (Ebbesen & Rienick, 1998; Odinot & Wolters, 2006; Turtle & Yuille, 1994).

In a crime investigation, witnesses are often questioned more than once. One of the reasons to question witnesses several times is the idea that witnesses may provide additional information during follow-up questioning. Roediger, McDermott and Goff (1997) concluded, however, that repeated recall can have both facilitating and detrimental effects on later retention. To understand these effects of repeated recall it is important to note that retrieval is not a neutral process, which leaves 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).

One characteristic of repeated recall is that it enhances the chances that various kinds of inconsistencies occur. Information remembered the first time may not be remembered later (omission error), or vice versa (commission error). Or information remembered the first time may be remembered differently a second time (distortion error). Especially the latter, but to some extend also the former types of inconsistencies in testimonies are considered as strong indicators for inaccuracies (Brewer, Potter, Fisher, Bond, & Luszcz, 1999; McNally, 2003; Talarico & Rubin, 2003). When eyewitnesses provide inconsistent information, concerns arise about the overall accuracy of the witnesses account. Fisher and Cutler (1995) have reported data from a survey, confirming that judges and lawyers strongly believe that inconsistency is predictive for inaccuracy. Conversely, consistency is often taken interpreted as an indicator or a proxy for accuracy. Accuracy and consistency, however, refer to different concept; reports of a witness can be consistent, without necessarily being accurate. Studies on the consistency-accuracy relationship are sparse. Brewer et al. (1999), Fisher et al. (1995) and Smeets, Candel and Merckelbach (2004) concluded that consistency is not a strong predictor of the accuracy of testimonies. Contrasting results are reported by Van Giezen, Arendsman and Spinhoven (2007), who found that consistency of reports for both neutral and emotional stimuli were significantly correlated with memory accuracy.

The present study was designed to assess the effects of repeatedly asking correct and suggestive questions on confidence, accuracy and consistency in the recall of an episodic memory. To that end, participants were shown a videotape of an extended complex 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 recall cues consisted of open-ended questions that needed not to be answered if the participant did not remember. The suggestive questions asked about information that might have been present but was not actually shown in the video. The initial test was given after 1 week and was repeated after 3 and 5 weeks. To get more insight in delay and repetition effects, control groups received the test twice (after 3 and 5 weeks) or only once (after 5 weeks). Relations between accuracy, confidence and consistency were determined for final recall in condition 1 (three repetitions) and 2 (two repetitions).



Odinot_4.indd 40 10-11-2008 13:19:02

## Method

### Participants

A group of 62 undergraduate students (52 female and 10 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 = 21) consisted of three recall sessions 1, 3 and 5 weeks after the video presentation, Condition 2 (N = 20) consisted of two recall sessions 3 and 5 weeks after the video presentation. Condition 3 (N = 21) 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. None of the participants indicated that they had seen the video before.

*Questionnaires*. For the recall sessions, a questionnaire was constructed consisting of 28 open-ended questions of which five contained suggestive information (see appendix 1). The questionnaire started with a very general question asking the participants to describe the two story lines in general terms. This question was included to reinstate and refresh the memory of the video before proceeding with the more specific questions. All other questions were open-ended question asking about specific details shown in the video. Some questions were more specific (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"). Five questions contained suggestive information, asking about aspects that were not shown in the video. For instance: "The driver of the car that was involved in the accident was injured. Where and how bad was he injured?", while the driver of the car was not injured at all. These five questions were randomly interspersed with the questions that were correct in nature. The same questionnaire was used in all recall sessions.

### 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 were the only witnesses, and that it was important therefore to report as accurate as possible about any details they remembered from the original video-presentation. It was also stressed, however, 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. To explain this to the participants, 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. After writing down the answer, 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).

After finishing the questionnaire, the experimenter judged if the information was given in small units and if confidence indications were made to every unit of information. Only in a few cases, participants were asked to do so afterwards. After completing the final session, participants were debriefed and paid or given credits.

The correct and suggestive questions were scored separately. All units of information provided by the participants on the correct questions 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 apparently incorrectly remembered or fantasized by the participant. The answers given on questions containing suggestion were scored in three categories. First, participants could avoid an answer by choosing the 'do not know' option. Second, the participant could respond with an answer in which the suggestion was rejected explicitly (e.g., "I haven not seen any blood"), this was scored as *correct*. Third, any other answer indicating acceptance of the suggestion (e.g., "there was blood on his arm") was scored as *incorrect*.

The responses were also scored on consistency by comparing the units of information for each participant across the recall sessions. Information was scored as *consistent* when it was recalled in all subsequent recall sessions (three times in condition 1 and two times in condition 2). Information did not have to be literally the same to be consistent, for instance, information as "the car didn't stop", provided during the first session and subsequently recalled as "the car kept moving" during the next recall session was scored as consistent. Information not recalled across all recall sessions or recalled differently was scored as *inconsistent*.



Odinot_4.indd 42 10-11-2008 13:19:02

All answers were scored on accuracy and consistency independently by two experimenters. Their scoring coincided in 94% of the cases. In case of disagreement, a third experimenter settled the dispute. Of all units generated 1.8% could not be classified as correct or incorrect; these units of information were discarded from further analysis. The mean number of information-units given to correct questions was 2.53 (with a minimum of 1 and a maximum of 29).

## Results

There are five main foci in this study: responses to suggestive questions, responses to correct questions, the consistency and confidence in the responses, and the relationship between accuracy, confidence and consistency. We will discuss these foci in separate sections.

### Suggestive Questions

The questionnaire contained five questions that were suggestive in nature. Basically there were three ways to respond: withholding a response by using the "I don't know" option, giving a correct response in which the suggestion is rejected ("I haven't seen any blood"), or giving an incorrect response in which the suggestion is accepted and results in a fantasized answer ("The man was bleeding at his arm"). Results were analyzed by comparing the numbers (or proportions) of each type of answer as a function of delay and repetition (see Table 1).

*Retention interval*. To analyze the effects of retention interval, the first recall sessions in the 1-, 3-, and 5-week interval conditions were compared and tested for differences in accuracy and confidence. Note that this is a comparison between groups.

The proportion of suggestive questions answered with an "I do not know" response (0.46 after 1 week, 0.35 after 3 weeks and 0.52 after 5 weeks), did not differ significantly as a function of delay ($F$ (2, 59) = 2.21, *NS*). However, the number of responses in which the suggestion was correctly rejected by the participants, showed a significant decrease with longer retention intervals ($F$ (2, 59) = 5.34, $p < .05$). The proportion of these answers dropped from 0.23 after a retention interval of one week, to 0.11 and 0.10, after three and five weeks, respectively. The confidence in these correct responses did not show a significant decrease with retention interval. The mean confidence was 5.64, 5.30 and 5.28, after 1, 3 and 5 weeks, respectively ($F$ (2, 34) = 0.30, *NS*).

**Table 1** Proportions of correctly rejected (C), incorrectly fantasized (I) and "Do not know" (D) answers provided on the suggestive questions, and corresponding average confidence ratings (sd in parentheses), as a function of retention interval and repeated recall.

| Suggestive questions | | Retention interval | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | 1 week | | | 3 weeks | | | 5 weeks | | |
| | | C | I | D | C | I | D | C | I | D |
| Condition 1 | Proportion | .23 | .30 | .46 | .20 | .34 | .46 | .20 | .34 | .46 |
| (N=21) | Confidence | 5.64 (1.2) | 4.23 (1.4) | | 5.84 (.85) | 3.92 (1.4) | | 5.90 (1.3) | 4.35 (1.4) | |
| Condition 2 | Proportion | | | | .11 | .54 | .35 | .10 | .59 | .31 |
| (N=20) | Confidence | | | | 5.30 (1.7) | 3.85 (1.4) | | 4.88 (1.88) | 3.82 (1.3) | |
| Condition 3 | Proportion | | | | | | | .09 | .38 | .52 |
| (N=21) | Confidence | | | | | | | 5.28 (1.1) | 3.89 (1.6) | |

The proportion of questions that were answered incorrectly was 0.30 after 1 week, 0.54 after 3 weeks and 0.38 after 5 weeks. These differences were significant ($F$ (2, 59) = 4.04, $p$ < .05), but there is no clear pattern relating incorrect answers and delay. Bonferroni post-hoc test showed a significant difference only between the interval of one and three weeks ($p$ < .05). The mean confidence in these incorrect answers (4.23, 3.85 and 3.89, after 1, 3 and 5 weeks, respectively) showed a decreasing, but not significant, trend ($F$ (2, 50) = .35, NS).

*Repeated recall*. The effect of repeated recall on accuracy and confidence is analyzed as a within subjects repeated measures variable, for condition 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). Repetition of recall did not affect the proportion of "do not know" responses. In condition 1 the proportions were 0.46, 0.46 and 0.46 for recall sessions 1, 2 and 3, respectively ($F$ (2, 19) = .00, *NS*), and in condition 2 these proportions were 0.35 and 0.31, in recall sessions 1 and 2, respectively ($t$ (19) = 1.01, *NS*). Also the proportions of correct rejection responses, and the mean confidence in these responses, were not significantly affected by repetition.

The effect of repetition on the proportion of incorrect answers showed a small, and insignificant increase, both in condition 1 (0.30, 0.34 and 0.34) and condition 2 (0.54 and 0.59). The mean confidence in the incorrect responses in condition 1 (4.15, 4.08 and 4.76 in the first, second and third recall session, respectively) showed a significant, but somewhat difficult to explain, difference between the second and third recall ($F$ (2, 12) = 6.15, p < .05, Bonferroni p < .05). The mean confidence of incorrect answers in condition 2 did not show an effect of repetition (M = 3.85 and 3.82, respectively, $t$ (16) = .14, *NS*).

It is of particular interest to compare the proportions and average confidence of correct (suggestion rejected) and incorrect (suggestion accepted) responses. From Table 1 it can be inferred that the proportion of incorrect answers was larger than the proportion of correct answers in all conditions. It is also clear, however, that the mean confidence of incor-



Odinot_4.indd 44 10-11-2008 13:19:02

rect answers was lower than the confidence of correctly answered questions. Bonferroni posthoc tests showed that all these differences in proportions and average confidence were significant ($p < .05$). Although on average confidence in incorrect answers is lower than in correct answers, still quite a few incorrect responses were given with a high level of confidence. The distribution of correct and incorrect answers to suggestive questions over confidence ratings is shown in Table 2.

Overall the results on the questions containing suggestive information indicate that suggestion had a substantial effect. In the first recall session, the subjects in conditions 1, 2 and 3 answered (either with an "I do not know" or a rejection answer) in 69%, 46% and 62% of the cases without falling for the suggestion, respectively. This means, however, that in 31%, 54% and 38% of the cases an incorrect answer was given, even when it was explicitly allowed to refrain from answering. Moreover, although on average confidence in incorrect answers was lower than for correct answers, a substantial proportion was given with a high level of confidence.

**Table 2** The number of correct and incorrect responses to suggestive questions at each confidence level as a function of conditions; the proportions of incorrect answers are presented in parentheses.

| Condition | Retention interval | Confidence Scale | | | | | | | Number of Do not know responses |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| Condition 1 N= 21 | 1 week | 3 (.66) | 3 (1.0) | 9 (.88) | 6 (.50) | 13 (.46) | 12 (.58) | 10 (.20) | 49 |
| | 3 weeks | 5 (1.0) | 4 (1.0) | 5 (1.0) | 7 (.86) | 14 (.57) | 11 (.45) | 10 (.20) | 48 |
| | 5 weeks | 3 (1.0) | 5 (1.00 | 4 (.75) | 8 (.75) | 12 (.83) | 10 (.60) | 15 (.26) | 48 |
| Condition 2 N=20 | 3 weeks | 7 (.78) | 9 (.88) | 11 (.90) | 11 (.91) | 9 (.77) | 10 (.70) | 8 (.63) | 35 |
| | 5 weeks | 4 (1.0) | 10 (.80) | 11 (1.0) | 13 (1.0) | 16 (.81) | 11 (.73) | 5 (.60) | 31 |
| Condition 3 N= 21 | 5 weeks | 5 (1.0) | 7 (1.0) | 4 (1.0) | 11 (.73) | 9 (.55) | 9 (.88) | 5 (.60) | 55 |

The 20 participants in condition 2 seem to be more prone to suggestion than the participants in conditions 1 and 3. Because there are no obvious differences in the presentation and testing conditions to explain this difference, it suggests that it may be caused by group differences. Therefore, we also looked at individual results. The individual participants, indeed, showed a wide variation in their responses to the suggestive questions. Of the 62 participants, nine refrained from answering any of the suggestive questions during first recall, whereas 10 participants answered at least four of the five suggestive questions. As the results already suggested, relatively more participants in condition 2 than in the other two conditions answered incorrectly to suggestive questions.

## Correct Questions

To analyze the effects of retention interval and repeated recall on accuracy and confidence, we determined the number of correct and incorrect units of information given by the participants in all conditions. The proportions correct and incorrect units of information and the corresponding mean confidence judgments are shown in Table 3.

First, we examined the effect of retention interval on the number of questions that were answered with an "I don't know" response and on the total number of units of information. During the first recall session after one week only 5.4% of the answers to all 23 correct questions were "I don't know" responses. After retention intervals of 3 and 5 weeks, this number was not significantly different (4.4 % and 7.9 %, respectively; ($F$ (2, 59) = 1.40, *NS*). The average total number of units of information provided per participant to the 23 correct questions also did not show a significant effect of delay ($F$ (2, 59) = 1.16, *NS*). The average total number of units of information, correct or incorrect, that was recalled after 1, 3 and 5-week intervals were 61.7, 56.2 and 55.9, respectively.

**Table 3** Proportions correct and incorrect units of information, and corresponding average confidence ratings (sd in parentheses), as a function of retention interval and repeated recall.

| **Correct questions** | | **Retention interval** | | | | | |
|---|---|---|---|---|---|---|---|
| | | **1 week** | | **3 weeks** | | **5 weeks** | |
| | | **Correct** | **Incorrect** | **Correct** | **Incorrect** | **Correct** | **Incorrect** |
| Condition 1 | Proportion | .88 | .12 | .85 | .15 | .85 | .15 |
| N=21 | Confidence | 6.37 (.66) | 5.50 (.68) | 6.34 (.53) | 5.75 (.72) | 6.46 (.44) | 5.55 (.84) |
| Condition 2 | Proportion | | | .80 | .20 | .79 | .21 |
| N=20 | Confidence | | | 6.25 (.43) | 5.12 (.93) | 6.20 (.53) | 5.23 (1.01) |
| Condition 3 | Proportion | | | | | .77 | .23 |
| N=21 | Confidence | | | | | 6.10 (.47) | 4.95 (.48) |

*Retention interval*. The effect of retention interval was analyzed by comparing results of the first recall sessions only. The number of correctly recalled units of information showed a significant decrease with retention interval, $F$ (2, 59) = 21.99, $p$ < 0.001. Bonferroni post-hoc tests showed significant differences between 1- and 3-week intervals (proportions correct 0.88 and 0.80, respectively, $p$ < .05), and between 1- and 5- week intervals (proportions correct 0.88 and 0.77, respectively, $p$ < .05). Confidence in the correct units of information also seems to decrease with longer intervals, but the difference after one, three and five weeks (M = 6.37, 6.25 and 6.10, respectively) was not significant ($F$ (2, 59) = 1.33, *NS*).

The number of incorrectly recalled units of information showed a significant increase with retention interval ($F$ (2, 59) = 21.99, $p$ < .001). Proportion incorrect after one week (0.12) was significantly lower than after three (0.20) or five weeks (0.23). Confidence in the incorrect units showed a significant decrease with longer intervals ($F$ (2, 59) = 3.22, $p$ < .05). Bonfer-



Odinot_4.indd 46 10-11-2008 13:19:02

roni post-hoc tests indicated that the mean level of confidence for incorrect recalled information was higher after 1 week (M = 5.50) than after 5 weeks (M = 4.95, $p < .05$).

In sum, the proportion of correctly recalled units of information decreased with longer retention intervals, while at the same time the proportion of incorrect information significantly increased. In the recall of both correct and incorrect units there is a trend that confidence decreases with increasing retention intervals. Interestingly, during initial recall participants were always significantly more confident about correctly than about incorrectly recalled units of information ($p < 0.01$, for all delay conditions).

*Repeated recall*. The effect of repeated recall was analyzed as a within subjects repeated measures variable, for condition 1 (3 recall sessions) and condition 2 (2 recall sessions). The proportion correctly recalled units of information remained almost the same across the subsequent recall sessions, both in condition 1 (0.88, 0.85 and 0.85, respectively) and in condition 2 (0.80 and 0.79, respectively). Also, the mean levels of confidence for these correctly recalled units of information were not significantly influenced by repeated recall in condition 1 (6.37, 6.34 and 6.46, respectively; $F$ (2, 40) = .43, *NS*), nor in condition 2 (6.25 and 6.20, respectively, $t$ (19) = .86, *NS*).

Also the proportion of incorrectly recalled units showed no effect of repeated recall ($F$ (2, 40) = 2.25, *NS*). Proportions incorrect units in condition 1 were 0.12, 0.15 and 0.15, in the first, second and third recall session, respectively. In condition 2 these proportions were 0.20 and 0.21, respectively. Also the mean confidence of the incorrectly recalled units was unaffected by repetition, both in condition 1 (5.50, 5.75 and 5.55 respectively; $F$ (2, 40) = 1.29, *NS*), and in condition 2 (5.12 and 5.23, respectively, $t$ (19) = -.60, *NS*).

Total units of information and proportions of inaccurate units as a function of confidence judgment levels are shown in Table 4. From this Table it is clear that in all conditions the proportion of incorrect units drops steadily with higher confidence levels. After a recall interval of one week, the proportion of errors with high confidence ratings is 0.06. With intervals of 3 and 5 weeks this proportion increases to 0.11. These figures suggest that confidence may be used as a modest indicator of accuracy. However, one should always be aware that high confidence is never a guarantee for accuracy.



Odinot_4.indd 47 10-11-2008 13:19:02

**Table 4** The total number of units of information at each confidence level per condition for the correct questions; proportions of incorrect units are presented in parentheses.

| | Retention interval | Confidence Scale | | | | | | | Total units of information |
|---|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| Condition 1 N = 21 | 1 week | 1 (1.00) | 12 (.75) | 23 (.430 | 60 (.33) | 151 (.23) | 209 (.18) | 813 (.06) | 1269 (.12) |
| | 3 weeks | 6 (.33) | 11 (.54) | 24 (.58) | 43 (.26) | 136 (.25) | 190 (.22) | 788 (.09) | 1198 (.15) |
| | 5 weeks | 3 (.67) | 17 (.53) | 30 (.33) | 43 (.37) | 119 (.29) | 176 (.21) | 805 (.09) | 1193 (.15) |
| Condition 2 N = 20 | 3 weeks | 7 (.86) | 40 (.52) | 32 (.56) | 74 (.42) | 141 (.30) | 227 (.21) | 603 (.10) | 1124 (.20) |
| | 5 weeks | 10 (.60) | 33 (.58) | 44 (.45) | 52 (.38) | 151 (.32) | 180 (.24) | 540 (.11) | 1010 (.21) |
| Condition 3 N = 21 | 5 weeks | 16 (.31) | 38 (.53) | 61 (.52) | 102 (.44) | 148 (.34) | 213 (.26) | 596 (.11) | 1174 (.23) |

## Consistency

The consistency in the responses across the recall sessions was determined for conditions 1 and 2, for the correct and suggestive questions separately.

*Suggestive questions*. A response to the same question was scored as consistent when the suggestion was correctly rejected across all recall sessions or when the same correct or incorrect answer was given. An answer was scored as inconsistent when an answer was not recalled consistently across the recall sessions or recalled differently.

In condition 1 a proportion of 0.74 of all answers to the suggestive questions was consistent across the three recall sessions. Of this proportion, 0.35 was a consistent "I don't know" answer, 0.18 were answers in which participants consistently rejected the suggestion, and 0.21 were recalled consistently but incorrectly. Of all answers to the suggestive questions a proportion of .26 was inconsistent. Of this proportion, 0.06 were inconsistencies in which a previously correct rejection or don't know answer changed into an incorrect answer. Mean confidence in consistent responses was significantly higher (M = 5.25) than confidence in inconsistent responses (M = 4.21; $t$ (13) = -2.77, $p$ < .05). Confidence did not increase significantly over repetitions (M = 5.78, 5.94 and 6.00 for recall 1, 2 and 3, respectively).

In condition 2 a proportion of 0.84 of all responses to the suggestive questions were consistent. Of this proportion, 0.26 were consistent "I don't know" responses, 0.08 were consistent correct rejection responses, and a remarkable 0.50 were consistently incorrect. Confidence was higher for all consistent responses (M = 4.63) than for the inconsistent responses (M = 3.82), but this difference just failed to reach significance ($t$ (6) = -2.3, *NS*). Again, repetition of consistent answers did not affect confidence significantly.



Odinot_4.indd 48 10-11-2008 13:19:02

*Correct Questions*. Consistencies in responses to the correct questions were analyzed by comparing the units of information across the recall sessions. In the responses to the correct questions in condition 1, the participants provided a total of 1683 single units of information. Almost half of all units, a proportion of 0.49, were consistently recalled during the three recall sessions. Of this proportion 0.43 was accurate and 0.06 was consistent but inaccurate. A proportion of 0.51 was answered inconsistently. In large part (0.43) these were correct units of information missing in one or two of the recalls sessions (omission or commission). Of the part that was incorrect (0.08), most were incorrect details given on one occasion but not on another (0.06), and only 0.02 were distortions (different answers on different occasions).

The proportion correct of consistently and inconsistently recalled information was the same (0.43). Also the proportion incorrect of consistently (0.06) and inconsistently (0.08) recalled information did not differ significantly ($t$ (20) = -1.24, *NS*). So consistently recalled information is not more correct than inconsistently recalled information. The participants were however, significantly more confident about consistently recalled information (M = 6.12) than about inconsistently recalled information (M = 5.89, $t$ (20) = 2.13, $p < .05$).

In condition 2, a total of 1416 units of information were provided to the correct questions. A portion of .51 of all these units was consistently recalled across the two recall sessions, of which 0.43 was correct and 0.08 was incorrect. Of the 0.49 inconsistently recalled units, 0.37 were correct (omissions and commissions), 0.09 changed from correct to incorrect, and 0.03 were distortions. Also here, the proportions correct and incorrect of consistently and inconsistently recalled information did not differ significantly. Although participants expressed more confidence in consistent information (M = 5.68) than inconsistent information (M = 5.51), this difference was not significant ($t$ (19) = 1.07, NS).

In sum, in both conditions roughly half of all the information provided was consistent across the recall sessions. Interestingly, the proportion correct of inconsistently recalled units of information was about the same as in consistent recall, but participants were more confident about consistently than inconsistently recalled information.

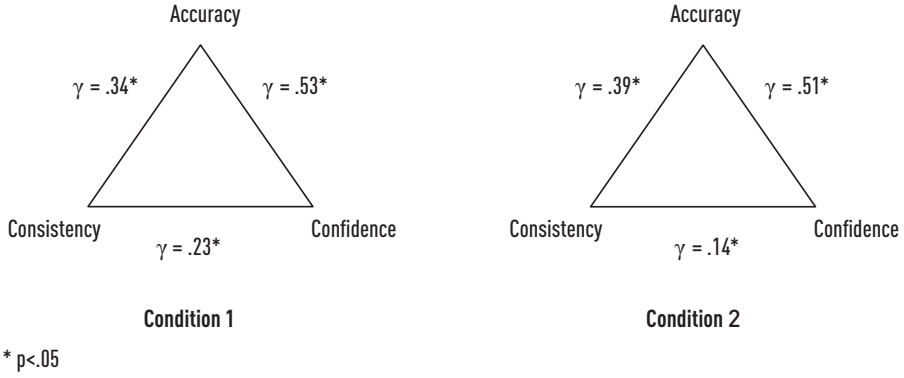
## Relation between accuracy, confidence and consistency

To quantify the relations between accuracy, confidence, and consistency for the correct questions, we calculated Goodman-Kruskal gamma correlation coefficients in the responses during final recall only. All units provided during final recall in conditions 1 and 2 were scored for accuracy, confidence and consistency. A unit was scored as consistent when it was recalled across all sessions; information was scored as inconsistent when it was recalled during final recall, but not at all or differently during previous sessions. Next, accuracy and confidence was determined for all consistent and inconsistent units to calculate gamma correlation coefficients, shown in Figure 1.

As can be inferred from this figure all correlations were positive and significant ($p < .05$). However, the relation between confidence and accuracy is stronger than consistency and accuracy. Relatively high correlations between accuracy and confidence were also found in the first and second recall in condition 1 ($\gamma = .61$ and $.51$, $p < .05$), the first recall in condition 2 ($\gamma = .55$, $p < .05$), and in condition 3 ($\gamma = .54$, $p < .05$).

The rather modest correlations between consistency and accuracy are in line with the previously discussed finding that accuracy rate was about the same for consistent and inconsistent answers. We also noted already that in the conditions as studied here, inconsistencies are mainly errors of omission or commission, i.e., incomplete recall in subsequent sessions. The results clearly indicate that this kind of incompleteness is not related to accuracy. In contrast, we found very few instances of the kind of inconsistency that would certainly predict inaccuracy, namely changes in the information provided to the same questions (distortions).

**Figure 1** Gamma correlations between accuracy, confidence and consistency for the correct questions during final recall, condition 1 and 2, respectively.




* p<.05





## Discussion

The purpose of this study was to examine the effect of repeated questioning and retention interval on the accuracy, confidence and consistency in the recall of a complex episodic event when memory is probed with questions that are suggestive or correct in nature.

The inclusion of questions containing suggestive information, inconspicuous interspersed with correct questions, led to particularly revealing findings. In total, almost half of the suggestive questions were answered with obviously incorrect details (depending on groups and conditions, proportions incorrect answers ranged from 0.30 and 0.59) and longer retention intervals resulted in less correct rejections of the suggestion. Moreover, a substantial proportion of the incorrect responses were consistently reported across the recall sessions.

Odinot_4.indd 50 10-11-2008 13:19:03

Although on average incorrect responses were given with lower confidence, still a substantial proportion was given with moderate to high degrees of confidence.

The main findings for the correct questions were that after longer intervals before first questioning less correct units of information are recalled and with lower confidence levels. Although the number of incorrect units of information increases with longer retention intervals, incorrect answers remain a minority compared to correct answers, and they are generally characterized by lower confidence levels. Repeated questioning did not enhance recall, nor did it influence the proportion of inaccurate recall. Moreover, we did not find any evidence for confidence inflation with repeatedly retrieving identical information from memory, replicating previous findings of Odinot and Wolters (2006).

Results with regard to consistency were also revealing. Of the units of information given to correct questions only about half was mentioned consistently over subsequent recall sessions. Interestingly, the percentage correct was about the same for consistently and inconsistently recalled units of information (e.g., in condition 1 consistent answers were 88% correct and inconsistent answers were 84% correct). This finding is due to the fact that most inconsistencies were errors of omission or commission, i.e., mentioning particular units of information in one recall session but not in another. Of the 16% inconsistent and incorrect units of information, only 5% were actual distortions (changes in the content of the information provided), the rest were incorrect units of information that were given on one occasion but not on another.

The conclusion for the relationship between accuracy, confidence and consistency is that in the conditions as studied all these relations are modest. Neither confidence nor consistency is very useful for predicting accuracy, but if anything, confidence is a better indicator than consistency.

The outcome of this study has some important practical implications. First, our findings illustrate that questions asked have to be free of suggestion. Obviously, this is not a new finding, but it emphasizes once again the importance of correctly interviewing witnesses. Any suggestion of misleading information strongly increases the chance that an incorrect answer is provided, and neither confidence nor consistency can be used to detect these inaccuracies.

Second, with correct questions our results show relatively high confidence-accuracy correlations. Especially with the shortest recall interval (one week), the large majority of answers given with a high confidence tend to be correct. With some caution, therefore, in event recall confidence may be used as a partial indicator of accuracy during the early stages of an investigation.

Third, in this study consistency is not strongly related with accuracy, a finding that is in line with most studies reported on this issue (Brewer et al., 1999; Fisher & Cutler, 1995; Smeets et al., 2004). In legal practice inconsistencies are seen as strongly predictive for accuracy. Although this is obviously true for contradictory inconsistencies (distortions), it is not true

for inconsistencies due to incomplete statements (errors of omission or commission). Witnesses should not be judged as unreliable when previously given information is omitted in later statements or when in later statements additional information is provided.

Fourth, repeatedly asking the same question did not seem to inflate the confidence expressed in the answers provided. Even when identical information was repeatedly and consistently retrieved from memory, confidence was not enhanced. This is not in line with the retrieval fluency hypothesis as suggested by Shaw (1996) and Shaw and McClure (1996). 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). In this study however, also no evidence for imagination inflation is found with the repetition of fantasized responses.

The general belief among memory experts is that we should be reluctant to use confidence as an indication of accuracy. This might be true for line-up identifications, but probably not for the recall of events. A witness who is very confident remembering some detail of a well-observed event is much more likely to be correct than incorrect, provides an important instrument. However, even a highly confident witness is likely to err in about 10% of the details provided. It is important, therefore, to be reluctant in applying the knowledge about the relationship between accuracy and confidence. Firstly, this knowledge is better used during police investigations, when possible errors are not crucial, than using it as evidence during a court trial. Secondly, our study shows that confidence as an indicator for accuracy becomes less strong after longer retention intervals. This is again an argument to restrict the use of confidence to the early stages of a police investigation and not to use it in later stages, i.e., in front of a judge or jurors. The time between witnessing an event and telling about it in court is much longer than the 5 weeks tested in this experiment. Moreover, with an increasing period between witnessing an event and testifying about it, there is an increasing possibility that a witness is exposed to suggestion and misinformation. As our results show, this clearly results in a less reliable relation between accuracy and confidence. Therefore, the conclusion remains that judges, lawyers and prosecutors have to be aware that the confidence expressed by a witness in court about their memories of an event is not a good predictor for accuracy.



## References

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.

Blagrove, M., & Akehurst, L. (2000). Effects of sleep loss on confidence-accuracy relationships for reasoning and eyewitness memory. *Journal of Experimental Psychology: Applied, 6*, 59-73.

Brewer, N., Potter, R., Fisher, R. P., Bond, N., & Luszcz, M. A. (1999). Beliefs and data on the relationship between consistency and accuracy of eyewitness testimony. *Applied Cognitive Psychology, 13*, 297-313.

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.

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., & Cutler, B. L. (1995). The relation between consistency and accuracy of eyewitness testimony. In G. Davies & S. lloyd Bostock & M. McMurran & C. Wilson (Eds.), *Psychology, law and criminal justice: International developments in research and practice* (pp. 21-28). Berlin: De Gruyter.

Gerrie, M. P., Belcher, L. E., & Garry, M. (2006). 'Mind the gap': False memories for missing aspects of an event. *Applied Cognitive Psychology, 20*, 689-696.

Giezen van, A. E., Arensman, E., & Spinhoven, P. (2007). Memory for emotionally arousing and neutral stimuli: An experimental study on accuracy and consistency of memory in victims of violent crime. *Manuscript submitted for publication*.

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

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.

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

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.

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.

Loftus, E. F. (1980). *Memory: Surprising new insights into how we remember and why we forget*. Addison-Wesley: Publishing Company, Inc.

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*, 361-366.

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.

McNally. (2003). *Remembering trauma*. Cambridge, MA: The Belknap Press of Harvard University Press.

Odinot, G., & Wolters, G. (2006). Repeated recall, retention interval and the accuracy-confidence relation in eyewitness memory. *Applied Cognitive Psychology, 20*, 973-985.

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

Pezdek, K., Sperry, K., & Shana, O. M. (2007). Interviewing witnesses: The effect of forced confabulation on event memory. *Law and Human Behavior, 31*, 463-478.

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

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.



Odinot_4.indd 53 10-11-2008 13:19:03

Roebers, C. M. (2002). Confidence judgments in children's and adults' events recall and suggestibility. *Developmental Psychology, 38*, 1052-1067.

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.

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.

Shaw, J. S., Garcia, L. A., & Robles, B. E. (1997). Cross-language postevent misinformation effects in spanish-english bilingual witnesses. *Journal of Applied Psychology, 82*, 889-899.

Shaw, J. S., Garven, S., & Wood, J. M. (1997). Co-witness information can have immediate effects on eyewitness memory reports. *Law and Human Behavior, 21*, 503-523.

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.

Shaw, J. S., 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. L. Lindsay (Eds.), *Handbook of eyewitness psychology* (Vol. 1, pp. 371-397). Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Shaw, J. S., McClure, K. A., & Wilkens, C. E. (2001). Recognition instructions and recognition practice can alter the confidence-response time relationship. *Journal of Applied Psychology, 86*, 93-103.

Smeets, T., Candel, I., & Merckelbach, H. (2004). Accuracy, completeness, and consistency of emotional memories. *American Journal of Psychology, 117*, 595-609.

Talarico, J. M., & Rubin, D. C. (2003). Confidence, not consistency, characterizes flashbulb memories. *Psychological Science, 14*, 455-461.

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.

Zaragoza, M. S., & Mitchell, K. J. (1996). Repeated exposure to suggestion and the creation of false memories. *Psychological Science, 7*, 294-300.



Odinot_4.indd 54 10-11-2008 13:19:03

# Appendix 1

## The suggestive questions

*"One of the boys tried to impress the girls by riding circles around them on his moped and making a lot of noise with the engine. What was the irritated reaction of the two girls?"*

In the video was shown that the two girls had to stop their bicycles because the boy was driving in circles around them on his moped, making a lot of noise. But the girls did not say anything, giggled a little bit and followed their way. There was no irritation shown by the girls.

*"At a certain moment two men had a disagreement while they were trading a car. What was this disagreement about?"*

In the video was shown that two men were trading a moped. The atmosphere between the men was friendly and there was no disagreement at all.

*"The minivan of the father was partly plastered with an advertisement. Can you describe the text and/ or illustration?"*

The minivan of the father was not covered with advertisement. However, this kind of minivan is often used for business purposes and advertisements.

*"The driver of the car that was involved in the accident was injured. Where and how bad was he injured?"*

The driver of the car was not injured at all. In the video was shown that the driver stepped out of his car, right after the accident to check the condition of the person who lies in front of the car.

*"The victim of the accident was seriously injured. Where was the boy bleeding?"*

In the video was shown that the boy was driving his moped and hit the back of the car in full speed. The boy flies over the car and rolls back on the ground. He lies unconsciously in front of the car however, he is not bleeding or visually wounded.

# 4

# Repeated partial eyewitness questioning causes confidence inflation but not retrieval-induced forgetting*

* A short version of this chapter is in press and published online: Odinot, G., Wolters, G., & Lavender, T. (in press). Repeated partial eyewitness questioning causes confidence inflation but not retrieval-induced forgetting. *Applied Cognitive Psychology*. http://dx.doi.org/10.1002/acp.1443

Odinot_4.indd 57 10-11-2008 13:19:03

## Summary

*During a crime investigation eyewitnesses are often interviewed more than once. Repeated post-event questioning offers an opportunity for retrieval practice. Practicing retrieval of a subset of memories may suppress access to related memories, a phenomenon known as retrieval-induced forgetting. In this paper we investigated the generalization of retrieval-induced forgetting to episodic eyewitness memory of a complex event. The results indicated that repeated retrieval improves future recall of practiced information, but does not induce forgetting of related information. Retrieval practice resulted in higher confidence ratings, both for correct and incorrect answers. The practical consequence of this latter finding is that repeated questioning should be avoided if possible, because it may lead to artificially high confidence levels.*

## Introduction

Repeated retrieval of memory traces can have consequences for confidence and the amount of information retrieved during later recall attempts. Typically, prior retrieval of information increases the probability of the same information being retrieved at a later recall attempt. Occasionally, also information may be retrieved that was not recalled previously, a phenomenon known as spontaneous recovery or hypermnesia (e.g., Roediger, McDermott, & Goff, 1997; Scrivner & Safer, 1988; Turtle & Yuille, 1994). Repeated retrieval of particular information, however, may also lead to diminished accessibility of other related information that was not retrieved in prior recall attempts. This phenomenon is called retrieval-induced forgetting (Anderson, Bjork, & Bjork, 1994; Anderson & McCulloch, 1999; Anderson & Spellman, 1995; Barnier, Hung, & Conway, 2004; MacLeod, 2002; Shaw, Bjork, & Handal, 1995)

The positive and negative effects of repeated retrieval or retrieval practice are highly relevant to the study of eyewitness reports. During the investigation of a crime, eyewitnesses may be asked to provide a description of the event. This initial interview is often followed by additional interviews during later stages of the investigation. 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 (Penrod & Cutler, 1995). However, repeated interviewing may also introduce distortion of memory, and it offers witnesses the opportunity to practice retrieval of their memories. Research suggests that such retrieval practice indeed may affect the amount of information recalled, as well as the level of confidence that is expressed by witnesses about the accuracy of their memory (e.g., Granhag, 1997; Roediger et al., 1997; Shaw, 1996; Shaw et al., 1995; Shaw & McClure, 1996).

Retrieval-induced forgetting was reported first in studies showing that recall of unpracticed exemplars of categories was impaired by practicing recall of other exemplars. This impair-



Odinot_4.indd 58 10-11-2008 13:19:03

ment occurs not only relative to recall of practiced exemplars, but also relative to a baseline of recall of exemplars from other unpracticed categories. These memory failures are attributed to an inhibition mechanism that might arise from the executive control processes that resolve interference between competing memory traces (see for more detailed discussions, e.g., Anderson et al., 1994; Anderson & Spellman, 1995; Levy & Anderson, 2002; MacLeod, 2002; MacLeod & Macrae, 2001).

Originally it was argued that retrieval-induced forgetting occurs because the strong semantic links in our memory required an active suppression of the non-practiced items, which resulted in robust effects of retrieval-induced forgetting in laboratory studies. However, retrieval-induced forgetting may also occur in the absence of pre-existing category-exemplar relationships. Such findings were reported for instance by Ciranni & Shimamura (1999) and Gomez-Ariza, Lechuga, Pelegrina, & Bajo (2005). Both studies investigated retrieval-induced forgetting for newly learned (episodic) associations. They showed that also newly formed associations between perceptual features, such as shape and color, are vulnerable to the inhibitory effects of retrieval practice, thus extending the domain of retrieval-induced forgetting from semantic memory to episodic memory.

Additional studies have shown that retrieval-induced forgetting also generalizes to conditions that more resemble real-life situations. For instance, Macrae & MacLeod, (1999) showed that retrieval-induced forgetting even occurs when explicit instructions to remember the stimulus material are absent. This is exactly what happens to real-world witnesses, as they do not know in advance the importance of the events that they perceive. And Barnier et al., (2004) reported retrieval-induced forgetting effects for autobiographical memories.

Shaw & McClure, (1996) tested the effect of retrieval-induced forgetting in eyewitness memory using more complex stimulus material than the simple word lists mostly used in research in this paradigm. In an initial phase, participants were shown color slides of the interior of a student's apartment and they were asked to memorize the objects shown. The slides contained two categories of objects (college textbooks and college sweatshirts). After viewing the slides the participants were questioned three times over a 20 minutes interval about half of the objects from one of the two categories. Following this repeated retrieval phase, participants were asked to recall as many of the target items as they could. Recall of practiced items was higher than that of unpracticed items from both the practiced and unpracticed categories, demonstrating a retrieval practice effect. More importantly, recall of unpracticed items of the practiced category was lower than that of unpracticed items of the unpracticed category, indicating that retrieval-induced forgetting indeed occurred. An interesting feature of this study was that also a control group was added that was not questioned at all prior to the final test. Recall in this control group appeared to be lower than recall of practiced items, but higher than recall of unpracticed items. In a similar vein, MacLeod (2002) reported that retrieval-induced forgetting can occur for the recall of details of the description of a suspect following repeated questioning on a subset of these details.

These findings have important practical implications because they suggest that repeatedly questioning witnesses can actually lead to poorer recall of previously neglected details. According to MacLeod (2002) the likelihood that retrieval-induced forgetting occurs in real-life settings, like interviewing eyewitnesses, may be quite high. Interviews of eyewitnesses often constitute incomplete retrieval tasks as police officers and other investigators tend to limit their questions to specific aspects of the incident (e.g., the burglar's weapon, see Shaw et al., 1995). Thereby, retrieval practice on a subset of the total memory might have an inhibiting effect on later recall of related information that was not part of the initial questioning. Details that became important during the course of an investigation may be difficult to remember because they are inhibited by information retrieved in earlier interviews that focused on other details.

Retrieval practice has been shown to prompt forgetting of related information in an increasing number of studies, and several authors have warned already for the potential problems for eyewitness reliability. However, the problem may not be as large as suggested, because also boundary conditions for the occurrence of retrieval-induced forgetting have been reported (see for an overview Levy & Anderson, 2002). Such boundary conditions may render the occurrence of retrieval-induced forgetting in real life eyewitness settings less likely. One of the boundary conditions is that it probably is a relatively short-lived phenomenon. MacLeod & Macrae (2001; see also Macrae & MacLeod, 1999) argued that retrieval-induced forgetting can be seen as an adaptive mechanism. By actively suppressing competing memories that share the same retrieval cue as the target memory, people can ensure that they only become aware of the recollection that is relevant for their current cognitive goals. This inhibition, however, should be relatively short-lived in order for it to be socially adaptive. Once people have achieved their current cognitive goals, the inhibitory effect should cease to operate.



The level of integration between competing memories has been suggested as another boundary condition. Anderson & McCulloch (1999) reported that instructions for participants to form interconnections between category exemplars reduced retrieval-induced forgetting. Highly integrated episodic memories therefore may be less susceptible to the disruptive effects of retrieval practice than poorly integrated memories.

In sum, the likelihood that retrieval-induced forgetting occurs in real life seems to be to quite high. Several studies indeed have shown robust retrieval-induced forgetting effects in eyewitness related settings. On the other hand, however, the effect of retrieval-induced forgetting is restricted by boundary conditions that make the occurrence of retrieval-induced forgetting in real life eyewitness situations less obvious. So far, no study has reported retrieval-induced forgetting in a complex dynamic event. It remains to be seen therefore whether retrieval-induced forgetting will occur in eyewitness memory of a complex episodic event and with relatively long retention intervals. This is the main purpose of this study.

Odinot_4.indd 60 10-11-2008 13:19:03

In addition, we will look at evidence for the opposite phenomenon of hypermnesia; i.e., remembering information on a later occasion that could not be remembered in previous recall attempts. The possibility that later retrieval attempts may produce new additional information is an important argument for repeated questioning of eyewitnesses. Therefore, we will also investigate if this presumed positive result of repeated questioning compensates for possible negative effects of retrieval-induced forgetting.

### Effects of retrieval practice on confidence

Experimental evidence suggests that retrieval practice does not only influence later recall in terms of the amount of information retrieved, but also in terms of the confidence people have in their memory reports. Shaw (1996) for example, conducted a series of experiments to investigate the effect of repeated post-event questioning on eyewitness confidence. He reported that repeated post-event questioning did not increase accuracy, but participants expressed greater confidence in their repeated responses, irrespective of whether responses were accurate or inaccurate. According to Shaw, Bjork and Handal (1995), such findings can be explained by assuming that repeated retrieval of an episodic memory leads to an increase in 'retrieval fluency' for that specific memory item and greater ease of future retrieval. Ease of retrieval has been shown to serve as a basis for confidence judgments (Kelley & Lindsay, 1993). The retrieval-fluency hypothesis can explain why both correct and incorrect memories suffer from confidence inflation due to repeated post-event questioning (Shaw & McClure, 1996).

Confidence inflation without a corresponding increase in accuracy is problematic in a legal context. There is a widely held intuitive belief that confidence expressed about a memory can be used to infer its 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). The relationship between confidence and accuracy depends on numerous factors, like kind of tests used (Robinson & Johnson, 1996), kind of material presented (Perfect, Watson, & Wagstaff, 1993; Roediger & McDermott, 1995), distribution of item difficulty (Kebbell, Wagstaff, & Covey, 1996), and personality characteristics (Bothwell, Brigham, & Pigott, 1987; Nolan & Markham, 1998). However, the general finding is that the relationship between accuracy and confidence is far from perfect. Therefore, it is obvious that any spurious inflation of confidence without a corresponding increase in accuracy is potentially harmful in a judicial context.

The goal of the present study is to investigate whether retrieval-induced forgetting, hypermnesia and confidence inflation generalize to highly complex stimuli and a time span that is more realistic for real-life eyewitness situations. Most studies addressing these questions used simple word lists or series of static slides as stimuli instead of complex dynamic events unfolding in a meaningful context like in a real life situation. Furthermore, the forced-choice recognition tasks that are often used in eyewitness research bear little resemblance



Odinot_4.indd 61 10-11-2008 13:19:03

to the free and cued recall questions asked in actual police interviews. Also time intervals between retrieval practice and final testing, especially in experiments studying retrieval-induced forgetting, are relatively short compared to what witnesses experience in real life. Therefore, in this experiment a 20 minute video is used as stimulus material during the study phase. This is followed by two retrieval practice sessions and a final test that are separated by intervals of several weeks. Participants are given cued recall tasks in which one half of the questions is repeatedly presented, and the other half is presented during final testing only. In these respects the design in this study differs substantially from previous research. Our aim is to see whether retrieval-induced forgetting occurs under these conditions. This issue is important because retrieval-induced forgetting could hamper criminal investigations. Confidence inflation for incorrect responses to practiced questions also poses a potential problem, because legal professionals often rely intuitively on confidence as an indicator of witness accuracy.

## Method

### Participants

Sixty-three students, 50 female and 10 male, participated in the experiment, either for credits or a financial reward. Their age ranged from 18 to 39 years, with a mean of 22 years.

### Materials and procedure

*Videotape*. Participants were individually shown a television programme previously broadcasted on Dutch television. The video was converted into a MPEG-file which was shown individually to the participants on a 17 inch monitor. The video depicts two different storylines that leads to an accident between two cars. The duration of the video was 20 min. Participants were not explicitly instructed to memorize the contents of the video but were aware participating an eyewitness experiment. None of the participants indicated that they had seen the video before.

*Questionnaires*. In the final test session, five weeks after viewing the video, all participants filled out a 30-item open-ended questionnaire about details in the video. For half of the participants, this final test session was preceded by two retrieval practice sessions (after one and three weeks) in which they had to answer half the questions of the complete list. Each questionnaire started with a general question wherein the participants were asked to describe the two story lines in global and general terms. As is common in police interviews, this general question is asked in order to reinstate and refresh the memory before proceeding with more specific questions. Next, participants answered in writing a series of questions that varied in specificity from relatively global (e.g., "Can you describe how the two cars ran into each other?") to highly detailed (e.g., "What was the colour of the car that



Odinot_4.indd 62 10-11-2008 13:19:03

the two brothers were planning to buy?"). It was stressed that answers should be as complete and detailed as possible, corresponding with their memory for details or scenes from the video. However, it was also stressed that if participants could not remember the answer they should refrain from answering by indicating "do not know".

To allow a fine-grained analysis of the responses, participants were instructed to give their answers in the form of small units of information. A unit was described as a single element or aspect of information. To explain this to participants, the following example was given. *Question:* 'What did the boy do after he was refused to enter a nightclub?'; *answer:* 'he went home'; 'on his green bicycle'; 'he took a silver coloured pistol'; 'out of the top drawer'; 'of his bedside table'; 'he went back to the nightclub'; 'were he shot the doorkeeper'. To encourage the subjects to give single units of information to the more global questions, the lines on the answering sheet were restricted in length. Participants could answer with as many units of information as they needed.

The complete 30-question list was composed of two 15 question blocks, labelled A and B. In each block, every question corresponded to a question in the other block by asking about a different aspect of the same scene from the video. In this way we created potential recall competition to enhance the possible occurrence of retrieval-induced forgetting. For example, the question "What was the colour of the car that the two brothers were planning to buy?" (block A) corresponded with "What was the colour of the car that the two brothers took for a test ride?" (block B). These details were shown briefly after one another in the same video-scene. The two blocks of the 15 questions were matched in terms of difficulty on the basis of the results from a pilot study.

During the final test session, participants were instructed to make a confidence judgment for each unit of information they provided. They were asked to indicate their confidence regarding the accuracy of each unit of information given on a 7-point Likert scale ranging from 1 (very uncertain) to 7 (absolutely certain). This was asked in the final session only, in order to avoid possible effects of repeated confidence judgments, such as remembering earlier ratings and being motivated to appear consistent.



## Design

Retrieval practice was manipulated between participants. Half of the participants received retrieval practice on a subset of their memories (either questions of block A or B), on two occasions; 1 and 3 weeks after seeing the video. So for each question that was practiced, the corresponding question in the other block was not practiced. The other half of the participants did not receive any retrieval practice. For all participants the final test session took place 5 weeks after the video. During the final test session, all participants filled out the complete 30-item questionnaire; each practiced question was directly followed by the corresponding but non-practiced question. This design allows to compare recall of practiced questions (Rp+) with recall of related but unpractised questions (Rp-) within subjects, and

Odinot_4.indd 63 10-11-2008 13:19:03

to compare between subjects recall of practised/unpractised questions with recall of the same questions without any retrieval practice (Nrp).

*Scoring*. All units of information recalled by the participants were scored as correct or incorrect. Information was scored *correct* when it matched the information in the video. *Incorrect* information consists of units of information not presented in the video, information that was either incorrectly remembered or fantasized by the participants. Two experimenters did the scoring, and in case of a disagreement, a third experimenter settled the dispute. Of all units generated less than 0.5% could not be classified as correct of incorrect; these were discarded from further analysis.

## Results

After removal of one outliers with very poor accuracy (11.4% correct), the data from the remaining sixty-two participants were analyzed. Per subject, the number correct units of information, the number incorrect units of information and the number of "do not know" answers were determined. Overall the average number of "do not know" answers per block of 15 questions in the final test was 2.1 (14 %). Although the average number of "do not know" answers was somewhat lower in the repeated retrieval condition (1.5) than in the unpractised (2.1) and in the no practice at all conditions (2.5), these differences were not significant. Because the number of correctly and incorrectly recalled units of information differs over questions and participants, we will present and analyse these data as the actual (average) numbers per block of 15 questions.

*Retrieval practice – correct recall*. The mean numbers of correctly and incorrectly recalled units of information per condition (i.e., averaged over blocks of 15 questions) are shown in Table 1. First, we assessed the effect of retrieval practice on correctly recalled units of information by comparing recall of the practiced (Rp+) and the unpractised but related question (Rp-) lists. A paired samples t-test showed a significant difference in recall between practiced (M = 22.5) and unpractised (M = 15.7) questions, $t$ (28) = 5.19, $p$ < 0.001. The effect of retrieval practice was confirmed by comparing recall of practiced questions (Rp+) with recall of the control group that did not receive practice at all (Nrp). An independent samples t-test showed a significant difference ($t$ (60) = 4.36, $p$ < .001). These results clearly show a positive effect of retrieval practice; it reduces forgetting with longer retention intervals.

To determine if retrieval-induced forgetting occurred, we compared recall of unpractised but related (Rp-) questions with recall in the control group (Nrp). This difference was not significant ($t$ (60) = 0.33, $p$ = n.s.). Actually recall of the Rp- lists was slightly better (M = 15.7) than recall in the control group (M = 15.2). This finding indicates that in the conditions studied here, there is no evidence for retrieval-induced forgetting.



Odinot_4.indd 64 10-11-2008 13:19:03

We also looked at the possible occurrence of hypermnesia in correctly recalled information units, i.e., correct recall of elements in later retrieval attempts that were not recalled in previous attempts. To determine the occurrence of hypermnesia, we counted the number of correctly recalled units of information that were absent in a previous recall attempt. Overall, we found only 49 instances of hypermnesia (an average of 1.6 per subject), 30 in the 1-3 weeks comparison, 19 in the 1-5 weeks comparison and no new correct items in the 3-5 weeks comparison.

**Table 1** Mean number of correctly recalled units of information per condition (i.e., blocks of 15 questions, the mean of the no retrieval control group is averaged over 2 blocks). Standard deviations in parentheses.

| | **Mean number of correct units of information** | | | |
|---|---|---|---|---|
| | **1 week** | **3 weeks** | **5 weeks** | |
| | **Rp+** | **Rp+** | **Rp+** | **Rp-** |
| **Retrieval practice** N=29 | 26.1 (10.0) | 24.2 (7.6) | 22.5 (7.5) | 15.7 (4.5) |
| | | | **Nrp** | |
| **No retrieval practice** N=33 | - | - | 15.2 (5.4) | |

*Retrieval practice – incorrect recall*. To test the possibility that retrieval practice may lead to an increase in errors, we analyzed recall errors in the same way as correct recall. The average numbers of incorrect units of information in each condition are presented in Table 2. Although there seem to be slightly more errors in the repeated retrieval conditions, none of these analyses showed any significant difference related to the retrieval practice manipulations (all *t* values < 1.0).



**Table 2** Mean number of incorrectly recalled units of information per condition (i.e., blocks of 15 questions). Standard deviations in parentheses.

| | **Mean number of incorrect units of information** | | | |
|---|---|---|---|---|
| | **1 week** | **3 weeks** | **5 weeks** | |
| | **Rp+** | **Rp+** | **Rp+** | **Rp-** |
| **Retrieval practice** N= 29 | 7.5 (4.2) | 7.4 (3.7) | 7.8 (4.7) | 6.8 (2.6) |
| | | | **Nrp** | |
| **No retrieval practice** N=33 | - | - | 6.9 (3.2) | |

*Confidence inflation*. Confidence ratings on a 7-point scale (from 1 = very uncertain to 7 = absolutely certain) were requested during the final test only. Mean confidence ratings were determined for correctly and incorrectly recalled information units in all conditions, and

Odinot_4.indd 65 10-11-2008 13:19:04

these are shown in Table 3. Overall, confidence was significantly higher for correct units of information (M = 5.4) than for incorrect units (M = 4.5), $t$ (61) = 11.2, $p$ < .001.

After retrieval practice, participants showed more confidence in their correct answers to practiced questions (M = 5.9) than in their correct answers to unpractised questions (M = 5.4), $t$ (28) = 4.84, $p$ < 0.001. Comparisons with confidence in accurate recall of the control group that did not receive retrieval practice (M = 5.2), showed a significant difference with practiced questions ($t$ (60) = 3.88, $p$ < .001), but no difference with unpractised questions ($t$ (60) = 1.01, $p$ = n.s.).

Analysis of confidence in incorrect answers yielded the same pattern of results as for correct answers. Participants who had received retrieval practice showed more confidence in their incorrect answers to practiced questions (M = 4.9) than in their incorrect answers to unpractised questions (M = 4.4), $t$ (28) = 2.1, $p$ < 0.05. Comparisons with the control group again showed a significant difference with practiced questions ($t$ (60) = 2.83, $p$ < .01), but no difference with unpractised questions ($t$ (60) = .92, $p$ = n.s.). The analyses of the confidence ratings provide converging evidence that indeed retrieval practice does result in confidence inflation, both for correct and incorrect answers.

**Table 3** Mean confidence ratings for correct and incorrect units of information in the retrieval practice conditions (scale values: 1=very uncertain to 7=absolutely certain). Standard deviations between parentheses.

| | **Mean confidence in final recall session** | | | |
|---|---|---|---|---|
| | **Correct** | | **Incorrect** | |
| | **Rp+** | **Rp+** | **Rp+** | **Rp-** |
| **Retrieval practice** N= 29 | 5.9 (.62) | 5.4 (.83) | 4.9 (1.02) | 4.4 (.99) |
| | **Nrp** | | **Nrp** | |
| **No retrieval practice** N= 33 | 5.2 (.82) | | 4.2 (1.09) | |

*Confidence – accuracy relationship*. To analyze the accuracy-confidence relation, we determined the number of correct and incorrect units of information for each confidence level over all participants, and we calculated Goodman-Kruskal gamma correlation coefficients. The correlation between accuracy and confidence in the group with no retrieval practice was G = 0.41. In the retrieval practice group, the Gamma correlation for the Rp+ items was G = 0.45, and for the Rp- items G = 0.29. All these correlations are highly significant (all $p$ < 0.001). To gain more insight in the confidence-accuracy relationship, we calculated the total number and the proportion of correct units of information as a function of confidence levels. As can be seen in Table 4, five weeks after seeing the video, answers given a confidence judgement at the lower end of the scale were as often correct as incorrect. Of the answers judged with the highest confidence score, 84% was correct. This shows that information recalled with the highest confidence has a high probability of being accurate, although even here chances of erroneous recall are not negligible.



Odinot_4.indd 66 10-11-2008 13:19:04

**Table 4** Distribution of units of information along the confidence scale.

| | Confidence scale | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | **1** | **2** | **3** | **4** | **5** | **6** | **7** | **Total** |
| # Answers | 129 | 214 | 306 | 343 | 478 | 541 | 1056 | 3067 |
| Prop. correct | 0.50 | 0.49 | 0.61 | 0.57 | 0.67 | 0.75 | 0.84 | 0.71 |

## Discussion

The main goal of the present study was to investigate whether retrieval-induced forgetting generalizes to real-life eyewitness situations. In addition, we looked at hypermnesia and confidence inflation. To create a more ecologically valid situation, we used complex stimuli, a cued recall format for repeated questioning, and retention intervals of several weeks. Under these conditions we found no evidence for retrieval-induced forgetting. Consistent with previously reported results, retrieval practice did have a positive effect on recall of practiced information, and it caused confidence inflation; participants became more confident, both in correct and incorrect answers. Hypermnesia did occur, but only to a very limited extent.

There are several explanations for the absence of retrieval-induced forgetting in the conditions as used in this study. First, the structure of the event memory created by our stimulus material obviously is much more complex than the hierarchical structure of word and picture categories that is often used in studies on retrieval-induced forgetting. It is likely that the memories from the video were highly integrated because they formed a meaningful, coherent narrative story. Interconnections between episodic memories ensure multiple retrieval pathways to find and give access to information. This memory integration may serve as a safeguard against retrieval-induced forgetting, as was shown by Anderson & McCulloch (1999). Highly integrated episodic memories, such as a coherent video narrative as used in this experiment, or a real-life event, are probably less vulnerable to the inhibitory effects of retrieval practice than recollections of isolated words and pictures which are organized by their categorical similarity.

Second, the recall procedure with item-specific cues could be a reason why retrieval-induced forgetting did not occur. According to MacLeod and Macrae (2001), retrieval-induced forgetting serves to suppress competing memories that share the same retrieval cue as a target memory. In our study, however, the overlap of retrieval cues for accessing a specific item in a highly integrated episodic memory was only partial. Although each pair of corresponding questions referred to the same scene in the video, different elements had to be recalled and the overlap of the retrieval cues was not complete. Therefore, recall competition may have been relatively small, reducing the need for retrieval inhibition.

A third explanation for the absence of retrieval-induced forgetting might be the retention interval between stimuli and recall sessions. Final testing took place 5 weeks after the



Odinot_4.indd 67 10-11-2008 13:19:04

study phase and two weeks after the second retrieval practice session. This exceeds by far the intervals that are typically used in studying retrieval-induced forgetting. Although MacLeod and Macrae (2001) still found some retrieval-induced forgetting after a 24 hour interval, they concluded that it probably is a relatively short-lived phenomenon. They argued that once the target memory has been retrieved, it is no longer necessary to block access to competing memories. Whichever explanation is correct, our results indicate that retrieval-induced forgetting either does not occur in the conditions used, or does not persist over repeated retrieval intervals of several weeks.

Our findings showed a significant inflation of witness confidence for the retrieval practice items, regardless whether the answers were correct or incorrect. This seems to be in conflict with results showing some, but not significant, confidence inflation with repeated recall (Odinot & Wolters, 2006). In that study, however, confidence levels were very high which probably precluded confidence inflation to show up because of a ceiling effect. The present result does replicate findings of Shaw and McClure (Shaw, 1996; Shaw & McClure, 1996) who also reported that repeated post-event questioning results in higher confidence levels. According to the 'retrieval fluency' hypothesis, post-event questioning about an episodic memory leads to strengthening and consequently an increase in retrieval fluency for the information that is recalled, which in turn may result in elevated levels of confidence. Because repeated post-event questioning affects retrieval fluency independent of the correctness of the response, confidence inflation occurs both for correct and incorrect memories.

The confidence-accuracy relations found in this study appeared to be independent of presence or absence of retrieval practice. The relationship suggests that confidence may be a moderately useful indicator for accuracy, because selecting only answers with the highest confidence rating results in filtering out a large proportion of incorrect information. Unfortunately, however, even with the highest confidence ratings there still remains a substantial proportion of incorrect information. Therefore, no single witness statement can be accepted as certainly correct on the basis of confidence alone.

We believe the result of this study have three important practical implications. First, witnesses should be questioned as soon as possible after an event, because delays reduce the amount of information that can be recalled. This, of course, is not a new finding, but it again emphasizes the importance of questioning witnesses as soon as possible after an event. Second, it seems unlikely that repeated questioning in real life conditions may cause retrieval-induced forgetting. Prior studies already suggested limiting factors that would make retrieval-induced forgetting less likely in everyday conditions, and the results of the present study bear evidence to this conclusion. We acknowledge, however, that such forgetting can not be completely ruled out in other conditions. Third, although repeated questioning probably does not cause retrieval-induced forgetting, it nevertheless should be avoided as much as possible. Our findings do not show strong indications that repeated



Odinot_4.indd 68 10-11-2008 13:19:04

recall attempts enhance memory (i.e., hypermnesia), but they do indicate that it does cause confidence inflation which is a potential problem in a legal context. Since confidence inflation occurs both for correct and incorrect answers, the tendency to rely on confidence as an indicator for accuracy enhances the probability that incorrect information is falsely accepted.

A number of questions remain regarding retrieval-induced forgetting in eyewitness memory. Several authors have suggested that retrieval-induced forgetting is highly likely to occur in some real life eyewitness situations (e.g., McLeod, 2002; Shaw et al. 1995). For instance, intensive repeated partial questioning of suspects is common practice in crime investigations. The Cognitive Interview (Fisher & Geiselman, 1992), a memory enhancing technique to help eyewitnesses to remember as much as possible during a police interview, involves deliberate repeated recall of the memory. How such repeated retrieval practices in a relative short time interval influence memory is still unclear. If retrieval-induced forgetting would occur in these situations, the question arises for how long this inhibition exists, and whether there is even a chance that the unpracticed items become unavailable permanently. Clearly, further research is needed to see what the specific conditions are for retrieval-induced forgetting to occur in real life eyewitness situations.

## References


Anderson, M. C., Bjork, R. A., & Bjork, E. A. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*, 1063-1087.

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.

Anderson, M. C., & Spellman, B. A. (1995). On the status of inhibitory mechanisms in cognition: Memory retrieval as a model case. *Psychological Review, 102*, 68-100.

Barnier, A. J., Hung, L., & Conway, M. A. (2004). Retrieval- induced forgetting of emotional and unemotional autobiographical memories. *Cognition and emotion, 18*, 457-477.

Bothwell, R. K., Brigham, J. C., & Pigott, M. A. (1987). An exploratory study of personality differences in eyewitness memory. *Journal of Social Behavior and Personality, 2*, 335-343.

Ciranni, M. A., & Shimamura, A. P. (1999). Retrieval-induced forgetting in episodic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 25*, 1403-1414.

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

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

Gomez-Ariza, C. J., Lechuga, M. T., Pelegrina, S., & Bajo, M. T. (2005). Retrieval-induced forgetting in recall and recognition of thematically related and unrelated sentences. *Memory & Cognition, 33*, 1431-1441.

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.




Odinot_4.indd 69 10-11-2008 13:19:04

Kebbell, M. R., Wagstaff, G. F., & Covey, J. A. (1996). The influence of item difficulty on the relationship between eyewitness confidence and accuracy. *British Journal of Psychology, 87*, 653-662.
Kelley, C. M., & Lindsay, D. S. (1993). Remembering mistaken for knowing: Ease of retrieval as a basic for confidence in answers to general knowledge questions. *Journal of Memory and Language, 32*, 1-24.
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, 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.
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. (2002). Retrieval-induced forgetting in eyewitness memory: Forgetting as a consequence of remembering. *Applied Cognitive Psychology, 16*, 135-149.
MacLeod, M. D., & Macrae, C. N. (2001). Gone but not forgotten: The transient nature of retrieval-induced forgetting. *Psychological Science, 12*, 148-152.
Macrae, C. N., & MacLeod, M. D. (1999). On recollections lost: When practice makes imperfect. *Journal of Personality and Social Psychology, 77*, 463-473.
Nolan, J., & Markham, R. (1998). The accuracy-confidence relationship in an eyewitness task: Anxiety as a modifier. *Applied Cognitive Psychology, 12*, 43-54.
Odinot, G., & Wolters, G. (2006). Repeated recall, retention interval and the accuracy-confidence relation in eyewitness memory. *Applied Cognitive Psychology, 20*, 973-985.
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.
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., & McDermott, K. B. (1995). Creating false memories: remembering words not presented in lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 803-814.
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.
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., Bjork, R. A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review, 2*, 249-253.
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.
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.
Wise, R. A., & Safer, M. A. (2004). What US judges know and believe about eyewitness testimony. *Applied Cognitive Psychology, 18*, 427-443.



Odinot_4.indd 70 10-11-2008 13:19:04

# Appendix

## Questionnaire

### Instructions for the refreshing phase;

*It has been a while since you have seen the video with the accident and the story of the persons involved. Before we start asking more specific questions we like you to go back to the video. Please give a short and global description of the two story lines you have seen in the video. Details will be discussed later.*

### Cued recall questions;

*One of the storylines in the video is about a male photo-model and his brother. The first scene of this storyline is about a photo shoot.*

1A At the scene of the photo shoot the photo-model had a telephone conversation with his brother. Do you remember what they talked about?

1B Do you remember which persons, besides the photo-model himself, were present at the photo shoot?

*The next scene shows the photo model and his brother at a car dealer.*

2A Can you give a description of the clothes the brother of the photo-model was wearing in the scene at the auto dealer?

2B Can you give a description of the clothes the photo-model was wearing in the scene at the auto dealer?

*The brother of the photo model invited his girlfriend for a diner at his house. The photo model arrives later on that evening.*

3A Do you remember what the brother of the photo-model gets from his girlfriend when she arrived for diner?

3B Do you remember what the brother of the photo-model was wearing when his girlfriend arrived?

4A Do you remember what the conversation was about during dinner?

4B Do you remember who after dinner left the house first and why?

*The photo model and his brother went back to the auto dealer for a test ride*.

5A Do you remember the color of the car that the brother of the photo-model wanted to buy?

5B Do you remember the color of the car the brother and the photo-model took for a test ride?

*The other storyline in the video shows an older couple. The first scene is at their house.*

6A Do you remember what the older woman was doing while her husband was listening to music in the living room?

6B Do you remember what kind of music the older man was listening to?

*The older man made reservations for a trip. His wife did not know about this.*

7A Do you remember why the older woman began to suspect her husband of doing something behind her back?

7B Do you remember the destination of the trip the older man and woman planned to make?

8A Can you describe the first reaction of the older woman when she saw the suitcases here husband had packed?

8B Can you describe how and where the older man told his wife about the trip he booked?

*The last part of the video shows the two brothers making a test ride en the older couple on their way to the airport by car.*

9A Do you remember what problems the car had that the brother and the photo-model were driving for a test ride?

9B Do you remember the topic of conversation of the older couple during their ride to the airport?

*During the test ride the car of the brothers passes two pedestrians narrowly.*

10A Do you remember what the photo-model did immediately after passing the pedestrians?

10B Can you give a description of the pedestrians?

11A Was the older woman wearing a seatbelt during the ride to the airport?

11B Was the husband of the older female wearing a seatbelt during the ride to the airport?

12A Who was driving the car; the older woman or her husband?

12B Who was driving the car; the photo-model or his brother?

13A Which car drove through the red light?

13B Who is badly injured and carried away on a stretcher?

14A Can you describe how the two cars ran into each other?

14B Can you describe where and how the car with the photo-model came to a halt after the accident?

15A Can you describe the damage of the car of the older couple?

15B Can you describe the damage of the car of the photo-model?



Odinot_4.indd 72 10-11-2008 13:19:04

# 5

# Eyewitness memory of a supermarket robbery: A case study of accuracy and confidence after 3 months*

* This chapter is in press and published online as: Odinot, G., Wolters, G., & Koppen, P. J. van (in press). Eyewitness memory of a supermarket robbery: A case study on accuracy and confidence after 3 months. *Law and Human Behavior.* http://dx.doi.org/10.1007/s10979-008-9152-x

Odinot_4.indd 73 10-11-2008 13:19:04

## Summary

*In this case study, 14 witnesses of an armed robbery were interviewed after three 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.*

## Introduction

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 as well as 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; Odinot & Wolters, 2006). Although there are some exceptions (for example, Christianson & Hubinette, 1993; Fisher, Geiselman, & Amador, 1989; Read, Tollesstrup, Hammersley, McFadzen, & 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 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.



Odinot_4.indd 74 10-11-2008 13:19:04

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 witnesses 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 & 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 internal. Woolnough & 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 (Koppen van & 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 for 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



Odinot_4.indd 75 10-11-2008 13:19:04

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 three 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 three months after the event took place. Accuracy-confidence correlations were calculated to see whether confidence is an indictor for memory accuracy in a real-life situation three 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 three months after witnessing a crime.

## The witnessed event

On a Friday night, 9 February 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 smallest 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 minutes to take what they wanted and leave the employees behind in shock and confusion. The police, called by the employees who had



Odinot_4.indd 76 10-11-2008 13:19:04

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, *Opsporing 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 were all inside the supermarket during the robbery, including the main victim 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.

Odinot_4.indd 77 10-11-2008 13:19:04

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 was 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 they had 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 minutes, with a mean of 28 minutes. 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 & 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, there-



Odinot_4.indd 78 10-11-2008 13:19:04

fore, 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, three 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.



### Number of Details

The witnesses reported a total of 1485 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%), ($x(1) = 39.1$, $p < 0.01$).

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,

Odinot_4.indd 79 10-11-2008 13:19:05

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).

An ANOVA on the number of recalled details over the categories showed a significant effect ($F$ (2, 39) = 3.5, p < .05). Posthoc 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*).

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

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

## 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 to 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.



Odinot_4.indd 80 10-11-2008 13:19:05

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 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 under-confidence 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 and Wells, 2006; Juslin, Olsson & Winman, 1996).

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

| | | Confidence scale | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | low | | | | | | high | Total units of information |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| **Central witnesses** | Person descriptions | 2 | 2 | 8 | 9 | 15 | 19 | 81 | 136 |
| *N = 9* | Prop. correct | .50 | .50 | .37 | .78 | .47 | .74 | .79 | .71 |
| | Object descriptions | 3 | 7 | 13 | 21 | 35 | 38 | 158 | 275 |
| | Prop. correct | .66 | .71 | .77 | .76 | .69 | .79 | .89 | .83 |
| | Action details | 0 | 3 | 6 | 9 | 22 | 35 | 87 | 162 |
| | Prop. correct | 0 | .67 | .50 | .67 | .82 | .63 | .82 | .76 |
| **Peripheral** | All details | 0 | 3 | 3 | 17 | 24 | 9 | 129 | 185 |
| *N = 5* | Prop. correct | 0 | 0 | .34 | .59 | .86 | 0 | .81 | .81 |
| **All witnesses** | All details | 5 | 14 | 32 | 56 | 96 | 101 | 455 | 759 |
| *N = 14* | Prop. correct | .67 | .72 | .63 | .70 | .83 | .74 | .84 | .78 |

## 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 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



Odinot_4.indd 81 10-11-2008 13:19:05

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



Odinot_4.indd 82 10-11-2008 13:19:05

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 interviewed 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 witness 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 question, 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 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 three 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



Odinot_4.indd 84 10-11-2008 13:19:05

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 to 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; Odinot & 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



Odinot_4.indd 85 10-11-2008 13:19:05

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 indicated 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., Odinot & Wolters, 2006), but it should never be allowed as evidence for memory accuracy in the courtroom.



## 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.

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.

Christianson, S.-Ä., & Hubinette, B. (1993). Hands up! A study of witnesses' emotional reactions and memories associated with bank robberies. *Applied Cognitive Psychology, 7*, 365-379.

Crombag, H. F. M. (1999). Collaborative story-telling: A hypothesis in need of experimental testing. *Psychology, Crime & Law, 5*, 279-289.

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.

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.

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.

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., & Amador, M. (1989). Field test of the cognitive interview: Enhancing the recollection of actual victims and witnesses of crime. *Journal of Applied Psychology, 74*, 722-727.

Friedman, W. J. (2004). Time in autobiographical memory. *Social Cognition, 22*, 591-605.

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.

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.

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.

Koppen, P. J. van & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior, 21*, 661-685.

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.

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.

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.

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.

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

Read, J. D., Tollesstrup, P. A., Hammersley, R., McFadzen, E., & Christensen, A. (1990). The unconscious transference effect: Are innocent bystanders ever misidentified? *Applied Cognitive Psychology, 4*, 3-31.

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, New Jersey: 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.

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.



Odinot_4.indd 87 10-11-2008 13:19:05

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.
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.
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.
Wagenaar, W. A. (1986). My memory: A study of autobiographical memory over six years. *Cognitive Psychology, 18*, 225-252.
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.
Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological science in the public interest, 7*, 45-75.
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.
Yuille, J. C., & Cutshall, J. L. (1986). A case study of eyewitness memory of a crime. *Journal of Applied Psychology, 71*, 291-301.



Odinot_4.indd 88 10-11-2008 13:19:05

# 6

# General discussion

Odinot_4.indd 89 10-11-2008 13:19:05

## Introduction

Eyewitnesses are the most important source of information in reporting criminal events. They are also an important source of evidence in legal cases, either by identifying suspects or by reporting details about criminal events, thereby often determining the outcome of trials (van Koppen & Malsch, 2008). As noted by Loftus and Ketcham (1991, p.16) "aside from a smoking pistol, nothing carries as much weight with a jury as the testimony of an actual witness". It is therefore crucial to know how accurate eyewitnesses are. Since the early work of Stern (1902) and Münsterberg (1908) it is already known that eyewitness reports are not always reliable and accurate. During the last 40 years psychological research has greatly enhanced our knowledge about the fallibility and the malleability of eyewitness memory (Loftus, 1979; 2005).

The central focus of this dissertation is on the accuracy of eyewitness memory, and especially on the relationship between accuracy and confidence (i.e., the subjective judgment of accuracy). We studied these factors both in laboratory studies (chapters two to four) and in a field study (chapter five). To enhance the ecological validity of the laboratory studies, we used a method that allowed us to determine accuracy and confidence scores for the recall of details of complex naturalistic events. Moreover, we examined the effects of some variables that are characteristic of actual eyewitness interviews, such as long (versus short) retention intervals, and repeated (versus non-repeated) questioning. This final chapter starts with a summary of the major findings followed by methodological considerations as well as recommendations for the legal practice.



## Accuracy

The results of the experiments discussed in this dissertation show that the overall level of memory accuracy for details of complex events is quite high. This is the result of the experimental design which is high in ecological validity by using open-ended questions and the option to respond with "I don't know". Accuracy rates range from 0.94 (after a brief interval for information recalled with high confidence) to 0.84 (after a long delay recalled with high confidence). Only a relative small proportion of all the information retrieved from memory appears to be incorrect, ranging from.12 (after a brief interval) to .29 (after a long interval). The finding that the participants in our studies appear to be quite accurate in their recall stands in contrast with the thought that eyewitness testimony is often unreliable. This idea has been created by the focus of most researchers on errors and false memories, resulting in a somewhat biased picture. In addition, the faith the legal system places in eyewitnesses has been shaken recently by the advent of forensic DNA testing (see, e.g., Innocence project,

2007; Wells, Memon, & Penrod, 2006). All together, it creates an image that suggests that the memory of witnesses constantly fails and produces errors.
The results of the experiments described in this dissertation show that indeed people do make mistakes and sometimes they are inaccurate in what they say they remember. However, the majority of information they recall is correct, especially the information that is provided with the maximum level of confidence.

## Accuracy in suggestive questioning

Things change dramatically though, when the questions are suggestive in nature, as shown in chapter 3. Subjects appear to be very prone to accept suggested incorrect details, even if they are given the opportunity not to answer. Especially after longer retention intervals, chances that incorrect (i.e., fantasized) answers are provided in response to a suggestive question become high, and correct rejections of the suggestion decreases fast. The conclusion that our participants were prone to suggestion obviously is not a new finding, but it emphasizes again the importance of interviewing witnesses in a correct manner, i.e., refraining form suggestive questions at all times.

## Confidence and the accuracy-confidence relationship

Given the importance of accuracy in eyewitness memory it is imperative to look for indicators that can be used to estimate accuracy. One such indicator is the subjective estimation of the accuracy of a memory expressed as a confidence judgment. It has been shown that both lay persons and legal professionals assume that confidence is a good indicator of accuracy (Cutler, Penrod, & Stuve, 1988; Leippe, 1980; Leippe, Manion, & Romanczyk, 1992; Lindsay, Wells, & O'Connor, 1989; Luus & Wells, 1994; Penrod & Cutler, 1995). Overall, the results of the experiments in this dissertation show that this intuition is correct to some extent. There is indeed a relation between confidence and accuracy in episodic memories. In all our studies we found that the distribution of correct and incorrect information as a function of confidence levels shows an increase in accuracy with increasing confidence. This relationship, however, is not perfect with within correlations ranging from 0.63 to 0.38, depending on the test conditions. Although these correlations are clearly higher than accuracy-confidence correlations found in person identification tasks (e.g., Bothwell, Deffenbacher, & Brigham, 1987; Sporer, Penrod, Read, & Cutler, 1995), they still indicate that inaccurate memories are sometimes given with a high level of confidence, and accurate memories with a low level of confidence. In calibration terms, the distributions of accurate and inaccurate recall as a



Odinot_4.indd 91 10-11-2008 13:19:05

function of confidence level invariably indicate overconfidence at the high end of the confidence scale and under-confidence at the low end of the confidence scale.

## Retention interval

As expected, the length of the retention interval (the delay between perceiving an event and first recall) has a large effect on the level of accuracy. Longer intervals not only cause more forgetting (i.e., more 'I do not know answers'), but also lead to more memory mistakes. These results emphasise the importance of questioning witnesses as soon as possible after an event. Any delay reduces the amount of information recalled and increases the chance of memory errors.

Confidence in recall is also negatively affected by longer retention intervals. On average, confidence decreases with longer retention intervals. However, because this decrease in confidence is almost proportional to the decrease in accuracy, the accuracy-confidence relationship remains almost the same.

## Repeated retrieval

A few effects of repeated retrieval are found. Repeated retrieval did not result in memory enhancement (chapters 2-5). Only in the study described in chapter 4 did repeatedly retrieving the same information from memory seem to inflate the confidence expressed in the answers provided. These results are somewhat surprising because other studies have reported that repeated recall attempts may increase recall, a phenomenon known as hypermnesia, (Roediger, McDermott, & Goff, 1997; Scrivner & Safer, 1988; Turtle & Yuille, 1994), and may also cause confidence inflation (Shaw, 1996; Shaw & McClure, 1996; Shaw, McClure, & Dykstra, 2007). It should be noted, however, that in our experiments exactly the same open-ended questions were asked in the subsequent retrieval sessions. This form of repeated questioning seems to do no more than to consolidate the information that was retrieved in previous attempts. This is also suggested by the fact that recall performance remained at about the same level in subsequent recall sessions. It is possible that also previously given confidence judgments were remembered, which would explain why repeated recall did not affect the subjective confidence in the accuracy of what is recalled.

It can be concluded that asking the same questions repeatedly is not effective in remembering additional information. Of course, it cannot be ruled out that changes in subsequent retrieval attempts, for example, by asking different questions or by using a cognitive interview to follow-up on a free recall attempt, would produce additional information.



Odinot_4.indd 92 10-11-2008 13:19:05

On the other hand, it is of importance to note that asking the same questions on subsequent occasions does not harm the witness report either, because no evidence is found of recall that is more incorrect or of inflated confidence. This is a forensically interesting finding, because repeatedly questioning witnesses about the same event is common practice in judicial investigations.

## Consistency

Consistency of recall, i.e., the same information being recalled at two different moments in time, is another indicator of memory accuracy. This indicator is often used in conditions were it is impossible to check the accuracy of recall against a record of the original event. Consistency of recall is used as a proxy for accuracy for instance in studies investigating so called 'flashbulb memories' (Brown & Kulik, 1977; Wolters & Goudsmit, 2005), and memory of traumatic experiences (Giezen van, 2007).

Consistency, however, is a necessary but not a sufficient condition for accuracy. Both accurate and inaccurate information can be recalled consistently. Moreover, inconsistencies in the form of omissions (information recalled at time 1 but not at time 2) or commissions (information not recalled at time 1 but recalled at time 2) may either be correct or incorrect.

The repeated retrieval paradigm allowed us to examine the relationship between consistency and accuracy. As was shown in chapter 3, consistency is not a good indicator of accuracy. Although consistency is related to accuracy, this relationship is even less than the accuracy-confidence relationship. The weak relationship between consistency and accuracy is understandable considering the nature of memory and the process of memory retrieval. Memories are not static entities but they can change over time. Moreover, the success of memory retrieval strongly depends on the quality of available retrieval cues and the retrieval operations performed by the person who tries to remember something. Therefore, it is likely that details that can be remembered at one point in time may not be remembered another time. Consequently, some inconsistencies between testimonies of the same person at different points in time are normal and testimonies of eyewitnesses should not be dismissed just because of these inconsistencies.

## Methodological issues

One of our aims was to study aspects of the accuracy of eyewitness memory in ecologically valid conditions that would allow generalization of the results to the real world. As a consequence we had to balance between experimental control and everyday variability. At one



Odinot_4.indd 93 10-11-2008 13:19:06

end of this balancing act we conducted a field study using as much methodological rigour as possible (thereby losing some of the ecological validity), and at the other end we performed laboratory studies in which we tried to induce conditions resembling real life situations (striving for, but not completely attaining, ecological validity).

There are methodological shortcomings of the laboratory studies that prevent us form generalizing our results to the real world. First, to allow comparisons of accuracy between conditions we had to use the same questions in subsequent retrieval attempts. This is quite unlike more natural situations where different questions will be asked in subsequent interviews. Using exactly the same questionnaire clearly introduces the possibility that subjects remembered the questions and the previously given answers. So the repeated retrieval condition used in our studies does not allow a generalization to the much more variable retrieval conditions that occur in the real life situations.

Second, participants in laboratory studies lack the emotion and stress that is experienced by eyewitnesses of actual criminal events. Also, laboratory experiments typically use students, acting as 'witnesses'. College students appear to be less suggestible and more accurate as eyewitness overall than are either children or the elderly (Cutler & Penrod, 1995). Therefore, we cannot be sure that this selective and homogeneous group of participants generates results that are generalizable to the average population.

It is interesting to note, however, that we found comparable results regarding accuracy and confidence of memory for event details in laboratory studies (with students watching attentively a video of complex event video) and in a field study (with supermarket employees who witnessed an actual robbery). This finding suggests that our laboratory results can reasonably well be generalized to real life situations.



## Is a confident witness a good witness?

Although the findings show a clear and consistent relationship between confidence judgements and the actual accuracy of memories, it has to be noted that the forensic usefulness of this finding is limited. Information remembered with high confidence are more often correct than details remembered with less confidence, but even a maximum level of confidence does not guarantee accuracy. This implies that in deciding about the accuracy of a particular memory on the basis of a confidence judgment there is always a margin of error. The proportion of inaccurate memory with the highest level of confidence varies between 0.06 with short intervals to 0.16 with longer intervals and depending on particular conditions. However, the diagnostic value (i.e., the ratio of the number of correct decisions and the number of false decisions) of the memories with the highest level of confidence after the shortest retention interval tested (1 week) is 16. This is above the criterion of 15 that is minimally required to be acceptable in judicial decisions (e.g., Wagenaar, van Koppen & Crombag,

1993). Therefore, with some caution, confidence may be used as a partial indicator of accuracy, especially during the early stages of an investigation. Unfortunately, however, there always remain incorrect items that are given the maximum confidence score. That is the reason why no single witness statement can be accepted as certainly correct, based on 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


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.

Brown, R., & Kulik, J. (1977). Flashbulb memories. *Cognition, 5*, 73-99.

Cutler, B. L., & Penrod, S. D. (1995). *Mistaken identification: The eyewitness, psychology, and the law*. New York: Cambridge University Press.

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

Giezen van, A. E. (2007). *A memory like an elephant? The consistency of memory for emotional events* (Doctoral dissertation). Leiden: University of Leiden.

Innocence project. (2007). *Facts on DNA exoneration cases*. www.innocenceproject.org.

Koppen, P. J. van & Malsch, M. (2008). *Het praktisch nut van de rechtspsychologie*. Deventer: Wolters Kluwer.

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.

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.

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.

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

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*, 361-366.

Loftus, E. F., & Ketcham, K. (Eds.). (1991). *Witness for the defense: The accused, the eyewitness, and the expert who puts memory on trial*. New York: St. Martin's 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.

Münsterberg, H. (1908). *On the witness stand: Essays on psychology and crime*. New York: Doubleday.

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

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.

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.






Odinot_4.indd 95 10-11-2008 13:19:06

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.

Shaw, J. S., 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. L. Lindsay (Eds.), *Handbook of eyewitness psychology* (Vol. 1, pp. 371-397). Mahwah, New Jersey: Erlbaum.

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.

Stern, W. (1902). Zur Psychologie der Aussage: Experimentelle Untersuchungen zur Erinnerungstreue. *Zeitschrift für die gesamte Rechtswissenschaft, 22*.

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.

Wagenaar, W. A., Koppen, P. J., van, & Crombag, H. F. M. (1993). *Anchored narratives: The psychology of criminal evidence*. Hertfordshire: Harvester Wheatsheaf.

Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological science in the public interest, 7*, 45-75.

Wolters, G., & Goudsmit, J. J. (2005). Flashbulb and event memory of September 11, 2001: Consistency, confidence and age effects. *Psychological Reports, 96*, 605-619.



Odinot_4.indd 96 10-11-2008 13:19:06

# Samenvatting

In veel strafzaken wordt de verdachte uitsluitend of vrijwel uitsluitend veroordeeld op grond van getuigenverklaringen. Objectieve middelen voor het toetsen van de juistheid van de getuigenverklaringen zijn er meestal niet. Daarom wordt gezocht naar indicatoren voor de juistheid van het verhaal van een getuige. Eén van die indicatoren is het veronderstelde verband tussen de zekerheid over een herinnering en de accuraatheid ervan. Uit onderzoek blijkt dat de zekerheid die een getuige aangeeft over een herinnering een sterke overtuigingskracht heeft, zowel op juridisch geschoolde als ongeschoolde toehoorders. Onderzoek naar de relatie tussen zekerheid en accuratesse bij het herkennen van personen laat echter zien dat het veronderstelde verband niet sterk is. Meta-analyses rapporteren lage correlaties in de orde van grootte van 0.25 – 0.30. Dat betekent dat de zekerheid waarmee een getuige een verdachte heeft herkend geen goede voorspeller is voor de accuraatheid. Een beslissing die wordt gebaseerd op de zekerheid van een getuige is dan ook niet gerechtvaardigd.

Relatief weinig onderzoek is gedaan naar de vraag of zekerheid een betrouwbare indicator is voor accuraatheid in episodische herinneringen (ervaringen en gebeurtenissen), terwijl het belang hiervan voor de rechtspraak evident is.

Studies over het verband tussen zekerheid en accuratesse in episodische herinneringen rapporteren vaak hogere correlaties dan bij identificaties van personen. Maar bij dit onderzoek wordt voornamelijk gebruik gemaakt van meerkeuzevragen en deze resultaten zijn daarom niet generaliseerbaar naar de dagelijkse praktijk. In een interviewsituatie vertelt de getuige uit zichzelf wat hij of zij zich kan herinneren en worden geen antwoordmogelijkheden aangereikt. Geheugenprocessen die worden gebruikt bij het beantwoorden van meerkeuzevragen komen dan ook niet overeen met geheugenprocessen die een getuige gebruikt tijdens een interview. Het gebruik van een onderzoeksmethode die een hoge ecologische validiteit heeft, is dan ook van groot belang voor de generaliseerbaarheid van de uitkomsten.

De vraag die centraal staat in dit proefschrift is of de subjectieve zekerheid over een *episodische* herinnering kan worden gebruikt als indicator voor de accuratesse. Om dit te onderzoeken zijn een aantal laboratoriumexperimenten en een veldstudie uitgevoerd. In de laboratoriumexperimenten kregen proefpersonen een videofilm met een gedramatiseerde maar realistische verhaallijn te zien (een aflevering uit de televisieserie '12 steden, 13 ongelukken' die ongeveer 10 jaar geleden werd uitgezonden). Na één of meer weken werd

Odinot_4.indd 97 10-11-2008 13:19:06

de proefpersonen gevraagd zich bepaalde details te herinneren en daarvan werd de accuratesse en de zekerheid gemeten. De herinnering werd getoetst met open vragen waarbij de informatie actief uit het geheugen moet worden opgehaald en waarbij altijd de mogelijkheid werd gegeven om te antwoorden met "ik weet het niet". In de empirische hoofdstukken 2 – 5 wordt van verschillende variabelen, zoals retentie intervallen, herhaald herinneren en suggestie, bekeken wat hun invloed is op de accuratesse en zekerheid van herinneringen en op de relatie tussen daartussen.

In hoofdstuk 2 wordt onderzocht wat het effect is van herhaald herinneren en van verschillende retentie intervallen op de accuraatheid en zekerheid in de herinneringen van proefpersonen. Tijdens het opsporingsproces is het niet ongebruikelijk dat getuigen pas enige tijd na het incident of gebeurtenis worden geïnterviewd. Naast langere retentie tijden komt het ook vaak voor dat belangrijke getuigen meerdere malen worden geïnterviewd. De uitkomsten van dit experiment laten zien dat een langere retentie tijd voor de eerste testsessie resulteerde in lagere accuraatheid en lagere zekerheid scores. Anders dan verwacht, bleek herhaald herinneren geen effect te hebben op accuraatheid en zekerheid. In alle condities werden relatief hoge correlaties gevonden tussen de subjectieve zekerheid en de accuratesse (0.49 – 0.63).

De studie beschreven in hoofdstuk 3 is vergelijkbaar met de vorige studie, maar in dit onderzoek werd ook gekeken naar het effect van suggestie en bovendien werd gekeken naar de consistentie van de antwoorden bij herhaald herinneren. Het herhaald ophalen van een herinnering biedt de gelegenheid voor oefening en ook voor het optreden van tegenstrijdigheden. Over het stellen van suggestieve vragen is bekend dat mensen eenvoudig te verleiden zijn tot het geven van een antwoord dat niet correct is.

De vragenlijst in dit experiment bevatte vijf suggestieve vragen die onopvallend waren verwerkt tussen vragen zonder suggestie. De proefpersonen beantwoordden dezelfde vragen één of meerdere malen op verschillende tijdsintervallen na het zien van een videofilm. De resultaten laten zien dat het bij langere tijdsintervallen steeds moeilijker wordt om de suggestie te negeren. Bij de suggestieve vragen werd in bijna de helft van de gevallen de (onjuiste) suggestie geaccepteerd en de zekerheidsoordelen waren vergelijkbaar met die van accurate herinneringen. Bij de correcte (niet suggestieve) vragen waren de effecten van herhaald herinneren en van de retentieperiode vergelijkbaar met die van de studie beschreven in hoofdstuk 2.

Wanneer in verschillende verklaringen van een getuige inconsistenties optreden gaat men twijfelen aan de betrouwbaarheid van de getuige. Dit is uiteraard terecht als iemand op verschillende momenten andere antwoorden geeft op de dezelfde vraag. Inconsistenties in de vorm van tegenstrijdige antwoorden op dezelfde vragen bleken echter zelden voor te komen. Wel kwamen tamelijk frequent inconsistenties voor in de vorm van omissies (later niet herinneren wat eerder wel werd herinnerd) en commissies (later herinneren wat eerder niet werd herinnerd). Deze vormen van inconsistentie bleken echter niets te zeggen



Odinot_4.indd 98 10-11-2008 13:19:06

over de betrouwbaarheid van de herinneringen. De gemiddelde accuratesse en zekerheid bij omissies en commissies was vergelijkbaar met die van consistente antwoorden. De correlatie van accuratesse met consistentie was lager dan de correlatie met zekerheid. Hieruit is te concluderen dat zekerheid een betere indicator is voor accuratesse dan consistentie.
In de literatuur is gerapporteerd dat het herhaaldelijk ophalen van dezelfde informatie het later ophalen van gerelateerde informatie bemoeilijkt. Dit inhibitie effect, dat bekend staat als 'retrieval induced forgetting', zou ook kunnen optreden wanneer getuigen meerdere malen worden geïnterviewd over slechts een deel van alles wat zij zich potentieel kunnen herinneren. In hoofdstuk 4 wordt getoetst of retrieval induced forgetting optreedt bij het herhaaldelijk beantwoorden van vragen die slechts een deel van een volledige episodische herinnering betreffen.
Voor dit experiment is een vragenlijst samengesteld die bestaat uit paren van vragen. Een vragenpaar was aan elkaar gerelateerd omdat ze betrekking hadden op dezelfde scene in de video. Van elk vragenpaar werd één vraag meerdere malen aangeboden. In de laatste sessie werden beide vragen voorgelegd. Het eerder beantwoorden van een deel van de vragen bleek geen negatief effect te hebben op het herinneren van het andere deel. Er werd dus geen bewijs gevonden voor het optreden van retrieval induced forgetting. Een mogelijke verklaring hiervoor is dat een coherente episodische herinnering op vele manieren toegankelijk is en dat het inhiberende effect van het gebruiken van één van die toegangen geen noemenswaardige invloed heeft op de andere toegangen.
In hoofdstuk 5 wordt een veldstudie beschreven naar de herinnering van medewerkers van een supermarkt die getuige zijn geweest van een overval. De getuigen zijn drie maanden na de overval geïnterviewd en de beelden van de beveiligingscamera's zijn gebruikt voor het vaststellen van de accuraatheid van de herinneringen. De meeste informatie die de getuigen zich herinnerden bleek correct, maar net als in de experimentele studies bleken sommige herinneringen onjuist. De correlatie tussen de zekerheid en de mate van accuratesse was slechts 0.38.
De getuigen werden ook gevraagd hoe vaak ze nog terugdachten aan de overval en hoe groot de emotionele impact van de overval is geweest. Getuigen die een hogere zelfgerapporteerde emotionele impact aangaven, bleken meer accuraat te zijn in hun herinneringen. Het vaker terugdenken aan de overval leidde tot een hogere zekerheidsscore. De betreffende overval is in het televisieprogramma "opsporing verzocht" belicht. De daar getoonde videoreconstructie bevatte informatie die niet correct was. Het zien van deze uitzending had echter geen invloed op de herinneringen van de getuigen.
Een opvallend resultaat in de bevindingen van dit proefschrift is dat de meeste details van complexe gebeurtenissen (een videofilm of een werkelijke overval) zelfs na meerdere weken nog accuraat herinnerd kunnen worden. Dit staat in contrast met de suggestie dat de verklaringen van getuigen over details per definitie onbetrouwbaar zijn. Deze sugges-

tie is ontstaan door psychologisch onderzoek waarbij de nadruk ligt op het falen van het geheugen of het implanteren van valse herinneringen. De condities zoals getoetst in dit proefschrift laten zien dat mensen inderdaad fouten maken, maar ook dat de overgrote meerderheid van de herinnerde informatie correct is.

Zoals verwacht verandert de situatie totaal wanneer de gestelde vragen suggestie bevatten. Vooral na langere tijd blijkt het moeilijk om suggestieve informatie te herkennen en af te wijzen. Deze bevinding is niet nieuw en benadrukt nogmaals het belang van het stellen van vragen die geheel vrij zijn van suggestie.

Het bestaan van een positief verband tussen zekerheid en accuratesse wordt bevestigd door de resultaten in dit proefschrift. De verdeling van correcte en incorrecte antwoorden als functie van zekerheid laat een toename zien van accuratesse bij een toenemende zekerheid. Maar 100% zekerheid is geen 100% accuraatheid en de gemiddelde correlatie tussen accuratesse en zekerheid schommelt rond de 0.50. Zeker bij langere retentie-intervallen neemt de hoeveelheid herinnerde informatie af en de kans op incorrect herinnerde informatie toe. Zekerheid over een herinnering is daarom slechts in beperkte mate bruikbaar als een indicatie voor de juistheid.

Herhaald herinneren bleek geen effect te hebben op de zekerheid. Dit is opmerkelijk omdat in de literatuur regelmatig melding wordt gemaakt van inflatie van de zekerheid die toegeschreven wordt aan het herhaaldelijk herinneren. Dit is mogelijk het gevolg van de in dit proefschrift gebruikte vragenlijsten met open vragen en de mogelijkheid te antwoorden met "weet niet", in tegenstelling tot de vaak gebruikte vragenlijsten met meerkeuzevragen.

Een belangrijk doel van dit proefschrift was het genereren van kennis over de relatie tussen accuratesse en zekerheid met een hoge ecologische validiteit waardoor de resultaten generaliseerbaar zouden zijn naar situaties buiten het laboratorium. Daartoe werd een veldstudie gedaan en werden laboratoriumexperimenten uitgevoerd met realistisch stimulusmateriaal, open vragen en lange retentie intervallen. Toch is bij het generaliseren van de resultaten van de laboratoriumexperimenten enige voorzichtigheid geboden door methodologische beperkingen. Er werden bijvoorbeeld identieke vragenlijsten gebruikt in opeenvolgende sessies, terwijl het bij het interviewen van getuigen meer voor de hand ligt dat bij herhaald interviewen verschillende vragen gesteld worden. Bij het herhaaldelijk antwoorden op dezelfde vragen is niet uitgesloten dat de voorgaande antwoorden worden herinnerd. Bovendien ontbreekt in een laboratoriumexperiment de stress en emotie die een getuige van een misdrijf wel ervaart. Ook zijn de studenten die gebruikt werden als proefpersonen natuurlijk niet representatief voor de gemiddelde populatie.

Ondanks deze beperkingen is het interessant te constateren dat vergelijkbare resultaten zijn gevonden betreffende de accuratesse en zekerheid van herinneringen in de laboratoriumexperimenten en in de veldstudie. Dit suggereert dat de bevindingen in het laboratorium wel degelijk generaliseerbaar zijn naar situaties in het echte leven.



Odinot_4.indd 100 10-11-2008 13:19:06

De resultaten in dit proefschrift laten zien dat er inderdaad een verband bestaat tussen de zekerheid en accuraatheid van herinneringen over een gebeurtenis. De forensische bruikbaarheid van dit verband is echter beperkt. Informatie herinnerd met hoge zekerheid is vaker correct dan informatie herinnerd met lage zekerheid. Maar een maximale zekerheid is geen garantie voor accuraatheid. Dit betekent dat in een beslissing over de accuratesse van een herinnering op basis van de zekerheid altijd een foutenmarge zit. De proportie incorrecte informatie gegeven met maximale zekerheid varieert van 0.06 na korte tijdsintervallen tot 0.16 na langere tijdsintervallen, mede afhankelijk van de geteste condities. De diagnostische waarde (proportie goed/ proportie fout) van de informatie die na korte tijd wordt herinnerd met maximale zekerheid is 16. Dit ligt boven het criterium van 15 dat minimaal gevraagd mag worden om acceptabel te zijn bij het nemen van beslissingen in een juridische context. Daarom is het mogelijk om, met de nodige voorzichtigheid, zekerheid te gebruiken als indicator voor accuraatheid in een episodische herinnering. Dit geldt zeker in het vroege stadium van een opsporingonderzoek. Wel moet worden benadrukt dat zekerheid over een herinnering, zelfs in optimale situaties, nooit een garantie biedt voor accuraatheid.



Odinot_4.indd 101 10-11-2008 13:19:06

Odinot_4.indd 102 10-11-2008 13:19:06

# Dankwoord

Het is alweer jaren geleden dat ik vol goede moed aan mijn promotieonderzoek begon. Vooral de laatste jaren heb ik uitgekeken naar dit moment, het schrijven van een dankwoord als afsluiting van mijn proefschrift maar het voelt nu als een anticlimax. Het betekent namelijk te gelijkertijd dat ik een tijdperk afsluit en afscheid neem van velen.

Het onderzoek vond plaats onder de vleugels van de sectie cognitieve psychologie van de Faculteit Sociale Wetenschappen, Universiteit Leiden. De feitelijke uitvoering van de studie vond plaats binnen het onderzoeksprogramma van het Nederlands Studiecentrum Criminaliteit en Rechtshandhaving (NSCR), waar ik was gedetacheerd. Op het NSCR en bij de sectie cognitieve psychologie heb ik altijd alle ruimte gekregen voor het opzetten en uitvoeren van het onderzoek.

Zonder de geweldige koffie van Lenie was dit boekje waarschijnlijk nooit tot stand gekomen. Maar nu het dan toch zover is wil ik een paar mensen expliciet noemen. Als eerste wil ik Gerben Bruinsma hartelijk danken voor de (inmiddels vele) jaren van warme gastvrijheid op het instituut! Peter van de Voort, bedankt voor al je hulp en de vele kopieën die je altijd weer voor me klaar had liggen. Albertien, dank voor het altijd warme welkom wanneer deze "spook"aio weer eens opdook. Ariena, bedankt voor alle verse Mare's, de gezellige kletspraatjes op de dode momenten en de onmisbare ondersteuning bij het plaatsen van die lastige d's en t's.

Verder heb ik altijd een beetje medelijden gehad met mijn kamergenoten die ik, als onrustig, rommelig en in mezelf pratend persoon, soms tot waanzin gedreven moet hebben. Ik ben jullie erg dankbaar voor de gezelligheid en zonder Albert, Floor, Ilse, Machteld, Kim, Barbra, Maurice en Hilde, was het een stuk minder leuk geweest! Enne, Machteld ik mis onze 15.00-uur-uit-het-raam-hang-pauze momenten ook heel erg....

Het op elk moment van de dag snel en ad rem moeten reageren op mogelijk rondvliegende objecten, vage woordkeuzes c.q. gebaren houdt de mens scherp en alert; Jan bedankt! Al reageerde ik wel eens te laat.... Zeker zo onmisbaar was de bijdrage van al mijn andere (epos) collega's, voor een kletspraatje, een beetje begrip en herkenning, een kopje koffie: jullie waren er.

De stagiaires die ik in de loop der jaren begeleid heb, weten zelf waarschijnlijk niet hoe waardevol ze zijn geweest voor mijn onderzoek. Ik heb het altijd ontzettend leuk en inspirerend gevonden om jullie te begeleiden. Bedankt, Marit, Ieke, Elger, Alexia, Karolijn, Evelijn, Tristan, Marleen, Estella en Lu-Ming!

Odinot_4.indd 103 10-11-2008 13:19:06

Further, I would like to thank Roy Mallpass, Chris Meissner, Laura Zimmerman and all the other (ex)members of the eyewitness identification laboratory at the University of Texas for their hospitality. My visit at the lab was an inspiring experience I really needed to overcome a dead point in my research. Nadja Schreiber, I still owe you a good night of sleep! Thanks!
Het dankwoord dreigt inmiddels een opsomming te worden van vele namen. En ik ben nog niet klaar. Lieve Damesch 8+ '96 dank voor de borrels, de lol, het delen, de vriendschap, kortom, bedankt voor het zijn wie jullie zijn. Alexa, het lege zaterdagmorgen strand met oer-Hollandse luchten en frisse zeewind genomen in volle galop staat in mijn geheugen gegrift. Het maakte de wereld even aangenaam klein en overzichtelijk; ik ben altijd in voor een vervolg! Ponyclub, bedankt voor de broodnodige onthaasting en mijn wekelijkse uitwaaimoment in het bos, in de bak en in de bar. Ik kom terug, zeker voor een snertrit! Marte, wat heerlijk dat ik de vele proefschriftdipjes, schrijfellende, twijfels en de toestand van de wereld altijd online met je kon en kan delen! Afstand doet er niet toe!
Lieve pa en ma, Karen en Ton bedankt voor het zijn wie jullie zijn. Ook de bonte verzameling van de familie Janssen wil ik niet vergeten te noemen.
Lieve Tijn, jouw rotsvaste overtuiging in mijn kunnen was echt onontbeerlijk de afgelopen jaren. Zonder je vele peptalks en de door jou gecreëerde ruimte was het nooit gelukt. En lieve lieve Niels, nu jij er bent is het leven nog leuker! Als ik alleen al aan je denk moet ik glimlachen.
Tot slot, vergeten is immers menselijk, bedank ik iedereen die ik vergeten ben.



Odinot_4.indd 104 10-11-2008 13:19:06

# Curriculum Vitae

Geralda Odinot was born on the 2th of January 1970 in Utrecht, the Netherlands. In 1993 she passed for the Colloquim Doctum of the University Leiden and started studying Psyhology in Leiden. She obtained her masters in cognitive psychology, with a special interest in memory, in 1999. After working as a consultant and as a purchasing agent of merchandise materials she started working on the present dissertation in December 2001. The research for her Ph.D. was conducted at the Netherlands Institue for the study of Crime and Law Enforcement (NSCR) and at the social science faculty of the University Leiden. Since October 2008 she is working as a post-doc research fellow at the Eyewitness Research Group at the University of Aberdeen.

Odinot_4.indd 105 10-11-2008 13:19:06

Odinot_4.indd 106 10-11-2008 13:19:06