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

Odinot, G.

1

Introduction

"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.

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

(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.

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.

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.

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).

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

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.

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.

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.

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.