

Obtaining information from eyewitnesses : effects of retrieval support in eyewitness interviews

Citation for published version (APA):

Krix, A. C. (2015). *Obtaining information from eyewitnesses : effects of retrieval support in eyewitness interviews*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20150226ak>

Document status and date:

Published: 01/01/2015

DOI:

[10.26481/dis.20150226ak](https://doi.org/10.26481/dis.20150226ak)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Obtaining information from eyewitnesses

Effects of retrieval support in eyewitness interviews

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Universiteit Maastricht,
op gezag van de Rector Magnificus, Prof. dr. L.L.G. Soete,
volgens het besluit van het College van Decanen,
in het openbaar te verdedigen op
donderdag 26 februari 2015 om 14.00 uur

door

Alana Christiane Krix



Promotor

Prof. dr. H.L.G.J. Merckelbach

Copromotor

Dr. M. Sauerland

Beoordelingscommissie

Prof. dr. G. Kok (voorzitter)

Dr. H. Otgaar

Prof. S.L. Sporer, PhD (Justus-Liebig-Universit t Gie en, Duitsland)

Dr. M. Vanderhallen

Prof. dr. R. Volbert (Charit -Universit tsmedizin Berlin, Duitsland)

The research presented in this dissertation was supported by a PhD scholarship from Studienstiftung des deutschen Volkes (German National Academic Foundation) awarded to Alana C. Krix.

CONTENTS

CHAPTER 1	General Introduction	7
CHAPTER 2	The Modality Effect in Eyewitness Accounts	59
CHAPTER 3	Retrieval Support with the SAI and Stress during Encoding	87
CHAPTER 4	Retrieval Support with the SAI and Divided Attention during Encoding	103
CHAPTER 5	Retrieval Support through Repeated Interviews	131
CHAPTER 6	Retrieval Support with the SAI and Working and Source Memory	155
CHAPTER 7	General Discussion	173
	References	191
	Summary	203
	Valorization Addendum	211
	Acknowledgments	215
	Curriculum Vitae	219

CHAPTER 1 **General Introduction**

This chapter is a translated and extended synthesis of the following articles:

Krix, A. C., & Sauerland, M. (2011). Im Zweifel für den Angeklagten: Über Fehlerquellen von Zeugenaussagen [Giving the defendant the benefit of the doubt: Sources of errors in eyewitness accounts]. *In-Mind Magazine*, 1/2011.

Retrieved from <http://de.in-mind.org/article/im-zweifel-fuer-den-angeklagten-ueber-fehlerquellen-von-zeugenaussagen>

Krix, A. C., & Sauerland, M. (2013). Wie lassen sich Zeugenaussagen verbessern? – Neue Entwicklungen und Methoden [How can eyewitness accounts be improved? – New developments and methods]. *Praxis der Rechtspsychologie*, 23, 136-150.

Krix, A. C., Sauerland, M., & Schreuder, M. J. (2013). Hoe getuigen zichzelf kunnen verhoren: de Zelfrapportage voor getuigen [How witnesses can interview themselves: The Self-Administered Interview]. *Expertise en Recht*, 5/6/2013, 180-184.

Two Eyewitness Cases

On July 6th, 1993, a taxi driver was found dead in her taxi in a small town in Austria. She had been robbed and shot dead.¹ The most important evidence in this case was the statements made by two witnesses. Specifically, they reported to the police that it was P.H. whom they had met close to the crime scene at the night of the crime and given him a lift from there to the station. They testified that the man they had picked up had shown them his military ID card. Moreover, they reported the demobilization date (September 30th, 1993) and the first name of the man, which they claimed had both been printed on the card. This first name was the same as P.H.'s who was a soldier at the time and who had been absent without leave from the military since July 6th. Also, the crime scene was only a few hundred meters away from the barracks where P.H. had been stationed. Eventually, he was charged with robbery and murder and sentenced to 20 years in prison. What had seemed such a straightforward case, however, soon turned out to be one of Austria's biggest judicial scandals. P.H. was not the man the witnesses had met that night. He was wrongfully convicted and spent eight years in prison for a crime he had not committed. In fact, P.H. had been at home during the time of the crime, which was confirmed by his mother. Furthermore, demobilization dates were never printed on Austrian military ID cards. What is more, the reported date did not even correspond with the date that was printed on P.H.'s military card (April 30th, 1993) and that referred to the date of his basic military training. Several years later the true perpetrator was convicted based on DNA evidence found at the crime scene. The witnesses were later charged with giving false statements, but were acquitted, because they were able to prove that they had been subject to an error and had not deliberately provided wrong accounts. P.H. was ultimately awarded a compensation of €950,000.

In another case, starting at the evening of September 3rd, 2010, in Germany, ten-year-old M. did not return to his parents' house after playing at a nearby skatepark.² He was reported missing and one of Germany's biggest search operations was initiated. Yet, while some of his clothes were found, the boy remained missing. Shortly after M.'s disappearance, an eyewitness came forward and reported to the police that he had observed a dark VW Passat B6 at the time in question

¹ <http://sbgv1.orf.at/stories/365816>; <http://www.nachrichten.at/nachrichten/chronik/Falschaussage-von-Zeugen-im-Taxi-Mord;art58,191272>; http://www.krone.at/Nachrichten/Freispruch_fuer_Kronzeugen_aus_Heidegger-Prozess-Mordfall_Deubler-Story-152955; <http://search.salzburg.com/display/41-107603527052004>. Retrieved on April 15th, 2014.

² http://www.t-online.de/nachrichten/panorama/kriminalitaet/id_43279022/fall-mirco-aus-greifath-polizei-sucht-passat-fahrer.html; <http://www.welt.de/vermishtes/weltgeschehen/article13455704/Anklage-offenbart-grausame-Details-bei-Mord-an-Mirco.html>; <http://www.rp-online.de/nrw/staedte/kempen/prozess-gegen-olaf-h-revisionsbegruendung-liegt-dem-gericht-vor-aid-1.2686043>. Retrieved on April 15th, 2014.

and close to the place where M. had last been seen. As it later turned out, this statement was the critical lead to find the perpetrator according to the police. Based on the statement, the police screened all owners of this particular Passat model in the area of the crime. After the screening, they apprehended O.H. who indeed owned such a Passat at the time. He was also incriminated by other evidence (e.g., a match in a fiber analysis comparing the fibers of the car seat and the fibers found on M.'s clothes) and finally confessed that he had sexually abused and killed M. Subsequently, he led the police to the body. O.H. was charged with sexual abuse and murder and received a life sentence.

Incorrect eyewitness evidence is the main cause for wrongful convictions and the case of P.H. is an example of this. Specifically, an analysis of 40 US cases in which miscarriages of justice had occurred showed that in more than 75% of the cases wrongful eyewitness accounts were the paramount cause (Wells et al., 1998). Different projects around the world seek to uncover miscarriages of justice and contribute to exonerating the wrongfully convicted. For example, in the USA, the "Innocence Project"³ has been involved in the exoneration of more than 300 wrongfully convicted since 1992 using DNA testing (Scheck, Neufeld, & Dwyer, 2003). Some of the convicts had been incarcerated for more than a decade, sometimes on death row or serving life sentences. Eyewitness errors played a decisive role in 71% of these miscarriages of justice (Saks & Koehler, 2005). In the Netherlands, the project "Gerede Twijfel"⁴ ("reasonable doubt") conducts thorough analyses of the court files of cases in which final judgments have been delivered and the convicts assert their innocence. This is done to evaluate whether in the course of the trial errors may have occurred and to examine whether there are new pieces of evidence pointing to the convict's innocence. The project subsequently files a report on every case, which can be used for a retrial. Until now, roughly 15 cases have been considered by the project.

In the last decades, research on eyewitness testimony has extensively demonstrated the error-proneness of eyewitness accounts. It has yielded important insights into what problems can occur during encoding, retention, and retrieval, and led to recommendations of what should be avoided when interviewing eyewitnesses. These factors have been divided into estimator and system variables based on whether they can be influenced by the legal system (Wells, 1978). That is, whereas system variables (e.g., the type of interview used) can be controlled by the legal system, estimator variables (e.g., viewing distance) cannot. During encoding, factors such as a long viewing distance (Busey & Loftus, 2007), poor lighting conditions (Wagenaar & van der Schrier, 1996), or the presence of a weapon (Fawcett, Russell, Peace, & Christie, 2013) have been identified to be responsible for impoverished encoding and subsequent poor memory performance. Regarding the retention interval, much research was conducted on the delay until the first interview

³ <http://www.innocenceproject.org>

⁴ <http://www.projectgeredetwijfel.nl>

and the associated forgetting of information (e.g., Odinet, Wolters, & van Giezen, 2013; Rubin & Wenzel, 1996; Tuckey & Brewer, 2003). Moreover, the misinformation effect, that is, the phenomenon that a witness' memory can be contaminated by incorrect post-event information obtained via media coverage or a discussion with a co-witness has been extensively studied (e.g., Gabbert, Memon, Allan, & Wright, 2004). Looking at the retrieval phase, research on suggestive interview techniques has yielded important findings on how eyewitness statements can be contaminated during the interview and what kinds of questions should be avoided (Loftus, 2005). Suggestive questioning was already the subject of Loftus and Palmer's (1974) seminal car accident study. Here, witnesses watched footage of a car accident and were subsequently asked about the speed of the cars involved in the collision. In doing so, the wording of the question was manipulated. Interestingly, witnesses who were asked how fast the cars were going when they *smashed* into each other estimated a higher speed of the cars than witnesses who were asked how fast the cars were going when they *hit* each other. Moreover, the former group was also more likely to affirm the question whether they had seen broken glass in the film than the latter, although there had been no broken glass at all.

Positive Legal Psychology

In spite of being error-prone and sometimes contributing to the wrongful conviction of innocent suspects, eyewitness statements are an important piece of evidence for police investigations, even in the era of DNA testing. DNA evidence is by no means available at all crime scenes. Moreover, in spite of its acclaimed objectivity, the interpretation of DNA samples, especially mixed samples containing material from several individuals, can be biased by context information (Dror & Hampikian, 2011). In many cases, eyewitness accounts are the only available evidence (Fisher, 2010). For example, evaluating Dutch police cases, de Poot, Bokhorst, van Koppen, and Muller (2004) found that while technical evidence, among which DNA, was present in only 19-41% of the cases, eyewitness evidence was available more often, that is, in 34-63% of the cases. Accordingly, as in the case of O.H. presented above, police officers indicate that eyewitness testimony often provides the critical leads for the investigations (Kebbell & Milne, 1998). Therefore, the quantity and accuracy⁵ of an account may determine, whether a case will be solved or not (Fisher, 2010). Providing both comprehensive and accurate accounts is also important, because witnesses often appear very compelling in court and their accounts receive a lot of weight (Boyce, Lindsay, & Brimacombe, 2008;

⁵ The terminology of this dissertation follows the output-bound correspondence conception of memory (Koriat & Goldsmith, 1994). Throughout, the terms quantity or comprehensiveness are used to designate the total number of details (correct plus incorrect) reported by a witness. The term accuracy is used to designate the number of correct details in relation to all details reported by a witness.

Wells, Memon, & Penrod, 2006). Hence, although the research line investigating the problems and shortcomings of eyewitness testimony has brought about invaluable insights, the above considerations illustrate the importance of a research approach that goes beyond addressing the pitfalls of eyewitness testimony and investigates what measures can be taken and what tools can be employed to improve the reliability of eyewitness accounts. Such a positive psychological approach comprises the investigation of all techniques that support the witness in providing a comprehensive and accurate account (Meissner, Hartwig, & Russano, 2010; Sauerland & Krix, 2011). Hence, research within the framework of this positive legal psychology aims to bring about processes and decisions in the legal system that are as accurate as possible. Thus, it seeks to increase the probative value of all tools and techniques used in the legal system and in doing so to ultimately reduce wrongful convictions. In addition to this, the approach encourages active cooperation with law enforcement practitioners, such as police investigators.

The Cognitive Interview

Indeed, in addition to the approach of uncovering problems of eyewitness evidence, research in the field of interviewing eyewitnesses has also traditionally accommodated a positive psychological approach. The *Cognitive Interview* (Fisher & Geiselman, 1992; Geiselman, Fisher, MacKinnon, & Holland, 1985) is an early and well-known example of this approach. It was developed after there had been requests from the police to improve the existing interviewing practice that often relied on deficient techniques at that time (Fisher, 2010). The original Cognitive Interview was particularly based on basic research findings in the field of memory retrieval and cognition and consisted of four core components. These are mental context reinstatement, the “report-everything” component, recall from different perspectives, and recall in reverse order. Mental context reinstatement relies on the principle of encoding specificity (Tulving & Thomson, 1973) according to which recall is improved if the context during retrieval resembles the context during encoding. Before providing their statements, witnesses are instructed to mentally reinstate the context of encoding by thinking back to what they saw or heard during the incident, but also to their thoughts and emotions they had at the time. The reason why this instruction facilitates retrieval is that memory is organized as a network (J. R. Anderson, 1983). In this network, the context cues (e.g., what the witness thought at the time) are stored parallel to and linked to the memory of the incident. Thinking of the context cues can thus provide additional pathways to get access to the recollections.

After having reinstated the context, witnesses are instructed to provide the most complete and accurate account possible and to also provide information even if it seems unimportant or out of chronological order. This “report-everything”

component fosters a comprehensive recall and serves to minimize witnesses' withholding of information in case they think that the information is not important for the police (Fisher & Schreiber, 2007). While a comprehensive account is encouraged, witnesses are reminded not to guess so as to avoid an overly increase in recall errors.

After providing a first account, witnesses may be asked to recall the incident from different perspectives and in reverse order. These two memory-enhancing components are in line with the principle of multiple and varied retrieval. According to this principle, details that are not accessible with one retrieval technique may well become accessible with another technique (R. C. Anderson & Pichert, 1978; Tulving & Watkins, 1975). This is because different retrieval attempts provide different retrieval cues that give access to different parts of a recollection.

Since its first application in the 1980s, the Cognitive Interview has been revised and is now available as an enhanced version (Fisher & Geiselman, 1992). Compared to the original version, the enhanced Cognitive Interview incorporated findings from social dynamics and communication (Fisher & Schreiber, 2007). Emphasis is put on creating a comfortable situation for the witness. For example, at the onset of the interview, rapport is developed between interviewer and witness. Rapport-building aims to build a relation between interviewer and witness that is characterized by empathy and sympathy (Collins, Lincoln, & Frank, 2002; Fisher, 2010). This serves to decrease nervousness or uneasiness in a situation in which witnesses may have to talk about very intimate or fearful situations with a person whom they hardly know (Wells et al., 2006). Rapport-building can occur both verbally and non-verbally. The non-verbal approach is aimed at the body language and the intonation of the voice (e.g., displaying a friendly and open demeanor and establishing eye contact). Verbal rapport-building can be achieved by addressing the witnesses by their names. It is also important to show a genuine interest in the witnesses, for example, by asking about their families or personal interests. At the same time, interviewers can also reveal personal information about themselves. Exchanging such personal information is known as self-disclosure.

In addition to rapport-building, control is transferred to the witness by making clear that it is only the witness who has valuable first-hand information about the incident. Hence, it is essential to create the expectation that the witness will do most of the talking and the interviewer will mostly listen and not interrupt the witness. The enhanced Cognitive Interview also relies on witness-compatible questioning, that is, the notion that questions should be individually tailored to the witness, instead of using a standardized list of questions, to accommodate for the uniqueness of each witness' mental record of the event (Fisher, 2010).

The Cognitive Interview has been examined in a multitude of studies and been the subject of two meta-analyses (Köhnken, Milne, Memon, & Bull, 1999; Memon, Meissner, & Fraser, 2010). These showed that relative to a control interview, the

Cognitive Interview elicits more correct details, the difference corresponding to a large effect (Köhnken et al.: $d = 0.87$; Memon et al.: $d = 1.20$). This is accompanied by a small, yet significant, increase in incorrect details, with the size of a small effect (Köhnken et al.: $d = 0.28$; Memon et al.: $d = 0.24$). As a result, the Cognitive Interview leads to more comprehensive accounts, without influencing accuracy. It has also been found to protect witnesses from reporting misinformation in a later interview when administered before (yet not after) the exposure to misinformation (Memon, Zaragoza, Clifford, & Kidd, 2010).

Notwithstanding the overall positive results, the two components recall from different perspectives and recall in reverse order have been the focus of some controversy, with some researchers arguing that they do not produce more information than simply another free recall attempt (Boon & Noon, 1994; Davis, McMahon, & Greenwood, 2005; but see Gilbert & Fisher, 2006). Moreover, they appear to be least frequently used by the police (Kebbell, Milne, & Wagstaff, 1999). It should be noted, however, that the Cognitive interview is not supposed to be an entity that must be followed in a strict manner. Rather, it can best be thought of as a toolbox of techniques that can be adapted for each interview (Fisher, 2010). One drawback of the Cognitive Interview, however, is that if conducted in full, it is fairly comprehensive and rather time-consuming. The police do not always have the time to conduct Cognitive Interviews and interviews may be delayed (Kebbell et al., 1999). Recently, researchers have started investigating shortened versions of the Cognitive Interview that are less time-consuming. Their results showed that shortened versions can be equally effective as the full version in terms of quantity and accuracy, at least when the context reinstatement and the “report-everything” component are included (e.g., Bensi, Nori, Gambetti, & Giusberti, 2011; Dando, Wilcock, Milne, & Henry, 2009). The issue of delayed interviewing due to time-constraints will be addressed in more detail later in this chapter.

In the following, new findings and developments in the field of eyewitness interviews will be presented that - like the Cognitive Interview - fit into the framework of a positive legal psychology and aim to provide retrieval support for eyewitnesses during recall. On the one hand, these new findings comprise a metamemory model (Goldsmith, Koriat, & Weinberg-Eliezer, 2002; Koriat & Goldsmith, 1996) and its implications for interviewing witnesses. On the other hand, they comprise techniques that, in spite of having been employed for quite some time, have only recently been evaluated in terms of their effectiveness and in terms of how they facilitate retrieval, such as rapport-building. Another part of this chapter will be dedicated to the development of new interview tools, such as the Self-Administered Interview© (Gabbert, Hope, & Fisher, 2009), and to the positive effects of repeated eyewitness interviews. Based on the reviewed literature, this chapter will conclude with the aims of this dissertation.

Regulation of Report Option and Grain Size in Eyewitness Accounts

In recent years, the metamemory model proposed by Koriat and co-workers (Goldsmith et al., 2002; Goldsmith, Koriat, & Pansky, 2005; Koriat & Goldsmith, 1996) has been applied to the case of eyewitness accounts (Evans & Fisher, 2011; Weber & Brewer, 2008). It serves as the theoretical rationale as to why it is essential to apply the instruction not to guess, as is done in the Cognitive Interview (Fisher & Geiselman, 1992). The model assumes two mechanisms, *monitoring* and *control*, that are active between retention and retrieval and that witnesses use to increase the accuracy of their statements. Specifically, for each response alternative available in memory, the monitoring mechanism assigns a likelihood that this response is correct. The response alternatives differ with regard to grain size or level of precision (e.g., “dark” vs. “navy blue”). Subsequently, the control mechanism puts a report criterion on this monitoring output. A response is only overtly provided, if its likelihood of being correct passes the report criterion. The exact level of the report criterion depends on personal goals and motivation, such as having to tell the truth as a witness. When being a witness, the report criterion is expected to be higher (i.e., the person is more cautious) than in a situation in which someone wants to impress their friends with a spectacular story. The response alternative with an assigned likelihood that just passes the report criterion will be provided.

If, however, a response alternative does not exceed the report criterion, there are two options to increase accuracy. The response can either be withheld altogether (i.e., responding with “I don’t know”) or a less precise response is volunteered (i.e., stating “dark” instead of “navy blue”). The first option is called *control of report option* and the second option *control of grain size*. The former implies a quantity-accuracy trade-off (Koriat & Goldsmith, 1996): People can increase their accuracy by withholding a response. The latter implies an accuracy-informativeness trade-off (Goldsmith et al., 2002, 2005): The more precise a response, the less likely it is correct. Accordingly, people can increase their accuracy by decreasing the level of precision.

Witnesses, indeed, do apply both types of control options to increase the accuracy of their statements (Evans & Fisher, 2011; Weber & Brewer, 2008). Hence, the model proposed by Koriat and co-workers and the supporting empirical evidence provide a rationale for the instruction often used in interviews that witnesses should not guess and that it is perfectly acceptable to provide a “don’t know” response. From the metamemory model, it becomes clear that a “don’t know” response is not necessarily a sign of a bad memory, but may instead indicate heightened caution to provide an accurate account in the courtroom. The model also provides an explanation for the finding that accuracy need not decline over time. Instead, witnesses may still provide accurate accounts if they reduce the precision of their recollections (e.g., Evans & Fisher, 2011; Goldsmith et al., 2005). In the

following, new findings regarding old techniques developed to support eyewitness recall, namely, rapport-building and eye-closure will be presented.

Rapport-Building

As described above, rapport-building occurs at the beginning of the interview and is part of the Cognitive Interview (Fisher & Geiselman, 1992). While the beneficial effects on the witness' well-being seem straightforward, only few studies have specifically addressed the question as to what the impact of rapport-building on eyewitness recall performance in particular is (Vallano & Schreiber Compo, 2011). These studies found that both verbal and non-verbal rapport-building improve witness accounts. Specifically, Collins et al. (2002) examined the effect of non-verbal rapport-building on recall performance of a videotaped incident involving a burning car. While incorrect recall was not influenced by rapport-building, the rapport group recalled more correct details than the no-rapport group. Moreover, in subsequent closed questions, the rapport group reported more new correct details (i.e., details that had not been recalled in the preceding free recall).

Vallano and Schreiber Compo (2011) looked into the impact of self-disclosure as a means of rapport-building. In this study, in addition to rapport-building, exposure to misinformation was manipulated. After watching a videotaped staged theft, participants read a summary of the theft. For half of the participants, this summary contained misinformation. Hereafter, the interviews were conducted, either with or without a rapport-building phase in which interviewer and interviewee exchanged personal information. As expected, the rapport group recalled fewer incorrect details and hence was more accurate than the no-rapport group. Interestingly, rapport-building not only improved recall performance, but also reduced the proneness to misinformation, as the rapport group reported fewer misinformation items than the no-rapport group. Given that rapport-building is both an easy and effective means to improve witness statements, it should become a regular part of all interviews.

Eye-Closure during Retrieval

Similar to rapport-building, eye-closure is already mentioned in the Cognitive Interview manual as a potential means to reduce distraction during the interview (Fisher & Geiselman, 1992) and therefore cannot be considered a new technique, either. Yet, researchers have only recently started to systematically examine the effect of eye-closure on witnesses' recall performance as well as the underlying memory mechanisms. Their results indeed support the notion that closing the eyes while remembering leads to better recall performance (e.g., Perfect et al., 2008;

Vredeveldt, Hitch, & Baddeley, 2011), even after a considerable delay (Vredeveldt, Baddeley, & Hitch, 2014).

The positive effects of eye-closure seem to be due to several mechanisms. Specifically, Vredeveldt et al. (2011) found support for two different mechanisms. First, in line with the cognitive load hypothesis that regards memory retrieval and monitoring the environment as two competing tasks, closing the eyes improves recall by freeing cognitive resources that would otherwise have been used to monitor the environment. Second, in line with the modality-specific interference hypothesis, eye-closure reduces visual distraction. This facilitates visualizing the to-be-remembered incident and subsequently recalling it. A third mechanism discussed in the context of eye-closure is that it causes spontaneous mental context reinstatement. Specifically, in one study eye-closure elicited more correct details than recall with open eyes when the retrieval context was dissimilar rather than similar to the encoding context (Vredeveldt & Penrod, 2013). This finding supports the notion that eye-closure improves recall through mental context reinstatement. That is, if the location of encoding and retrieval are identical, the location itself serves as the physical context. Because the latter outshines mental context reinstatement that is caused by eye-closure (Smith, 1988, 1994), closing the eyes has no advantage to recall with open eyes. Eye-closure only improves recall performance, when physical context cues are absent (e.g., because the location of encoding and retrieval are different).

Eye-closure is also rather promising, because it is a simple yet effective technique that is cheap and time-efficient to conduct. It may be a viable procedure in cases in which due to time-constraints, comprehensive interviews as the Cognitive Interview would be delayed. Still, eye-closure should not be seen as a replacement, but rather a supplement of the Cognitive Interview (Vredeveldt & Penrod, 2013). We now turn to two new interview tools designed to facilitate memory retrieval.

Timeline Technique

Recalling a crime that was committed by multiple perpetrators poses higher cognitive demands onto witnesses as it involves correctly assigning the various actions to the different perpetrators and putting the actions into correct temporal order. The so called timeline technique was developed to support eyewitnesses' recall in exactly such situations (Hope, Mullis, & Gabbert, 2013). This technique makes use of the importance of temporal context for retrieval (Unsworth, 2008). The tool consists of a cardboard on which a timeline is printed that represents the temporal space along which the incident took place. Additionally, person description and action cards are provided onto which witnesses write their descriptions. The cards can subsequently be attached to the timeline. Witnesses may start at any point in the timeline and reorder the cards. Moreover, they mark on the cards as to who did

what and when. Compared to a report in which witnesses are asked to describe everything from beginning to end, the timeline technique lowers cognitive load, because it renders the decision as to what to report first as less important and reduces the necessity to switch between different storylines.

First results showed that the timeline technique elicited more correct details than a free recall (Hope et al., 2013), while overall accuracy was unaffected. Furthermore, the timeline technique led to better recall performance precisely for those details that are most relevant when there are multiple perpetrators, namely linking an action to a person and sequencing of actions. It remains to be examined how the technique could be incorporated into already existing interview tools.

The Self-Administered Interview

The second new interview tool to be presented in this chapter is the Self-Administered Interview© (SAI; Gabbert et al., 2009; Hope, Gabbert, & Fisher, 2011). It is based on the Cognitive Interview and adopts some of its memory-enhancing components (Gabbert et al., 2009). Put in simple terms, the SAI may be considered a shortened written version of the Cognitive Interview. As described above, one problem of interviews such as the Cognitive Interview is their comprehensiveness. Due to limited police resources, it is often not possible to interview witnesses immediately after the crime (Kebbell et al., 1999). This is especially the case when there are multiple witnesses. However, delaying the interview can be quite problematic, because memory research has shown that with increasing delay more information is forgotten (e.g., Rubin & Wenzel, 1996; Tuckey & Brewer, 2003). Especially precise or verbatim information is affected (e.g., Goldsmith et al., 2005). Moreover, with increasing delay, it also becomes more likely that witnesses are exposed to misinformation (e.g., Ayers & Reder, 1998).

From these considerations, it becomes clear that delaying the interview can pose a hindrance to successful investigations. To tackle the problems associated with delayed interviewing and to ensure an early comprehensive interview, Gabbert et al. (2009) developed the SAI in close cooperation with the police (see Hope et al., 2011, for a review of the development process). The SAI comes in the form of a booklet consisting of 16 pages and seven sections that will be described below. Each section is completed, if it is relevant for the case at hand. Otherwise, the witness can directly move on to the next section (e.g., if there were no vehicles present, the witness can just skip this section). The SAI is distributed to witnesses at the crime scene and witnesses independently complete it directly after the crime. If completion at the crime scene is not possible, witnesses should complete it at a place nearby or at home and return it as quickly as possible to the police. As a generic recall tool, it can be used for all different types of crimes and does not need to be adjusted for each case. Its development process was guided by the idea to iden-

tify retrieval-facilitating techniques known from the Cognitive Interview that can be applied in a written interview. For this purpose, especially the cognitive components seemed suitable.

Indeed, just as in the Cognitive Interview, the most important memory-enhancing component of the SAI is the mental reinstatement of context. That is, before writing down their recollections, witnesses picture the incident in their minds. Moreover, witnesses are instructed to provide the most complete and accurate report possible, but not to guess. This corresponds to the “report-everything” component of the Cognitive Interview. After receiving these two sets of instructions (i.e., context reinstatement and request for a complete recall), witnesses provide a description of the course of events and all persons involved (section A). The second section (B) focuses on the perpetrator description. Witnesses are required to describe the appearance and clothing of all persons who committed the crime. For this purpose, witnesses also receive non-leading recall prompts (e.g., gender, hair, facial details, build, clothing, etc.) to cue recall. Moreover, two figures are provided showing the front and back side of a person. These may be used by witnesses to indicate, for example, the position of accessories or to draw a tattoo and its position. In section C, witnesses draw a sketch of the scene to receive as much spatial information (e.g., positions and directions) as possible, because this type of information can be more easily retrieved non-verbally than verbally. As some witnesses may feel uncomfortable drawing, the instructions reassure witnesses that the sketch is not supposed to be a test of their drawing abilities. By using several retrieval attempts (viz., the different sections) and by querying both verbal and non-verbal information, the SAI, just like the Cognitive Interview, relies on multiple and varied retrieval (Tulving & Watkins, 1975). In the next section (section D), other potential witnesses can be described. This may help identify further witnesses if the police have not been aware of all witnesses present at the crime scene. If vehicles (i.e., cars, but also bikes or scooters) were present at the crime scene or involved in the crime (e.g., as a get-away car), witnesses can describe those in section E. As with the section on the perpetrator’s appearance, retrieval cues (e.g., color, brand, model, license plate etc.) are provided. The witnessing conditions (e.g., viewing distance, obstructions to view, and lighting conditions) are prompted in section F. In the last section (G), witnesses receive the opportunity to write down any other information that has not been asked about. In sum, then, the SAI incorporates a wealth of recall-facilitating mechanisms and as such provides a vast amount of retrieval support. It should be noted, though, that similar to eye-closure, the SAI is supposed to be a supplement, not a replacement, of a comprehensive personal interview (Hope et al., 2011) to support the police in the immediate aftermath of a crime.

As the SAI is relatively new, there are only a few published studies that have looked at its effectiveness. Tables 1.1 and 1.2 provide an overview of the findings of these studies (published up to and including August 26th, 2014). Specifically,

Table 1.1 summarizes the studies that examined the effects of SAI completion on recall performance immediately after witnessing an incident. Table 1.2 summarizes the studies that examined how completion of the SAI immediately after the incident (T1) influences recall performance in a subsequent interview (T2) conducted at a one- or two-week delay.

Table 1.1

Results of the Published Studies Examining Performance with the SAI vs. Control Interviews Immediately after the Incident for Correct Details, Incorrect Details, and Accuracy

Study	N	Control group	d^1		
			Correct	Incorrect	Accuracy
Gabbert, Hope, & Fisher (2009), Exp. 1	55	CI	-0.66	0.71	-0.97
		FR	1.67	1.22	-0.49
Gawrylowicz, Memon, & Scoboria (2014)	84	FR	1.20	-.2	0.10
Hope, Gabbert, Fisher, & Jamieson (2014)	60	FR	2.65	1.20	-0.38
Maras, Mulcahy, Memon, Picariello, & Bowler, (2014; non-clinical subsample)	35	SR	0.96	0.82	0.02

Note. CI = Cognitive Interview; FR = free recall; SR = structured recall interview.

¹Positive values indicate increases in the respective dependent variable in the SAI relative to the control group. ²No effect size could be calculated, because the respective paper does not report the necessary statistical values.

In the studies examining recall performance immediately after witnessing an incident (see Table 1.1), the SAI was mostly compared with free recall (FR) as the control interview. The FR provides only little retrieval support, as it does not contain any memory-facilitating components or retrieval cues and just asks the witness to provide an account of what happened and of descriptions of the people involved. Relative to an FR, the SAI elicited significantly more correct details, with the increases ranging from 53.3% to 86.4% (Gabbert et al., 2009, Experiment 1; Gawrylowicz, Memon, & Scoboria, 2014; Hope, Gabbert, Fisher, & Jamieson, 2014). At the same time, the accuracy rates remained unaffected. In spite of the large increase in correct details and although the accuracy rates obtained with the SAI were high (ranging from 89% to 96%), the results deserve a critical note. Specifically, the SAI did elicit more incorrect details than the FR (Gabbert et al., 2009, Experiment 1; Hope et al., 2014), with the increase corresponding to large effects. This may be problematic for real cases, because, unlike in laboratory studies, for the police it is unknown whether a given detail is correct or incorrect. This issue should be monitored and will again be addressed in the general discussion (chapter 7). Interestingly, the pattern of increased recall quantity and unaffected accuracy is also usually found in Cognitive Interview studies (Memon et al., 2010) on which the SAI is based. Also similar to results obtained with the Cognitive Interview, completing the SAI immediately after the witnessed incident has been found to protect against reporting misinformation in a subsequent interview (Gabbert, Hope, Fisher, & Jamieson, 2012; Paterson, Eijkemans, & Kemp, 2014).

Table 1.2

Results of the Published Studies Examining Recall in a Subsequent (T2) Interview as a Function of Initial (T1) SAI Completion for Correct Details, Incorrect Details, Accuracy, and Misinformation Reported

Study	N	Control group at T1	Type of T2-interview ¹	<i>d</i> ²			
				Correct	Incorrect	Accuracy	Misinfo
Gabbert, Hope, & Fisher (2009), Exp. 2	42	No interview	FR	0.71	-0.56	0.91	
Gabbert, Hope, Fisher, & Jamieson (2012), Exp. 1	40	No interview	FR	1.69	0.42	0.26	-0.70
Gabbert, Hope, Fisher, & Jamieson (2012), Exp. 2	62	No interview	Cued questions	1.35	-0.62	0.66	-0.48
Gawrylowicz, Memon, & Scoboria (2014) ³	84	FR	FR	0.68	- ⁴	0.68	
Hope, Gabbert, Fisher, & Jamieson (2014)	60	FR No interview	CI	0.89	- ⁴	0.57	
			CI	0.66	- ⁴	0.61	
Maras, Mulcahy, Memon, Picariello, & Bowler, (2014; non-clinical subsample)	29	SR	SR	0.66	- ⁴	- ⁴	
Paterson, Eijkemans, & Kemp (2014) ⁵	52	No interview	FR	0.64	- ⁴	- ⁴	-0.38

Note. FR = free recall; CI = Cognitive Interview; SR = structured recall interview.

¹The delay until the second interview was one week, except for Paterson et al.'s (2014) study, in which the delay was two weeks. ²Positive values indicate increases in the respective dependent variable in the SAI relative to the control group. ³The study dealt with the so called transfer effect, that is, participants testified about a different event in their second interview than in the first interview. ⁴No effect size could be calculated, because the respective paper does not report the necessary statistical values. ⁵For reasons of comparability with the other studies, only the comparison between the group immediately completing the SAI and the no-initial-interview control group is reported in the table.

More importantly (see Table 1.2), completing the SAI seems to preserve memory for a subsequent personal interview (Gabbert et al., 2009), at least when the SAI is completed no later than 24 hours after the incident (Paterson et al., 2014). The first studies that investigated this memory-preserving effect (Gabbert et al., 2009; Gabbert et al., 2012; Paterson et al., 2014), however, employed a no-initial-interview control group. Consequently, it is unclear whether the results were due to an early recall opportunity per se or to the SAI in particular. A more thorough test was conducted in a study by Hope et al. (2014) in which participants watched

a film showing a staged crime. Hereafter, they completed an SAI, an FR, or no initial interview at all. One week later, the participants were invited to a second session in which all underwent a Cognitive Interview. Here, those participants who had completed an SAI earlier produced more accurate statements than participants who had filled in an FR. Interestingly, recall performance at the second session did not differ between the FR group and the group that had not received an interview at session 1, showing that memory is not protected for a subsequent interview by having an initial recall attempt *per se*. Hope et al. explained these results by drawing on network models of memory (e.g., J. R. Anderson, 1983). According to these models, memory consists of a network of nodes and links between the nodes. The more elaborate retrieval attempt obtained with the SAI strengthens the network more than a less elaborate retrieval attempt with the FR. Hence, the memory of the incident is more likely to be preserved.

In addition to yielding more comprehensive statements, SAI completion also seems to teach witnesses retrieval skills which they can apply to other witnessed events. In a study by Gawrylowicz et al. (2014), participants completed either an SAI or an FR directly after watching a videotaped staged crime. One week later, they witnessed a second, completely different, videotaped crime and again underwent an interview which was an FR for all participants. Regarding recall performance of the second crime, those participants who had filled in the SAI for the first scenario recalled both more correct details and more accurately than participants who had previously filled in an FR.

In the meantime, the SAI has been taken from the laboratory to the field and used by the police in real cases. Field trials in the United Kingdom suggest a high acceptance and satisfaction with the tool among police officers (Hope et al., 2011). In addition to the United Kingdom, the SAI is regularly used in Norway and the Netherlands, where a field trial has recently been completed. In Germany, field trials are underway.

Since the SAI is a new tool, some important research questions are unanswered. For example, it has not been examined yet, how the SAI fares in different witness populations, such as adolescents, and how individuals with reading difficulties come to terms with the SAI. Moreover, the influence of individual differences in, for example, working memory capacity, known to influence recall performance (e.g., Peters, Jelicic, Verbeek, & Merckelbach, 2007) could be examined during recall with the SAI. Finally, perhaps with the exception of the misinformation studies (Gabbert et al., 2012; Paterson et al., 2014), little is known at this stage about whether the vast amount of retrieval support provided in the SAI can help witnesses recall when the witnessing conditions were suboptimal (e.g., because the witness' attention was divided while witnessing the crime).

Positive Effects of Repeated Interviews – Reminiscence

Witnesses are likely to be interviewed several times during police investigations (Bornstein, Liebel, & Scarberry, 1998), not only in cases involving the SAI. When being interviewed on two or more occasions, witnesses can report details consistently in all interviews or they can forget details they had reported earlier. Moreover, inconsistencies can occur across several retrieval attempts that comprise direct contradictions and reminiscence. The latter refers to the recall of new (i.e., previously unrecalled) details. Both contradictions and reminiscence are usually considered to be a sign of low overall accuracy of the witness' statement by legal professionals (e.g., Brewer, Potter, Fisher, Bond, & Luszcz, 1999). Presumably, reminiscence is mistrusted, because it seems to contradict the notion of fading memory over time (Fisher, Brewer, & Mitchell, 2009). Reminiscence, however, is a normal memory phenomenon that occurs concurrently with forgetting (Erdelyi, 2010). Almost every participant who is interviewed twice about the same incident experiences reminiscence (Brock, Fisher, & Cutler, 1999; Gilbert & Fisher, 2006). It is assumed that reminiscence is the result of a change of retrieval cues from the first to the subsequent interview that makes different parts of a recollection accessible (Fisher et al., 2009; Tulving & Watkins, 1975; see the principle of multiple retrieval in the Cognitive Interview; Fisher & Geiselman, 1992). The amount of reminiscent correct information elicited in a second interview can be substantial. Relative to the first interview, a second interview has been found to yield an additional 25% or more correct new information (Hope et al., 2014; Odinot, Memon, La Rooy, & Millen, 2013). Moreover, the accuracy of reminiscent details can be quite high and almost as high as the accuracy of consistent details (e.g., Brock et al., 1999; Gilbert & Fisher, 2006; Odinot, Wolters et al., 2013). Finally, the number of inconsistencies is not or only weakly correlated with overall recall accuracy (e.g., Brewer et al., 1999; Odinot, Memon et al., 2013; Smeets, Candel, & Merckelbach, 2004). As a result, the concerns about reminiscence seem unwarranted from an empirical perspective. Instead, repeated interviews can be considered a means to support witnesses to provide comprehensive statements. In a recent experiment (Oeberst, 2012), the huge discrepancy between the assumptions held by laypeople on the one hand and empirical findings on the other hand regarding the accuracy of reminiscent details was nicely illustrated. In this study, memory performance across two recall attempts was directly compared with laypersons' expectations (Experiment 2). Specifically, one group of participants, the memory group, watched a stimulus film showing an altercation and hereafter provided two statements about what they had witnessed, one immediately after the incident and one after one week. A second group of participants, the estimation group, also watched the film but subsequently estimated the performance of the memory group. The estimation group assumed that overall recall accuracy of the memory group would drop over time and that the accuracy of reminiscent details would be lower than

the accuracy of consistent and forgotten details. However, the estimation group could not have been more wrong. Instead of dropping over time, actual recall accuracy remained stable. The accuracy of reminiscent details was high and did not differ from the accuracy of both consistent and forgotten details. What is more, actual recall accuracy was underestimated without exception. The degree of underestimation was especially striking for the reminiscent details. While the memory group achieved a reminiscence accuracy rate of 84%, the estimation group estimated their performance at just 19%. The results also showed that reminiscence was no exception, but occurred in the vast majority of the participants of the memory group. Hence, legal professionals' distrust regarding reminiscent details is unjustified. Even though Oeberst's study yielded interesting results, it has some limitations. Specifically, the data were collected in an undergraduate student sample, so it is unclear whether the findings apply to professionals who are experienced with witnesses, such as police officers. Moreover, the study did not uncover potential reasons as to why reminiscence is mistrusted. Consequently, some important questions remain open. Nevertheless, it can be stated that repeated interviews support eyewitnesses and are a means to elicit new correct information. It goes without saying that this is only the case when interviewers refrain from using suggestive questioning techniques (Loftus, 2005).

Aims and Outline of the Dissertation

In line with a positive legal psychology (Meissner et al., 2010), it is the aim of the present dissertation to examine the effects of retrieval support on recall performance in eyewitness interviews. To this end, the SAI (Gabbert et al., 2009; Hope et al., 2011) will receive a greater focus and be considered a proxy tool for providing eyewitnesses with ample retrieval support. Prior to conducting the experiments, we translated the SAI into German (Eigenständiges Vernehmungsprotokoll für Augenzeugen; EVA) and Dutch (Zelfrapportage voor getuigen; ZeG). Both translations can be found in Appendix A and Appendix B at the end of this chapter. In several experiments, we investigated whether retrieval support can help witnesses achieve comprehensive and accurate accounts when the witnessing situations during encoding or retrieval were suboptimal. Moreover, the effect of retrieval support through repeated interviews was examined. The first chapters will mainly be concerned with factors emerging during encoding, while the last chapters will predominantly be concerned with factors during retrieval. Throughout the chapters, the above described metamemory model by Koriat and co-workers (Goldsmith et al., 2002; Koriat & Goldsmith, 1996) as well as the Cognitive Interview and its components, such as rapport-building, will be picked up on. The following questions served as the starting points of each chapter:

Chapter 2: Does the modality of an eyewitness interview influence the quantity and accuracy of the statement?

While classic interviews as the Cognitive Interview are conducted in spoken, the SAI is a written interview. However, little is known about the impact of the modality (writing vs. speaking) on eyewitness performance. Hence, before conducting further in-depth research on the SAI, the question arises whether either written or spoken interviews are more suitable to facilitate eyewitness recall. This question was addressed in two experiments. Moreover, potential interactions with individual differences in executive functioning (e.g., working memory capacity) and comprehensiveness (i.e., amount of retrieval support) of the recall instructions were investigated.

Chapter 3: Can retrieval support with the Self-Administered Interview© facilitate recall when the witness was exposed to stress?

Chapter 3 focuses on a suboptimal witnessing situation originating in the encoding phase. Specifically, previous research has shown that stress while witnessing an incident can negatively influence witnesses' recall performance (Deffenbacher, Bornstein, Penrod, & McGorty, 2004). However, the police are likely to interview all witnesses, irrespective of their stress levels, to obtain eyewitness evidence. Hence, it is important to know whether interviews offering a high amount of retrieval support can facilitate recall, irrespective of the stress level experienced.

Chapter 4: Can initial retrieval support with the Self-Administered Interview© facilitate recall when the witness' attention was divided during the crime? When does initial retrieval support preserve memory for a subsequent interview the best?

The aim of chapter 4 is twofold. First, it deals with another suboptimal situation during encoding that leads to reduced memory performance, namely divided attention (e.g., N. D. Anderson, Craik, & Naveh-Benjamin, 1998). Indeed, witnesses' attention may not always be fully focused on the incident (e.g., because they are looking for an escape route). Thus, we examined whether retrieval support can enhance recall performance irrespective of the amount of attention paid. Second, a factor relevant in the retrieval phase, namely the type of the subsequent interview after initial retrieval support, is considered. Drawing from the findings obtained by Hope et al. (2014), the question was addressed as to what kind of a second interview is needed to preserve memory after witnesses have been provided with initial retrieval support.

Chapter 5: Are repeated interviews a suitable means to improve recall performance? Do police officers see the merit of repeated interviews?

Chapter 5 also deals with the retrieval phase and repeated interviews, to wit, the provision of retrieval support via repeated eyewitness interviews. With this in mind, recall consistency and especially the recall of reminiscent (i.e., previously

unrecalled) information were investigated. We also explored whether the amount of retrieval support provided in the initial interview has an impact on the (in)consistency of recall. Moreover, we examined whether experienced police detectives see the merit of repeated interviews, by comparing their estimates of recall performance across repeated interviews with actual recall performance.

Chapter 6: Is retrieval support equally effective for witnesses with different levels of working and source memory?

Chapter 6 is concerned with individual differences in witnesses that have an impact during retrieval. Specifically, reduced working and poor source memory have been found to lead to more recall errors (e.g., Unsworth & Brewer, 2010b). We examined whether retrieval support can help witnesses with low working or poor source memory to obtain comprehensive and accurate accounts.

The thesis concludes with a general discussion (*chapter 7*) in which a summary and (meta-analytical) integration of all results are presented. Moreover, their implications for interviewing practice are discussed and suggestions for future research are provided.

Appendix A: Eigenständiges Vernehmungsprotokoll für Augenzeugen



Polizeisiegel
hier einfügen

Dieses Heft beinhaltet Fragen zu dem Vorfall, den Sie als Zeuge beobachtet haben. Bitte beantworten Sie diese Fragen. Dazu ist es wichtig, dass Sie die Fragen sorgfältig durchlesen und die Anweisungen befolgen.

Das Heft besteht aus mehreren Abschnitten, die Sie der Reihenfolge entsprechend (d. h. von vorne nach hinten) ausfüllen sollen. Es kann sein, dass nicht alle Abschnitte für Sie relevant sind. Wenn das der Fall ist, können Sie dann einfach zum nächsten Abschnitt übergehen. Wenn ein Abschnitt jedoch für Sie relevant ist, sollten Sie ihn so vollständig wie möglich ausfüllen.

Bitte füllen Sie diesen Fragebogen alleine und ohne die Hilfe Dritter aus.

Sollten Sie Fragen haben, setzen Sie sich bitte mit dem folgenden Beamten, der Ihnen dieses Heft ausgehändigt hat, in Verbindung.

DIESES HEFT WURDE IHNEN AUSGEHÄNDIGT VON:

Name: _____

Polizeistation: _____

Telefon-Nummer: _____

Kennziffer: _____

Ausgegeben am (Datum und Uhrzeit): _____

BITTE TRAGEN SIE HIER IHRE KONTAKTDATEN EIN:

Name: _____

Adresse: _____

Telefon-Nummer: _____

Handy-Nummer: _____

Email: _____

Ausgefüllt am (Datum und Uhrzeit): _____

BLÄTTERN SIE ERST UM, WENN SIE DIESE SEITE VOLLSTÄNDIG AUSGEFÜLLT HABEN

Copyright © 2006, Gabbert, Hope & Fisher.

Agreements of use: All rights reserved, including translation. No part of this publication may be photocopied, reproduced or transmitted in any form or by any means electronic or mechanical, recording or duplication in any information storage or retrieval system without permission in writing from the above named Authors even within the terms granted by the Copyright Licensing Agency Ltd.

Permission is not granted for partial or alternative use of this publication other than that specified by the Authors.

 **EVA** Eigenständiges
Vernehmungsprotokoll für
Augenzeugen

HIER KEINE EINTRAGUNGEN



ABSCHNITT A: Was ist passiert?

Lesen Sie bitte die nachfolgenden Informationen durch und befolgen Sie die Anweisungen.

Bevor Sie Ihre Erinnerungen an den Vorfall in diesem Heft aufschreiben, nehmen Sie sich etwas Zeit, um sich gedanklich vorzustellen, wo Sie waren, was Sie sahen, was Sie gerade dachten und wie Sie sich zu dem Zeitpunkt fühlten, als Sie Zeuge des Ereignisses wurden.

Warum sollen Sie das tun?

Über das Ereignis nachzudenken, bevor Sie Ihre Erinnerungen daran aufschreiben, hilft Ihnen dabei, sich an mehr Details zu erinnern.

Was sollen Sie jetzt tun?

Nehmen Sie sich viel Zeit, um sich zu konzentrieren und sich in Gedanken bildlich vorzustellen, was passiert ist.

Denken Sie dabei an die folgenden Punkte:

- Wo sie waren
- Was Sie gerade taten
- Mit wem Sie sich aufgehalten haben
- Wie Sie sich fühlten
- Was gerade passierte
- Wer alles beteiligt war
- Was Sie sehen konnten
- Was sie hören konnten

Dabei kann es hilfreich sein, Ihre Augen zu schließen, während Sie sich an den Vorfall erinnern.

Blättern Sie erst um, wenn es Ihnen gelungen ist, den Vorfall vollständig zu erinnern.

 **EVA** Eigenständiges
Vernehmungsprotokoll für
Augenzeugen

Lesen Sie bitte die nachfolgenden Informationen durch und befolgen Sie die Anweisungen.

Schreiben Sie bitte in den dafür vorgesehenen Feldern alle Details auf, die Sie über den Vorfall und die beteiligten Personen erinnern können.

Schreiben Sie alles so auf, wie es Ihnen einfällt. Es ist nicht schlimm, wenn Sie die Dinge nicht genau in der Reihenfolge erinnern, in der sie sich ereigneten.

Lassen Sie keine Details aus, aber raten Sie nicht in Bezug auf Details, an die Sie sich nicht erinnern können.

Sie können ganze Sätze oder auch in Stichpunkten schreiben – achten Sie jedoch darauf, dass Ihr Bericht so **vollständig** und **richtig** wie möglich ist.

Bitte füllen Sie diesen Fragebogen alleine und ohne die Hilfe Dritter aus. Uns interessiert nur Ihre eigene Erinnerung an den Vorfall.

Beginnen Sie hier:



Hier weiter:

A large, empty rectangular box with a thin black border, intended for the user to write or paste content.



Hier weiter:

A large, empty rectangular box with a thin black border, intended for the continuation of the text.



Hier weiter:

A large, empty rectangular box with a thin black border, intended for the user to write or paste content.

EVA **Eigenständiges Vernehmungsprotokoll für Augenzeugen**

ABSCHNITT B: Wer hat die Tat begangen?

Lesen Sie bitte die nachfolgenden Informationen durch und befolgen Sie die Anweisungen.

Bitte geben Sie auf der nächsten Seite so viele Merkmale wie möglich über den/die Täter/-in (die Person, welche die Straftat begangen hat oder die in den Vorfall verwickelt war) an. Wenn es mehr als eine/-n Täter/-in gab, beginnen Sie mit derjenigen Person, auf die Sie die beste Sicht hatten oder die Sie am deutlichsten erinnern.

Wenn Sie in diesem Heft bereits an anderer Stelle eine Beschreibung abgegeben haben, nehmen Sie sich etwas Zeit und überlegen Sie, ob es zusätzliche Merkmale gibt, an die Sie sich jetzt erinnern können.

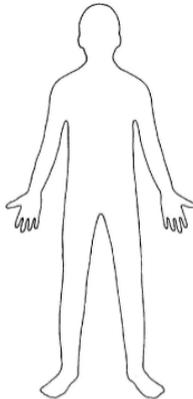
Geben Sie, wenn möglich, die folgenden Informationen über jede Person an. **Raten Sie nicht** in Bezug auf Details, über die Sie sich nicht sicher sind.

- Geschlecht
- Alter
- Größe
- Ethnie
- Gewicht / Statur
- Augen / Ohren / Mund / Nase / etc.
- Haarfarbe
- Gesichtsbehaarung
- Gesichtsfarbe / Teint
- Bekleidung / Schuhe
- Dialekt / Akzent
- Brille
- Schmuck
- Accessoires
- Narben / (Mutter-)Male / Tätowierungen

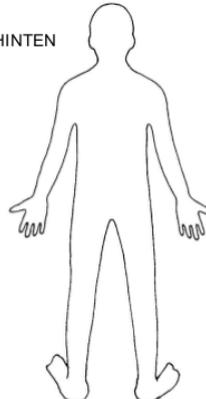
• Gibt es weitere Merkmale des/der Täter/-in bzw. der Täter/-innen, die noch nicht abgefragt wurden?

Sie können die schematische Abbildung unten auf dieser Seite verwenden, um weitere Informationen hinzuzufügen. Sie können auf den Abbildungen schreiben und/oder zeichnen, um zusätzliche Merkmale oder Informationen darzustellen (z.B. die Position möglicher Accessoires etc.).

VORNE



HINTEN



 **EVA** Eigenständiges
Vernehmungsprotokoll für
Augenzeugen

Beschreibung(en) bitte hier abgeben

A large empty rectangular box intended for the witness's description.

**ABSCHNITT C: Der Ort des Geschehens**

Lesen Sie bitte die nachfolgenden Informationen durch und befolgen Sie die Anweisungen.

Wenn Sie eine Skizze vom Ort des Geschehens anfertigen, hilft Ihnen dies dabei, zusätzliche Details zu erinnern und somit schließlich weitere Details über den Vorfall zu berichten.

Bitte benutzen Sie das Feld weiter unten und/oder auf der nächsten Seite, um eine Skizze des Orts anzufertigen, wie Sie den Ort in Erinnerung haben.

Sie sollen dabei angeben, wo Sie waren und wo andere Personen waren. Sie können Pfeile benutzen, um Ihre eigenen Bewegungen und die Bewegungen der anderen Personen, die Sie gesehen haben, deutlich zu machen. Sie können auch Pfeile verwenden, um die Bewegungsrichtung von möglichen Fahrzeugen oder Gegenständen, die Sie vielleicht gesehen haben, anzugeben.

Sie können in Ihrer Skizze Beschriftungen oder Notizen verwenden, um Merkmale des Orts anzugeben oder um deutlich zu machen, dass Sie sich hinsichtlich einer Sache nicht sicher sind.

Dies ist keine Überprüfung Ihrer zeichnerischen Fähigkeiten – wir sind lediglich an der Anordnung des Orts interessiert, z.B. was und wo Sie etwas gesehen haben.



Skizze hier einfügen:

A large, empty rectangular box with a thin black border, intended for the user to draw a sketch.



ABSCHNITT D: Waren eine oder mehrere Personen anwesend, die gesehen haben, was passiert ist?

JA (Füllen Sie diesen Abschnitt aus)

NEIN (Gehen Sie zum nächsten Abschnitt E)

Lesen Sie bitte die nachfolgenden Informationen durch und befolgen Sie die Anweisungen.

Geben Sie bitte eine Beschreibung von jeder Person ab, die anwesend war und gesehen haben könnte, was passiert ist, die aber nicht in den Vorfall verwickelt war (z.B. weitere mögliche Zeugen/-innen)?

Wenn Sie in diesem Heft bereits an anderer Stelle eine Beschreibung abgegeben haben, nehmen Sie sich etwas Zeit und überlegen Sie, ob es zusätzliche Merkmale gibt, an die Sie sich jetzt erinnern können.

Beginnen Sie hier:



ABSCHNITT E: Waren ein oder mehrere Fahrzeuge beteiligt?

JA (Füllen Sie diesen Abschnitt aus)

NEIN (Gehen Sie zum nächsten Abschnitt F)

Lesen Sie bitte die nachfolgenden Informationen durch und befolgen Sie die Anweisungen.

Geben Sie so viele Details wie möglich über jedes Fahrzeug an.

Sollten Sie bereits an anderer Stelle in diesem Heft eine Beschreibung abgegeben haben, nehmen Sie sich etwas Zeit und überlegen Sie, ob es zusätzliche Details gibt, an die Sie sich jetzt erinnern können.

Geben Sie, wenn möglich, die folgenden Informationen über jedes Fahrzeug an. **Raten Sie nicht** in Bezug auf Details, über die Sie sich nicht sicher sind.

- Größe
- Form
- Farbe
- Fabrikat / Modell
- Anzahl Türen
- Kennzeichen
- Fahrstil
- Geschwindigkeit
- Weitere Details?

Beginnen Sie hier:



Hier weiter:

A large, empty rectangular box with a thin black border, intended for the continuation of the text.



ABSCHNITT F: Wie gut haben Sie den Vorfall gesehen?

Beantworten Sie bitte die folgenden Fragen.

- Wie lange haben Sie den Vorfall als Zeuge gesehen?
- Wie waren die Wetterverhältnisse zu diesem Zeitpunkt?
- Zu welcher Tageszeit hat sich der Vorfall ereignet?
- Haben Sie den Vorfall bei Tageslicht oder bei künstlichem Licht gesehen? (Falls möglich, bitte näher beschreiben.)
- Gab es irgendwelche dauerhaften oder vorübergehenden Beeinträchtigungen Ihrer Sicht? (Falls möglich, bitte näher beschreiben.)
- Gibt es bestimmte Gründe, warum Sie sich den Vorfall oder den/die Täter/-in bzw. Täter/-innen gemerkt haben? (Falls möglich, bitte näher beschreiben.)
- War jemand beteiligt, den Sie kannten oder den Sie schon einmal gesehen haben? (Wenn ja, wo und wann?)



ABSCHNITT G: Gibt es noch weitere Informationen, die Sie über den Vorfall berichten wollen, die aber noch nicht abgefragt wurden? Falls ja, schreiben Sie diese Informationen in das Feld unten auf dieser Seite.

Beginnen Sie hier:

BITTE ZURÜCK AN:

Appendix B: Zelfrapportage voor getuigen



Plak hier
politielogo

Dit formulier bevat vragen die gaan over het incident waarvan u getuige was. Zou u deze vragen willen beantwoorden? Het is belangrijk dat u de vragen goed leest en de instructies volgt.

Er zijn verschillende onderdelen die u moet invullen in chronologische volgorde (beginnend van voor naar achter). Sommige onderdelen zullen voor u niet relevant zijn om in te vullen. In dat geval kunt u doorgaan naar het volgende onderdeel. Als een onderdeel wel relevant is, moet u het zo volledig mogelijk invullen.

Het is belangrijk dat u de vragen zelfstandig beantwoordt, zonder hulp aan anderen te vragen.

Wanneer u nog vragen heeft, kunt u contact opnemen met de politieambtenaar die u dit formulier heeft verstrekt.

DIT FORMULIER WERD U VERSTREKT DOOR:

Naam: _____
Politiebureau: _____
Telefoonnummer: _____
Zaak: _____
Datum en tijd van uitreiken: _____

VUL DE VOLGENDE DETAILS IN:

Naam: _____
Adres: _____
Telefoonnummer: _____
Mobiele telefoon: _____
Datum en tijd van invullen: _____

GA PAS VERDER ALS U DE HIER BOVEN GEVRAAGDE INFORMATIE HEEFT INGEVULD

Copyright © 2006, Gabbert, Hope & Fisher.

Agreements of use: All rights reserved, including translation. No part of this publication may be photocopied, reproduced or transmitted in any form or by any means electronic or mechanical, recording or duplication in any information storage or retrieval system without permission in writing from the above named Authors even within the terms granted by the Copyright Licensing Agency Ltd.

Permission is not granted for partial or alternative use of this publication other than that specified by the Authors.

Versie 27.11.2012

ZeG Zelfrapportage voor
Getuigen

GELIEVE HIER NIET TE SCHRIJVEN

ZeG Zelfrapportage voor Getuigen

ONDERDEEL A: Wat is er gebeurd?

Het is belangrijk dat u de volgende informatie leest en de instructies volgt.

Neem, voordat u begint met het opschrijven van uw herinneringen in dit formulier, enkele momenten de tijd om u voor de geest te halen waar u was, wat u zag, wat u dacht en hoe u zich voelde op het moment dat u getuige was van de gebeurtenis.

Waarom zou u dit moeten doen?

Nadenken over de gebeurtenis voordat u uw herinneringen gaat opschrijven, helpt u om zich meer details te herinneren.

Wat zou u nu moeten doen?

Neem ruim de tijd om u te concentreren en de herinnering van het incident weer voor de geest te halen.

Denk na over de volgende punten:

- Waar u was
- Wat u aan het doen was
- Met wie u samen was
- Hoe u zich voelde
- Wat er gebeurde
- Wie er betrokken was
- Wat u kon zien
- Wat u kon horen

Het kan helpen om de ogen te sluiten terwijl u zich het incident herinnert.

Sla de pagina pas om als het u gelukt is de gebeurtenis volledig te herinneren.

ZeG Zelfrapportage voor Getuigen

Het is belangrijk dat u de volgende informatie leest en de instructies volgt.

In de ruimte hieronder kunt u alle details opschrijven die u zich kunt herinneren over het incident en de mensen die erbij betrokken waren.

Schrijf de details op zoals u ze zich herinnert. Het maakt niet uit of u zich de details in een andere volgorde herinnert dan de volgorde waarin ze plaatsvonden.

Laat geen details weg, maar **ga niet gokken** over details die u zich niet kunt herinneren.

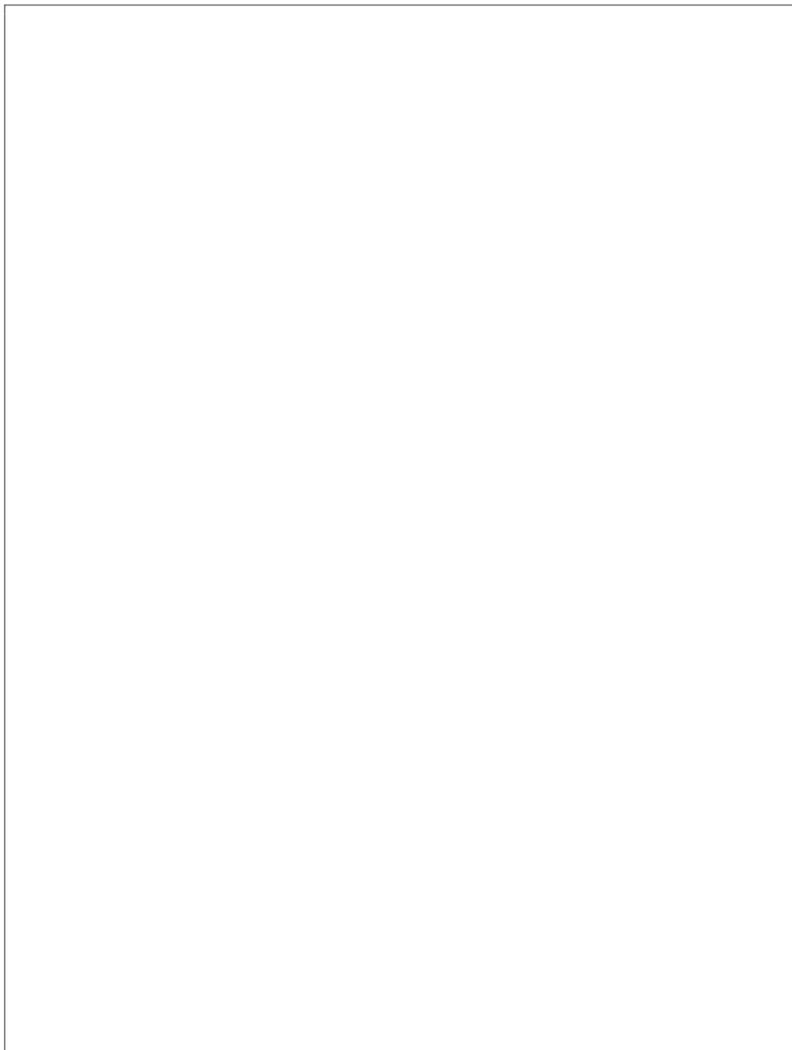
Voel u vrij om volle zinnen of steekwoorden te gebruiken – maar zorg er alstublieft voor dat uw verhaal zo **compleet** en **correct** mogelijk is.

Het is belangrijk dat u dit formulier invult zonder hulp aan anderen te vragen. Het gaat er om wat ú heeft waargenomen.

Start hier:

 **ZeG** Zelfrapportage voor
Getuigen

Ga hier verder:



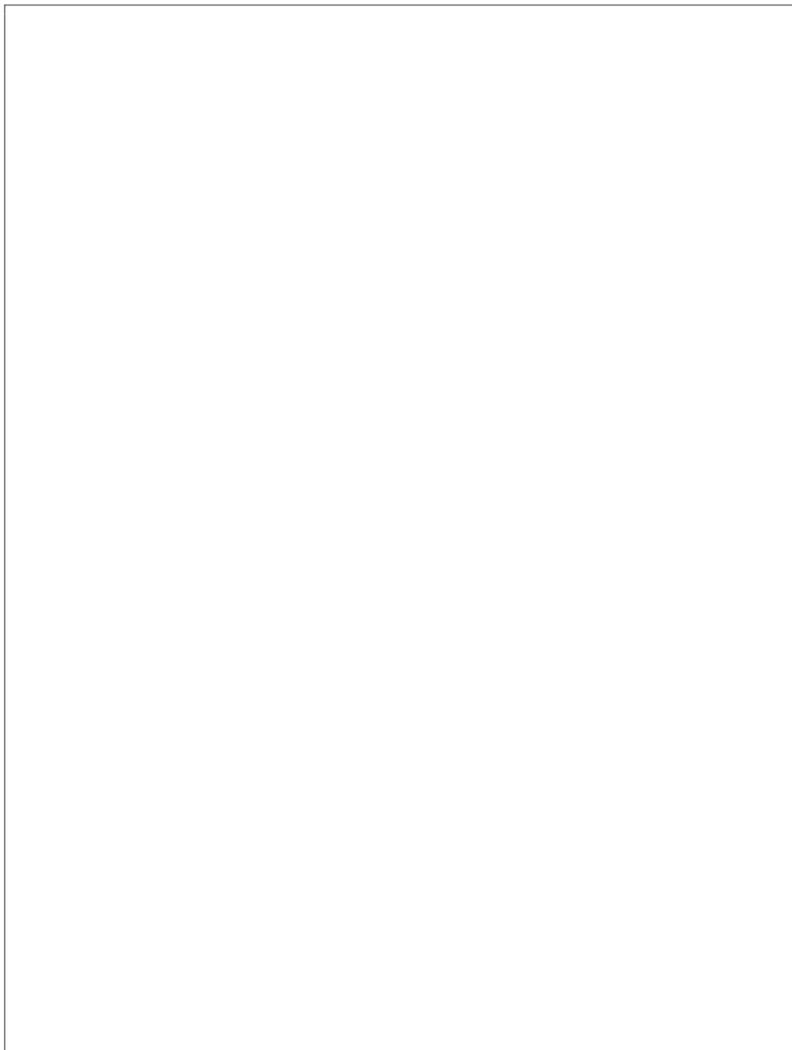
 **ZeG** Zelfrapportage voor
Getuigen

Ga hier verder:



 **ZeG** Zelfrapportage voor
Getuigen

Ga hier verder:



ZeG Zelfrapportage voor Getuigen

ONDERDEEL B: Wie beging het misdrijf?

Het is belangrijk dat u de volgende informatie leest en de instructies volgt.

Geef op de volgende pagina zo veel mogelijk fysieke details over de dader (de persoon die het misdrijf beging of die bij het incident betrokken was). Wanneer er meer dan één dader was, begin dan met de persoon waarop u het beste zicht had of die u zich het duidelijkst kunt herinneren.

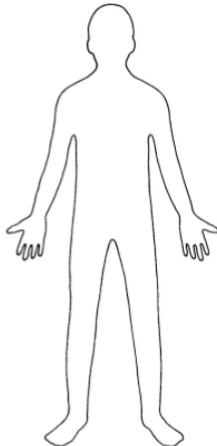
Als u in dit formulier al ergens een omschrijving heeft gegeven, neem dan de tijd om na te denken of er meer details zijn die u zich nu kunt herinneren.

Wanneer mogelijk, geef de volgende informatie van elke persoon. **Ga niet gokken** over dingen waarover u niet zeker bent.

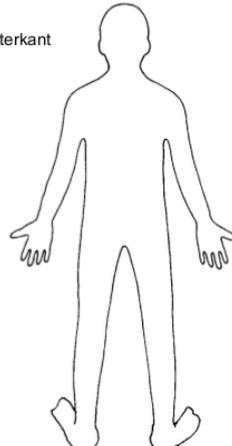
- Geslacht
- Leeftijd
- Lengte
- Etnische achtergrond
- Gewicht / Lichaamsbouw
- Ogen / Oren / Mond / Neus / enz.
- Haarkleur
- Gezichtsbehaarung
- Gelaatskleur
- Kleding / Schoenen
- Accent
- Bril
- Juwelen
- Accessoires
- Littekens / (Moeder-)Vlekken / Tatoeages
- Andere details van de dader(s) waar we niet naar hebben gevraagd?

U kunt onderstaande figuren gebruiken om extra informatie aan te geven. U kunt op de figuren schrijven en/of tekenen om bijkomende details of informatie te geven (bijvoorbeeld positie van potentiële accessoires).

Voor kant



Achterkant



 **ZeG** Zelfrapportage voor
Getuigen

Geef hier de omschrijving(en):

 **ZeG** Zelfrapportage voor
Getuigen**ONDERDEEL C: De situatie**

Het is belangrijk dat u de volgende informatie leest en de instructies volgt.

Een schets maken van de situatie kan u helpen om zich meer details te herinneren en het kan verdere informatie over het incident opleveren.

U kunt de ruimte hieronder en/of op de volgende pagina gebruiken om de situatie zo te schetsen als u zich die herinnert.

U dient details in te voegen over waar u was en waar andere personen waren. U kunt pijlen gebruiken om de bewegingen van u en de anderen die u gezien en gehoord heeft, aan te geven. U kunt ook pijlen gebruiken om de bewegingsrichting aan te geven van voertuigen of andere objecten als u die gezien heeft.

Verder kunt u ook aantekeningen en notities in uw schets gebruiken om eigenschappen van de situatie aan te geven, of om aan te geven dat u ergens niet zeker over bent.

**Dit is geen test van uw tekenkunsten – We zijn enkel geïnteresseerd in de situatie.
Bijvoorbeeld, wát u zag en wáár.**

 **ZeG** Zelfrapportage voor
Getuigen

Schets hier:



ZeG Zelfrapportage voor Getuigen

**ONDERDEEL D: Waren er andere personen aanwezig die gezien hebben wat er is
gebeurd?**

JA (Vul dit onderdeel in)

NEE (Ga door naar onderdeel E)

Het is belangrijk dat u de volgende informatie leest en de instructies volgt.

Geef een omschrijving van iedereen die aanwezig was en die ook gezien kan hebben wat er is
gebeurd maar die niet betrokken was bij het incident (andere potentiële getuigen).

Als u in dit formulier al ergens een omschrijving heeft gegeven, neem dan de tijd om na te denken of
er meer details zijn die u zich nu kunt herinneren.

Start hier:

ZeG Zelfrapportage voor Getuigen

ONDERDEEL E: Waren er voertuigen betrokken?

JA (Vul dit onderdeel in)

NEE (Ga door naar onderdeel F)

Het is belangrijk dat u de volgende informatie leest en de instructies volgt.

Geef zo veel mogelijk details over elk voertuig dat bij het incident betrokken was.

Als u al een omschrijving heeft gegeven in dit formulier, denk dan even na of er verdere details zijn die u zich nu kunt herinneren.

Wanneer mogelijk, geef de volgende informatie over elk voertuig. **Ga niet gokken** over dingen waarover u niet zeker bent.

- Grootte
- Merk / Model
- Rijstijl
- Vorm
- Aantal deuren
- Snelheid
- Kleur
- Nummerplaat
- Andere details?

Start hier:

 **ZeG** Zelfrapportage voor
Getuigen

Ga hier verder:



ZeG Zelfrapportage voor Getuigen

ONDERDEEL F: Hoe heeft u het incident waargenomen?

Beantwoord de volgende vragen

- Hoe lang heeft u het incident waargenomen?
- Wat waren de weersomstandigheden ten tijde van het incident?
- Op welk moment van de dag vond het incident plaats?
- Heeft u het incident waargenomen bij daglicht of kunstlicht? (Beschrijf indien mogelijk.)
- Was er sprake van permanente of tijdelijke obstakels in uw zicht? (Beschrijf indien mogelijk.)
- Zijn er bijzondere oorzaken waardoor u het incident of de dader(s) heeft onthouden? (Beschrijf indien mogelijk.)
- Was er een bekende van u bij betrokken, of iemand die u eerder gezien heeft? (Zo ja, waar en wanneer?)

 **ZeG** Zelfrapportage voor
Getuigen

ONDERDEEL G: Is er verder nog informatie die u wilt vermelden over het incident waarover we u nog niets hebben gevraagd? Zo ja, noteer deze informatie in de ruimte hieronder.

Start hier:

GEEF DIT FORMULIER TERUG AAN:

CHAPTER 2 **The Modality Effect in Eyewitness Accounts**

This chapter is an adapted version of the following article:

Sauerland, M., Krix, A. C., Kan, N., Glunz, S., & Sak, A. (2014). Speaking is silver, writing is golden? The role of cognitive and social factors in written versus spoken witness accounts. *Memory & Cognition*, *42*, 978-992. doi:10.3758/s13421-014-0401-6

Abstract

Contradictory empirical findings and theoretical accounts exist that are in favor of either a written or a spoken superiority effect. In this paper, we present two experiments that put the recall modality effect in the context of eyewitness reports to another test. More specifically, we investigated the role of cognitive and social factors in the effect. In both experiments, participants watched a videotaped staged crime and then gave spoken or written accounts of the event and the people involved. In Experiment 1, $N = 135$ participants were assigned to written, spoken-videotaped, spoken-distracted, or spoken-voice recorded conditions to test for the impact of cognitive demand and social factors in the form of interviewer presence. Experiment 2 ($N = 124$) tested the idea that instruction comprehensiveness differentially impacts recall performance in written vs. spoken accounts. While there was no evidence for a spoken superiority effect, we found some support for a written superiority effect for description quantity, but not accuracy. Furthermore, any differences found in description quantity as a function of recall modality could be traced back to participants' free reports. Following up with cued open-ended questions compensated for this effect, although at the expense of description accuracy. This suggests that current police practice of arbitrarily obtaining written or spoken accounts is mostly unproblematic.

While silence is certainly not golden when it comes to obtaining eyewitness evidence, the question of whether a witness should preferably speak or write when testifying is a more difficult one. When the police obtain information from eyewitnesses, they can ask for a written description of the course of events and the perpetrators or they can conduct a personal investigative interview. It appears that the modality of an eyewitness report depends on the seriousness of the crime and the importance of a witness for the case, with proceedings in civil cases often requiring written accounts and more serious crimes predominantly involving oral police interviews (Sauerland & Sporer, 2011). The literature on modality effects in various fields suggests, however, that whether an eyewitness report is given in writing or orally may have a significant effect on the amount and accuracy of the information obtained.

In an early study on the impact of modality on speech production, participants had to discuss one of two topics either in writing or orally (Horowitz & Newman, 1964). The results showed that spoken expression was more productive than written expression in terms of expressed ideas, expansion of previously stated ideas, but also in terms of irrelevant ideas, indicating that speaking was more productive, but somewhat less efficient than writing. Similarly, Kellogg (2007) found that spoken renarration of a story was more complete and more accurate (proportion correct), but elicited more distortions than written renarration. In a survey of medical history, Bergmann, Jacobs, Hoffmann, and Boeing (2004) found that participants underreported some diseases in a written questionnaire they had previously reported in a personal interview.

In the eyewitness field, only two studies have addressed the modality issue so far.⁶ The first one (Bekerian & Dennett, 1990) presented participants with 16 color slides that depicted a visual narrative of a car accident. Consistent with the literature discussed above, spoken reports were more detailed and more accurate than written reports. More recently, Sauerland and Sporer (2011) tested participants' memory of a staged theft presented to them in a video fragment. Analyses of participants' crime and perpetrator descriptions revealed a clear advantage of spoken descriptions, both in terms of description quantity and accuracy.

This spoken superiority effect can be explained by means of physical, cognitive, and social factors. Compared to writing, speaking demands less muscular energy, is acquired earlier in life (Horowitz & Newman, 1964), is more practiced, and does not require the activation of graphemic representations for spelling words (Kellogg, 2007). Furthermore, speaking puts fewer demands on working memory in the sense that it is faster, and thus ideas have to be stored in memory for a shorter period of time before they are expressed. Finally, the speaking condi-

⁶ Another early witness study on the modality effect exists that found no effect (Lipton, 1977). However, neither effect sizes, nor means and standard deviations to establish them were reported to assess the statistical power of the tests conducted. Furthermore, tests of interactions with the other independent variables investigated (sex, retention interval between encoding and recall, structure of recall) were missing. Hence, no conclusions can be drawn from this study (cf. Sauerland & Sporer, 2011).

tions in the described studies involved the presence of an interviewer, which may have a facilitating effect by improving motivation to perform and by providing prompts and encouraging non-verbal cues (Bergmann et al., 2004; Rosenthal, 2002).

A different line of research, however, challenges the idea of a *spoken superiority* effect and rather suggests a *written superiority* effect. Contrary to Kellogg (2007), Grabowski (2007) argues that speaking, not writing, puts higher demands on working memory, such that it impedes output monitoring, lacks self-pacing, and is associated with a larger ratio of produced information units per time interval, thus increasing cognitive load. These assumptions were tested in a series of experiments with different stimulus materials. In an attempt to test the impact of each of the three factors, four conditions were employed, namely written, invisible written, spoken-voice recorded, and spoken. While writing allows for monitoring previously produced information, control over the produced information units per time interval as well as self-pacing, invisible writing disables monitoring. Voice recording additionally disables control over time per unit and normal speaking also disables self-pacing. Note that participants in the voice recording condition were allowed to pause recording, but not to rewind and revise. No differences across conditions were found when recalling European states and capitals (Experiment 1), while the two written conditions outperformed the two spoken conditions when participants recalled 40 objects they had studied earlier (Experiment 2). The results of a third experiment in which participants watched a filmed theft somewhat resembled those of Horowitz and Newman (1964), with speaking participants reporting more episodes in total, but also more repetitions. When eliminating repetitions, the difference between the two groups in terms of number of episodes recalled disappeared. Writing, however, elicited fewer errors than speaking. Again, no differences occurred within the two written and the two spoken conditions, indicating that the smaller ratio of produced information units per time interval might be the most crucial factor driving the written superiority effect found here (Grabowski, 2007). Taken together, these results speak to the idea that writing imposes less cognitive load than speaking.

Note, however, that this account remains silent as to the possible facilitating or inhibiting influences of interviewer-presence. While Bergmann et al. (2004) emphasized possible positive effects, Wagstaff et al. (2008) focused on potential negative effects. Specifically, they proposed a cognitive-neuropsychological model of social inhibition that postulates that the presence of others places demands on the frontal and executive systems, including working memory, source monitoring, and cognitive inhibition (Kane & Engle, 2002; Mitchell, Johnson, Raye, & Green, 2004). These systems determine recall success (Engle & Kane, 2004). If the mere presence of others increased cognitive load, this would result in decreased processing capacity for recall. Wagstaff et al. (2008) tested this idea in an experiment implementing three different conditions: interviewer only (control), interviewer

and one observer, and interviewer and two observers. In agreement with the postulated model, the number of correct responses was significantly smaller in the two experimental conditions than in the control condition, and participants in the two-observers condition performed worse than participants in the one-observer condition. Although Wagstaff et al. did not include a no-interviewer condition, such a condition should be superior to conditions with interviewers according to the model.

To summarize, different explanatory accounts exist in favor of both a spoken and a written superiority effect and both have been supported by empirical evidence. It is the aim of the current paper to put the recall modality effect in the context of eyewitness testimony to another test, and to investigate possible underlying mechanisms. In Experiment 1, we investigated the role of cognitive demand and interviewer presence in written, spoken-voice recorded, spoken-distracted, and spoken-videotaped conditions. Following Grabowski (2007), cognitive load should be lowest in the written condition (output monitoring, self-pacing, and control over output production per time interval), followed by the spoken-voice recorded (only self-pacing), the spoken-videotaped, and the spoken-distracted conditions (additional cognitive load). Accordingly, recall performance should be expected to decrease in this order.

If, however, writing puts higher demands on working memory than speaking (Kellogg, 2007), the written condition should be inferior to the spoken-videotaped condition, while no clear predictions can be made for the relation between the written and spoken-distracted condition. According to this account, performance in the spoken-voice recorded condition should be comparable to the spoken-videotaped condition. Taking into account the possible role of an interviewer (Wagstaff et al., 2008), however, the spoken-voice recorded condition should outperform the spoken-videotaped condition. More specifically, if the presence of others increases cognitive load, the written and the spoken-voice recorded conditions should outperform the spoken-videotaped and spoken-distracted conditions.

We also tested the impact of different levels of executive functioning (so to speak an individual threshold that marks the point when cognitive load is experienced as high) on performance in written relative to spoken statements. For this purpose, we administered several tests that measure different aspects of executive functioning, such as working memory capacity, source memory, and cognitive inhibition. We expected that executive functioning should predict recall performance to a greater extent in conditions imposing higher rather than lower cognitive load, because when tasks are less demanding, participants can fully allocate resources to the recall task at hand. In contrast, when performing a recall task with high cognitive load, individuals with higher relative to lower executive functioning should be an advantage.

EXPERIMENT 1

Method

Participants

One-hundred-and-thirty-five women⁷ (age: $M = 21.0$, $SD = 2.3$, range: 18-28 years) participated in return to course credit or a €10 voucher. One participant was excluded, because it was unclear which film condition she had been assigned to. Participants were Bachelor (84.4%) and Master students (14.1%), or were members of the general public (1.5%). All participants were tested in their mother tongue (Dutch: $n = 87$; German: $n = 48$). The study was approved by the local ethical committee.

Design

Participants were randomly assigned within a 4 (interviewing condition: written vs. spoken-voice recorded vs. spoken-videotaped vs. spoken-distracted) x 4 (film version: 1 through 4) between-participants design. The number of Dutch and Germans was counterbalanced across conditions.

Materials

Stimulus films

To avoid a possible influence of features innate to the actor in the perpetrator-role (e.g., facial distinctiveness, nature of clothing, or number of clothing items), we created four different film versions using the same four female actors, rotating their roles. Each film version was edited to last approximately 3:20 min. The action in all films can be described as follows:

Two women (the later thief and accomplice) meet in a bar and order drinks. Then the later victim enters and takes a seat at the bar. While talking to the bar-keeper, the thief inspects the bag of the victim and pushes it off a stool. Either the thief or the victim picks it up, depending on the film version, and puts it back onto the stool. The thief walks back to the table. Eventually, the thief and the accomplice approach the bar to pay their drinks. While the accomplice is distracting the victim by paying the bill, the thief steals the victim's wallet from her bag. The thief and the accomplice leave. When the victim wants to pay, she cannot find her wallet.

⁷ Only females were included for testing an unrelated research question.

*Tasks measuring executive functioning**Operation span (Ospan) task*

The Ospan task is a complex span task to measure working memory capacity (Engle, Cantor, & Carullo, 1992). That is, participants are required to pursue a secondary task (solve arithmetic problems), while remembering words. Specifically, participants are presented with equation-word pairs (e.g., “Is $(10/5)-3=2?$ PAINT”). After reading out the equation and determining its accuracy, participants read out the to-be-remembered word. Following each trial (ranging from two to five equation-word pairs), participants are prompted to write down the to-be-remembered words. In total, there are twelve trials. According to the partial-credit unit scoring (Conway et al., 2005), a correctly recalled word is considered a correct response, irrespective of whether the word was recalled in the correct order. Then the accuracy across trials is calculated.

Source monitoring tests

We used two different source monitoring tests (see Unsworth & Brewer, 2010a) that were taken from another study (see chapter 6). In the picture source recognition test, participants are shown 30 pictures which appear for 1 s one at a time in one of four quadrants on screen. At test, participants are presented with 30 old and 30 new pictures. They indicate whether a picture is old or new. If considered old, they are asked in which quadrant it appeared.

In the gender source recognition test, participants hear 30 English one-syllable nouns that are spoken by a female or male speaker. At test, participants are presented with 30 old and 30 new words. Participants indicate, whether a word is old or new. If considered old, they specify whether it was spoken by the male or female speaker. The scores are the proportions of correct responses.

Random number generation (RNG)

In the RNG task that measures cognitive inhibition (Ginsburg & Karpiuk, 1994, 1995), participants randomly generate numbers ranging from 0 to 9 at a pace of one number per second, as indicated by a metronome. Scoring was based on the indices described by Peters, Giesbrecht, Jelacic, and Merckelbach (2007). We will focus on repetition (i.e., identical pairs; e.g., 2, 2), seriation (i.e., consecutive digrams; e.g., 1, 2), poker (i.e., repetitions within 20 sequences of 5 successive responses), and variance of digits.

Forward and backward digit span

As simple span measures that do not include a secondary task (Unsworth & Engle, 2007), we employed the forward and backward digit span tasks (Wechsler, 1997). Here, strings of digits are read out by the experimenter. The string length increases, ranging from three to eight and from two to seven in the forward and backward

versions, respectively. Participants have to repeat the strings, either in the same (forward span) or reversed order (backward span). The tasks are stopped when two consecutive errors occur. The length of the last correctly recalled string constitutes the score.

Distractor task

A distractor task was used to increase cognitive load during retrieval in the spoken-distracted condition. Participants were shown a presentation on a screen that consisted of 90% green and 10% red balls. They were instructed to put up their hand every time a red ball was presented. When they forgot to raise their hand, the experimenter reminded them to do so.

Procedure

To ensure that participants were unaware of the fact that they were expected to act as witnesses they were told that it would be their task to judge social situations. After giving consent, participants watched one of four stimulus films and were instructed to pay close attention as they would be questioned about it afterwards. If participants recognized one of the actors, they were debriefed and excluded. Otherwise, participants proceeded with the executive functioning tasks, followed by the free recall (FR) instructions regarding the sequence of events. Specifically, participants were asked to report everything they could remember about the actions and surroundings as completely, detailed, and accurately as possible. Participants were discouraged from guessing. Thirteen cued open-ended questions (CQ) followed (see Appendix C). Next, participants described all persons (thief, accomplice, victim, and barkeeper). For each person, participants received FR instructions first. Again, participants were asked to be as complete, detailed, and accurate as possible. The description should be so specific that the described person could be recognized in a crowd. Again, guessing was discouraged. After all four FRs had been completed, participants answered 12 more CQs about the appearance of each actor (see Appendix D). Finally, participants were thanked for their participation and debriefed via e-mail after testing had been concluded.

Coding

For coding the quantity (sum of correct, incorrect, and confabulated details) and accuracy (number of correct details divided by quantity) of participants' descriptions, different coding schemes were developed for each film version. For each of the four films, two of three trained coders (ABC) independently coded all details reported by 10 participants (i.e., 40 statements in total). Specifically, 10 statements

each were coded by two coders referring to film 1 (coders A and B), 2 (AC), 3 (BC), and 4 (AC).⁸

Details were coded as correct or incorrect, if they did or did not match the content of the stimulus film, respectively. Details were considered confabulated when they were both incorrect and non-existent (e.g., describing a hat in the absence of headgear; see Dando, Wilcock, & Milne, 2009, for a similar approach). A statement such as “the woman (1) wore a black (2) t-shirt (3)” yielded three details. Although we did not explicitly code for the precision of details, our coding scheme was so extensive that it accommodated for responses of varying levels of precision (e.g., dark vs. navy blue). Subjective details (e.g., ‘beautiful shirt’) were not coded. If participants refrained from responding to a cued question (i.e., a “don’t know” response), this was accepted as such (after all, participants were instructed not to guess). It was regarded as a sign that the participants exerted their report option (i.e., the freedom to withhold details; Koriat & Goldsmith, 1996) and not considered an omission error. To code the accuracy of age, height, and weight estimates, we accepted deviations of 2 years, 4 cm, or 3 kg from the true value. If an interval was given (mostly in FRs) and the range was smaller or analogous to the ranges we used (4 years, 8 cm, or 6 kg) the answer was coded as correct. If the range was larger, however, the answer was coded as incorrect. We are aware that this approach is not in line with grain size theory (Goldsmith, Koriat, & Pansky, 2005), according to which any interval, irrespective of its width, would be considered correct as long as it contains the true value. We chose our approach for the following reasons. First, when a very wide interval is given, this may be uninformative at best, but adverse at worst, as it entails that a wider range of innocent people could become the focus of the investigations. Second, our coding procedure serves to achieve higher consistency when scoring point versus interval estimates. If, for example, the true age is 30, a point estimate of 20 would be coded as incorrect. It is not reasonable why an interval estimate of 20 to 40 should be coded as correct. Third, since the instructions required participants to provide point values, interval estimates were provided very seldom. As a result, the precise nature of the coding procedure of the interval estimates is unlikely to influence our overall results.

⁸ In hindsight, we realized that the coders were not formally blind to the conditions in this experiment as they could infer the spoken-videotaped or spoken-distracted conditions, if an interviewer remark was transcribed or if the transcript contained fragments that could be attributed to spoken language. The written vs. spoken-voice recorded conditions and the spoken videotaped vs. spoken distracted conditions were much less distinguishable. We changed our procedure for Experiment 2 to ensure coders’ blindness. Although the procedure in Experiment 1 was suboptimal, the replication of the general pattern of results in Experiment 2 seems to suggest that there is no reason to believe that the procedure employed in Experiment 1 compromised the results.

Inter-coder reliability

Inter-rater reliability was found to be *substantial* to *almost perfect* (Landis & Koch, 1977). Specifically, for correct recall, Cohen's κ were .89, .83, .89, and .87 for Films 1 to 4, $ps < .001$, respectively. For incorrect recall, κ coefficients were .89, .77, .93, and .75, $ps < .001$, respectively.

Results

For both experiments, we report Cohen's d (Cohen, 1988) for main effects with $df = 1$ in the numerator, and η_p^2 for main effects with $df > 1$ in the numerator and interaction effects (see Sporer & Cohn, 2011). To investigate recall performance as a function of modality (written vs. spoken-voice recorded vs. spoken-videotaped vs. spoken-distracted) and film version (1 through 4), we calculated two-way ANOVAs. Since there were no significant interactions between film version and modality, we collapsed data across film versions for the following analyses, $F_s \leq 1.51$, $ps \geq .153$, $\eta_p^2s \leq 0.10$. Thus, we computed one-factorial ANOVAs with modality as independent variable and description quantity and accuracy as dependent variables.

An alpha level of .05 was applied for the statistical tests of main effects and interactions. For pairwise comparisons of the modality groups, we made use of an adjusted Bonferroni correction (Shaffer, 1986) and set the critical $\alpha = .017$. Note that post-hoc contrasts that are not mentioned were non-significant. Results are reported for FRs and the combination of FR and CQ (FR-CQ). Table 2.1 displays the mean description quantity and accuracy observed for FRs and FR-CQs.

Quantity and Accuracy of Event Descriptions

We found a significant main effect of modality on FR quantity, $F(3, 130) = 3.67$, $p = .014$, $\eta_p^2 = 0.08$. Post-hoc test analyses revealed that written FRs were more detailed ($M = 88.88$) than spoken-voice recorded statements ($M = 69.50$), $p = .002$, $d = 0.80$, and tended to be more detailed than spoken-distracted statements ($M = 76.06$), $p = .037$, $d = 0.54$. Spoken-videotaped statements also tended to contain more details ($M = 81.73$) than spoken-voice recorded statements, $p = .047$, $d = 0.47$.

For FR-CQs, the main effect of modality on quantity was marginally significant, $F(3, 130) = 2.48$, $p = .064$, $\eta_p^2 = 0.05$. Written statements were more detailed ($M = 101.52$) than spoken-voice recorded statements ($M = 85.53$), $p = .011$, $d = 0.65$. Modality had no impact on the accuracy of FRs and FR-CQs describing the event, $F_s(3, 129) \leq 0.62$, $ps \geq .324$, $\eta_p^2s \leq 0.01$.

Table 2.1

Mean Event and Person Description Quantity and Accuracy as a Function of Modality Condition (Experiment 1)

		Written		Spoken-videotaped		Spoken-voice recorded		Spoken-distracted	
Event description		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>SD</i>	<i>SD</i>
Quantity	FR	88.88 ^{a°}	24.53	81.73 [†]	27.91	69.50 ^{††}	23.85	76.06 [°]	23.20
	FR-CQ	101.52 ^a	24.75	97.37	26.80	85.53 ^a	24.15	92.09	25.76
Accuracy (%)	FR	98.59	2.04	98.72	2.40	98.81	1.83	98.62	1.86
	FR-CQ	96.42	3.40	95.56	3.08	95.38	3.08	95.58	3.60
Person descriptions									
Quantity	FR	27.67	7.54	25.12	9.41	23.32	7.08	27.03	7.94
	FR-CQ	53.58	9.39	52.82	10.48	51.56	9.12	55.76	9.63
Accuracy (%)	FR	82.22	8.03	87.02	8.23	85.36	10.17	83.21	10.54
	FR-CQ	73.75	5.90	73.75	6.11	72.40	7.45	71.87	7.47

Note. FR = free recall; FR-CQ = combination of free recall and cued questions. Means sharing the same superscript letter within a row differ at $p < .017$. Means sharing the same superscript symbol differ at $p < .05$.

Quantity and Accuracy of Person Descriptions

Modality had no impact on the quantity or accuracy of FRs and FR-CQs referring to person descriptions, $F_s(3, 133) \leq 1.77, p_s \geq .156, \eta_p^2s \leq 0.04$.

Executive Functioning and Interview Performance

To reduce the number of predictors to be entered into the regression analyses, we ran a factor analysis (rotation method: Varimax). For this purpose, repetition, seriation, and poker scores were inverted so that higher scores would be associated with higher executive functioning. Using the Kaiser criterion, this yielded a four factor solution that explained 65.75% of the variance. The factor loadings can be found in Table 2.2.

Table 2.2

Factor Loadings for Factor Analysis of Measures of Executive Functioning for Experiments 1 and 2

	Experiment 1				Experiment 2	
	Factor 1	Factor 2 (WMC)	Factor 3 (SM)	Factor 4	Factor 1 (WMC)	Factor 2 (SM)
(A)Ospan	-.09	.65	.05	.42	.59	.09
Forward span	.10	.68	-.01	-.04	.85	.01
Backward span	< .01	.82	.03	-.08	.73	-.01
Picture source	.13	.01	.82	-.07	.07	.82
Gender source	-.14	.03	.79	.12	.01	.81
Repetition (RNG) inverted	.93	-.04	-.02	-.03		
Seriation (RNG) inverted	.08	.23	-.02	.64		
Poker (RNG) inverted	.93	.11	< .01	.02		
Digit variance (RNG)	-.09	-.29	.07	.70		

Note. WMC = working memory capacity; SM = source monitoring. In Experiment 1, the Ospan task was used, in Experiment 2, the AOsapn task. Factor loadings in bold indicate the highest factor loading for a given measure of executive functioning.

Although all four factors are relevant to eyewitness testimony, we could not enter all of them into the regression equation, as that would have required a much larger sample size (i.e., $N = 186$; Tabachnik & Fidell, 2007). Following theoretical considerations on the contribution of each factor to eyewitness memory, we selected factors 2 and 3 on which source and working memory capacity loaded the highest. Especially the latter has been found to be associated with both correct and incorrect recall (Unsworth & Brewer, 2010b). In contrast, cognitive inhibition (as measured by RNG and mainly loading on factors 1 and 4) is mostly associated with intrusions (but not correct recall; e.g., Peters, Jelicic, Haas, & Merckelbach, 2006). Hence, we considered our factor selection to yield a more complete picture of the relationship between modality, executive functioning, and recall performance than any other possible selection, taking sample size restraints into account. Note, however, that additional tentative analyses with the scores of factors 1 and 4 yielded no significant effects, F -changes ≤ 1.66 , $ps \geq .179$, $F_s \leq 1.43$, $ps \geq .228$.

To code for modality, we added dummy variables to the regression equation. We calculated the regression equations, once determining the spoken-videotaped group and once the written group as control group. Note that while this approach does not influence the results of the regression analyses in terms of interaction and main effects, it allows for pairwise comparisons between the interview conditions. The predictors were entered into the regression in two blocks. To analyze the interaction with the source monitoring factor, in the first block, all main effects and the interactions involving the working memory capacity factor were entered, followed by the interaction terms of the source monitoring factor. To analyze the

interaction with the working memory capacity factor, this was done the other way around (i.e., for every dependent variable, two regression analyses were conducted). When the interactions were non-significant, we re-ran the analyses with the main effects only. As dependent variables, we again used description quantity and accuracy. Analyses were run separately for event and person descriptions. For the sake of brevity, we report the FR-CQ analyses only.

Event descriptions

No significant interaction effects were found between modality and the source monitoring factor, F -changes ≤ 0.67 , $ps \geq .570$. That is, the relationship between source monitoring and recall performance (quantity and accuracy) was not moderated by modality. However, as one would expect, a higher source monitoring factor score was associated with higher event description quantity and accuracy ($R^2 = .12$, $F[5, 120] = 3.30$, $p = .008$, $\beta = .26$, $p = .004$; $R^2 = .18$, $F[8, 117] = 3.25$, $p = .002$, $\beta = .34$, $p < .001$).

For the working memory capacity factor, there was no significant interaction with modality, F -change = 2.39, $ps \geq .073$, and no main effect ($R^2 = .12$, $F[5, 120] = 3.30$, $p = .008$, $\beta = .11$, $p = .192$) for description quantity.

The interaction between modality and the working memory capacity factor score on the accuracy of the reported event details was significant. However, comparisons of all other interviewing conditions with the spoken-videotaped group yielded no significant interaction. When the written group was determined as the control group, the interaction term of the working memory factor score and the spoken-voice recorded group was significant ($\beta = -.28$, $p = .022$). The post-hoc analyses indicated a non-significant relationship for the written group ($R^2 = .03$, $F[1, 29] = 0.96$, $p = .335$, $\beta = .18$), and a significant negative relationship for the spoken-voice recorded group, indicating that a lower working memory factor score was associated with higher accuracy ($R^2 = .18$, $F[1, 30] = 6.70$, $p = .015$, $\beta = -.43$). Table 2.3 displays the referring statistics.

Person descriptions

For the person descriptions, there were no significant interactions between modality and the source monitoring or the working memory capacity factors, F -changes ≤ 1.05 , $ps \geq .376$. Models without interaction terms also returned no significant results, $F_s \leq 2.11$, $p = .068$.

Table 2.3

Interaction between Modality and Working Memory Capacity Factor Score for Event Descriptions (Regression) of Experiment 1

	Event Accuracy	
R^2	.20	
F -change	2.77	
p	.045	
	WMC (control group = spoken- videotaped)	WMC (control group = written)
B	0.01	0.01
$SE B$	0.01	0.01
β	.15	.20
p	.467	.219
	Interaction WMC x written	Interaction WMC x spoken-videotaped
B	< .01	< -.01
$SE B$.01	.01
β	.03	-.02
p	.837	.837
	Interaction WMC x spoken-voice recorded	Interaction WMC x spoken-voice recorded
B	-0.02	-0.02
$SE B$.01	.01
β	-.25	-.28
p	.068	.022
	Interaction WMC x spoken-distracted	Interaction WMC x spoken-distracted
B	< 0.01	< 0.01
$SE B$.01	.01
β	.05	.03
p	.680	.815

Note. WMC = working memory capacity.

Discussion

In Experiment 1, we investigated the role of cognitive demand and interviewer presence in the recall modality effect in eyewitness performance. To this end, four different interviewing conditions were tested. Two conclusions can be drawn from the findings: First, if anything, our results provide support for a written superiority effect (Grabowski, 2007) for description quantity, but not accuracy. Specifically, written FRs describing the event were more detailed than spoken-voice recorded reports and tended to be more detailed than spoken-distracted reports. Additionally, when looking at the means of the remaining non-significant comparisons, there was a tendency of the written condition to elicit the most detailed reports, compared to all spoken conditions. Note, however, that not a single comparison between the standard written and spoken conditions (i.e., written vs. spoken-videotaped) produced a significant result.

Second, on a descriptive level, the spoken-voice recorded condition consistently elicited the least detailed reports. This contradicts the cognitive-neuropsychological model of social inhibition which postulates that the presence of an interviewer should adversely affect recall performance (Wagstaff et al., 2008). According to this account, the spoken-voice recorded condition should have outperformed the spoken-videotaped condition. Possible *positive* effects of interviewer-presence might be able to explain the results. For example, interviewers may prompt witnesses to make continued efforts to retrieve more details (Bergman et al., 2004; Sauerland & Sporer, 2011). It is difficult, however, to explain the current results with a lack of social interaction alone, given that written statements were also made without social interaction. Although speculative, it could also be that students generally lack dictating experience, causing uneasiness while making statements in the spoken-voice recorded condition (Gould, 1978; Gould & Boies, 1978). It would be interesting to include a spoken-voice recorded group that is more practiced in the use of voice recorders to test this notion.

The interactions between executive functioning measures and modality were mostly non-significant, making it difficult to conclusively assess the impact of executive functioning as a function of recall modality. The only significant result we obtained in this regard was contrary to earlier findings. Usually, lower working memory capacity is associated with more recall errors (Engle & Kane, 2004). However, we found lower working memory capacity to be associated with higher recall accuracy within the spoken-voice recorded group. Note, however, that we did not find this pattern for any of the other groups. We have no ready explanation for this unanticipated finding and refrain from speculating about it, as it might simply constitute an outlier.

Further exploring the modality effect in witness statements, in Experiment 2, we sought to examine the effect of instruction comprehensiveness on recall performance in written vs. spoken accounts. Differences in the employed instructions

could also serve as an explanation for the contradictory results found in Experiment 1 and by Sauerland and Sporer (2011) regarding the modality effect. Although both studies used FR instructions, Sauerland and Sporer's (2011) were less specific. For the event descriptions, they only instructed participants to describe the crime as "detailed as possible". In Experiment 1, however, we gave participants more guidance by asking to give a "complete, detailed, and accurate" description of "actions, events, and the surrounding". Guessing was discouraged. For the description of the perpetrator, the participants of both studies were instructed to give person descriptions that were specific enough to identify the referring persons from a crowd. Nevertheless, in Experiment 1 we added the terms "complete", "detailed", and "accurate" and again discouraged guessing. Possibly, these more specific instructions gave participants the cues they needed to recall more relevant information, thereby eliminating the difference between the two modalities found earlier. Experiment 2 followed up on this idea by varying the supportive nature of the recall instructions for written and spoken-videotaped modality conditions. While the scarce instructions merely included the prompt to describe all details that came to mind about the event and the appearance of the persons involved, the comprehensive instructions contained several components that are known to support retrieval and facilitate recall. First, borrowing elements from rapport-building (e.g., Vallano & Schreiber Compo, 2011), the instructions transferred control to the participants and pointed out that only they have the crucial information that is needed for solving the crime. Second, using the "report-everything" component that, as rapport-building, also comes from the Cognitive Interview (Fisher & Geiselman, 1992), the comprehensive instructions invited an account that was as complete and accurate as possible. Third, in order to reduce cognitive load, it was pointed out that recall order did not matter and that details should be recalled as they come to mind. Finally, the comprehensive instructions contained non-leading recall cues, prompting participants to recall details referring to actions, objects, surrounding, facial details, clothing etc. (see Gabbert, Hope, & Fisher, 2009). The idea was that these components should facilitate recall and free resources for retrieval. We hypothesized that comprehensive instructions should be especially beneficial under recall conditions that are highly demanding. If the differences between the results of Sauerland and Sporer (2011) and Experiment 1 originate primarily in differences in the instructions, one would expect that comprehensive instructions would be more beneficial for writing participants than for speaking ones. If, however, writing is less demanding than speaking, as suggested by Grabowski (2007) and the results of Experiment 1, the beneficial value of the comprehensive instructions should be greater for speaking than for writing participants. Furthermore, we expected comprehensive instructions and less demanding description conditions to elicit more detailed reports than scarce instructions and more demanding description conditions.

EXPERIMENT 2

Method

Participants

One-hundred-and-twenty-four participants (95 women; age: $M = 21.5$, $SD = 4.0$, range: 18-57 years) took part in exchange for course credit or a €15 voucher. Participants were Bachelor (90.3%) and Master students (6.5%), or were members of the general public (3.2%). All participants were tested in their native language (Dutch: $n = 78$; German: $n = 46$). The study was approved by the local ethical committee. Four participants were excluded from analysis, because they were outliers recalling only very few details. This could be attributed to misunderstanding the instructions. Specifically, these participants erroneously understood that their task was to report only the beginning of the film (instead of the complete film). In fact, they were instructed to testify about the “film they had seen in the beginning”.

Design

Participants were randomly assigned within a 2 (modality: written vs. spoken) x 2 (comprehensiveness of instructions: scarce vs. comprehensive) x 2 (film version: 1 vs. 2) between-participants design. The number of Dutch and Germans was counterbalanced across conditions.

Procedure

The procedure of Experiment 2 was very similar to Experiment 1, with the following exceptions: Participants saw one of two film versions (films 1 and 2 from Experiment 1), rather than one of four. After watching the stimulus film, participants performed the same executive functioning tasks, however, the RNG task was omitted and the Ospan task was replaced with the AOspan task for practical reasons (Unsworth, Heitz, Schrock, & Engle, 2005). Participants received the following instructions for the event description. The instructions provided in the written and spoken conditions were identical.

Scarce: “We would now like to ask you to answer some questions about the film you saw in the beginning. For this purpose, imagine being a witness reporting to the police about the incident shown in the film. If you have any questions, please turn to the experimenter. Please describe all details about the incident that you can remember. Do not guess about details that you cannot remember.”

Comprehensive: “We would now like to ask you to answer some questions about the film you saw in the beginning. For this purpose, imagine being a witness

reporting to the police about the incident shown in the film. Note that only you as a witness have valuable first-hand information that can contribute to solving the case. If you have any questions, please turn to the experimenter. Please describe all details about the incident that you can remember. For this purpose, think of all the persons involved, the actions and events, but also of objects and their positions as well as of the surrounding. Mention things as you remember them. It does not matter whether you remember the details in a different order than the one in which they occurred. Your description should be as complete, detailed, and precise, but also as accurate as possible. Do not leave out any details, but do not guess about details that you cannot remember.”

The person descriptions were as follows (depending on who should be described, thief was replaced with accomplice, victim, or barkeeper):

Scarce: “Please describe all details that you can remember about the thief’s appearance and clothing. Do not guess about details that you cannot remember.”

Comprehensive: “Please describe all details that you can remember about the thief’s appearance and clothing. Your description should be as complete, detailed, and precise, but also as accurate as possible. For this purpose, think of facial details, hair/hairstyle, build, and clothing. Your description should be so specific that reading the description would enable someone to pick the described person from a crowd. Do not leave out any details, but do not guess about details that you cannot remember.”

Automated Ospan Task (AOspan)

The automated version of the Ospan task (Unsworth et al., 2005) is a computerized version of the Ospan task (Engle et al., 1992) used in Experiment 1. Here, participants are presented with a set of to-be-memorized letters after solving an arithmetic problem. At test, the letters have to be recalled in the order of presentation by clicking on the appropriate letters. The partial-credit unit scoring was used to obtain participants’ scores.

Coding

The coding procedure was analogous to Experiment 1. To ensure coders’ blindness to the conditions, the transcriptions were cleaned by a person uninvolved in the coding process, so that no inferences about the conditions could be made (e.g., deleting clarifications made by the interviewer if the participants had a question or removing hesitation sounds like “uhm”).

Inter-coder reliability

For establishing inter-rater reliability, 12 randomly selected statements were coded by three independent coders. Following the criterion of Landis and Koch (1977), inter-rater reliability was *almost perfect*. Specifically, Fleiss' (1971) κ was 0.92 for correct details and 0.84 for incorrect details.

Results

To investigate recall performance as a function of modality (written vs. spoken), instructions (scarce vs. comprehensive), and film version (Film 1 vs. Film 2), we calculated three-way ANOVAs. When there were no significant interactions between film version and the other variables, we report analyses collapsed across film versions. Table 2.4 displays the mean description quantity and accuracy observed for FRs and FR-CQs for event and person descriptions. All non-reported main effects and interactions were non-significant.

Quantity and Accuracy of Event Descriptions

For FRs, the main effects of instructions, $F(1, 112) = 8.18, p = .005, d = 0.51$ was significant, indicating that comprehensive instructions elicited more details ($M = 83.57$) than scarce instructions ($M = 72.97$). The main effect of modality, $F(1, 112) = 11.18, p = .001, d = 0.58$, was qualified by a significant Modality by Film interaction, $F(1, 112) = 4.03, p = .047, \eta_p^2 = 0.04$. Specifically, in Film 1 there was no difference in FR quantity as a function of modality, $F(1, 116) = 0.69, p = .407, d = 0.21$, while in Film 2, written FR-statements ($M = 90.63$) were more detailed than spoken statements ($M = 71.13$), $F(1, 116) = 14.06, p < .001, d = 1.02$.

For FR-CQ quantity, the main effect of instructions, $F(1, 116) = 4.38, p = .039, d = 0.37$, was qualified by a marginally significant interaction between Modality and Instructions, $F(1, 116) = 3.90, p = .051, \eta_p^2 = 0.03$. Simple main effects analyses revealed no modality effect for scarce instructions (written: $M = 86.07$; spoken: $M = 86.81$), $F(1, 116) = 0.02, p = .894, d = -0.04$. With comprehensive instructions, however, writing ($M = 102.00$) elicited more details than speaking ($M = 87.27$), $F(1, 116) = 7.07, p = .009, d = 0.65$.

Event description accuracy of FRs and FR-CQs did not differ as a function of modality or instructions, and there was no interaction, $F_s \leq 2.11, p_s \geq .149, \eta_p^2_s \leq 0.01, |d|_s \leq 0.27$.

Table 2.4

Mean Event and Person Description Quantity and Accuracy as a Function of Modality and Instructions (Experiment 2)

		Written		Spoken-videotaped		Total		
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Event description								
Quantity	FR*	Scarce	76.07	22.11	70.06	14.07	72.97 ^a	18.48
		Comprehensive	92.47	23.01	74.67	19.53	83.57 ^a	22.98
		Total	84.41 ^b	23.85	72.33 ^b	16.99	78.27	21.44
	FR-CQ	Scarce	86.07	22.51	86.81	17.88	86.45 ^c	20.08
		Comprehensive	102.00 ^d	22.50	87.27 ^d	22.70	94.63 ^c	23.60
		Total	94.17	23.71	87.03	20.22	90.54	22.20
Accuracy (%)	FR	Scarce	95.14	4.36	93.65	4.68	94.37	4.55
		Comprehensive	95.25	3.25	95.63	3.25	95.44	3.23
		Total	95.20	3.80	94.62	4.13	94.90	3.97
	FR-CQ	Scarce	94.40	4.25	92.92	4.37	93.63	4.34
		Comprehensive	94.62	3.27	94.16	3.43	94.39	3.33
		Total	94.51	3.47	93.53	3.96	94.01	3.87
Person descriptions								
Quantity	FR	Scarce	27.34	8.75	26.29	7.65	26.80 ^e	8.15
		Comprehensive	31.47	7.68	30.60	7.91	31.03 ^e	7.74
		Total	29.44	8.41	28.41	8.02	28.92	8.19
	FR-CQ	Scarce	50.17	9.16	52.19	9.70	51.22	9.42
		Comprehensive	51.30	10.13	52.40	10.36	51.85	10.17
		Total	50.75	9.60	52.30	9.95	51.53	9.77
Accuracy (%)	FR	Scarce	84.05	8.57	81.67	10.17	82.82	9.43
		Comprehensive	81.61	9.75	83.72	7.66	82.67	8.76
		Total	82.81	9.19	82.68	9.01	82.74	9.06
	FR-CQ	Scarce	70.83	9.01	66.62	10.20	68.66	9.79
		Comprehensive	71.08	10.38	72.02	7.56	71.55	9.01
		Total	70.96	9.65	69.28	9.33	70.10	9.48

Note. FR = free recall; FR-CQ = combination of free recall and cued questions. Variables marked with an asterisk indicate a significant interaction with the film version. Means sharing the same superscript letter within a row or column indicate significant main effects or simple main effects with $p < .05$.

Quantity and Accuracy of Person Descriptions

For quantity of person descriptions obtained in the FR, only the main effect of instructions became significant, with comprehensive instructions leading to more detailed FRs ($M = 31.03$) than scarce instructions ($M = 26.80$), $F(1, 116) = 8.33$, $p = .005$, $d = 0.53$.

Person description accuracy established for FRs and FR-CQs did not differ as a function of modality or instructions, and there was no interaction, $F_s \leq 2.73$, $p_s \geq .101$, $\eta_p^2_s \leq 0.02$, $|d|_s \leq 0.30$.

Executive Functioning and Interview Performance

As in Experiment 1, we ran a factor analysis (rotation method: Varimax) on the executive functioning measures in order to reduce the number of predictors that would be entered into a regression analysis. Using the Kaiser criterion, this yielded a two factor solution (one source monitoring and one working memory capacity factor) that explained 58.79% of the variance (see Table 2.2 for the factor loadings). Note that these two factors precisely correspond to the working memory capacity and source memory factors obtained in Experiment 1.

These two factors, modality and instruction conditions, as well as the interaction terms were entered into separate regression equations (enter method) for person and event quantity and accuracy. Again, we report FR-CQ data only. When the interaction terms were non-significant, they were removed one at a time and the analyses rerun until only main effects or significant interaction terms remained. Table 2.5 shows the results of the final regression equations, after non-significant predictors were removed.

Event descriptions

Analogous to the above reported marginally significant interaction between modality and instructions, this interaction was significant (for details, see the ANOVA reported above). The regression equation regarding accuracy of the event descriptions was non-significant.

Person descriptions

The regression equation regarding quantity of the person descriptions was not significant. For person description accuracy, a higher source memory factor score was associated with increased accuracy.

Table 2.5

Results of Regression Analyses Predicting Recall Performance from Modality, Instructions, and Executive Functioning Factor Scores (Experiment 2)

Variable	<i>B</i>	<i>SE</i>	β	<i>t</i>	<i>p</i>	<i>R</i> ²	<i>F</i>
Quantity of event description						.10	2.40*
Modality	2.34	5.65	.05	0.42	.679		
Instructions	17.00	5.72	.39	2.97	.004		
WMC	2.22	2.06	.10	1.08	.284		
SM	0.71	2.04	.03	0.35	.728		
Modality x Instructions	-16.38	8.04	-.32	-2.04	.044		
Accuracy of event description						.03	0.99
Modality	-0.01	0.01	-.13	-1.38	.171		
Instructions	0.01	0.01	.12	1.33	.188		
WMC	< 0.01	< 0.01	.06	0.59	.559		
SM	< 0.01	< 0.01	-.01	-0.08	.933		
Quantity of person descriptions						.03	1.00
Modality	1.90	1.85	.10	1.03	.306		
Instructions	0.77	1.82	.04	0.43	.671		
WMC	1.14	0.93	.11	1.23	.222		
SM	1.26	0.93	.13	1.36	.178		
Accuracy of person descriptions						.08	2.61*
Modality	< -0.01	0.02	-.02	-0.21	.832		
Instructions	0.03	0.02	.15	1.66	.099		
WMC	0.01	0.01	.11	1.26	.210		
SM	0.02	0.01	.22	2.40	.018		

Note. WMC = working memory capacity factor score; SM = source monitoring factor score. * $p < .05$.

Discussion

In Experiment 2, we studied the effect of recall instructions on the recall modality effect in eyewitness accounts. To this end, we included a spoken-videotaped and a written condition and employed either scarce or comprehensive instructions. While Sauerland and Sporer's (2011) results suggested that writing is more demanding than speaking, the results of Experiment 1 and Grabowski (2007) suggest the opposite. We expected the less demanding description condition (i.e., writing or speaking) to produce more detailed accounts and comprehensive instructions to elicit more detailed reports than scarce instructions. Furthermore, we expected that the beneficial value of comprehensive instructions would be greater under more rather than less demanding conditions (i.e., an interaction).

Our results provide only limited support for our hypotheses. First, replicating the well-established finding that recall instructions are an important determinant

of recall performance (e.g., Fisher & Geiselman, 1992; Gabbert et al., 2009), we found that comprehensive instructions led to more detailed accounts than scarce instructions for FRs in general as well as the event FR-CQs (but not the person FR-CQs). Second, a main effect of modality such that writing was more beneficial than speaking was apparent only for the event FR, but not the person FR or the FR-CQs in general. Third and similarly, an interaction between instructions and modality became significant only for the event FR-CQ (but not the person FR-CQ and FRs in general). Finally, similar to Experiment 1, the level of executive functioning did not interact with the modality to predict recall performance.

Our findings lend some support to Grabowski's (2007) hypothesis that writing is less demanding than speaking. They are, however, in direct conflict with the idea that differences in the instructions used in Experiment 1 and by Sauerland and Sporer (2011) might serve as an explanation for the contradictory results found in these two experiments. What is more, the interaction between instructions and modality took a different form than expected. We had expected both written and spoken conditions to profit from the comprehensive instructions, with a greater beneficial value for the more demanding condition. However, speaking participants did not profit from the comprehensive instructions at all, while writing participants did. Thus, the findings are neither in line with the assumption that speaking is more demanding than writing (Grabowski, 2007), nor with the opposing view that writing is more demanding than speaking (Kellogg, 2007). Either way, both conditions should have profited from the comprehensive instructions, only with a greater beneficial value for the one compared to the other condition. Thus, the confusing finding is that speaking participants did not profit from the comprehensive instructions at all when it came to providing event descriptions. One possible explanation for this finding could be that participants might have had difficulties remembering the instructions in the spoken condition. On the other hand, the significant effect of instruction comprehensiveness for person description quantity contradicts this argument.

General Discussion

Across two experiments, we investigated the relevance of recall modality for eyewitness performance. First, given that theoretical explanations and supporting empirical data exist for both a possible written and a possible spoken superiority effect (e.g., Grabowski, 2007; Kellogg, 2007; Sauerland & Sporer, 2011), we were interested in the direction of the effect. Second, we wanted to test the mechanisms that drive the effect. To this end, we assessed the role of cognitive demand, including executive functioning, the presence of an interviewer, and the role of recall instruction comprehensiveness.

Turning to the first question, our data lend no support to the notion of a spoken superiority effect. This is contrary to the idea that speaking demands fewer cognitive resources (Kellogg, 2007) and contradicts earlier findings in the eyewitness literature (Bekerian & Dennett, 1990; Sauerland & Sporer, 2011) and other fields (Bergmann et al., 2004; Horowitz & Newman, 1964). In line with Grabowski (2007), however, the data provide some, albeit limited, evidence for a written superiority effect. Across two experiments, written event FRs were consistently somewhat superior to spoken FRs (note, though, that a significant difference between written and spoken-videotaped conditions was only found in Experiment 2, but not Experiment 1). However, this effect disappeared when looking at the complete event descriptions (i.e., FR-CQs) and at person descriptions. Furthermore, the accuracy of reports provided in writing or orally did not differ at all. Altogether, our results suggest that any differences in description quantity occur in the FRs and that following up with cued open-ended questions seems to be an effective tool to level possible differences between spoken and written recall. Unfortunately, however, as Tables 2.1 and 2.4 indicate, the use of cued questions comes with a decreased level of accuracy, as is known from previous research (Lipton, 1977; Powell, Fisher, & Wright, 2005; Sauerland & Sporer, 2011).

More support for the view that writing might be superior to speaking evolves when taking into the account interviewer presence. In this context, the literature on the effects of the presence of others suggests a facilitative effect on performance for simple tasks, but an inhibiting effect for more complex tasks (Bond & Titus, 1983; Zajonc, 1965). Accordingly, performance under more cognitively demanding and hence complex recall conditions might deteriorate with the inclusion of an interviewer, while less demanding recall conditions might benefit from it. Contrasting the interviewer-present and interviewer-absent spoken conditions in Experiment 1 yields that interviewer presence has a facilitating rather than an inhibiting effect on eyewitness recall performance. A direct comparison of the two interviewer-absent conditions in Experiment 1 furthermore suggests that, if anything, participants perform better when writing than speaking. A direct comparison of written and spoken interviewer-present conditions is not possible in the current design, as we did not include a written condition with an interviewer (or rather an observer). Based on the data available here, we reason that writing is not more complex than speaking. From this follows that if the presence of others facilitates spoken witness reports, it should also facilitate written ones. The inclusion of a written condition both with and without an observer present (in addition to interviewer-present and interviewer absent spoken conditions) is essential to make an informed judgment on this issue. Ideally, the spoken interviewer-absent condition in such a follow-up study should include participants who are practiced in the use of voice recorders in order to avoid the possible negative effects of operating voice recorders (Gould, 1978; Gould & Boies, 1978).

Moving to the role of cognitive demand in the recall modality effect, little can be said on the basis of the current data. If, for example, output production per time interval was crucial by itself, as suggested by Grabowski (2007), the written condition should consistently have outperformed the spoken ones. It is possible, however, that this factor interacts with interviewer or observer presence, as discussed above.

As for the influence of recall instructions, comprehensive instructions were mostly beneficial *across* modalities, while fluctuations as a function of recall modality were either absent or inconsistent. Similarly, differential effects of executive functioning on recall according to recall modality were absent or inconsistent. From a legal perspective, it is desirable to facilitate all witness reports, including those of individuals with various degrees of executive functioning. Given, however that no differential effects were found for the written compared with the spoken-videotaped conditions, which could be considered the standard police interviewing conditions, this point may not need to be a concern in the current police practice.

A limitation of the present study is the fact that we tested our participants for visual material only (i.e., no conversations could be heard in the stimulus film). Hence, it is unclear whether our findings transfer to auditory details, and our conclusions must be limited to visual details. This point also pertains to the use of a visual (instead of a verbal) distraction task in Experiment 1. We do not know, if the visual distraction task we used would have inhibited recall of auditory material, too. Previous research indicates that it might not have, as visual distraction during recall selectively inhibits recall of visual material and auditory distraction selectively inhibits recall of auditory material (Vredeveldt, Hitch, & Baddeley, 2011). However, as we tested both Dutch and German speaking participants, we wanted to keep the influence of language to a minimum and hence, refrained from including auditory material in the stimulus film. Based on Vredeveldt et al.'s findings and given the nature of the stimulus film (i.e., containing visual details only), the use of a visual instead of an auditory distraction task seemed most appropriate.

Overall, the current results suggest that writing might be superior to speaking when making a witness statement, at least in highly educated samples. This effect might be even stronger when introducing an observer to the written condition. Given, however, that the deficit in spoken descriptions that might occur in FRs can be compensated by the use of cued open-ended follow up questions, a general recommendation for obtaining written rather than spoken testimony is unwarranted based on the current data. This is reassuring insofar as it suggests that current police practice of obtaining written versus spoken accounts as a function of crime seriousness and witness centrality is unproblematic. Two notes of caution are in order, however. First, meaningful differences between recall modalities might arise as a function of (lower) educational background, increased task difficulty, personal preference, or other factors. Second, the current findings directly contradict the findings of a recent eyewitness study in which strong effects in sup-

port of a spoken superiority effect in witness accounts were found (Sauerland & Sporer, 2011). The methodology used here and by Sauerland and Sporer was highly similar in terms of the sample (mainly university students), experimenters (female master students), stimulus material (staged crime videos with four actors, depicting the theft of an object: a pair of sunglasses or a wallet, comparable in length), and coding procedure. We can only speculate about the reasons underlying these conflicting findings. Possibly, the effect is subject to the decline effect (Lehrer, 2010; Schooler, 2011) or one of the findings constitutes a false alarm. The chance for such a result is 5% after all. In either case, these contradictory findings in the light of no obvious differences in methodology give reason for caution and should be cause for further investigation into the recall modality effect in eyewitness testimony. For now, we conclude that both writing and speaking seem to be appropriate tools when obtaining eyewitness reports. Which of them is golden rather than silver, however, cannot be decided based on the data available at present.

Appendix C: Cued Questions for the Event Description

1. Where did the event take place? Describe the location of the event.
2. How long was the duration of the film?
3. How many persons were at the crime scene?
4. How many persons were involved in the theft?
5. How was the wallet stolen? Describe the manner in which the theft occurred.
6. Did the thief have (an) accomplice(s)?
If yes, continue with question 7.
If no, continue with question 9.
7. What did the accomplice(s) do during the theft?
8. Did the accomplice(s) stay at the crime scene after the wallet was stolen? If yes, for how long?
9. What did the bystander(s) do during the theft of the wallet?
10. How long did the theft of the wallet take?
11. Did the thief stay at the crime scene after the wallet had been stolen? If yes, for how long?
12. Where did the thief put the stolen wallet?
13. How long did it take until the victim realized that her wallet had been stolen?

Appendix D: Cued Questions for the Person Descriptions

1. How old was the thief? Give a specific age (e.g., x years).
2. Describe the hair color of the thief.
3. Describe the hair length of the thief.
4. Describe the hairstyle of the thief.
5. Describe the build of the thief.
6. Describe the body proportions of the thief.
7. Estimate the height of the thief (cm). Give a specific height (e.g., x cm).
8. Estimate the weight of the thief (kg). Give a specific weight (e.g., x kg).
9. Did the thief have any distinctive features? If yes, describe them.
10. Describe the clothing of the thief.
11. Did the thief wear jewelry or accessories? If yes, describe them.
12. Are there additional details that you remember, but that you have not been asked about yet? If yes, please describe them.

The questions referring to the other actors were identical.

CHAPTER 3 Retrieval Support with the SAI and Stress during Encoding

Krix, A. C., Sauerland, M., Raymaekers, L. H. C., Memon, A., & Smeets, T. (2014). Eyewitness evidence obtained with the Self-Administered Interview© is unaffected by stress. Manuscript submitted for publication.

Abstract

The Self-Administered Interview© (SAI) was designed to elicit an early eyewitness statement. It is based on the Cognitive Interview and adopts some of its memory-enhancing components. Because it is completed directly after the crime, witnesses may still be experiencing stress that was caused by the incident at the time of retrieving information. This is important given that it is widely known that stress during retrieval produces memory-impairing effects. The aim of the present study was therefore to investigate the effectiveness of the SAI under different levels of stress. We expected the SAI to elicit more comprehensive accounts than free recall (FR). In line with findings from the neurobiological field, stress was expected to lead to reduced description quantity and accuracy. One-hundred-and-twenty-seven participants either underwent a stress task, known to reliably elicit autonomic and glucocorticoid stress responses, or a control task. Salivary cortisol levels were measured to check the effectiveness of the stress induction. Moreover, all participants were exposed to a live staged crime committed by a confederate. Hereafter, they completed either an SAI or an FR. As hypothesized, the SAI elicited more perpetrator details and verifiable event details than the FR, while there was no effect on accuracy. Unexpectedly, in spite of leading to stress-induced cortisol elevations, stress exposure had no effects on memory performance, nor did it interact with interview type. The results suggest that the SAI can be safely used under different levels of stress in the immediate aftermath of a crime.

Observing a crime is often a highly stressful experience for witnesses. During the investigations, the police are likely to approach all witnesses to obtain eyewitness evidence, irrespective of their stress level. Hence, investigating the effects of stress on witnesses' recall performance is important.

In the past, various theoretical accounts have been proposed to address the influence of stress on memory performance. Christianson (1992) argued that memory for emotional events, relative to neutral events, was enhanced for central details, but impaired for peripheral details, because during emotional events attention would be focused on the central details. Such a memory advantage for central details (Christianson & Loftus, 1987, 1991), however, did not always emerge (Heuer & Reisberg, 1990; Wessel, van der Kooy, & Merckelbach, 2000). Contributing to these inconsistent findings, no unitary definition of central details exists, which have been defined either as spatially central (Christianson & Loftus, 1991; Wessel et al., 2000) or as central to the plot (Heuer & Reisberg, 1990; Ibabe & Sporer, 2004). In a recent attempt to combine these two definitions, Houston, Clifford, Phillips, and Memon (2013) hypothesized that for emotional events, memory for the perpetrator may be enhanced, because the perpetrator is a central element both spatially and thematically. In line with this prediction, Houston et al. observed increased quantity for perpetrator details, but reduced quantity for event details, after participants watched a crime relative to a neutral event.

Instead of focusing on different memorability of central versus peripheral details of emotional events, Deffenbacher (1994) proposed a model addressing memory for emotional events as a whole. According to this model, highly stressful incidents, such as witnessing a crime, trigger an activation mode. In this activation mode, increases in somatic anxiety would initially moderately enhance memory performance, but further increases would result in a massive drop of memory performance. In line with these predictions, a meta-analysis showed that stress may indeed negatively affect eyewitnesses' recall accuracy (Deffenbacher, Bornstein, Penrod, & McGorty, 2004). The corresponding moderator analyses showed that especially responses to specific questions relative to free recall were affected and the effect of stress was larger when a staged crime (live or videotaped) was shown rather than a different stressor. Yet, in a more recent study, confrontation with a high relative to low arousal stimulus event resulted in lower quantity, but higher accuracy of event details, while recall of person details was unaffected (Hulse & Memon, 2006).

Given these overall inconsistent results, inspection of the findings from a different research area may be informative. Neurobiological studies have emphasized the importance of the timing of the stressor for memory performance. Specifically, stress during encoding generally leads to memory-enhancing effects (Cahill, Gorski, & Le, 2003; Roozendaal, 2002; Smeets, Otgaar, Candel, & Wolf, 2008). This is likely due to cortisol acting on the amygdala, thereby modulating the activity of the hippocampus during memory consolidation (Roozendaal & McGaugh, 2011;

Schwabe, Joëls, Roozendaal, Wolf, & Oitzl, 2012). Stress during retrieval, however, causes memory impairment (de Quervain et al., 2000; Kuhlmann, Piel, & Wolf, 2005; Smeets et al., 2008). Here, cortisol is also likely to be involved by inhibiting the activity of hippocampal parts involved in memory retrieval (Buchanan, Tranel, & Adolphs, 2006; Roozendaal, 2002).

These results regarding the timing of the stressor on recall performance may be especially relevant for interviews conducted immediately after the crime. Specifically, an eyewitness interview that is likely to be employed shortly after the incident when the witness may still experience distress from observing the crime is the recently developed Self-Administered Interview© (Gabbert, Hope, & Fisher, 2009; see Hope, Gabbert, & Fisher, 2011, for a detailed description). To avoid problems of delayed interviewing (e.g., forgetting, Rubin & Wenzel, 1996, or misinformation effects, Gabbert, Memon, Allan, & Wright, 2004) this self-administered booklet is independently completed by the witnesses at the crime scene. Its purpose is to secure an early witness statement when limited police resources do not allow conducting an immediate personal interview. The SAI, which is based on the Cognitive Interview (CI; Fisher & Geiselman, 1992), adopts some of the memory-enhancing techniques of the CI, such as mental reinstatement of context, and hence provides ample retrieval support. The SAI elicits more comprehensive accounts than free recall (FR), without either decreasing or increasing accuracy (Gabbert et al., 2009; Gawrylowicz, Memon, & Scoboria, 2014), and preserves memory for a subsequent personal interview (Hope, Gabbert, Fisher, & Jamieson, 2014; see chapter 4).

Although the SAI is usually completed immediately after the crime when stress levels are still high, to our knowledge, no study so far has tested the effectiveness of the SAI relative to a control interview under different stress levels. We found only one study that dealt with the related CI under different levels of stress during encoding (Ginet & Verkamp, 2007). In this study, all participants watched a videotaped car accident. Stress was manipulated by telling half of the participants that they would be exposed to electroshocks during the film. Two weeks later, participants were interviewed about the film with a CI or a control interview. Here, the high-stress group made fewer errors and was therefore more accurate than the low-stress group. Moreover, the CI group outperformed the control group in terms of quantity (but not accuracy). However, this study only relied on self-report measures to control for induced stress, which do not always correspond with objective physiological responses (e.g., Hellhammer & Schubert, 2012). It is also unclear as to whether the results of this experiment are transferable to the SAI, because the stressor was present only during encoding. When using the SAI, encoding and retrieval are likely to occur under stress. As stress during retrieval may impair recall performance, it is unclear whether interviews with ample retrieval support, such as the SAI, are still more effective than interviews that do not provide retrieval support (e.g., FR).

The aim of this study was to compare the SAI with FR under different levels of stress. By exposing all participants to a live staged crime, a high level of ecological validity was attained. To our knowledge, this is the first laboratory experiment on the SAI that uses a live event. To induce stress, we used a powerful, standardized procedure, the Maastricht Acute Stress Test (MAST; Smeets et al., 2012), which is known to elicit robust subjective, autonomic, and glucocorticoid stress responses (e.g., Meyer, Smeets, Giesbrecht, Quaedflieg, & Merckelbach, 2013; Quaedflieg, Meyer, Schwabe, & Smeets, 2013). Participants' salivary cortisol stress levels were measured to control for the effectiveness of the stress manipulation in stimulating the hypothalamic-pituitary-adrenal (HPA) axis that is responsible for stress reactions (e.g., cortisol secretion).

In line with Deffenbacher's (1994) model of the impact of stress on recall performance and findings from the neurobiological field (Kuhlmann et al., 2005; Smeets et al., 2008), we hypothesized that participants exposed to stress during retrieval would recall fewer details and be less accurate in their reports than participants not exposed to stress. In line with previous SAI research (Gabbert et al., 2009; Gawrylowicz et al., 2014; see chapter 4), we hypothesized that participants completing the SAI would report more details, compared to those completing an FR. No differences were expected regarding accuracy. Also, no interactions between stress and interview type were expected.

Method

Participants and Design

In total, $N = 127$ participants (21 men; age: $M = 22.2$, $SD = 4.9$, range 18-63 years) took part in the experiment in exchange for course credit or a €10 voucher. Participants were mostly students (81.1%) or recruited in the vicinity of the university. The study was approved by the local ethical committee. Participants were randomly assigned within a 2 (stress: control vs. stress) \times 2 (interview type: FR vs. SAI) between-participants design. Participants interacted with one of two confederates, which was counterbalanced across experimental conditions.

Interviews

Self-Administered Interview

The SAI is an interview booklet that provides a high level of retrieval support (Gabbert et al., 2009; Hope et al., 2011). It is a standardized, yet generic recall tool that can be used for different types of crimes. Before writing down their recollections, witnesses first mentally reinstate the context. That is, they are instructed to

think back to the witnessed incident and picture in their minds what they saw or heard, what they were thinking and how they were feeling at the time that the incident took place. In the following separate sections, they are asked to provide descriptions of the course of events, the appearance of the perpetrator(s), and, if applicable, of potential other witnesses or vehicles involved. The perpetrator and vehicle sections also contain non-leading prompts to cue recall. In another section, witnesses are requested to draw a sketch of the scene to facilitate the retrieval of spatial information (e.g., locations). In this way, the SAI relies on multiple and varied retrieval. In the final section, the witnessing conditions are prompted to assess how well the witness saw the incident (e.g., Were there any obstructions in your view?). Throughout the interview, witnesses receive the instruction to provide the most complete and accurate account possible and to refrain from guessing.

Free recall

The written FR form simply instructs participants to report all details they can remember about the sequence of actions and events, and about all persons involved, including the perpetrator(s) and potential other witnesses (see Gabbert et al., 2009, for a similar approach). Analogous to the SAI, participants are reminded to provide the most complete and accurate account possible, but not to guess. The FR provides only little retrieval support and differs in several ways from the SAI. First, it does not feature memory-enhancing components (e.g., mental context reinstatement). Second, it involves only one instead of multiple and varied retrieval attempts. Finally, the FR does not feature prompts to cue recall.

Stress Induction versus No-Stress Control Manipulation

The MAST (Smeets et al., 2012) is a concise procedure to reliably elicit robust subjective, autonomic, and glucocorticoid stress responses. It involves a 5 min preparation phase in which the task is explained and a 10 min acute stress phase that includes repeated exposure to cold pressor stress and mental arithmetic challenge. Specifically, participants immerse their hands into ice water (4°C; plexiglas box with an electrical cooler and a circulation pump from JULABO Labortechnik, Seelbach, Germany) during 5 trials lasting 60 to 90 sec. Alternating with the hand immersion trials, participants engage in mental arithmetic challenges during which they have to count backwards as fast and accurately as possible in steps of 17 starting at 2043 for 45, 60, or 90 sec. Whenever they count too slowly or make a mistake, they receive negative feedback (i.e., to count faster and/or recommence at 2043). To increase task unpredictability, participants are told that the order and duration of the hand immersion and mental arithmetic trials are randomly chosen by the computer and that they are videotaped so that their facial expressions can be analyzed afterwards.

The no-stress control condition (see Smeets et al., 2012, Experiment 3) also comprises a 5 min preparation phase and a 10 min hand immersion phase, albeit in lukewarm water (25 °C), alternated with a simple counting task during which participants have to repeatedly count from 1 to 25 at their own pace. The experimenter remains in the room to check participants' compliance with the instructions, but participants are not given any feedback on their performance nor are they videotaped. The duration and order of hand immersion and arithmetic trials parallels that of the MAST.

Salivary Cortisol Responses

Cortisol stress measures prior and in response to the (control) MAST were obtained with synthetic Salivette (Sarstedt®, Etten-Leur, The Netherlands) devices 5 min before ($t_{\text{pre-stress}}$) and three times after the MAST ($t_{+0\text{min}}$, $t_{+10\text{min}}$, $t_{+20\text{min}}$ with reference to the end of the stress or control procedure). Samples were stored at -20°C until cortisol levels were determined by a commercially available luminescence immune assay kit (IBL, Hamburg, Germany). Mean intra- and inter-assay coefficients of variation are typically less than 5%, and the lower and upper detection limits were 0.015 mg/dl (0.41 nmol/l) and 4.0 mg/dl (110.4 nmol/l), respectively.

Positive and Negative Affect Schedule

The negative affect subscale of the Positive and Negative Affect Schedule (PANAS; Watson, Clark, & Tellegen, 1988) was used to measure subjective distress. PANAS consists of two mood scales each containing 10 items that measure positive or negative affect. In each item, participants are asked to rate the extent to which they experience a certain emotion on a 5-point scale (1 '*very slightly or not at all*' to 5 '*very much*').

Procedure

Participants were tested individually and instructed not to consume any foods or drinks, or to engage in physical exercise for at least 2 hours before testing. After signing the informed consent form, the first saliva sample ($t_{\text{pre-stress}}$) was taken. Participants were told that the aim of the study was to examine cognitive and physical reactions to a stressful experience. They were not informed about the upcoming mock crime. After engaging in the MAST for 15 minutes, participants were informed that there would be a short break and that a saliva sample ($t_{+0\text{min}}$) had to be taken. The information that the MAST would be continued was false to avoid decreases of stress levels at this time (Smeets et al., 2012). Immediately before and after the MAST, participants completed the PANAS (Watson et al., 1988).

The experimenter told the participants that s/he had to put the saliva samples into the freezer and left. Moments later, one of two confederates (a woman with long blonde hair or a man with short brown hair, both 22 years old) entered the room for about one minute ($M = 50.8$ sec, $SD = 12.6$) and interacted with the participants according to a script. Before data collection, the confederates had practiced the script to ensure a standardized procedure. The confederate inquired whether the experimenter was there and introduced him-/herself as the previous participant or, if the current participant was the first one of the day, as a colleague of the experimenter. The confederate informed the participants that s/he had forgotten his/her cell phone and now wanted to pick it up. After taking the phone from a lab table the confederate left the room. When participants were reluctant to let the confederate go with the phone, the confederate had to come up with excuses why s/he could not wait for the experimenter to return (e.g., stating that his/her class would start soon). In six cases, the participants did not allow the confederate to take the phone, so s/he had to leave the lab without it.⁹ After the confederate had left, the experimenter re-entered the room and looked for his/her phone. When the participants mentioned that the phone had been picked up, the experimenter informed them that the phone had been stolen and that they had witnessed a theft.

Participants were then taken to another room where they completed an FR form or an SAI (Gabbert et al., 2009), depending on the interview condition. No time limits were imposed. While writing down their accounts, the third ($t_{+10\text{min}}$) and fourth ($t_{+20\text{min}}$) saliva samples were taken 10 and 20 min after termination of the MAST (approximately in the first and second half of the interviews). This interfered little with retrieval, as participants could continue writing while providing the samples. One week later, participants returned for a second session to identify the confederate in a lineup. The eyewitness identification results are unrelated to our research question and reported elsewhere (see Sauerland et al., 2014). Participants were thanked for their participation and fully debriefed upon completion of data collection.

Coding

Participants' statements were transcribed and coded against coding schemes. In line with Houston et al. (2013), perpetrator and event details were analyzed separately. Pictures of the confederates were taken on every test day. This allowed scoring the accuracy of the appearance and the clothing details. Although the number of details in the perpetrator schemes slightly varied across test days (i.e., depending on the outfits worn), they comprised at least 30 details. Because the interactions between confederate and participant could not be videotaped, verifiable and unverifiable details were differentiated for coding the event descriptions. For-

⁹ Data analyses yielded the same pattern of results with and without the participants who did not let the confederate take the phone, so their data were included.

ty-eight verifiable event details were part of the script (e.g., thief took the phone) and were coded for accuracy. In contrast, unverifiable event details were details not defined in the script (e.g., thief pointed to the phone) and, because it is unknown whether they actually happened, we coded only their quantity.

For the perpetrator descriptions, details pertaining to the face, hair, body, and clothes were coded. For the event descriptions, identity (i.e., who did something?), action, object, setting (i.e., locations and directions), and conversation details were coded. The latter were defined as the smallest unit of information still containing a subject and a predicate (Campos & Alonso-Quecuty, 2008). The statement “The female (1) thief (2) who wore a black (3) shirt (4) took (5) the cell phone (6).” would yield six details (see also chapter 2). Subjective responses (e.g., “He was ugly”) were not scored. As to the perpetrator descriptions and verifiable event details, details were considered correct or incorrect if they did or did not match the appearance of the confederate or the script. Confabulations were both incorrect and non-existent (e.g., a hat when no headgear was worn; see chapter 2).

Inter-coder reliability

For both the perpetrator and event descriptions, thirty randomly selected statements were independently coded by two coders. For accuracy of the perpetrator descriptions and of the verifiable event details, Cohen's κ was .91 and .92, $ps < .001$, respectively. For the quantity of the unverifiable event details, the single-measures intra-class correlation (absolute agreement) was $ICC = 0.84$, $p < .001$, indicating *excellent* inter-coder reliability (Landis & Koch, 1977).

Results

An *alpha* level of .05 was used for all statistical tests. In case of multiple comparisons, Bonferroni-adjusted *alpha* values are reported. We report Cohen's *d* (Cohen, 1988) for dependent or independent samples for the main effects with $df = 1$ in the numerator and η_p^2 for the interaction effects and main effects with $df > 1$ in the numerator (see Sporer & Cohn, 2011).

Manipulation Check

Cortisol stress responses

Cortisol data were log-transformed before analysis as Shapiro-Wilk tests of normality showed typical skewness of the data. Cortisol responses were analyzed with a 2 (condition: stress vs. control) x 4 (measurement: $t_{\text{pre-stress}}$ vs. $t_{+0\text{min}}$ vs. $t_{+10\text{min}}$ vs.

$t_{+20\text{min}}$) analysis of variance (ANOVA). The cortisol levels over time in the stress and control group are displayed in Figure 3.1.

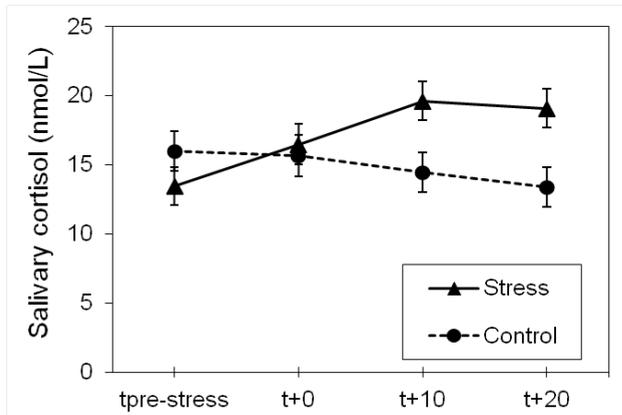


Figure 3.1. Salivary cortisol levels (nmol/L) over time in the stress and in the control group. tpre-stress = measurement before administration of the MAST/control task; t+0 = measurement upon termination of the MAST/control task; t+10 = measurement 10 min after termination of the MAST/control task; t+20 = measurement 20 min after termination of the MAST/control task.

At base line, the cortisol concentration did not differ between the stress and the control condition, $F(1, 123) = 1.47, p = .228, d = 0.22$. There was a significant Measurement by Stress interaction, $F(3, 363) = 43.95, p < .001, \eta_p^2 = 0.27$. Follow-up tests for the stress group showed a significant main effect of measurement, $F(3, 189) = 32.91, p < .001, \eta_p^2 = 0.34$, with cortisol increases from $t_{\text{pre-stress}}$ to $t_{+0\text{min}}$ ($p < .001, d = 0.42$) and from $t_{+0\text{min}}$ to $t_{+10\text{min}}$ ($p < .001, d = 0.36$), but not from $t_{+10\text{min}}$ to $t_{+20\text{min}}$ ($p = .227, d = -0.11$). Follow-up tests for the control group also yielded a significant main effect of measurement, $F(3, 174) = 20.42, p < .001, \eta_p^2 = 0.26$, with cortisol decreases from $t_{+0\text{min}}$ to $t_{+10\text{min}}$ ($p < .001, d = -0.11$) and from $t_{+10\text{min}}$ to $t_{+20\text{min}}$ ($p < .001, d = -0.10$), but not from $t_{\text{pre-stress}}$ to $t_{+0\text{min}}$ ($p = .999, d = -0.01$). The results of these analyses are in line with the general finding in the literature that cortisol increase occurs with temporary delay after a person has been exposed to a stressor (Dickerson & Kemeny, 2004). On the other hand, the controls' cortisol levels dropped due to the circadian rhythm of cortisol decreasing throughout the day (Nicolson, 2008).

Negative affect (PANAS)

Negative affect induced by the stress manipulation was analyzed in a 2 (stress) \times 2 (measurement: pre vs. post stress manipulation) ANOVA. The significant main effects of stress, $F(1, 122) = 26.91, p < .001, d = 0.93$, and time, $F(1, 122) = 9.43, p =$

.003, $d = 0.26$, were qualified by a significant interaction, $F(1, 122) = 25.56, p < .001, \eta_p^2 = 0.17$. While negative affect did not change over time in the control group ($M_{pre} = 11.87, SD_{pre} = 2.40; M_{post} = 11.07, SD_{post} = 2.17$), $F(1, 122) = 1.94, p = .167, d = -0.35$, it significantly increased in the stress group ($M_{pre} = 13.00, SD_{pre} = 3.15; M_{post} = 16.29, SD_{post} = 6.76$), $F(1, 122) = 33.56, p < .001, d = 0.56$.

Impact of Stress and Interview Type on Recall Performance

To analyze recall performance as a function of confederate (1 vs. 2), interview type (FR vs. SAI), and stress (stress vs. control group), we ran 2 x 2 x 2 ANOVAs. Because there were no significant interactions with confederate, $F_s \leq 3.56, p_s \geq .062, \eta_p^2_s \leq 0.03$, we collapsed the data across confederates. Quantity and accuracy (correct details divided by quantity; see Meissner, Sporer, & Susa, 2008) of perpetrator descriptions and verifiable event details and quantity of unverifiable event details were the dependent variables (see Table 3.1 for means and standard deviations).

The data of one participant could not be analyzed, because she only described the MAST, but not the interaction with the confederate. For the perpetrator descriptions, the data of an additional ten participants (SAI/control: $n = 3$; FR/stress: $n = 3$; FR/control: $n = 4$) could not be analyzed, because the corresponding pictures of the confederate were missing, or because it was unclear which picture the participant had been assigned to.

Unexpectedly, none of the main effects of stress, $F_s \leq 1.46, p_s \geq .230, d_s \leq 0.17$, and interactions between stress and interview, $F_s \leq 1.33, p_s \geq .252, \eta_p^2_s \leq 0.01$, were significant. Hence, in the following, only the effects of interview type will be described.

Quantity

Replicating earlier findings (Gabbert et al., 2009; Gawrylowicz et al., 2014; see chapter 4), the SAI group ($M = 21.77, SD = 5.13$) recalled more verifiable event details than the FR group ($M = 13.64, SD = 5.06$), $F(1, 122) = 79.70, p < .001, d = 1.60$.¹⁰ The former also ($M_{SAI} = 13.73, SD_{SAI} = 4.52$) recalled more perpetrator details than the latter ($M_{FR} = 6.96, SD_{FR} = 3.12$), $F(1, 112) = 84.43, p < .001, d = 1.72$. The quantity of the unverifiable details was not influenced by interview type, $F(1, 122) = 2.31, p = .131, d = 0.27$.

Accuracy

Both the accuracy of the perpetrator descriptions, $F(1, 112) = 0.01, p = .916, d = 0.03$, and the accuracy of the verifiable event details, $F(1, 122) = 0.05, p = .822, d =$

¹⁰ When verifiable and unverifiable event details were analyzed together, analogous results emerged.

0.04, were unaffected by interview type.¹¹ Note that due to the small number of recall errors in the verifiable event descriptions (SAI: $M = 0.48$, $SD = 0.73$; FR: $M = 0.30$, $SD = 0.59$) accuracy rates were very high, which probably led to a ceiling effect making it difficult to detect differences.

Table 3.1

Quantity and Accuracy of Event and Person Descriptions as a Function of Interview Type and Stress Condition

		SAI		FR		Across interview types	
		<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Verifiable event description							
Quantity	Control	22.13	4.93	14.40	5.57	18.39	6.50
	Stress	21.42	5.37	12.90	4.48	17.30	6.53
	Across stress conditions	21.77 ^a	5.13	13.64 ^a	5.06	17.83	6.51
Accuracy (%)	Control	97.79	3.05	98.16	3.87	97.97	3.45
	Stress	98.05	3.36	97.98	4.45	98.02	3.89
	Across stress conditions	97.92	3.19	98.07	4.14	97.99	3.67
Unverifiable event description							
Quantity	Control	11.06	8.47	9.83	6.11	10.47	7.39
	Stress	11.58	5.64	9.23	5.73	10.44	5.76
	Across stress conditions	11.32	7.12	9.52	5.88	10.45	6.59
Perpetrator description							
Quantity	Control	13.66	4.73	7.77	2.83	10.87	4.91
	Stress	13.79	4.40	6.21	3.24	10.31	5.43
	Across stress conditions	13.73 ^b	4.52	6.96 ^b	3.12	10.58	5.18
Accuracy (%)	Control	78.59	13.76	76.53	13.00	77.62	13.32
	Stress	78.93	11.37	81.55	18.19	80.13	14.82
	Across stress conditions	78.77	12.44	79.13	15.96	78.94	14.13

Note. Means sharing the same superscript letter within a row indicate significant main effects with $p < .05$.

Discussion

The aim of the present study was to examine the effectiveness of the SAI when the witness was exposed to stress. By employing a live staged crime, a high level of ecological validity was achieved. Moreover, we were able to apply relatively high

¹¹ When perpetrator and verifiable event details were analyzed together, analogous results emerged.

levels of stress, by using a powerful, standardized procedure for stress induction in the lab (MAST; Smeets et al., 2012). Another asset of the present study is that we directly controlled for the effectiveness of the stress manipulation, by measuring participants' salivary cortisol levels.

Replicating the results of previous studies (Gabbert et al., 2009; Gawrylowicz et al., 2014; see chapter 4) and as hypothesized, we found that the SAI with its ample retrieval support elicited more comprehensive accounts than FR, without compromising accuracy. Specifically, relative to the FR, the SAI elicited 60% more verifiable event details and 97% more perpetrator details. This also demonstrates that higher recall quantity with the SAI not only emerges with stimulus films, but also when more realistic events are used.

From participants' cortisol levels, we can conclude that our stress manipulation was successful. Specifically, in the stress group, cortisol levels increased after exposure to the stressor, while the control group experienced a decrease of cortisol levels over time in line with the circadian rhythm. Moreover, the MAST significantly increased negative affect in the stress group, as indicated by the PANAS results. Nonetheless, stress did not influence memory performance. Hence, our hypotheses pertaining to stress were not confirmed and our findings contradict previous research showing both memory-enhancing (e.g., Ginet & Verkampt, 2007; Houston et al., 2013; Hulse & Memon, 2006) and memory-impairing effects of stress on recall performance (e.g., Deffenbacher et al., 2004).

Different from real situations in which the stress is likely to be caused by the crime, the stress induction in our study occurred prior to witnessing the staged theft. The stress induction may not have been successful, because participants did not attribute their arousal to the witnessed crime. However, this reasoning is not supported by the results of previous studies (Buchanan et al., 2006; de Quervain et al., 2000; Smeets et al., 2008). In these studies, immediately before retrieving previously studied wordlists, participants underwent a similar stress induction as in the current study. Because wordlists are not stressful themselves, it is safe to assume that the participants attributed their arousal to the stress task and not to retrieving the words. Still, reliable memory-impairing effects occurred in the stress groups. Interestingly, the eyewitness study that found memory-enhancing effects of stress during encoding (Ginet & Verkampt, 2007) also introduced the stress manipulation prior to presenting the stimulus film to participants. Specifically, participants in the stress group were told beforehand that they would receive electroshocks while watching the film. Hence, they had good reason to attribute their arousal to the threat of receiving electroshocks and not to the content of the film.

In our view, the cause for the null findings may rather be that stress was present (or absent) during both encoding and retrieval. Accordingly, the memory-enhancing effects during encoding (Cahill et al., 2003; Roozendaal, 2002) and memory-impairing effects during retrieval (de Quervain et al., 2000; Kuhlmann et al., 2005) may have compensated each other, yielding null results. The SAI, howev-

er, was developed to be administered in exactly such situations, that is, immediately after a crime. Although it would be interesting from a theoretical perspective to examine the influence of stress on recall performance when the SAI is administered after a delay, from a practical perspective, the approach of this experiment is most relevant, as it modeled the situation for which the SAI was designed.

Importantly, there were no significant interactions between stress and interview type. Hence, the SAI is equally effective under different levels of stress and more effective than FR when the interview is distributed directly after the crime. The present findings are comparable to the results obtained by Ginet and Verkamp (2007) who examined the related CI. The absence of an interaction is good news, because it implies that the SAI can be safely used under different stress levels in the immediate aftermath of a crime. Even if the police need not overly consider the amount of stress the witness has been experiencing before distributing the SAI, they should nevertheless try to set the witness at ease for the sake of the witness' psychological health.

Some limitations of the present study are noteworthy. First, given the nature of the research paradigm (i.e., staged live event for each participant individually), there was no straightforward way to determine "ground truth" for the encounter. Prior to data collection, both experimenter and confederates received extensive training sessions regarding the script for the interaction with the participant so that a standardized procedure was ensured. Therefore, we deemed it justified to differentiate between verifiable and unverifiable event details when scoring the event descriptions. Significant deviations from the script (e.g., the participant did not let the confederate take the phone) were recorded and considered during coding. Second, from a theoretical perspective, the design of the study did not allow a separate inspection of the effects of stress at encoding and at retrieval. However, our approach followed applied considerations, considering the SAI was devised to be administered directly after the crime (Gabbert et al., 2009) when the stress response is still likely to exert an impact. Finally, no information was collected about the female participants' menstrual cycle phases, although the latter can affect cortisol reactivity (Kirschbaum, Kudielka, Gaab, Schommer, & Hellhammer, 1999). Although we cannot exclude that some female participants' cortisol reactivity was reduced, cortisol levels were significantly elevated in the stress group. Hence, not controlling for the menstrual phase was deemed unproblematic.

Overall, the present study should serve as a starting point for further investigating the effects of stress on recall performance with the SAI. Future studies should address the following three research question. First, given that our participants were bystanders rather than victims does the pattern of results also hold for victim witnesses? Second, what is the effectiveness of the SAI when only encoding but not retrieval occurred under stress? Third, does stress during encoding impact the previously found memory-preserving effect of the SAI (Hope et al., 2014; see chapter 4)? For now, however, we can conclude that the SAI elicits more compre-

hensive statements than FR immediately after a crime, irrespective of the stress level experienced.

CHAPTER 4 Retrieval Support with the SAI and Divided Attention during Encoding

This chapter is an adapted version of the following article:

Krix, A. C., Sauerland, M., Gabbert, F., & Hope, L. (2014). Providing eyewitnesses with initial retrieval support: What works at immediate and subsequent recall? *Psychology, Crime & Law* 20, 1005-1027. doi:10.1080/1068316x.2014.902456

Abstract

The effect of retrieval support on eyewitness recall was investigated in two experiments. Based on the outshining hypothesis, Experiment 1 tested whether retrieval support enhances witness performance (compared to free recall) especially when witnessing conditions are suboptimal (e.g., because witnesses were distracted during the crime). Eighty-eight participants watched a videotaped crime with either full or divided attention and subsequently received retrieval support with the Self-Administered Interview© (SAI) or completed a free recall (FR). One week later (T2) all participants completed a second FR. Unexpectedly, retrieval support did not lead to better memory performance than FR under divided attention conditions suggesting that retrieval support is not effective to overcome adverse effects of divided attention. Moreover, presence of retrieval support at Time 1 (T1) had no effect on memory performance at T2. Experiment 2 ($N = 81$) tested the hypothesis that these T2-results were due to a reporting issue undermining the memory-preserving effect of T1-retrieval support by manipulating retrieval support (SAI vs. FR) at T1 and T2. As expected, T1-retrieval support led to increased accuracy at T2, corresponding to a moderate effect. Thus, the beneficial value of T1-retrieval support seems greatest with high-quality T2-interviews. Interviewers should consider this when planning a subsequent interview.

Eyewitness testimony is an important element for successful police investigations, as it is often the only available evidence (Kebbell & Milne, 1998). Hence, obtaining the most comprehensive and accurate statements from eyewitnesses is essential. In an effort to help witnesses remember, research on eyewitness testimony has identified various techniques that provide retrieval support. One such technique is the mental reinstatement of context, which is part of the Cognitive Interview (CI, Fisher & Geiselman, 1992). The mental reinstatement of context mnemonic is based on the principle of encoding specificity, that is, the idea that retrieval is facilitated when encoding context is recreated during retrieval (Tulving & Thomson, 1973). Specifically, witnesses are instructed to mentally recreate the context of encoding by thinking back to what they saw, heard, thought or felt during the incident. Such context cues are stored alongside the memory of the incident and facilitate retrieval by providing additional access pathways to memory.

Another example of an interview tool that provides ample retrieval support is the recently developed Self-Administered Interview© (SAI; Gabbert, Hope, & Fisher, 2009; see Hope, Gabbert, & Fisher, 2011, for a more detailed description of its structure and development process). It is a self-administered booklet designed to be completed by the witnesses directly after the crime to ensure an early interview when the police do not have the time and resources to conduct a timely personal interview. Thereby, it serves as a supplement, not a replacement, of a comprehensive subsequent personal interview. As a generic recall tool, it can be used for different kinds of crimes. The SAI is based upon the CI and adopts some of the CI's memory-enhancing components, such as the aforementioned mental context reinstatement and the "report-everything" component. The latter comprises the instruction to provide the most complete and accurate account possible. Moreover, the SAI fosters a high-quality statement by discouraging witnesses from guessing and by using non-leading questions (Hope et al., 2011). The interview provides a structure for recall and consists of several sections each focusing on a different aspect of the testimony (e.g., course of events, or appearance of the perpetrator). Additionally, the retrieval of spatial information is supported, as witnesses are required to draw a sketch of the scene. To summarize, the SAI relies on multiple and varied (i.e., verbal and non-verbal) retrieval attempts, which is beneficial for recall, because information that cannot be retrieved with one technique or retrieval attempt may well become accessible with another one (Tulving & Watkins, 1975).

Since the SAI is a relatively new tool, so far, only few studies have been conducted to examine it, but those that have yielded promising results. Specifically, the SAI has been shown to elicit a more comprehensive recall than a free recall (FR) without compromising accuracy (Gabbert et al., 2009, Experiment 1; Gawrylowicz, Memon, & Scoboria, 2014). FR is generally understood as a method that simply instructs people to provide a free narrative of what they can recall, without impeding, but also not supporting retrieval from memory. It lacks the

memory-enhancing components of the SAI, such as the mental reinstatement of context and multiple and varied retrieval, and hence, provides only little retrieval support. More importantly, the SAI can preserve memory for a subsequent interview (Hope, Gabbert, Fisher, & Jamieson, 2014). In Hope et al.'s study, participants were assigned to one of three groups. Participants initially completed either an SAI, an FR, or no interview at all (i.e., a no-initial-interview control group). In the second session one week later, all participants underwent a CI. Here the reports of the SAI group were more accurate than those of the other two groups. Interestingly, the FR group and the no-initial-interview control group did not differ in accuracy. This indicates that an early recall opportunity without additional retrieval support does not produce memory-enhancing effects at a subsequent retrieval attempt. Hope et al. (2014) explain these results with associative network models of memory (e.g., J. R. Anderson, 1983) according to which memory traces are represented as a network consisting of nodes and links between them. A high-quality retrieval attempt as implemented by the SAI with its ample retrieval support strengthens both the nodes and links of the network and thus facilitates retrieval in a subsequent recall attempt. The memory trace should be relatively less strengthened if the initial retrieval is less effortful, such as occurs with an FR.

It is the aim of the present study to examine the effect of retrieval support on eyewitness performance more closely in two experiments. Assuming a positive psychological perspective instead of focusing on deficits, the present paper seeks to increase knowledge regarding how to promote the reliability of eyewitness statements. Increasing this knowledge base further is important, because of the critical role eyewitnesses often play during criminal investigations and can lead to an adjustment of the existing interview practice. In Experiment 1, the potential positive effect of retrieval support on recall performance was examined when attention was divided during encoding. Experiment 2 took on the mixed results that were obtained in Experiment 1 and in previous research regarding the effect of providing retrieval support in the first interview on recall performance in subsequent interviews. Hence, we examined the conditions as to when providing witnesses with retrieval support at T1 can preserve memory at T2 and investigated whether the type of T2-interview matters in this regard. In both experiments, the SAI was selected as a proxy tool for providing witnesses with retrieval support. In doing so, we additionally contribute to increasing the available data base regarding the SAI. The FR was used as an interview format that does not provide retrieval support.

While observing a crime, witnesses may pursue a secondary task, such as speaking on the phone, paying a bill, driving a car, or looking for an escape route. As a consequence, their attention may not be fully allocated to the incident. Divided attention during encoding is suboptimal, as it has disruptive effects on encoding and can lead to reduced memory performance (e.g., N. D. Anderson, Craik, & Naveh-Benjamin, 1998; Craik, Govoni, Naveh-Benjamin, & Anderson, 1996). In the

context of eyewitness testimony, divided attention has been found to increase witnesses' suggestibility (Lane, 2006) and is thought to play a role in the weapon focus effect, which refers to a reduced memory performance if a weapon was present during the crime (Fawcett, Russell, Peace, & Christie, 2013). The negative effects of divided attention have clearly been documented. Yet, to our knowledge, the potential positive effect of providing witnesses with retrieval support on recall performance when they were distracted during the crime has not been investigated.

Observations in the domain of context-dependent memory suggest that retrieval support should be especially effective when suboptimal conditions, such as divided attention, were prevalent during the crime. The outshining hypothesis (Smith, 1988, 1994) predicts that environmental context cues are not always equally effective for memory retrieval. Specifically, if there are strong non-contextual memory cues during retrieval, the meaning of environmental context cues is relatively diminished (i.e., outshone). On the other hand, if memory cues are weak, witnesses may profit relatively more from contextual cues when attempting to retrieve information. For example, Smith, Glenberg, and Bjork (1978; but see Smith & Vela, 2001) found that administering a recall test in the very room where the stimuli had been encoded, instead of in a different room, led to better memory performance. As predicted by the outshining hypothesis, no such difference was found for recognition performance: Here the to-be-recognized word was a strong enough non-contextual cue that outshone the context cues provided by the room. More support for the outshining hypothesis comes from Fernandez and Glenberg (1985). In an associative-processing task, their participants were presented with pairs of words and asked to generate sentences that included those word pairs. In a subsequent memory test, environmental context (same vs. different room) did not have an effect on memory performance. Hence, associative processing during encoding reduced the influence of context cues during recall (see also Smith & Vela, 2001).

Divided attention, relative to full attention, also diminishes associative processing during encoding (e.g., Castel & Craik, 2003; but see Naveh-Benjamin, Guez, & Marom, 2003). Because pursuing a secondary task consumes additional cognitive resources, building of meaningful associations between the information seen may be impeded. In line with the outshining hypothesis (Smith, 1988, 1994), the meaning of context cues for recall should be heightened in this suboptimal situation. Consequently, retrieval support (e.g., with the SAI) may enhance recall more than an FR in suboptimal conditions, because it provides mental context cues for recall.

With these considerations in mind, the aim of Experiment 1 was to test whether providing witnesses with retrieval support can limit the detrimental effects of divided attention conditions. Participants completed two recall attempts, one immediately after witnessing a staged crime (T1) and one a week later (T2).

For T1, we expected the reports of retrieval support participants to be more comprehensive than those of FR participants, without compromising accuracy. In line with the outshining hypothesis (Smith, 1988, 1994), we expected an interaction between the amount of attention paid and the presence of retrieval support. Specifically, the decline in memory performance from full to divided attention should be less marked in the retrieval support group, due to the beneficial effect of the mental context reinstatement instruction. Therefore, the difference between the retrieval support and the FR group should be greater under divided than under full attention conditions. For T2, the same interaction between initial retrieval support and attention was expected as for T1. Apart from that and in line with network models of memory (e.g., J. R. Anderson, 1983), we expected to replicate the memory preserving-effect of the SAI, that is, of an interview providing ample retrieval support (Hope et al., 2014). In other words, for T2, we expected that participants who were given retrieval support at T1 would provide both a more comprehensive and more accurate account than participants who completed an FR at T1.

EXPERIMENT 1

Method

Participants and Design

In total, 93 participants took part in Experiment 1. Five were excluded, because they did not attend the T2-session, leaving $N = 88$ participants for analysis (65 women; 41 German native speakers, 47 Dutch native speakers; age: $M = 21.9$, $Mdn = 20.5$ years, range: 18 to 64 years). Sample size was based on the size in previous SAI research (Gabbert et al., 2009) and on a power analysis with G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007). This analysis yielded a required total sample size of $N = 90$, given $\beta = .80$, $\alpha = .05$, and a medium to large effect size. Participants were mainly (95.5%) psychology undergraduates receiving course credit in exchange for participation or members of the general public. Performance did not differ between students and non-students. Participants were randomly assigned within a 2 (presence of retrieval support at T1: FR vs. SAI) x 2 (attention during encoding: divided vs. full) between-participants design and tested individually.

Materials

Film

The non-violent stimulus film, presented without audio, lasted 3:14 min and depicted the staged theft of a laptop. Six amateur actors (4 men, 2 women; ages 21 - 36) appeared in the film. The incident took place in a university environment and showed a number of students studying in a communal area. One student left his laptop unattended while taking a phone call, and a thief, along with an accomplice, stole the laptop.

Divided attention task

An arithmetic verification task adopted from Logie, Maylor, Della Sala, and Smith (2004) was selected as the divided attention task, as solving arithmetic problems has been shown to consume working-memory resources (Logie, Gilhooly, & Wynn, 1994). Twenty arithmetic problems and their corresponding solutions were presented via loudspeakers. The arithmetic problems, containing an equal number of addition, subtraction, multiplication, and division problems, were taken from a previous study (Jamieson, Gabbert, Allan, & Carson, 2009) and were presented at a rate of one problem every 10 s. One incorrect solution was presented for each of the four basic arithmetic operations. The participants' task was to report if they detected that a solution to an arithmetic problem was incorrect. The onset of the presentation of the arithmetic problems was synchronized with the beginning of the film. In this task, participants made $M = 3.20$ ($SD = 1.76$, $Mdn = 3$) errors, including both commission and omission errors.

Interviews

Self-Administered Interview

German and Dutch translations of the original English version of the SAI (Gabbert et al., 2009; see Hope et al., 2011, for a detailed description) were prepared using a back-translation process. In the SAI, before writing down their recollections, witnesses first mentally reinstated the context. That is, they were instructed to think back to the witnessed incident and picture in their minds what they saw, what they were thinking, and how they were feeling at the time. In separate sections, they were asked to provide descriptions of the course of events, the appearance of the perpetrator(s), and, if applicable, of potential other witnesses or vehicles involved. The perpetrator and vehicle sections featured non-leading recall prompts to cue recall. Witnesses were also requested to draw a sketch of the scene, so as to facilitate the retrieval of spatial information. Thereby, the SAI relied on multiple and varied retrieval. In a final section, the witnessing conditions were prompted to

assess how well the witness had seen the incident (e.g., What time of day did the event occur? Were there any obstructions in your view?). Throughout the whole interview, witnesses were instructed to provide the most complete and accurate account of the witnessed incident possible, but not to guess.

Free recall

Following Gabbert et al. (2009), the FR instructed participants to report all details that they could remember about the sequence of actions and events, and of all persons involved, including the perpetrator(s) and other witnesses. Analogous to the SAI, participants were instructed to provide the most complete and accurate account possible, but were discouraged from guessing. Note that the FR differed in significant ways from the SAI. First, it lacked significant memory-enhancing components (e.g., mental context reinstatement). Second, it entailed only one instead of multiple and varied retrieval attempts. Finally, unlike the SAI, the FR did not feature recall prompts to cue recall.

Procedure

The experiment consisted of two sessions. Using a cover story, we told participants that the study was about social perception. Upon arrival to the T1-session, participants signed an informed consent and provided their demographic data. In the full attention condition, participants were told that they would watch a film about a social situation. They were asked to watch the film carefully, because they would be required to answer some questions about it. No information was revealed about the nature of these questions. Participants in the divided attention condition were additionally instructed to simultaneously listen to arithmetic problems carefully. Participants were to say “false” aloud if they considered a solution to be incorrect and to remain silent if they considered a solution to be correct. The film was then shown to participants on a screen with a screen size of 22 in. (55.88 cm) and a resolution of 1920 x 1080. The participants in the divided attention group simultaneously listened to the arithmetic problems and responses were recorded by the experimenter. After the film, participants completed a number of unrelated filler tasks for 30 min.

To reduce any effects of physical context on recall performance, participants were led to a different room prior to providing information on what they had seen in the film. Depending on the retrieval support condition, the participants were either handed the SAI or the FR. Participants provided the statements in their native language. No time limits were imposed on participants in any condition.

The T2-session was scheduled one week later. Participants were not told beforehand about the content of this session. As in Gabbert et al. (2009, Experiment 2), all participants received an FR form for providing their second statement. Instructions reminded them that they had become a witness to an incident in the

previous week and that they would be asked to give a second statement about the incident. Hereafter, the participants were thanked for participation and dismissed. After the completion of data collection, all participants were fully debriefed.

Coding

Participants' statements were transcribed and coded employing a comprehensive coding scheme consisting of 560 details, of which approximately 58% were person, 17% action, 11% object, and 13% setting details. A detail was entered into the coding system when the coders agreed on its inclusion. The statement "The thief (1) sat (2) at the rightmost (3) table (4)." would yield four details (see Wright & Holliday, 2007, for a similar approach). Subjective responses, such as "He was ugly", were not scored. A detail was considered correct, if it matched the content of the stimulus film, and considered incorrect, if it did not match the content of the film. Details were considered confabulated when they were both incorrect and non-existent (e.g., the thief attacked the victim; see Dando, Wilcock, & Milne, 2009, for a similar approach). To code the accuracy of age, height, and weight estimates, we accepted deviations of plus or minus 2 years, 4 cm, or 3 kg, respectively, from the true value (see Fahsing, Ask, & Granhag, 2004, for a similar approach). The Dutch and German statements were coded by native speakers of each language.

Inter-coder reliability

To establish inter-coder reliability, the randomly selected statements of ten Dutch and ten German participants (i.e., 20 participants or 23% of the total sample) were independently coded by two coders. For correct recall, Cohen's κ were .99 and .98 for the German and Dutch statements, $ps < .001$, respectively. For incorrect recall, κ coefficients were .98 and .94, $ps < .001$, respectively.

Results and Discussion

An α level of .05 was used for all statistical tests. We report Cohen's d (Cohen, 1988) for dependent or independent samples for the main effects (with $df = 1$ in the numerator) and η_p^2 for the interaction effects (see Sporer & Cohn, 2011). To investigate recall performance as a function of presence of retrieval support at T1 (FR vs. retrieval support with the SAI) and attention (divided vs. full), we calculated two-way ANOVAs. The number of correct and incorrect details and accuracy (number of correct details divided by all reported details; see Meissner, Sporer, &

Susa, 2008) were the dependent variables. Table 4.1 displays the means and standard deviations of all dependent variables for T1 and T2.¹²

Table 4.1

Means and Standard Deviations for Number of Correct and Incorrect Details, and Accuracy at Time 1 and Time 2 (Experiment 1)

	Time 1						Time 2					
	Divided attention		Full attention		Across attention conditions		Divided attention		Full attention		Across attention conditions	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Number of correct details												
FR	39.00	24.49	78.32	26.64	58.66	32.17	38.00	19.46	72.00	30.02	55.00	30.35
Retrieval support	49.50	25.30	113.41	32.38	81.45	43.24	31.73	16.73	79.59	30.84	55.66	34.46
Across interviews	44.25	25.17	95.86	34.26	70.06	39.58	34.86	18.21	75.80	30.32	55.33	32.28
Number of incorrect details												
FR	7.55	4.90	7.55	4.48	7.55	4.64	8.14	4.99	7.27	6.32	7.70	5.64
Retrieval support	13.68	7.77	13.77	6.02	13.73	6.87	7.72	5.47	7.50	6.22	7.61	5.79
Across interviews	10.61	7.13	10.66	6.11	10.64	6.60	7.93	5.18	7.39	6.20	7.66	5.69
Accuracy (%)												
FR	81.01	14.39	91.01	4.75	86.01	11.74	80.55	14.29	90.82	6.51	85.69	12.14
Retrieval support	78.22	7.48	88.73	5.39	83.47	8.35	79.97	10.67	91.84	5.52	85.91	10.32
Across interviews	79.62	11.42	89.87	5.15	84.74	10.21	80.26	12.46	91.33	5.99	85.80	11.20

Recall at T1

Number of correct details

The main effects of retrieval support, $F(1, 84) = 15.25, p < .001, d = 0.60$, and attention, $F(1, 84) = 78.20, p < .001, d = 1.72$, were qualified by a significant interaction, $F(1, 84) = 4.44, p = .038, \eta_p^2 = 0.05$. Unexpectedly, the simple main effects analyses revealed that only within the full attention group did the retrieval support participants ($M = 113.41$) recall significantly more correct details than the FR participants ($M = 78.32$), $F(1, 84) = 18.07, p < .001, d = 1.18$. This was not the case when attention was divided (retrieval support: $M = 49.50$; FR: $M = 39.00$), $F(1, 84) = 1.62, p = .207, d = 0.42$.

¹² The distributions of the number of incorrect details and accuracy of T1 and the number of correct and incorrect details, and accuracy of T2 were significantly skewed. However, the results of the analyses with the transformed variables yielded the same results. Therefore, we will report the analyses with the untransformed variables.

Number of incorrect details

At T1 the retrieval support group ($M = 13.73$) recalled more incorrect details than the FR group ($M = 7.55$), $F(1, 84) = 23.91$, $p < .001$, $d = -1.05$. The main effect of attention, $F(1, 84) < 0.01$, $p = .971$, $d = -0.01$, and the interaction, $F(1, 84) < 0.01$, $p = .971$, $\eta_p^2 < 0.01$, were non-significant.

Accuracy

Accuracy of the T1-statements did not differ as a function of retrieval support, $F(1, 84) = 1.80$, $p = .184$, $d = -0.25$. However, full attention led to greater accuracy ($M = 89.87\%$) than divided attention ($M = 79.62\%$), $F(1, 84) = 29.41$, $p < .001$, $d = 1.16$. The interaction was non-significant, $F(1, 84) = 0.02$, $p = .895$, $\eta_p^2 < 0.01$.

Replicating previous findings (Gabbert et al., 2009; Hope et al., 2014), at T1 the group receiving retrieval support with the SAI recalled more correct details (at least in the full attention condition), but also more incorrect details than the FR group. As previously found, the accuracy rates did not differ. The expected interaction pattern between attention and presence of retrieval support did not emerge. For the number of correct details, the retrieval support group only performed better than the FR group when participants had watched the film with full attention.

Recall at T2

Contrary to our expectations, retrieval support at T1 did not have any effects on recall performance at T2, $F_s(1, 84) \leq 0.02$, $p_s \geq .902$, $d_s = 0.02$. The interactions between retrieval support and attention were also non-significant, $F_s(1, 84) \leq 1.68$, $p_s \geq .198$, $\eta_p^2_s \leq 0.02$. The full attention group ($M = 75.80$) recalled more correct details at T2 than the divided attention group ($M = 34.86$), $F(1, 84) = 58.72$, $p < .001$, $d = 1.64$. Their reports were also more accurate ($M = 91.33\%$) than in the divided attention group ($M = 80.26\%$), $F(1, 84) = 27.60$, $p < .001$, $d = 1.13$. For incorrect recall, the main effect of attention was non-significant, $F(1, 84) = 0.20$, $p = .659$, $d = 0.10$. Hence, we did not obtain a memory-preserving effect of initial retrieval support that was found earlier (Hope et al., 2014) and that would have been predicted by network models of memory (e.g., J. R. Anderson, 1983). To explore why we did not detect any differences at T2, we investigated memory performance over time.

Comparing T1 and T2 Recall

To analyze performance over time, we included Time (T1 vs. T2) as a within-participant factor, yielding a $2 \times 2 \times 2$ mixed-factorial analysis. For reasons of brevity, we will not report the main effects of attention. All other main effects and in-

teractions not mentioned in the following were non-significant, $F_s(1, 84) \leq 3.27$, $p_s \geq .074$, $|d|_s \leq 0.11$, $\eta_p^2_s \leq 0.04$. For testing why there were no effects of T1-retrieval support on performance at T2, the interaction effects between time and presence of retrieval support are most relevant.

Number of correct details over time

For the correct details, there were significant main effects of time, $F(1, 84) = 98.64$, $p < .001$, $d = 0.38$, and retrieval support, $F(1, 84) = 4.73$, $p = .033$, $d = 0.34$, which were qualified by a significant interaction, $F(1, 84) = 55.71$, $p < .001$, $\eta_p^2 = 0.40$. While the number of correct details significantly dropped over time within the retrieval support group (T1: $M = 81.45$; T2: $M = 55.66$), $F(1, 86) = 129.81$, $p < .001$, $d = 0.59$, it remained stable within the FR group (T1: $M = 58.66$; T2: $M = 55.00$), $F(1, 86) = 2.61$, $p = .110$, $d = 0.12$. The results show that while the number of correct details was much higher in the retrieval support group than in the FR group at T1, over time the number of correct details of the retrieval support group fell to the level of the FR group.

Number of incorrect details over time

For the incorrect details, there were also significant main effects of time, $F(1, 84) = 28.75$, $p < .001$, $d = 0.48$, and retrieval support, $F(1, 84) = 7.43$, $p = .008$, $d = -0.59$, qualified by a significant interaction, $F(1, 84) = 31.91$, $p < .001$, $\eta_p^2 = 0.28$. While the number of incorrect details significantly dropped over time within the retrieval support group (T1: $M = 13.73$; T2: $M = 7.61$), $F(1, 86) = 61.81$, $p < .001$, $d = 0.95$, it remained stable within the FR group (T1: $M = 7.55$; T2: $M = 7.70$), $F(1, 86) = 0.04$, $p = .838$, $d = -0.03$.

Accuracy over time

The interaction between time and retrieval support was marginally significant, $F(1, 84) = 3.75$, $p = .056$, $\eta_p^2 = 0.04$. Accuracy of the retrieval support group significantly increased from T1 ($M = 83.47\%$) to T2 ($M = 85.91\%$), $F(1, 86) = 5.95$, $p = .017$, $d = -0.25$, but remained stable in the FR group (T1: $M = 86.01\%$; T2: $M = 85.69\%$), $F(1, 86) = 0.11$, $p = .746$, $d = 0.03$.

The aim of Experiment 1 was to examine the effect of retrieval support on recall performance when the witness' attention was divided during encoding. Based on the outshining hypothesis (Smith, 1988, 1994), we hypothesized that the retrieval support group would always show better recall performance than the FR group. More importantly, this advantage should be more marked when attention was divided than when it was full. Unexpectedly, this pattern of results did not emerge for either T1 or T2. Note, however, that when the interactions were non-

significant, accepting the null hypothesis cannot be attributed to a lack of power, even though the observed statistical power was low (power ≤ 0.06 , $f_s \leq 0.032$, $\alpha = .05$). Indeed, none of the means displayed the descriptive trend in the expected direction, as can be seen in Table 4.1. The same was true in the case where the interaction reached significance (power = 0.52, $f_s = 0.228$, $\alpha = .004$). For the number of correct details at T1, the retrieval support group recalled more correct details than the FR group within the full attention condition. Descriptively, this difference was much smaller within the divided attention condition and recall performance did not differ as a function of the presence of retrieval support. Even though our hypothesis was not supported, it should be noted that the retrieval support group never performed worse than the FR group in terms of recall accuracy. It seems thus that providing witnesses with retrieval support is not an effective means to overcome the adverse effects of divided attention during encoding.

Regarding both the number of correct and incorrect details, the retrieval support group experienced a significant loss of reported details and fell precisely to the level of the FR group over time. The differences on a descriptive level were so small that at T2, no differences in recall performance as a function of the presence of T1-retrieval support occurred. Hence, accepting the null hypothesis cannot be attributed to a lack of power (power = 0.05, $f_s \leq 0.014$, $\alpha = .05$). It is striking that the retrieval support group was apparently unable to uphold their level of performance at T2. This contradicts the findings obtained by Hope et al. (2014) and shows that retrieval support at T1 does not always lead to carry-over effects at T2.

Interestingly, previous studies examining recall performance with the CI, which also provides ample retrieval support, report results similar to ours. In these studies, CI performance was compared with standard interview (SI; i.e., an interview with less retrieval support) performance on several recall attempts (e.g., Brock, Fisher, & Cutler, 1999; McCauley & Fisher, 1995; Memon, Wark, Bull, & Köhnken, 1997). Completing the CI at T1, relative to the SI, did not influence memory performance at T2 (i.e., there were no carry-over effects from T1 to T2), irrespective of the type of T2-interview used. Brock et al. (1999) orthogonally manipulated type of interview at T1 and T2 (CI vs. SI). Over time, the greatest number of details was forgotten when participants first received the CI and then the SI. Strikingly, we also found the greatest loss of reported information over time in the retrieval support/FR condition. Possible reasons given for the failure to find carry-over effects included that memory may become less context-dependent over time (Memon et al., 1997) or that the CI facilitates non-retrieval processes (e.g., the communication of recollections) instead of memory retrieval itself (Brock et al., 1999). Yet, none of the studies followed up on these assumptions, so the reasons as to why no memory-preserving effect occurred still remain unclear.

To explain the difference in results regarding the current and Hope et al.'s (2014) experiment, we will focus on the main difference between the two studies, that is, the interviews used at T2. Hope et al., who found a memory-preserving

effect after retrieval support at T1, used a CI at T2, an interview with ample retrieval support, while we employed an FR, an interview with little retrieval support. The memory-preserving effect of initially providing retrieval support may only be detectable when the T2-interview possesses certain features. It should be noted, though, that the use of an FR at T2 in the present study seemed justified based on both previous research results and applied considerations. Indeed, Gabbert et al. (2009, Experiment 2) also used an FR at T2 and found that participants who received initial retrieval support by completing an SAI recalled more correct details and provided more accurate reports in a T2-FR than a no-initial-interview control group. Moreover, there are countries, in which the CI is not used at all (e.g., in the Netherlands) or used very infrequently (e.g., in Germany) by the police. Therefore, we considered examining the effectiveness of initial retrieval support on the performance in T2-interviews with varying degrees of retrieval support (i.e., FR instead of CI as the T2-interview) to be highly relevant.

The question that remains to be answered is, whether a memory issue (i.e., actual forgetting of information), or a reporting issue (i.e., failure to overtly report details from memory) is accountable for the results of Experiment 1. Previous findings and theory render the first explanation unlikely. Retrieving information from memory increases the probability that this information will later be recalled (e.g., Ebbesen & Rienick, 1998; Foos & Fisher, 1988; McDaniel & Fisher, 1991), a phenomenon known as the *testing effect* (Roediger & Karpicke, 2006). According to network models of memory (e.g., J. R. Anderson, 1983), an early recall strengthens the network which results in a stronger memory trace. Moreover, it has been found that especially a high-quality initial recall can aid later recall attempts (e.g., Foos & Fisher, 1988; Marsh, Tversky, & Hutson, 2005; Tversky & Marsh, 2000). In Experiment 1, the SAI with its ample retrieval support elicited more correct details than the FR at T1 without affecting accuracy. Therefore, it should have led to better recall performance at T2, regardless of the T2-interview type.

Instead of a memory issue, we believe it likely that a reporting issue is responsible for the results. In their metamemory model, Koriat and Goldsmith (1996) distinguish between recall of information and overt reporting of this information as two separate components. People do not report everything they can recall. Instead, reporting details from memory is considered an active process which is influenced by one's current personal and situational goals as well as task demands (Bartlett, 1932; Neisser, 1996). In free report conditions, as in the present experiment, participants have the freedom to report or withhold information (Koriat & Goldsmith, 1994). Koriat and Goldsmith (1996) termed this the *control of report option*. In line with this, the retrieval support group may have withheld relatively more information than the FR group and, as a consequence, may have underreported at T2. The reason for this might be that the retrieval support group chose not to report certain details, because they "only" received FR instructions and did not assume that an elaborate account was expected of them. The sparse FR at T2

that provided little structure and retrieval support (relative to the SAI) may thus have led to a provision of only a bare minimum of details.

At this time, results regarding the effect of early retrieval support on subsequent recall performance are mixed (Brock et al., 1999; Hope et al., 2014; Memon et al., 1997) and knowledge about the conditions regarding when early retrieval support preserves memory is limited. Yet, witnesses are likely to be interviewed on several occasions in the course of the investigations (Bornstein, Liebel, & Scarborough, 1998). Therefore, it is important to study which measures need to be taken to ensure that recall is as comprehensive and accurate as possible in both the first and subsequent interviews. Hence, we followed the results of Experiment 1 up to shed light onto the conditions during which providing retrieval support at T1 can preserve memory for a subsequent interview. Here we examined the effect of initial retrieval support on subsequent recall by orthogonally manipulating the presence of retrieval support (i.e., FR vs. retrieval support with the SAI) both at T1 and at T2. The FR at T2 was supplemented with non-leading cued questions. This interview (i.e., FR and subsequent cued questions) is quite common in cases where a CI is not regularly administered. Both types of interviews used at T2 were meant to incite the expectation to provide an elaborate account. Moreover, analyses of the FR at T2 group with and without the responses to the cued questions enabled us to make direct comparisons with the findings of Experiment 1. This approach allowed us to disentangle, whether a memory issue or reporting issue was responsible for the results found in Experiment 1.

For T1-recall, we hypothesized that the retrieval support group would recall more details than the FR group, while there would be no influence on accuracy. For T2-recall, we expected that the retrieval support group would recall more details and at a higher accuracy than the supplemented FR (FR_{sup}) group. Furthermore, we predicted carry-over effects from T1 to T2-recall (i.e., a memory-preserving effect of initial retrieval support on subsequent recall). Specifically, participants who received retrieval support at T1, relative to the FR, were expected to report more details and at a higher accuracy one week later. Moreover, due to its beneficial effect, participants who received retrieval support at both T1 and T2 should be most accurate at T2. Participants who first had an FR and then an FR_{sup} should recall at the lowest accuracy, with the other two interview groups (retrieval support-FR_{sup}, FR-retrieval support) falling in-between.

EXPERIMENT 2

Method

Participants and Design

Eighty-seven native German participants took part in Experiment 2. Six participants were excluded from analyses due to non-attendance at T2, or a failure to comply with instructions, leaving 81 participants for analysis (59 women; age: $M = 22.22$, $Mdn = 21$, range: 19 to 33 years). The majority (80.2%) were psychology undergraduates. Remaining participants were recruited in the vicinity of the university. Psychology students and the other participants did not differ in their performance. In exchange for participation, participants either received course credit or a €15 voucher. Experiment 2 employed a 2×2 between-participants design. Presence of retrieval support at T1 (FR vs. retrieval support with the SAI) and presence of retrieval support at T2 (FR_{sup} vs. retrieval support with the SAI) were the independent variables.

Materials

The stimulus film, filler tasks, and interviews were identical to Experiment 1. The non-leading cued questions that supplemented the FR at T2 comprised eight questions regarding the incident, 14 questions regarding the appearance of the persons involved in the crime and one question concerning the appearance of other persons pictured (see Appendix E).

Procedure

In Experiment 2, all participants watched the film with full attention and, depending on the T1-retrieval support condition, completed either an SAI or an FR at T1. One week later (T2) participants either received an SAI or an FR_{sup}. In the FR_{sup} condition, after providing their free reports, participants answered cued questions regarding the incident, the perpetrator(s) and other persons involved. Participants were instructed to answer the cued questions as completely and accurately as possible, but were discouraged from guessing. They should skip a question or respond with “I don’t know” if they did not know the answer. We administered the questions in the FR_{sup} at T2 group only, because we assumed that, unlike an SAI, an ordinary FR would not create the expectation to provide an elaborate recall account. The T1-FR was not followed by cued questions, because we wanted to use the same T1-interviews as in Experiment 1 for the sake of comparability.

Coding

For the FR_{sup} at T2 group all details mentioned in the free report were used to calculate the dependent variables. This was done both when a detail was consistently reported in the free report and the cued questions and when there were contradictions. Details elicited by the cued questions were only added when the respective details had not been mentioned in the preceding free report. Following this approach, $M = 23.63\%$ ($SD = 7.24$) of the correct details and $M = 47.69\%$ ($SD = 19.24$) of the incorrect details reported at T2 came from the cued questions. Hence, unless stated otherwise, participants' performance in the FR at T2 includes the responses to the cued questions.

Results and Discussion

Table 4.2 displays the means and standard deviations of all dependent variables for T1 and T2.¹³

Recall at T1

To investigate performance at T1, we calculated one-way ANOVAs with presence of T1-retrieval support as the independent variable. Correct and incorrect details and accuracy were the dependent variables.

Analogous to Experiment 1, at T1 the retrieval support group ($M = 119.17$) recalled more correct details than the FR group ($M = 90.05$), $F(1, 79) = 29.09$, $p < .001$, $d = 1.20$. The number of incorrect details at T1 was also larger for the retrieval support group ($M = 14.78$) than for the FR group ($M = 11.45$), $F(1, 79) = 7.92$, $p = .006$, $d = -0.63$. Accuracy did not differ as a function of T1-retrieval support, $F(1, 79) = 0.03$, $p = .867$, $d = 0.04$. Thus, in line with our hypotheses and similar to Experiment 1 and previous findings (Gabbert et al., 2009; Gawrylowicz et al., 2014; Hope et al., 2014), we found that at T1 initial retrieval support elicited a more comprehensive recall than an initial FR, without compromising accuracy. Next, we examined recall performance at T2.

¹³ The distribution of the number of incorrect details at T2 was significantly skewed. However, the results of the analyses with the transformed variable yielded the same results. Therefore, we will report the analyses with the untransformed variable.

Table 4.2

Means and Standard Deviations for Number of Correct Details and Incorrect Details, and Accuracy at Time 1 and at Time 2 (Experiment 2)

	Time 1		FR Time 2		FR _{sup} Time 2		Retrieval support Time 2		Across interviews Time 2 ^a	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Number of correct details										
FR Time 1	90.05	22.55	80.50	14.96	115.35	17.57	117.95	22.66	116.65	20.06
Retrieval support Time 1	119.17	25.89	88.70	25.46	114.70	25.91	104.43	24.83	109.44	25.58
Across interviews Time 1	104.79	28.24	84.60	21.02	115.03	21.85	111.02	24.48	113.00	23.16
Number of incorrect details										
FR Time 1	11.45	5.03	10.65	4.90	25.30	6.62	20.30	6.91	22.80	7.14
Retrieval support Time 1	14.78	5.60	12.10	6.90	22.30	8.23	13.57	5.29	17.83	8.11
Across interviews Time 1	13.14	5.55	11.38	5.95	23.80	7.53	16.85	6.95	20.28	8.00
Accuracy (%)										
FR Time 1	88.72	4.27	88.45	4.60	82.09	3.45	85.41	3.49	83.75	3.81
Retrieval support Time 1	88.87	3.78	88.48	5.07	83.72	4.60	88.13	4.85	85.98	5.18
Across interviews Time 1	88.80	4.01	88.46	4.78	82.90	4.10	86.80	4.41	84.88	4.66

Note. FR = free recall data including details from free report, but not cued questions; FR_{sup} = free recall data including both details from free report and cued questions.

^aThe data in the column "Across interviews Time 2" is the mean of FR_{sup} at Time 2 and retrieval support at Time 2.

Recall at T2

T2 recall performance was analyzed using two-way ANOVAs with presence of T1-retrieval support and presence of T2-retrieval support as the independent variables. All analyses regarding the number of correct details at T2 were non-significant, $F_s(1, 77) \leq 1.92$, $ps \geq .170$, $|d|s \leq 0.31$, $\eta_p^2 = 0.02$. Other main effects or interactions not mentioned in the following were non-significant, $F_s(1, 77) \leq 3.84$, $ps \geq .054$, $|d|s \leq 0.38$, $\eta_p^2s \leq 0.02$. The same holds for non-significant post-hoc comparisons with alpha adjustments, $ps \geq .039$, $ds \leq 0.64$.

Number of incorrect details

Participants who received retrieval support at T1 ($M = 17.83$) recalled significantly fewer incorrect details one week later than participants who completed an FR at T1 ($M = 22.80$), $F(1, 77) = 10.28$, $p = .002$, $d = 0.65$. The T2-retrieval support group ($M = 16.85$) also recalled fewer incorrect details than the T2-FR_{sup} group ($M = 23.80$), $F(1, 77) = 20.47$, $p < .001$, $d = 0.96$.

Accuracy

Participants who received retrieval support at T1 ($M = 85.98\%$) recalled more accurately at T2 than participants who completed an FR at T1 ($M = 83.75\%$), $F(1, 77) = 5.53$, $p = .021$, $d = 0.49$. Likewise, the T2-retrieval support group ($M = 86.80\%$) provided a more accurate account than the T2-FR_{sup} group ($M = 82.90\%$), $F(1, 77) = 17.49$, $p < .001$, $d = 0.92$.

We also examined the effect of providing participants with retrieval support both at T1 and at T2 relative to the other interview combinations on recall accuracy. Descriptively, accuracy was highest in the group receiving retrieval support at both T1 and T2 ($M = 88.13\%$, $SD = 4.85$), followed by the FR/retrieval support group ($M = 85.41\%$, $SD = 3.49$), the retrieval support/FR_{sup} group ($M = 83.72\%$, $SD = 4.60$), and the FR/FR_{sup} group ($M = 82.09\%$, $SD = 3.45$). Overall, the difference between the groups was significant, $F(3, 77) = 7.91$, $p < .001$, $\eta_p^2 = 0.24$. As expected, post-hoc comparisons (with an adjusted Bonferroni correction of $\alpha = .017$; Shaffer, 1986) revealed that at T2, the group receiving retrieval support at both T1 and T2 recalled significantly more accurately than both the retrieval support/FR_{sup}, $p = .001$, $d = 0.93$, and the FR/FR_{sup} group, $p < .001$, $d = 1.43$. The FR/retrieval support group recalled more accurately than the FR/FR_{sup} group, $p = .014$, $d = 0.96$.

Consistent with Hope et al. (2014), we found a memory-preserving effect of initially providing participants with retrieval support: Participants who initially received retrieval support achieved a higher accuracy at T2 than participants who initially completed an FR. Somewhat unexpected, this was rather driven by the number of incorrect details than by the number of correct details. The beneficial effect of retrieval support is further revealed in the finding that participants who had retrieval support at both recall attempts recalled more accurately than participants who first completed an FR and then an FR_{sup}. The type of T2-interview is also important which is indicated by the recall of fewer incorrect details and higher accuracy of the T2-retrieval support group than the T2-FR_{sup} group.

Comparing T1 and T2 Recall

To further explore the pattern of results found at T2, we investigated memory performance over time, adding Time (T1 vs. T2) as a within-participants factor to the ANOVAs. In the following, we will only describe the significant results. The results of the ANOVAs can be found in Table 4.3.

Number of correct details over time

For correct recall, there were significant Time by T1-retrieval support and Time by T2-retrieval support interactions. Irrespective of the presence of T2-retrieval sup-

port, the number of correct details declined over time within the T1-retrieval support group (T1: $M = 119.17$; T2: $M = 109.44$), $F(1, 79) = 21.25$, $p < .001$, $d = 0.38$. In contrast, the number of correct details increased over time within the T1-FR group (T1: $M = 90.05$; T2: $M = 116.65$), $F(1, 79) = 154.91$, $p < .001$, $d = -1.23$. Irrespective of the presence of T1-retrieval support, providing participants with retrieval support at T2 did not have an influence on the number of correct details over time, $F(1, 79) = 1.48$, $p = .227$, $d = -0.16$. In contrast, completing an FR_{sup} at T2 resulted in an increase in the number of correct details (T1: $M = 102.78$, $SD = 29.11$; T2: $M = 115.03$, $SD = 21.85$), $F(1, 79) = 11.89$, $p = .001$, $d = -0.46$.

Table 4.3
Results of the ANOVAs of the Analysis over Time (Experiment 2)

		<i>F</i>	<i>p</i>	Effect size	
				η_p^2	<i>d</i>
Number of correct details					
Main effects	Time	33.78	< .001	-	-0.31
	RS-T1	4.76	.032	-	0.48
	RS-T2	< 0.01	.989	-	< -0.01
Interactions	Time x RS-T1	154.10	< .001	0.67	-
	Time x RS-T2	6.66	.012	0.08	-
	RS-T1 x RS-T2	1.51	.223	0.02	-
	Time x RS-T1 x RS-T2	0.02	.882	< 0.01	-
Number of incorrect details					
Main effects	Time	134.23	< .001	-	-1.02
	RS-T1	0.39	.534	-	0.14
	RS-T2	8.61	.004	-	0.65
Interactions	Time x RS-T1	43.34	< .001	0.36	-
	Time x RS-T2	28.55	< .001	0.27	-
	RS-T1 x RS-T2	2.98	.088	0.04	-
	Time x RS-T1 x RS-T2	0.12	.735	< 0.01	-
Accuracy					
Main effects	Time	107.75	< .001	-	0.90
	RS-T1	1.88	.174	-	0.31
	RS-T2	6.14	.015	-	0.56
Interactions	Time x RS-T1	7.27	.009	0.09	-
	Time x RS-T2	22.20	< .001	0.22	-
	RS-T1 x RS-T2	1.06	.307	0.01	-
	Time x RS-T1 x RS-T2	0.63	.428	0.01	-

Note. $df_{num} = 1$; $df_{denom} = 77$. RS-T1 = factor presence of retrieval support at T1; RS-T2 = factor presence of retrieval support at T2.

Number of incorrect details over time

For incorrect recall, there were also significant Time by T1-retrieval support and Time by T2-retrieval support interactions. Irrespective of the presence of T2-retrieval support, the number of incorrect details increased over time within the T1-retrieval support group (T1: $M = 14.78$; T2: $M = 17.83$), $F(1, 79) = 9.02$, $p = .004$, $d = -0.41$. Yet, within the T1-FR group the number of incorrect details doubled over time (T1: $M = 11.45$; T2: $M = 22.80$), $F(1, 79) = 121.99$, $p < .001$, $d = -1.80$.

Providing retrieval support at T2 resulted in an increase in the number of incorrect details over time, irrespective of the presence of T1-retrieval support (T1: $M = 13.05$, $SD = 4.93$; T2: $M = 16.85$, $SD = 6.95$), $F(1, 79) = 12.33$, $p = .001$, $d = -0.62$. Completing an FR_{sup} at T2 led to an even greater increase in incorrect details over time (T1: $M = 13.23$, $SD = 6.19$; T2: $M = 23.80$, $SD = 7.53$), $F(1, 79) = 92.89$, $p < .001$, $d = -1.52$.

Accuracy over time

For accuracy, there were also significant Time by T1-retrieval support and Time by T2-retrieval support interactions. Irrespective of the presence of T2-retrieval support, accuracy decreased over time within the T1-retrieval support group (T1: $M = 88.87\%$; T2: $M = 85.98\%$), $F(1, 79) = 23.10$, $p < .001$, $d = 0.60$. An even more marked decrease over time occurred within the T1-FR group (T1: $M = 88.72\%$; T2: $M = 83.75\%$), $F(1, 79) = 66.90$, $p < .001$, $d = 1.22$.

Similar results were obtained regarding the presence of T2-retrieval support. Irrespective of the presence of T1-retrieval support, accuracy decreased over time within the T2-retrieval support group (T1: $M = 88.95\%$, $SD = 3.87$; T2: $M = 86.80\%$, $SD = 4.41$), $F(1, 79) = 14.90$, $p < .001$, $d = 0.51$. The decrease of accuracy over time was even stronger within the T2-FR_{sup} group (T1: $M = 88.68\%$, $SD = 4.28$; T2: $M = 82.90\%$, $SD = 4.10$), $F(1, 79) = 104.76$, $p < .001$, $d = 1.38$.

In summary, although memory performance declined from T1 to T2, retrieval support at T1 relative to an FR seemed to limit memory deterioration: Providing participants with retrieval support at T1 led to a smaller increase in incorrect details and a smaller decrease of accuracy over time than an FR at T1. This is in line with a memory-preserving effect of providing initial retrieval support (Hope et al., 2014).

Recall Performance at T2 without the Cued Questions

To further clarify the question, whether a reporting issue was responsible for the results of Experiment 1, we analyzed recall performance at T2 without the responses to the cued questions that supplemented the FR at T2. Taking away those responses yields an FR that precisely corresponds to the T2-interview used in

Experiment 1. A replication of the null findings in these analyses would support the reporting issue explanation of Experiment 1. Non-replication of the findings would support the memory issue explanation.

Apart from a marginally significant main effect of presence of T1-retrieval support on the number of incorrect details, $F(1, 77) = 3.84, p = .054, d = 0.38$, presence of T1-retrieval support did not have significant effects on recall performance at T2, $F_s(1, 77) \leq 1.85, p_s \geq .178, d_s \leq 0.30$. Note that this differs from the analyses reported above that included the cued recall. The main effect of the presence of T2-retrieval support was significant for both correct, $F(1, 77) = 28.50, p < .001, d = 1.16$, and incorrect details, $F(1, 77) = 17.04, p < .001, d = -0.85$, but not accuracy, $F(1, 77) = 2.82, p > .097, d = -0.36$. The main effects were qualified by significant T1-Retrieval Support by T2-Retrieval Support interactions (correct details: $F[1, 77] = 4.76, p = .032, \eta_p^2 = 0.06$; incorrect details: $F[1, 77] = 9.22, p = .003, \eta_p^2 = 0.11$). Importantly, simple main effects analyses revealed that providing participants with retrieval support at T1 only had an influence on T2-recall performance when there was also retrieval support at T2. Compared to the group receiving retrieval support at both T1 and T2 ($M = 104.43$), the FR/retrieval support group ($M = 117.95$) tended to recall more correct details, $F(1, 77) = 3.73, p = .057, d = -0.57$. The FR/retrieval support group also recalled more incorrect details at T2 ($M = 20.30$) than the group receiving retrieval support at both T1 and T2 ($M = 13.57$), $F(1, 77) = 12.63, p = .001, d = 1.10$. Interestingly, as in Experiment 1, there was no effect of the presence of T1-retrieval support when the interview at T2 was an FR, $F_s(1, 77) \leq 1.34, p_s \geq .251, |d|s \leq 0.39$. As can be seen in Table 4.2, after the removal of the responses to the cued questions, a similar decline of the number of reported details as in Experiment 1 was apparent for the participants who first received retrieval support and subsequently completed the FR.

To sum up the results of Experiment 2, as expected, we found that initially providing witnesses with retrieval support preserves memory for a subsequent recall. That is, participants who received retrieval support relative to an FR at T1 recalled fewer incorrect details and hence provided more accurate accounts one week later. Moreover, the deterioration of memory over time was decelerated for the T1-retrieval support group relative to the T1-FR group. Note that these findings differ from the results obtained in Experiment 1. They are, however, consistent with the results obtained by Hope et al. (2014) and with network models of memory (e.g., J. R. Anderson, 1983), according to which an early recall opportunity preserves memory. This especially holds for high-quality initial recall opportunities (e.g., Ebbesen & Rienick, 1998; Foos & Fisher, 1988; Tversky & Marsh, 2000) that support retrieval, such as the SAI.

Because we used the same types of interview at T1 across experiments, the discrepant findings are very likely to be related to the type of interviews used at T2. Such an interpretation is supported by additional data analyses of T2-recall in Experiment 2 that took into account the details mentioned in the FR only and ex-

cluded details elicited by the cued questions. By removing those details, the T2-FR of Experiment 2 was matched the T2-interview of Experiment 1. In these additional analyses, we replicated the results from Experiment 1. Specifically, there was no significant effect of the presence of T1-retrieval support on subsequent recall when an FR was applied at T2.

These results support our hypothesis that a reporting issue was responsible for the failure to detect any differences at T2 across retrieval support conditions in Experiment 1. We assume that the different interviews used at T2 in Experiment 1 vs. 2 produced different expectations about reporting. In free recall conditions, people can decide which information to withhold and which to report (Koriat & Goldsmith, 1994, 1996). This *control of report option* is influenced by current goals and task demands, such as the assumption that a comprehensive account of what one has witnessed is expected. Accordingly, in Experiment 1 the retrieval support group may have withheld relatively more information than the FR group, because the scant free recall instructions did not generate the expectation that they should report a detailed and comprehensive account at T2. Consequently, the retrieval support group may have underreported. This was not the case in Experiment 2 when participants received more elaborate recall instructions.

General Discussion

In two experiments, seeking to increase knowledge about ways to promote the reliability of eyewitness statements, we investigated the effect of initially providing witnesses with retrieval support on recall performance. For this purpose, we contrasted an interview with a high amount of retrieval support, the recently developed SAI (Gabbert et al., 2009), with an interview that does not provide retrieval support, the FR. The aim of Experiment 1 was to test the effect of retrieval support when the witnessing conditions were suboptimal, because the witness was distracted during the crime. Previous research has comprehensively shown the negative effects of divided attention on memory performance (e.g., Craik et al., 1996; Lane, 2006), but little is known about ways in which to compensate its negative effects. Based on the outshining hypothesis (Smith, 1988, 1994), we expected retrieval support to be especially effective relative to an FR when the participants watched the stimulus film with divided vs. full attention. However, the predicted interaction pattern did not emerge. Note, however, that the retrieval support group did not perform worse in terms of recall accuracy than the FR group in any condition. These findings have relevance for interview practice in real cases. Specifically, the police should inquire whether the witness may have been distracted during the crime before administering the initial interview, because the advantage of providing retrieval support could be diminished. Nevertheless, it is not harmful to provide witnesses with retrieval support in the initial interview when they were dis-

tracted during the crime. As earlier findings indicate (Gabbert et al., 2009, Experiment 2), this is preferable to a situation where no initial interview is carried out at all.

In Experiment 1, we observed an unexpected tremendous loss of details over time within the retrieval support group rather than a memory-preserving effect of initial retrieval support on T2-recall performance. Since witnesses are likely to be interviewed several times (Bornstein et al., 1998), it is important to know when initial retrieval support can preserve memory for subsequent recall. Therefore, we addressed this question in Experiment 2 by orthogonally manipulating the presence of retrieval support at T1 and T2. Here we found a memory-preserving effect of initial retrieval support. The pattern of results supported our hypothesis that a reporting issue could explain the non-significant results of Experiment 1 (Koriat & Goldsmith, 1996). Apparently, the sparse FR at T2 used in Experiment 1 led to a provision of only a bare minimum of details, thus undermining the memory-preserving effect of initial retrieval support. From an applied perspective, the results clearly show that police investigators should be aware that witnesses about to testify need to know that a comprehensive account is expected from them, otherwise an initial interview with ample retrieval support is unlikely to have the desired effect on the subsequent recall attempt. This could, for example, be established during the rapport building phase between interviewer and interviewee (Collins, Lincoln, & Frank, 2002; Vallano & Schreiber Compo, 2011). Our data also show that the type of interview at T2 is important for recall performance, even after initial retrieval support was provided, as the T2-retrieval support group recalled more accurately than the T2-FR_{sup} group. In summary, this means that an initial comprehensive retrieval support interview assists the police during the investigations immediately after a crime by preserving memory while this does not exempt police from applying best practice during the subsequent personal interview.

One remark on the difference in accuracy at T2 as a function of T1-retrieval support in Experiment 2 seems in order. Although inspection of the means suggests a rather small difference between the interview groups (about 2.2%) it should be noted that the respective effect size was moderate. Moreover, the difference in T2-accuracy as a function of presence of T1-retrieval support was driven by a large increase in incorrect details in the T1-FR group relative to the T1-retrieval support group. Specifically, the former group recalled almost 30% more incorrect details at T2 than the latter. Given that every single incorrect detail can entail dire consequences, this again highlights the importance of initial retrieval support.

Another issue worth mentioning concerns the large standard deviations we observed for both retrieval support and FR groups in both experiments, especially regarding the number of correct details. This indicates that there are substantial differences in the potential of the interviews to elicit information from partici-

pants. Given the homogeneity of each sample, this is unlikely to reflect individual differences in memory performance alone, because undergraduates are not likely to differ much in this cognitive task (see Peters, Jelicic, Verbeek, & Merckelbach, 2007). Rather, this could reflect differences in the control of the report option, that is, the readiness to report details from memory (Koriat & Goldsmith, 1996). That is, some participants may have a stricter response criterion than others. As a consequence, the former will provide fewer details than the latter. Future research should look into this issue.

Turning to the limitations of the present studies, one limitation pertains to the use of a stimulus film instead of a staged crime and the fact that it showed a rather non-arousing (i.e., not stress-inducing) crime. Another limitation applies to the use of an undergraduate sample rather than participants from the general public, with undergraduates most likely being more skilled at writing than the average witness. However, we do not think that any of these issues poses a threat to the validity of our results for the following reasons. In meta-analyses on the CI which is closely related to the SAI, medium of presentation (live vs. video) was either not found to be a moderator of recall performance (Memon, Meissner, & Fraser, 2010) or, contrary to what one would expect, effect sizes were larger when the event was staged than when a video was shown (Köhnken, Milne, Memon, & Bull, 1999). Therefore, we conclude that using a film probably did not influence our results or, if anything, led to an underestimation of the true effects. Meta-analytically, type of event (neutral vs. arousing) had a small, albeit significant, influence for correct recall such that arousing events yielded smaller effect sizes than neutral events (Memon et al., 2010). However, for both neutral and arousing events, the effect sizes in favor of the CI were very large ($d > 1$), suggesting that the use of a neutral event was unproblematic in the present study. Moreover, not all types of crime are likely to induce stress in the witness. Also, especially when there are multiple witnesses, there are probably bystanders who are not directly involved or watching the crime from a distance and hence, may not be stressed by the crime. Note that crimes with multiple witnesses are also the cases in which the SAI is most likely to be used (Hope et al., 2011). Still, we encourage future projects examining whether more arousing events interact with retrieval support with the SAI which would also increase and broaden the existing SAI database. Finally, results of a different SAI study suggest that the composition of the sample may not be decisive for the effectiveness of the SAI, at least for T1-recall. Using a sample recruited from the general public, Gawrylowicz et al. (2014) found that at T1 the SAI group recalled more correct details than the FR group, whereas accuracy did not differ. In the present studies we found the same results for T1-recall performance and obtained similar effect sizes with an undergraduate sample. Future studies should nevertheless examine recall performance with the SAI across different witness populations.

Taken together, the results of the current experiments replicate and extend earlier findings concerning the effect of early retrieval support and are highly rele-

vant for police interview practice. Specifically, like Gabbert et al. (2009) and Hope et al. (2014), we detected that providing retrieval support during the initial interview elicits a more comprehensive, albeit not more accurate, initial account than an FR. Unexpectedly, an interview with retrieval support is not more effective than an FR when the witness was distracted during the incident. Finally, reluctance to report was identified as an important determinant regarding whether a memory-preserving effect after conducting an initial interview with retrieval support emerges. The impact of an initial interview with ample retrieval support on a subsequent interview is greatest when the latter is a high-quality interview and when the witness assumes that a comprehensive account is expected.

Appendix E: Cued Questions Following the Free Recall at T2 (Experiment 2)

Questions relating to the sequence of events.

1. How many people in total could be seen in the film?
2. Describe the crime scene!
3. What was stolen?
4. What precisely did the victim do so that the perpetrator/s could commit the theft?
5. Did the perpetrator/s have (an) accomplice/s? If yes, what did the accomplice/s do?
6. Describe how the perpetrator/s executed the theft!
7. Where did the perpetrator/s put the stolen object(s)?
8. If other people were present, what did they do during the theft?
9. Are there any further details regarding the course of events which you remember but which we have not asked you about? If yes, which details are these?

Questions relating to the appearance of the people involved in the crime (i.e., perpetrator and accomplice).

1. What was the sex of the person?
2. How old was the person?
3. How tall was the person in cm?
4. How heavy was the person in kg?
5. What was the ethnicity of the person?
6. Describe the figure of the person!
7. Describe the clothing of the person!
8. Did the person wear any jewelry or accessories? If yes, please describe!
9. What was the hair color of the person?
10. Describe the hair style and the hair length of the person!
11. Did the person have any distinctive features (e.g., scars, moles, or tattoos)? If yes, how did they look like?
12. Did the person wear glasses? If yes, how did they look like?
13. Did the person have facial hair? If yes, how did it look like?
14. Are there any other details about the person which you remember but which have not been asked about? If yes, which are these?

Question relating to all other people present in the film.

1. Were there any other people present in the film? If yes, please describe them.

CHAPTER 5 Retrieval Support through Repeated Interviews

Krix, A. C., Sauerland, M., Lorei, C., & Rispens, I. (2014). Consistency across repeated eyewitness interviews: Contrasting police detectives' beliefs with actual eyewitness performance. Manuscript submitted for publication.

Abstract

In the legal system, inconsistencies in eyewitness accounts are often used to discredit witnesses' credibility. This is at odds with research findings showing that witnesses frequently report reminiscent details (details previously unrecalled) at an accuracy rate that is nearly as high as for consistently recalled information. The present study sought to put the validity of beliefs about recall consistency to a test by directly comparing them with actual memory performance in two recall attempts. All participants watched a film of a staged theft. Subsequently, the memory group ($N = 84$) provided one statement immediately after the film (either with the Self-Administered Interview© or free recall) and one after a one-week delay. The estimation group ($N = 81$) consisting of experienced police detectives estimated the recall performance of the memory group. The results showed that actual recall performance was consistently underestimated. Also, a sharp decline of memory performance between recall attempts was assumed by the estimation group, whereas actual accuracy remained stable. While reminiscent details were almost as accurate as consistent details, they were estimated to be much less accurate than consistent information and as inaccurate as direct contradictions. The police detectives expressed a great concern that reminiscence was the result of suggestive external influences. In conclusion, it seems that experienced police detectives hold many implicit beliefs about recall consistency that do not correspond with actual recall performance. Recommendations for police trainings are provided. These aim at fostering a differentiated view on eyewitness performance and the inclusion of more comprehensive classes on human memory structure.

Imagine a witness who is interviewed by the police at the crime scene. He describes the perpetrator as middle-aged, slim, brown-haired, and wearing jeans. Several weeks later, the witness is invited to a second police interview after being identified as a key witness. Here, he provides another perpetrator description that is mostly consistent with the earlier description. However, he now remembers the perpetrator to be blonde (instead of brown-haired) and comes up with new information, that is, a black cardigan.

In fact, witnesses are likely to be interviewed several times during the investigations (Bornstein, Liebel, & Scarberry, 1998) and consequently, it is possible that contradictory or new details are recalled. Chances are that the witness of this example will not be considered very credible in court, as jurors are explicitly instructed to attend to inconsistencies in the statements to determine the witness' credibility (Sixth Circuit Criminal Pattern Jury Instructions, 2013). Inconsistent statements are regarded as an indicator of overall inaccuracy by police officers, lawyers, and mock-jurors alike (Brewer, Potter, Fisher, Bond, & Luszcz, 1999; Uviller, 1993). This belief, however, is in direct contrast to research findings showing that (in)consistency of recall is not or only weakly related to overall recall accuracy (Brewer et al., 1999; Odinot, Memon, La Rooy, & Millen, 2013; Oeberst, 2012; Smeets, Candel, & Merckelbach, 2004).

When it comes to inconsistencies, it is important to distinguish the different types of inconsistency. While direct contradictions generally imply that at least one of the reported details is incorrect, this is not applicable to reminiscent details (i.e., details that are recalled only in a later, but not in the first interview). Even though reminiscence seems at odds with the idea of declining memory over time, previous research has shown that it is a widespread phenomenon and that forgetting does not preclude reminiscence (Erdelyi, 2010; La Rooy, Katz, Malloy, & Lamb, 2010; Scrivner & Safer, 1988). In fact, when recalling the same incident twice, almost all participants exhibit reminiscence (Brock, Fisher, & Cutler, 1999; Gilbert & Fisher, 2006; Hope, Gabbert, Fisher, & Jamieson, 2014) and accuracy rates of reminiscent details can be quite high. Although the accuracy of the latter may not always be as high as the accuracy of consistent details (Brock et al., 1999; Hope et al., 2014; Odinot, Memon, et al., 2013), it is much higher than the accuracy of contradictory details (Brock et al., 1999; Gilbert & Fisher, 2006).

Recently, research has shown that the consistency of recall is influenced by the type of interview used at the first recall attempt (T1). Specifically, Hope et al. (2014) compared recall performance between two groups who completed either a Self-Administered Interview© (SAI) or a free recall (FR) as the T1-interview and underwent a Cognitive Interview one week later (T2). The SAI is based on the Cognitive Interview (Fisher & Geiselman, 1992) and adopts some of its memory-enhancing components, such as mental context reinstatement, and unlike the FR provides ample retrieval support. It was developed to elicit an early comprehensive statement (Gabbert, Hope, & Fisher, 2009). Recall with the SAI elicited a great-

er proportion of consistent details, but a smaller proportion of reminiscent details than recall using FR (Hope et al., 2014). The proportion of forgotten and contradictory details and the accuracy rates of all different consistency categories did not differ as a function of T1-interview type. The authors explained the results by pointing out that at T1 the SAI group reported much more information than the FR group. As a consequence, there may have been less opportunity to report reminiscent details in the SAI relative to the FR group.

Given that reminiscence is a common phenomenon, the question arises how it can be accounted for from a scientific perspective. Drawing on cognitive theory of memory, Fisher, Brewer, and Mitchell (2009) argue that reminiscence may be caused by a change of retrieval cues. This reasoning is based on the well-known principle of varied retrieval (Tulving & Watkins, 1975). According to this principle, details that cannot be retrieved with one technique may well become accessible with another one. Hence, it is assumed that reminiscence occurs if a retrieval cue is present in the subsequent but not in the first interview. The more disparate the two retrieval attempts the more reminiscence should emerge. The second important principle identified by Fisher et al. (2009) that is of relevance for reminiscence is the independence of components. More specifically, complex events consist of many independent components (Brewer et al., 1999; Fisher & Chandler, 1991). As a result, incorrectly recalling one component should not have an influence on recall of other components. Cognitive theory thus predicts that reminiscence should occur frequently. Moreover, accuracy of reminiscent details should not be low in principle, but among other things depend on the nature of the second interview. Finally, due to the independence of components, any type of inconsistency should be unrelated to overall accuracy.

The question remains as to why laypersons are so skeptical about reminiscence. In an attempt to capture laypersons' beliefs about recall consistency, Fisher et al. (2009) formulated the courtroom approach of memory. It should be made very clear that this approach is not a scientific theory, but a conglomerate of beliefs assumed to be held by laypersons about memory performance across several retrieval attempts. In the courtroom approach, the central belief is that memories fade over time. Reminiscence constitutes a violation of this principle. As a consequence, laypersons should hold the following assumptions about recall of reminiscent and contradictory statements: 1) Both reminiscent and contradictory details should be inaccurate (relative to consistent details) and indicators of overall inaccuracy. 2) Reminiscence should occur infrequently. 3) Reminiscence should be the result of external influences (e.g., co-witness information). Fisher et al. (2009) derived these assumptions from observations of how judges and attorneys behave in the courtroom but did not directly test them. Pointing out that the existing database is "not robust" (p. 133), the authors encouraged further empirical studies to test laypersons' assumptions summarized in the courtroom approach and extend the existing database on recall consistency.

To our knowledge, there has only been one study that directly tested the beliefs summarized in the courtroom approach and compared the assumptions held by laypersons with regards to performance across two recall attempts with actual recall performance. In this study, all participants watched a film about an altercation (Oeberst, 2012, Experiment 2). Subsequently, a memory group provided two statements about the content of the film, one immediately and one after a week. A second group of participants who had also seen the film estimated the recall performance of the memory group. This estimation group assumed that overall recall accuracy would decline over time and that the accuracy of reminiscent details would be lower than the accuracy of consistent and forgotten details. However, this did not correspond with the performance of the memory group at all. In fact, actual recall accuracy remained stable over time and the accuracy of reminiscent, consistent, and forgotten details did not differ from each other. Moreover, recall accuracy was underestimated invariably. This was especially striking regarding the reminiscent details: While the actual obtained accuracy rate was 84%, estimated performance was as low as 19%.

Even though this study by Oeberst (2012) yielded interesting findings and found support for both the existence of the beliefs summarized in the courtroom approach (Fisher et al., 2009) and their discrepancy with reality, it has three important limitations. The first limitation pertains to the ecological validity of the sample, which was comprised of students who are unlikely to have experience with witnesses. Hence, it is unclear whether the findings transfer to individuals involved in the legal process, such as police detectives. Second, the study was ignorant of two elements contained in the courtroom approach. Specifically, beliefs about the accuracy of contradictions were not examined. A comparison of the estimated accuracy of contradictory details (of which at least one detail is incorrect) and reminiscent details, however, could yield interesting insights into the credibility attributed to reminiscence. Furthermore, while the accuracy rates of consistent, reminiscent, and forgotten details were collected in previous research (Oeberst, 2012), the observed and estimated prevalence of the consistency categories were not. Obtaining this information would be illuminative as to whether the frequency of occurrence of the different consistency categories is over- or underestimated and would also shed light onto the attributed credibility of reminiscence. The third limitation is that the study remained silent concerning potential reasons why reminiscent details are considered incorrect. Knowledge about these reasons could serve as the starting point to revise the content of police training courses so as to increase knowledge on the principles of human memory and to reduce potential bias against reminiscence.

It was the aim of the present study to test laypersons' assumptions regarding recall performance over time and consistency of recall that are summarized in the courtroom approach of memory (Fisher et al., 2009). As in previous research (Oeberst, 2012), this was done by contrasting estimates of recall performance with

actual recall performance across two recall attempts. More specifically, we sought to extend the findings of Oeberst's (2012) study and to address the limitations of the study, using the following means. First, instead of a student sample, we relied on an ecologically valid sample of experienced police detectives as the estimation group. Second, to test the assumptions about the frequency of reminiscence, we collected estimates of the frequencies of the different consistency categories in addition to the estimates of their accuracy rates. Third, going beyond documenting the discrepancies between laypersons' beliefs and actual recall performance, we inquired about reasons for mistrust in reminiscence. Fourth, we considered the recent finding that consistency rates vary as function of T1-interview type (Hope et al., 2014), because this may influence the extent of the discrepancy between laypersons' beliefs and actual recall performance. Therefore, we varied the T1-interview type (SAI vs. FR) both in the memory and in the estimation group.

Based on the beliefs summarized in the courtroom approach and previous findings (Oeberst, 2012), we hypothesized that recall accuracy would be underestimated without exception. Second, we expected an interaction such that the estimated decrease of accuracy over time would be stronger than actual decrease of accuracy. Third, the discrepancy between the accuracy of reminiscent details and the accuracy of consistent and forgotten details would be greater in the estimation than in the memory group. Fourth, the discrepancy between the accuracy of reminiscent and contradictory details would be smaller in the estimation than in the memory group. Finally, it was hypothesized that the proportion of consistent and forgotten details would be overestimated, and the proportion of reminiscent and contradictory details underestimated.

Method

Ethics Statement

This study was approved by the ethics committee of the Faculty of Psychology and Neuroscience of Maastricht University and follows the rules stated in the Declaration of Helsinki. All participants read and signed a written informed consent.

Participants and Design

The memory group consisted of $N = 84$ German police students (22 women; $n = 23$ members of the criminal police and $n = 61$ of the uniformed police; age: $M = 24.6$, $Mdn = 23$, range: 20 to 39 years). None of the participants of the memory group indicated having previous knowledge of or practical experience with the SAI (note that the SAI was neither part of the students' training nor in use by this German police force at the time of testing).

The estimation group consisted of $N = 93$ detectives of the Dutch police. Twelve participants of the estimation group had to be excluded, because they were shown a wrong film version, leaving $N = 81$ for analysis (26 women; age: $M = 42.2$, $Mdn = 42$, range: 26 to 58 years) with an average work experience of $M = 19.7$ years ($SD = 9.4$; range 5 to 38 years) and an average experience of $M = 12.8$ years ($SD = 8.9$; range 0 to 38 years) in interviewing eyewitnesses. Of these, 35% had heard of the SAI prior to the study, mostly from interview courses (75%) or colleagues (18%), but no one had used it. Participation occurred on a voluntary basis.

The study employed a 2 (group: memory vs. estimation) \times 2 (interview type at the first recall attempt: SAI vs. FR) between-participants design. The dependent variables were overall accuracy at T1 and T2, accuracy and proportion of details mentioned at T1 but forgotten at T2, as well as accuracy and proportion of consistent, reminiscent, and contradictory details at T2. Figure 5.1 provides an overview of the design.

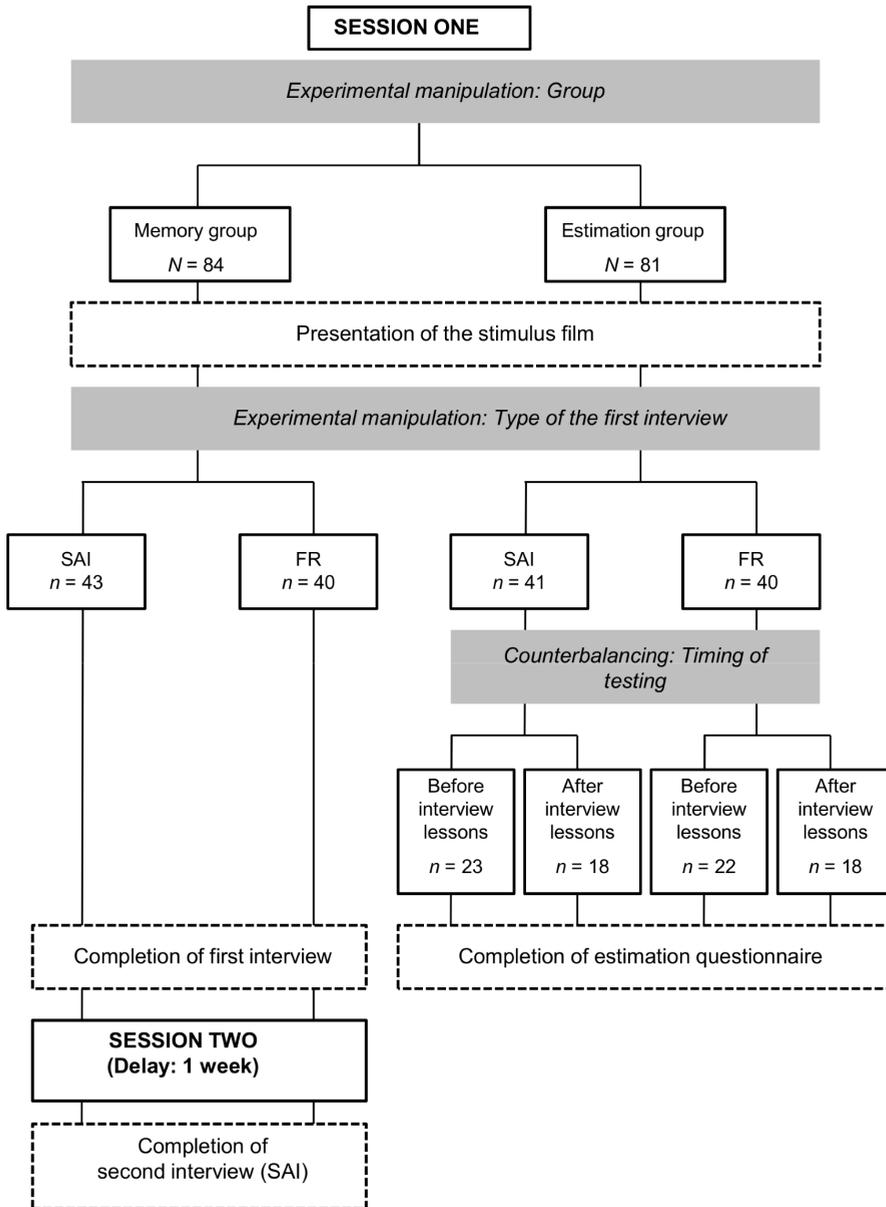


Figure 5.1. Design and procedure of the study.

Materials

Film

To test an unrelated research question, we created two different film versions (A and B) from the same video footage that showed the non-violent theft of a wallet. The films were identical apart from a few short sequences that were shown in one film version, but were removed in the other version and vice versa (e.g., thief and accomplice point to the victim or thief disposes of the emptied wallet). Note, however, that the act of stealing was present in both versions. Both film versions were edited to last approximately 2:10 min. In both films, the female victim chats with two friends (male and female) and is observed from a distance by the thief and the accomplice (both male). After the friends have left the scene, the accomplice approaches and distracts the victim by pretending to ask for directions. At the same time, the thief steals the wallet from her purse and subsequently both perpetrators leave the scene. The two friends return and the victim realizes that her wallet is missing. In a last scene, the thief is shown inspecting the content of the wallet.

Interviews used in the memory group

Self-Administered Interview

In the SAI (Gabbert et al., 2009; see Hope, Gabbert, & Fisher, 2011, for a more detailed description), participants first mentally reinstated the context by thinking back to the incident, picturing in their minds what they saw and recreating the thoughts and emotions they had at the time. Supported by non-leading recall cues, in separate sections, they provided descriptions of the course of events, the appearance of the perpetrator(s), and, if applicable, of potential other witnesses or vehicles involved. To promote multiple and varied retrieval, participants were also requested to draw a sketch of the scene, so as to facilitate the retrieval of spatial information. In a final section, the witnessing conditions were prompted (e.g., What time of day did the event occur?). Throughout the interview, participants were instructed to provide the most complete and accurate account of the witnessed incident possible, but not to guess.

Free recall

The FR comprised only one retrieval attempt and requested participants to report all details they could remember about the sequence of actions and events, and of all persons involved, including the perpetrator(s) and other witnesses (see Gabbert et al., 2009). Analogous to the SAI, participants were instructed to provide the most complete and accurate account possible, but were discouraged from guessing.

Questionnaire used in the estimation group

In the estimation questionnaire, participants of the estimation group were informed that another group (i.e., the memory group) had also watched the same film and that it would be their task to estimate the recall performance of this group. No information was disclosed as to who the memory group was to prevent any potential biases. Participants were informed that the interview took place directly after the film and were provided with a brief description of either the SAI or FR, depending on the T1-interview condition. Participants were then requested to estimate the overall accuracy of the memory group at T1. For this purpose, the term accuracy and how accuracy is calculated were explained in simple terms. After providing this estimate, participants were informed that one week later there was a second interview about the film in the form of the SAI. For the participants in the FR condition, the description of the SAI was provided at this point. Subsequently, participants estimated the accuracy and the proportion of the details that were mentioned at T1 but forgotten at T2, the overall accuracy, and the accuracy and proportion of the consistent, reminiscent, and contradictory details in the second interview. For the sake of comprehensibility, examples of all consistency categories were provided. Participants were also asked to briefly explain in their own words why they regarded reminiscent details as accurate or inaccurate, respectively.

Procedure*Memory group*

The participants of the memory group were tested in groups of up to 20 people (i.e., one class of students) during their lessons at the police university. In the first session, participants were told that as part of an eyewitness experiment they would view a film and subsequently answer some questions about it. After signing the informed consent and providing demographic information, participants of each class watched either film version A or B determined by random assignment. After a delay of about 30 minutes during which the lesson was continued, participants completed either an SAI or an FR. They were informed that there would be a second session one week later. However, nothing was disclosed about the content of this session. In this second session, all participants filled in an SAI as the T2-interview. Finally, participants received a recognition test to address an unrelated research question. In both sessions, individual seating was ensured and, to prevent motivation loss among the slower participants, the group was only dismissed after the last student had finished.

Estimation group

The participants of the estimation group were also tested in groups. Testing occurred in the framework of an extensive course offered to detectives by the Dutch police academy on how to investigate comprehensive cases. This course consisted of several units, one of which dealt with interviewing witnesses and the trustworthiness of witness accounts. Note that the course covered the SAI, but not the issue of repeated interviewing and reminiscence. Timing of testing was counterbalanced so that half of the classes were tested before and half after the lessons on interviewing witnesses (see Figure 5.1). Participants were informed that they would take part in an experiment on eyewitness accounts and that they would watch a short film that would be followed by a questionnaire. After signing the informed consent and providing demographic information, they were shown film A¹⁴ and subsequently completed one of two versions of the estimation questionnaire (i.e., either the version describing the SAI as the first interview or the version describing the FR as the first interview). Upon completion of data collection all participants were fully debriefed.

Coding

The statements of the memory group were transcribed and coded employing a comprehensive coding scheme consisting of 685 details, of which approximately 56% were person, 19% setting, 13% action, and 12% object details. The statement “The thief (1) sat (2) at the rightmost (3) table (4).” would yield four details (see chapter 2, for a similar approach). Subjective responses, such as “He was ugly”, were not scored. A detail was considered correct, if it matched the content of the film, and considered incorrect, if it did not match the content of the film. Details were considered confabulated when they were both incorrect and non-existent (e.g., recalling a hat when no headgear was worn; see chapter 2).

Details were also coded for consistency. Specifically, consistent details were recalled in both interviews (e.g., describing the thief as blonde in both interviews). Details mentioned in the second interview that were in direct conflict with details mentioned in the first interview were considered contradictory (e.g., describing the hair color as brown in the second and as blonde in the first interview). For determining the accuracy of contradictions, the detail reported in the second interview was used. Forgotten details were only reported in the first, but not the

¹⁴ We used film A for the estimation group only, because we considered the two film versions to be equivalent. Note that within the memory group, there were no significant main effects of film on any of the dependent variables. For T2-accuracy and the proportion of forgotten details, there were significant Film by T1-Interview interactions. However, as there was no apparent reason or theoretical rationale for the interactions to occur, we decided not to consider the film version in the estimation group.

second interview and reminiscent details were only reported in the second interview.

Inter-coder reliability

To establish inter-coder reliability, the randomly selected statements of 15 participants from the memory group were independently coded by two coders. Across all categories, Cohen's κ was 0.88, $p < .001$, indicating *almost perfect* agreement (Landis & Koch, 1977).

Results

An *alpha* level of .05 was used for all statistical tests. We report Cohen's *d* (Cohen, 1988) for dependent or independent samples for main effects with $df = 1$ in the numerator and η_p^2 for main effects with $df > 1$ and interaction effects (Sporer & Cohn, 2011). When the estimates of the estimation group did not differ as a function of timing of testing (before vs. after the eyewitness lessons), $ps \geq .168$, they were collapsed across timing. Otherwise, analyses were run separately for the two subgroups. When the dependent variables did not differ as a function of T1-interview type, $ps \geq .108$, T1-interview type was removed and the analyses rerun.

Accuracy

Table 5.1 displays the means and standard deviations of T1 and T2 recall accuracy, as well as accuracy of the different consistency categories of the memory group and the corresponding estimates of the estimation group.

Recall performance over time

To test whether the courses of actual and estimated accuracy rates over time differed, we calculated ANOVAs with group (memory vs. estimation) and recall attempt (T1 vs. T2) as the independent variables. Accuracy at T1 and T2 were the dependent variables. The significant main effects of group, $F(1, 163) = 781.06$, $p < .001$, $d = 4.35$, and time, $F(1, 163) = 161.77$, $p < .001$, $d = 0.31$, were qualified by a significant interaction, $F(1, 163) = 117.89$, $p < .001$, $\eta_p^2 = 0.42$. As expected, the analysis of the simple main effects revealed that within the memory group, accuracy remained stable from T1 ($M = 87.47\%$, $SD = 4.70$) to T2 ($M = 85.97\%$, $SD = 4.81$), $F(1, 163) = 1.76$, $p = .186$, $d = 0.32$. In contrast, the estimation group assumed that accuracy would sharply decline from T1 ($M = 46.19\%$, $SD = 18.31$) to T2 ($M = 27.10\%$, $SD = 16.36$), $F(1, 163) = 272.96$, $p < .001$, $d = 1.09$. Also as expected, the

accuracy of the memory group was significantly underestimated both at T1, $F(1, 163) = 399.90, p < .001, d = -3.11$, and at T2, $F(1, 163) = 998.84, p < .001, d = -4.92$.

Table 5.1

Accuracy of Overall T1 and T2 Recall and of the Consistency Categories for the Memory Group and the Corresponding Estimates of the Estimation Group

	Memory group		Estimation group	
	<i>M (%)</i>	<i>SD</i>	<i>M (%)</i>	<i>SD</i>
Overall T1	87.47	4.70	46.19 ^c	18.31
Overall T2	85.97	4.81	27.10 ^c	16.36
Consistent	91.96 ^{ab}	3.61	55.00 ^{de}	21.85
Forgotten	86.16 ^a	8.01	37.90 ^{de}	20.14
Reminiscent	85.53 ^b	8.32	28.59 ^d	19.86
Contradictory	39.87 ^{ab}	30.48	25.34 ^e	15.17

Note. Means in a column sharing the same superscript differ at $p < .05$.

Consistency categories

To analyze whether accuracy rates of the consistency categories differed within the memory and the estimation group, we calculated mixed-model ANOVAs with group (memory vs. estimation) and consistency category as the independent variables. Accuracy of the four consistency categories were the dependent variables. The significant main effects of group, $F(1, 156) = 502.77, p < .001, d = 3.58$, and consistency category, $F(3, 468) = 175.70, p < .001, \eta_p^2 = 0.53$, were qualified by a significant interaction, $F(3, 468) = 49.94, p < .001, \eta_p^2 = 0.24$. Post-hoc analyses with Bonferroni tests¹⁵ showed that within the memory group, accuracy of the consistent details ($M = 91.96\%, SD = 3.61$) was significantly higher than accuracy of the forgotten ($M = 86.16\%, SD = 8.01$), $p = .024, d = 0.89$, reminiscent ($M = 85.53\%, SD = 8.32$), $p = .011, d = 1.03$, and contradictory details ($M = 39.87\%, SD = 30.48$), $p < .001, d = 2.54$. Both forgotten, $p < .001, d = 2.07$, and reminiscent details, $p < .001, d = 2.12$, were more accurate than the contradictory details. Accuracy of forgotten and reminiscent details did not differ, $p = .999, d = 0.12$.

Within the estimation group and analogous to the memory group, estimated accuracy of the consistent details ($M = 55.00\%, SD = 21.85$) was significantly higher than estimated accuracy of the forgotten ($M = 37.90\%, SD = 20.14$), $p < .001, d = 0.79$, reminiscent ($M = 28.59\%, SD = 19.86$), $p < .001, d = 1.26$, and contradictory details ($M = 25.34\%, SD = 15.17$), $p < .001, d = 1.53$. However, in contrast to the results found in the memory group, estimated accuracy of the forgotten details was higher than estimated accuracy of both reminiscent, $p < .001, d = 0.47$, and contra-

¹⁵ As the test yields adjusted significance levels, the usual level of .05 to determine statistical significance can be applied.

dictory details, $p < .001$, $d = 0.70$. Interestingly, the estimated accuracy of the reminiscent and contradictory details did not differ, $p = .999$, $d = 0.18$.

Also in line with our hypotheses, the actual accuracy of all consistency categories was underestimated (consistent: $F[1, 163] = 233.73$, $p < .001$, $d = -2.38$; forgotten: $F[1, 162] = 413.57$, $p < .001$, $d = -3.18$; reminiscent: $F[1, 163] = 584.33$, $p < .001$, $d = -3.76$; contradictory: $F[1, 157] = 14.55$, $p < .001$, $d = -0.60$).

To summarize the results regarding actual and estimated accuracy of recall, all our hypotheses were confirmed. First, recall accuracy was underestimated without exception. Second, while actual accuracy remained stable from T1 to T2, a sharp decline was assumed by the estimation group. Third, the estimation group expected that reminiscent details would be as inaccurate as contradictory details and much less accurate than consistent details. In reality, even though consistent details were most accurate, accuracy of reminiscent details was at a high level and substantially higher than accuracy of contradictions. The results are in line with previous findings (Oeberst, 2012) and confirm the existence of the beliefs summarized in the courtroom approach of memory (Fisher et al., 2009) that do not correspond with actual recall performance. In the following, the results on the proportion of the consistency categories will be reported.

Proportion of the Different Consistency Categories

To compare the estimated and actual proportions of the consistency categories, we calculated one-way ANOVAs with group (memory vs. estimation) or two-way ANOVAs with group and T1-interview type as the independent variables. For $n = 13$ participants, the sum of the proportions of consistent, reminiscent, and contradictory details of all details recalled at T2 was unequal to 100% (although participants were reminded to provide estimates that add up to 100%), indicating that they had misunderstood the question. Hence, they were excluded from the analysis. For consistent and contradictory details, timing of the questionnaire had an influence. Specifically, in the group that had not yet followed the eyewitness lessons ($M_{consistent} = 60.08\%$, $SD_{consistent} = 16.53$; $M_{contradictory} = 18.92\%$, $SD_{contradictory} = 11.64$), the estimated proportion of consistent details was higher, $F(1, 63) = 5.65$, $p = .020$, $d = 0.59$, and the estimated proportion of contradictory statements lower, $F(1, 63) = 8.81$, $p = .004$, $d = -0.74$, than in the group that had already followed the lessons ($M_{consistent} = 49.64\%$, $SD_{consistent} = 19.19$; $M_{contradictory} = 30.00\%$, $SD_{contradictory} = 18.56$). Hence, for the analyses of these variables, the estimation group was split.

Forgotten details

The proportion of T1-information that was forgotten at T2 was significantly overestimated (estimation group: $M = 42.85\%$, $SD = 20.91$; memory group: $M = 24.20\%$, $SD = 8.80$), $F(1, 163) = 56.48$, $p < .001$, $d = 1.17$.

Consistent details

When the memory group was compared to the participants of the estimation group who had already received the eyewitness interview lessons, both main effects (group: $F[1, 108] = 52.55, p < .001, d = 1.40$; interview type: $F[1, 108] = 7.70, p = .007, d = 0.77$), and the interaction, $F(1, 108) = 9.24, p = .003, \eta_p^2 = 0.08$, were significant. The proportion of consistent information was underestimated both when the T1-interview was an SAI (estimation group: $M = 49.29\%, SD = 21.29$; memory group: $M = 76.94\%, SD = 8.60$), $F(1, 108) = 53.25, p < .001, d = -2.16$, and when it was an FR, although to a lesser degree (estimation group: $M = 50.00\%, SD = 17.65$; memory group: $M = 61.31\%, SD = 9.13$), $F(1, 108) = 8.81, p = .004, d = -0.96$. When the memory group was compared to the participants of the estimation group who had not yet received the eyewitness interview lessons, analogous results were obtained with one exception. Specifically, when the T1-interview was an FR, actual and estimated proportion of consistent details did not differ (estimation group: $M = 62.90\%, SD = 12.40$; memory group: $M = 61.31\%, SD = 9.13$), $F(1, 119) = 0.23, p = .629, d = 0.15$. Table 5.2 displays the means and standard deviations of the actual and estimated proportion of consistent details.

Table 5.2

Proportion of Consistent Details as a Function of T1-Interview Type and Group

Type of T1-interview	Memory group		Estimation group before eyewitness lessons		Estimation group after eyewitness lessons	
	<i>M</i> (%)	<i>SD</i>	<i>M</i> (%)	<i>SD</i>	<i>M</i> (%)	<i>SD</i>
SAI	76.94 ^{ab}	8.60	57.40 ^a	19.64	49.29 ^b	21.29
FR	61.31 ^c	9.13	62.89	12.40	50.00 ^c	17.65
Total	69.31	11.80	60.08	16.53	49.64	19.19

Note. The estimation group is split as a function of whether they had already followed the eyewitness lessons. Means in a row sharing the same superscript differ at $p < .05$.

Reminiscent details

For the proportion of reminiscent details, both main effects (group: $F[1, 147] = 16.99, p < .001, d = 0.56$; interview type: $F[1, 147] = 18.69, p < .001, d = 0.70$) and the interaction, $F(1, 147) = 32.93, p < .001, \eta_p^2 = 0.18$, were significant. Analysis of the simple main effects revealed that when the first interview was an SAI, the actual and estimated proportions of reminiscent details did not differ (estimation group: $M = 21.88\%, SD = 13.37$; memory group: $M = 19.21\%, SD = 8.24$), $F(1, 147) = 1.33, p = .250, d = 0.25$. However, when the first interview was an FR, the proportion of reminiscent details was significantly underestimated (estimation group: $M = 19.55\%, SD = 8.69$; memory group: $M = 35.82\%, SD = 9.70$), $F(1, 147) = 47.72, p <$

.001, $d = -1.75$. Table 5.3 displays the means and standard deviations of the actual and estimated proportion of reminiscent details.

Table 5.3
Proportion of Reminiscent Details as a Function of T1-Interview Type and Group

Type of T1-interview	Memory group		Estimation group	
	<i>M (%)</i>	<i>SD</i>	<i>M (%)</i>	<i>SD</i>
SAI	19.21	8.24	21.88	13.37
FR	35.82 ^a	9.70	19.55 ^a	8.69
Total	27.32	12.22	20.73	11.28

Note. Means in a row sharing the same superscript differ at $p < .05$.

Contradictory details

The proportion of contradictory details found in the memory group ($M = 3.37\%$, $SD = 2.20$) was significantly overestimated both by the participants of the estimation group who had already followed the eyewitness interview lessons ($M = 30.00\%$, $SD = 18.56$), $F(1, 108) = 166.35$, $p < .001$, $d = 2.84$, and by those who had not yet followed the eyewitness lessons ($M = 18.92\%$, $SD = 11.64$), $F(1, 119) = 139.80$, $p < .001$, $d = 2.30$. All other main effects and interactions were non-significant, $F_s \leq 1.78$, $p_s \geq .185$, $d_s \leq 0.15$, η_p^2 s < 0.01 .

To summarize, we found mixed support for our hypotheses regarding the proportions of the consistency categories. As expected, the amount of forgetting was overestimated. The frequency of reminiscence was underestimated only when the T1-interview was an FR, but not when it was an SAI, showing that it may be important to take into consideration the T1-interview type when examining lay-persons' beliefs. This result was due to a higher reminiscence rate in the FR memory group, as previously found by Hope et al. (2014). In direct contrast to our hypotheses, the amount of consistency was underestimated and the amount of contradictions overestimated.

Reasons for High or Low Accuracy Rates of Reminiscent Details

The reasons provided by the estimation group as to why they believe that reminiscent details are characterized by high or low accuracy rates, respectively, can be found in Table 5.4 (note that some participants provided several reasons). The majority of the responses reflect a skeptical attitude towards reminiscent details, mostly out of concern that they have been contaminated by external influence. We will address the responses in more detail in the discussion.

Table 5.4

Reasons Provided by the Estimation Group for a Low or High Accuracy of Reminiscent Details

Reasons provided for a low accuracy of reminiscent details	% (n)
New details are the result of compromising external influence (including media reports, co-witness discussions, and information leaked by the police)	48.1 (39)
The witness fills in gaps (also to make the story more coherent)	13.6 (11)
Memory decreases over time	9.9 (8)
All stored details can already be retrieved in the first interview	3.7 (3)
Replaying the incident in one's mind distorts the mental picture of the incident	3.7 (3)
The witness makes details up	2.5 (2)
Witnesses do not change their stories	2.5 (2)
The witness just wants to please the interviewing police officer	2.5 (2)
Witnesses have a bad memory in general	1.2 (1)
The recollection during the second interview is based on what was said in the first interview and not on the actual incident	1.2 (1)
Reminiscent details are the result of a source memory error	1.2 (1)
Reasons provided for a high accuracy of reminiscent details	
Replaying the incident in one's mind can elicit relatively accurate reminiscent details	9.9 (8)
The stress is reduced in the second interview	3.7 (3)
Reminiscent details can be quite accurate if decent interview techniques/good questions are used	2.5 (2)
The witness remembered the detail all along, but did not consider it important at first	1.2 (1)
Reminiscent details have not been influenced by external factors	1.2 (1)
Reminiscent details have been influenced by external factors	1.2 (1)
Not all details are accessible in the first interview	1.2 (1)

Additional Analyses

In an additional analysis, we examined whether inconsistencies in recall were related to overall accuracy within the memory group. The number of reminiscent details was not related to T2-recall accuracy, $r(82) = -.03$, $p = .793$. However, there was a large negative correlation between number of contradictions and T2-recall accuracy, $r(82) = -.50$, $p < .001$. Members of the legal system have been found to use the number of contradictions to challenge the accuracy of the *rest* of the statement (i.e., excluding contradictions; Fisher et al., 2009). Hence, the correlation between the number of contradictions and accuracy of the rest of the statement may be more relevant. This correlation was smaller, but still of moderate size, $r(82) = -.32$, $p = .003$.

In contrast to the belief of the estimation group, accuracy of recall did not decline over time in the memory group (see above). It may be the case, however, that in line with a quantity-accuracy trade-off (Koriat & Goldsmith, 1996), the deterioration of memory manifested itself in reduced recall quantity rather than in

reduced accuracy. To test this idea, the quantity of recall of the memory group was analyzed as a function of T1-interview type (SAI vs. FR) and recall attempt (T1 vs. T2). Comparing actual recall performance and estimations is not possible for this variable, because the quantity of recall was not collected in the estimation group. The main effect of T1-interview type was non-significant, $F(1, 82) = 1.54, p = .219, d = 0.27$. The significant main effect of time, $F(1, 82) = 6.57, p = .012, d = 0.15$, was qualified by a significant Time by T1-Interview Type interaction, $F(1, 82) = 67.25, p < .001, \eta_p^2 = 0.45$. The simple main effects analysis revealed that within the SAI-group, the quantity indeed significantly decreased from T1 ($M = 130.67, SD = 28.47$) to T2 ($M = 119.54, SD = 30.12$), $F(1, 82) = 16.28, p < .001, d = -0.38$. In contrast, within the FR-group, the quantity significantly increased from T1 ($M = 106.44, SD = 31.72$) to T2 ($M = 127.71, SD = 33.80$), $F(1, 82) = 56.58, p < .001, d = 0.65$.

Finally, because the SAI is a relatively new instrument and only few studies have examined its effectiveness, we compared recall performance with the SAI vs. with FR in the memory group. At T1, the SAI elicited both more correct ($M = 112.95, SD = 24.01$), $F(1, 82) = 14.19, p < .001, d = 0.82$, and more incorrect details ($M = 17.72, SD = 6.88$), $F(1, 82) = 5.78, p = .018, d = 0.52$, than the FR (correct: $M = 93.10, SD = 24.29$; incorrect: $M = 13.34, SD = 9.65$). Accuracy was unaffected (SAI: $M = 86.58\%, SD = 3.66$; FR: $M = 88.42\%, SD = 5.48$), $F(1, 82) = 3.31, p = .073, d = -0.40$. These results are in line with previous research that compared the SAI with FR (e.g., Gabbert et al., 2009; Hope et al., 2014; see chapter 4).

Regarding recall performance at T2, initial SAI vs. FR completion did not influence the number of correct (SAI: $M = 102.35, SD = 26.11$; FR: $M = 109.59, SD = 26.14$), $F(1, 82) = 1.61, p = .208, d = -0.28$, or incorrect details at T2 (SAI: $M = 17.19, SD = 6.39$; FR: $M = 18.12, SD = 10.22$), $F(1, 82) = 0.26, p = .614, d = -0.11$. For accuracy at T2, there was a significant interaction between T1-interview type and film version, $F(1, 80) = 6.05, p = .016, \eta_p^2 = 0.07$ (the main effects were non-significant, $F_s \leq 0.93, p_s \geq .338, |d_s| \leq 0.19$). While accuracy rates did not differ for film A (SAI: $M = 86.81\%, SD = 3.55$; FR: $M = 85.28\%, SD = 5.28$), $F(1, 80) = 1.14, p = .288, d = 0.34$, for film B, the SAI group ($M = 84.16\%, SD = 5.35$) provided less accurate statements than the FR group ($M = 87.67\%, SD = 4.44$), $F(1, 80) = 5.72, p = .019, d = -0.71$. To sum up, in contrast to previous findings (Hope et al., 2014; see chapter 4, Experiment 2) there was no memory-preserving effect of initial SAI completion.

Discussion

In the present study, we compared police detectives' estimates of eyewitnesses' performance across two recall attempts with actual recall performance. In doing so, the present study aimed to test layperson's beliefs about the consistency of recall that are summarized in the courtroom approach of memory (Fisher et al.,

2009). At the same time, we extended earlier findings in this field and addressed the limitations of previous research (Oeberst, 2012) by use of the following means. First, to our knowledge, this is the first study that directly compared frequency estimates with actual frequencies of the consistency categories. Second, for the estimations, we relied on an ecologically valid sample of police detectives who had on average 19 years of work experience and 13 years of investigative interviewing experience. Their estimates were compared to the recall performance of police students who only recently started their career at the police and are therefore unlikely to differ from non-police witnesses. The police students' recall performance was indeed comparable to the performance of non-police witnesses in other experiments conducted in our lab that used the same interview types (see chapter 4). Third, taking into account recent findings (Hope et al., 2014), the interview type at T1 was varied, as this may have an impact on the discrepancy between layperson's beliefs and actual recall performance. Fourth, by investigating the reasons for the mistrust in reminiscence this study went beyond merely documenting discrepancies between layperson's beliefs and witnesses' actual recall performance.

Before addressing the results on layperson's beliefs, a general note on reminiscence as a result of repeated interviews seems in order. Like many previous studies, the present study identified reminiscence as a common phenomenon (Brock et al., 1999; Erdelyi, 2010; Hope et al., 2014; Odinot, Memon, et al., 2013). Every participant of the memory group experienced reminiscence. Moreover, the amount of reminiscence was no significant indicator of overall accuracy, showing that its occurrence should not serve to discredit witnesses in court (note that this is different for the number of contradictions that were negatively related to accuracy of the rest of the testimony). Rather, the present findings support the notion that repeated interviews can be a means to improve recall performance, as they present the opportunity to report new, that is, reminiscent details at a reasonably high accuracy rate (Gilbert & Fisher, 2006; Hope et al., 2014; Odinot, Wolters, & van Giezen, 2013). Moreover, they illustrate the importance to not indiscriminately use the term inconsistency but to differentiate between contradictions and reminiscence (Fisher et al., 2009; Oeberst, 2012).

Turning to the results on the beliefs, recall that the courtroom approach summarizes beliefs about recall consistency that are supposedly held by laypersons (Fisher et al., 2009). Based on the assumption of fading memory over time, laypersons should assume both reminiscent and contradictory details to be inaccurate and indicators of overall inaccuracy. Moreover, according to laypersons, reminiscence should occur infrequently and be the result of external influences (e.g., co-witness information). Indeed, the present study revealed that experienced police detectives hold these implicit beliefs that do not correspond with actual recall performance. In the following, we will first discuss the results regarding accuracy, followed by the results on the frequencies.

In agreement with the belief of memory fading over time, the estimation group assumed a sharp decline of accuracy, whereas it actually remained stable. Note, however, that this discrepancy does not preclude actual memory fading. The estimation group may have erroneously assumed that memory degradation inevitably leads to reduced accuracy. Yet, from research dealing with the metamemory model by Koriat and colleagues it is known that memory degradation may manifest itself in reduced quantity (Koriat & Goldsmith, 1996) or reduced level of specificity (Goldsmith, Koriat, & Pansky, 2005; Goldsmith, Koriat, & Weinberg-Eliezer, 2002), while accuracy remains stable (Evans & Fisher, 2011). A decrease of quantity from T1 to T2 was observed in the memory group, at least for those who completed an SAI at T1. In any case, it can be stated that concerns about an excessive decline of accuracy seem unwarranted.

The detectives also displayed a considerable degree of mistrust in the accuracy of reminiscence as predicted in the courtroom approach of memory (Fisher et al., 2009). Specifically, the estimation group assumed that reminiscent details would be as inaccurate as contradictory details. This is striking, because the probability of a contradiction at T2 being correct cannot exceed 50%. In reality, even though consistent details were most accurate, accuracy of reminiscent details was high and much higher than accuracy of contradictions, as found previously (Brock et al., 1999; Gilbert & Fisher, 2006; Odinot, Memon, et al., 2013).

Overall, the results regarding estimated and actual accuracy are analogous to those obtained by Oeberst (2012). Thus, it seems that experienced police detectives hold similar implicit beliefs about memory performance across two retrieval attempts and exhibit the same mistrust in reminiscence than students who are unlikely to have experience with witnesses. As in Oeberst's (2012) study, the accuracy rates were always underestimated. This general pattern of underestimation is difficult to reconcile with the notion that police officers usually consider eyewitness evidence to be crucial and to frequently provide critical investigation leads (Kebbell & Milne, 1998). Regarding overall accuracy, the detectives seemed to be more skeptical than students, as their estimates were even lower than those provided by a student sample (Oeberst, 2012). Interestingly, estimated accuracy of consistent and reminiscent details received similar estimates by both samples.

As to the frequencies of the consistency categories, the amount of forgetting was overestimated, which is in line with our hypotheses. Reflecting the belief that reminiscence is uncommon (Fisher et al., 2009), the frequency of reminiscence was underestimated in the FR group, while this was not the case in the SAI group. The pattern emerged not because the beliefs of the estimation group differed across interview types (the estimates of the frequency of reminiscence were approximately the same for FR and SAI). Rather, as reported previously (Hope et al., 2014), the actual reminiscence rate was much higher in the FR than in the SAI group. This shows that it is important to consider the type of T1-interview when examining layperson's beliefs about recall consistency. Also unexpectedly, due to the dramatic

overestimation of the frequency of contradictions, the frequency of consistent recall was underestimated. The latter also reflects skepticism towards eyewitness accounts in the sample of detectives. The eyewitness lessons made the detectives even more critical as indicated by the larger deviance of the estimates in the subgroup that had already followed the lessons. This finding also demonstrates the need to revise existing interview trainings, which will be addressed in more detail further below.

Turning to the reasons provided for an assumed high or low accuracy of reminiscence, it was most frequently stated that reminiscence was the product of negative external influences, such as co-witness discussions or media coverage. This reflects the belief that reminiscence is the result of an unusual non-cognitive mechanism, which would be expected by the courtroom approach (Fisher et al., 2009). In addition, alleged cognitive mechanisms were specified, such as an assumed regular tendency among witnesses to fill in gaps, the ability to exhaustively recall all details in the first interview, and a decrease of memory over time. As Erdelyi (2010) pointed out, however, forgetting does not preclude reminiscence. It should be noted, though, that there were a few detectives who acknowledged that reminiscent details can be quite accurate. Yet, only one specified that not all details may be accessible in the first interview and two participants acknowledged that reminiscence may be related to the kind of questioning in the T2-interview, but did not explicate how. The reasons identified in the present study provide important insights into the beliefs and misconceptions held by laypersons about human memory functioning and reveal substantial knowledge gaps in this field. As such, they provide leads concerning how police detectives' training can be improved. These will be addressed below.

Finally, replicating previous results (Gabbert et al., 2009; see chapter 4), we found that in the memory group the SAI elicited more correct details (approximately 21%) than FR, which amounted to a large effect, but it also moderately increased incorrect recall. Again, accuracy was unaffected. There was no memory-preserving effect of initial SAI completion on recall performance at T2. This is surprising given that, following the recommendations of chapter 4 (Experiment 2), we employed a T2-interview that provided ample retrieval support (i.e., the SAI). Although speculative, it may not only be essential to use an interview during the second recall attempt that fosters comprehensive accounts, but to also use an interview that slightly differs from the first one to keep witnesses motivated. This issue will be addressed in more detail in the general discussion in chapter 7.

Turning to potential limitations of the study, it should be noted that the estimation group provided the estimates of the first and second recall attempt during a single session and not in two sessions with a one week interval as the memory group. However, we do not believe that this poses a threat to the validity of the results. Specifically, Oeberst (2012, Experiment 2) included two estimation groups, one that provided all estimates in one session and one that provided the estimates

at the same time as the memory group. Apart from one dependent variable, there were no differences in results between the two estimation groups. Hence, it is not decisive whether the estimation group actually experienced the retention interval. Another limitation may pertain to the procedure that before providing the estimates, the estimation group watched the stimulus film. This was done to give the estimation group a better idea of what they had to assess. This approach differs from real cases where police officers do not have the opportunity to see the crime prior to assessing the witnesses' statement. Yet, we do not believe that this deviation influenced the findings, because Oeberst (2012) obtained similar results across two experiments, irrespective of whether the estimation group had inspected the to-be-remembered material or not.

Even though the detectives examined in the present study were right in principle to evaluate reminiscent details as less accurate than consistent details and to point out the potential deleterious influence of incorrect post-event information, this should not give rise to a general mistrust in reminiscence or overgeneralization, respectively. Note that overgeneralization is also apparent in other areas of eyewitness memory, such as in the supposition that child eyewitnesses are generally less credible than adults, while recent research has shown that in some cases, children's recollections may be *less* vulnerable to suggestive influence than adults' recollections (see Brainerd, 2013, for an overview; Otgaar, Howe, Peters, Sauerland, & Raymaekers, 2013). Such misjudgments may lead to unjustified discrediting of witnesses. This is unfortunate given the central role of the police for gathering evidence. More specifically, if the police are not convinced of the reliability of witnesses' statements because of inconsistencies, they are likely not to invest additional resources into the investigation or may not follow up on certain details mentioned by the witness. Ultimately, this may jeopardize the success of the investigations. This demonstrates the importance to revise trainings for members of the legal system regarding the functioning of human memory.

It is clear that police officers are but one of several occupational groups in the legal system concerned with assessing the accuracy of eyewitness statements. We therefore encourage replication with other groups, that is, attorneys and judges.

From the present findings recommendations can be derived on how trainings for police officers on the structure of memory and interviewing eyewitness can be improved. First, such trainings should foster a differentiated view on eyewitnesses' performance to prevent the emergence of overgeneralization. Second, a greater emphasis should be put on teaching the structure of human memory. While the course in which the estimation group took part covered network models of memory (J. R. Anderson, 1983), our results suggest that this may not be sufficient. Building on these models, courses should in addition deal with the principle of the independence of components of a complex event (Fisher et al., 2009) and the notion that multiple and varied retrieval attempts make different aspects of the memory accessible (Tulving & Watkins, 1975). Discussing these two memory prin-

ciples should illustrate that reminiscence is a normal phenomenon that occurs concurrent with forgetting (Erdelyi, 2010) and is largely independent of overall accuracy. In this regard, the importance of differentiating between reminiscence and contradictions should be made clear. Moreover, to account for reminiscence in a given eyewitness account, it is important to convey strategies how to inquire whether the witness may have been exposed to post-event information and whether the interviews conducted provided different retrieval cues. As our findings have shown, memory decay does not necessarily manifest itself in reduced accuracy over time. Instead, the quantity or the level of precision may deteriorate, while accuracy remains stable. Presenting the metamemory framework by Koriat and Goldsmith (Goldsmith et al., 2002; Goldsmith et al., 2005; Koriat & Goldsmith, 1996) that provides an explanation for these trade-offs between quantity or level of precision, respectively, and accuracy would be important to illustrate this point. Finally and perhaps more importantly, reminiscence should not only be dealt with in the theoretical elements of the courses, but also in the practical elements. Specifically, in addition to improving participants' interview skills, practice interviews, which are a routine part of interview courses, could be used for teaching the concept of reminiscence. That is, the participants could interview each other on several occasions about the same incident and in the process directly experience reminiscence themselves. Such a personal experience may increase comprehension and acceptance of reminiscence. Incorporating these principles into police training courses may help reduce the bias against reminiscence and subsequent misjudgments of the witnesses' credibility. Ultimately, this may lead to more accurate decisions in court.

CHAPTER 6 Retrieval Support with the SAI and Working and Source Memory

Krix, A. C., Sauerland, M., Merckelbach, H., Gabbert, F., & Hope, L. (2014). How effective is retrieval support for witnesses with different levels of working and source memory? Manuscript submitted for publication.

Abstract

The present study examined the effectiveness of providing retrieval support to witnesses who differ in working memory capacity and source monitoring abilities. We hypothesized that the provision of retrieval support, relative to free recall (FR), would compensate deficits linked to reduced working memory and source monitoring by providing more structure and context cues for retrieval. Thus, we expected no associations between recall performance and working memory capacity and source monitoring abilities in the retrieval support group, but significant positive associations in the FR group. This study combined data from two experiments ($N = 125$) in which participants either received retrieval support with the Self-Administered Interview© ($n = 63$) or completed an FR ($n = 62$) along with working and source memory tests. Contrary to our expectations, presence of retrieval support did not moderate the relationship between working memory capacity and recall performance. In one source memory test, better source memory was associated with more accurate accounts in the retrieval support group. In the FR group, lower source memory was associated with higher accuracy. This suggests that individuals with reduced source memory may not benefit from retrieval-support with mental context reinstatement. Further research in this field is encouraged.

Eyewitnesses are of crucial importance to the police and the courts, but their accounts can be incomplete or, even worse, inaccurate. There are many factors that influence the quantity and accuracy of eyewitness statements. These include situational factors, such as distance or lighting conditions, but also cognitive factors (Wells, 1978). In the current paper, we focus on two cognitive factors that are especially relevant for successful retrieval, namely working memory and source monitoring.

During retrieval from long-term memory, working memory is involved in maintenance and manipulation of information (e.g., Baddeley, 2000). In Baddeley's (Baddeley, 2010; Baddeley & Hitch, 1974) classic model, working memory comprises multiple short-term memory stores for verbal and visuo-spatial material. The model also includes the central executive that is responsible for attentional control and the supervision of the short-term memory stores. Similarly, Engle and colleagues (Engle & Kane, 2004; Engle, Tuholski, Laughlin, & Conway, 1999) conceptualize working memory as consisting of domain-specific memory caches with rehearsal processes and domain-general executive attention. Individual differences in working memory capacity (WMC) are considered to reflect differences in executive attention. In Engle and colleagues' (Engle et al., 1999; Engle & Kane, 2004) model, executive attention is important for accessing information in long-term memory traces and holding this information together with current task goals active, especially during the simultaneous completion of complex tasks. Executive attention is also considered important for resolving interference through cognitive inhibition (i.e., attending to relevant information and suppressing irrelevant information) during retrieval, and monitoring of previous recall output (Baddeley, Emslie, Kolodny, & Duncan, 1998; Jahanshahi, Saleem, Ho, Dirnberger, & Fuller, 2006). To date, there is no consensus as to whether WMC is part of, or independent to, executive functioning. While working memory is sometimes subsumed under executive functioning (e.g., Elliott, 2003), other researchers consider it to be independent, yet closely related to executive functioning (McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010). Research has found both working memory and executive functioning to be highly correlated and to share the underlying component of executive attention (McCabe et al., 2010). Indeed, WMC is correlated with performance in executive attention tasks, such as the Stroop task, as well as with higher-level cognition, such as fluid intelligence (Kane & Engle, 2002). Moreover, both WMC and executive functioning are associated with prefrontal cortex activity (Kane & Engle, 2002).

Another important aspect for successful retrieval is source monitoring (Johnson, Hashtroudi, & Lindsay, 1993). Source monitoring occurs during working memory tasks and, like working memory, is intimately linked to prefrontal activity (Mitchell, Johnson, Raye, & Green, 2004). Source monitoring refers to evaluations of the origins of memory material (e.g., Does a remembered detail originate from the witnessed incident or from a different occasion?). Source monitoring evalua-

tions are thought to rely on characteristics of the memory material, such as perceptual (e.g., sounds) and affective details, as well as the cognitive operations that are involved (e.g., deductions). Thus, material that stems from experienced events is thought to contain more perceptual, contextual, and affective information, and fewer cognitive operations than material that originates from imagination or fantasy. Errors in source monitoring can occur for several reasons (Johnson et al., 1993; see also the Constructive Memory Framework; Schacter, Norman, & Koutstaal, 1998). Since source monitoring entails the evaluation of activated memory records, it is dependent on the quality of the encoded information, including the binding of the different features of an event to form a coherent representation. Source monitoring is also dependent on the quality of the retrieved information. Here, working memory comes into play, which is responsible for accessing and holding active information from long-term memory. Finally, source monitoring relies on the quality of the judgment processes (e.g., use of appropriate criteria for the source decision).

WMC and source monitoring can influence both correct recall and recall errors (e.g., McCabe et al., 2010; Unsworth & Brewer, 2010b; Unsworth & Engle, 2007; Zhu et al., 2010a). Individuals with lower WMC typically recall fewer correct items than individuals with higher WMC (e.g., Rosen & Engle, 1997; Unsworth & Brewer, 2010b). Moreover, reduced WMC and deficits in source monitoring are associated with more recall errors (e.g., Unsworth & Brewer, 2010a, 2010b) and more false recognitions (Peters, Jelicic, Verbeek, & Merckelbach, 2007).

The relationship between WMC and recall performance has also been examined in the context of eyewitness memory. For example, in an experiment by Jaschinski and Wentura (2002), participants watched a film depicting a staged crime. Next, they read a narrative about the film that contained both correct information and misleading information. The authors found that individuals with higher WMC (as measured with the Operation span task; Turner & Engle, 1989) were less susceptible to the misinformation effect (i.e., were less likely to report the misleading information). Similar results were obtained when memory for misinformation was assessed in a recognition test (Zhu et al., 2010a). Another eyewitness study examined proneness to intrusions (i.e., the reporting of non-pictured details) as a function of WMC (Gerrie & Garry, 2007). In that study, participants with high WMC exhibited a lower false recognition rate of crucial details than participants with low WMC.

So far, eyewitness studies examining the relationship between cognitive factors and memory performance have focused on how reduced WMC is related to memory errors. From a practical point of view and in line with a positive psychological approach (see also Meissner, Hartwig, & Russano, 2010), an important question is whether there are ways to remedy the performance deficit of people with reduced WMC and source monitoring. In fact, there are many groups of potential witnesses who are characterized by reduced levels of WMC and source moni-

toring, including individuals with lower intelligence (Engle et al., 1999), the elderly (Johnson et al., 1993; Salthouse, 1994), or individuals with particular mental disorders (e.g., Abraham, Windmann, McKenna, & Güntürkün, 2007; Marchetta, Hurks, Krabbendam, & Jolles, 2008). The aim of the present study was to test the effectiveness of retrieval support in individuals with different levels of WMC and source monitoring. Retrieval support refers to techniques that facilitate retrieval and help the witness during recall. One such technique is mental context reinstatement, which is part of the Cognitive Interview (CI; Fisher & Geiselman, 1992). It is based on the principle of encoding specificity (Tulving & Thomson, 1973), which proposes that recall is improved to the extent that the encoding context is recreated, or reinstated, during retrieval. To this end, witnesses are instructed to think back to what they saw, heard, thought, and felt while witnessing the incident. These contextual cues are stored parallel to the memory of the incident and promote retrieval by providing additional access pathways to the memory (see also network models of memory; e.g., J. R. Anderson, 1983).

Another interview that provides extensive retrieval support is the recently developed Self-Administered Interview© (SAI; Gabbert, Hope, & Fisher, 2009; Hope, Gabbert, & Fisher, 2011). This interview format is independently completed by witnesses at the crime scene so as to elicit an early comprehensive statement when the police do not have the time to conduct an immediate personal interview. The SAI is based on the Cognitive Interview (Fisher & Geiselman, 1992) and uses some of its memory-enhancing components, such as the mental context reinstatement and the “report-everything” instruction (i.e., the instruction to provide the most complete and accurate account possible). The SAI includes non-leading questions and discourages witnesses from guessing. It consists of different sections each focusing on a different topic (e.g., course of events, appearance of the perpetrator) and hence, provides a strong structure for recall. Furthermore, recall of spatial information is facilitated by asking witnesses to provide a sketch of the scene. Thus, the SAI makes use of multiple and varied retrieval, which has been proven beneficial for recall performance, because details that cannot be accessed with one method may well become accessible with another one (Tulving & Watkins, 1975). It has been found that the SAI induces a more comprehensive account than a control interview (i.e., a free recall, FR), without compromising accuracy (Gabbert et al., 2009; Gawryłowicz, Memon, & Scoboria, 2014; see chapter 4), and can preserve memory for a subsequent interview (Hope, Gabbert, Fisher, & Jamieson, 2014; see chapter 4).

Previous eyewitness research that took into account individual differences in cognitive measures, such as WMC, either focused on only a small part of eyewitness testimony (i.e., reporting of misinformation; Jaschinski & Wentura, 2002; Zhu et al., 2010a) or used recognition tests (Gerrie & Garry, 2007; Zhu et al., 2010a) that are far removed from eyewitness interviews employed in normal police practice. With this in mind, the present study relied on an eyewitness paradigm that

resembles real-life situations more closely to examine the relationship between executive functioning and memory performance.

We considered the SAI as a proxy tool for providing witnesses with retrieval support. The FR, which simply instructs people to provide a free narrative of the incident and does not feature memory-enhancing components, was selected as a format that does not provide retrieval support. We examined whether relative to individuals with high WMC and source monitoring abilities, individuals with reduced WMC and source monitoring abilities would benefit more from retrieval support than from FR. We hypothesized that interviews providing high levels of retrieval support (such as the SAI) would provide more structure during recall than FR interviews, which may help focus attention on the recall task and suppress irrelevant information. In doing so, retrieval support may compensate for attentional deficits associated with reduced WMC. Source monitoring is dependent on the quality of the encoded information, as well as on the quality of the retrieved information and of the judgment processes (Johnson et al., 1993). While the quality of the encoded information is beyond the influence of manipulations during the retrieval phase, at least the quality of the retrieval and judgment parts of source monitoring should be accessible for manipulations of the interview type. Unlike the FR, the SAI provides context cues that facilitate the retrieval process (Tulving & Thomson, 1973). Context cues are stored alongside the target event and provide additional pathways to get access to the recollection (Smith, 1994). Indeed, it has been suggested that context reinstatement may therefore facilitate source monitoring (Memon, Zaragoza, Clifford, & Kidd, 2010). Because improved source monitoring aids retrieval, especially individuals who have difficulties with source monitoring should benefit from context reinstatement. Thus, by freeing attentional resources and facilitating source monitoring, retrieval support should be beneficial to individuals with reduced WMC and source monitoring abilities when they try to recall complex information. In line with this, in the FR group, we expected a positive association between WMC and source monitoring performance and both number of correct details and accuracy. A negative association was expected between WMC and source monitoring performance and number of incorrect details. For the group receiving retrieval support with the SAI, we expected lower (or null) correlations between WMC and source monitoring performance and number of correct details, number of incorrect details, and accuracy.

Method

Participants and Design

The present study is an overarching analysis and combines data from the two experiments (Experiment 1: $n = 88$; Experiment 2: $n = 81$) reported in chapter 4.

Both in Experiment 1 and Experiment 2 participants were randomly assigned to the interview conditions, receiving either an SAI or an FR. The WMC and source memory measures (which have not been reported previously) were consistently collected in both experiments. In Experiment 1, half of the participants ($n = 44$) watched the stimulus film with divided attention. This manipulation impaired subsequent recall performance (see chapter 4, for a comprehensive description of methodology and findings). As these findings may obscure the relationship between recall performance and scores in the cognitive tasks, the participants who had watched the stimulus film with divided attention were excluded from the current analyses. This left $N = 125$ participants in total for analysis, with $n = 44$ (SAI: $n = 22$, FR: $n = 22$) participants from Experiment 1 and $n = 81$ (SAI: $n = 41$, FR: $n = 40$) participants from Experiment 2. The resulting sample comprised $n = 91$ women, as well as $n = 101$ German native speakers and $n = 24$ Dutch native speakers. Mean age was $M = 22.51$ years (range: 18 to 64 years, $SD = 5.84$, $Mdn = 21$ years). Participants were psychology undergraduates (85%), students of other areas of study (12%), or were recruited from the general public (3%). Note that the samples of Experiment 1 and 2 were quite similar, as they did not differ with regard to mean age, $F(1, 123) = 0.56$, $p = .454$, gender, $p = .999$ (two-sided; Fisher's exact test), and occupation (student vs. non-student), $p = .125$ (two-sided; Fisher's exact test). Participants received either course credit or a €15 voucher in exchange for participation. Inclusion criteria were German or Dutch as the native language and an age of 18 to 65 years. Participants were randomly assigned to experimental conditions and tested individually. A between-participants design was employed with presence or absence of retrieval support (retrieval support with the SAI vs. FR) as the independent variable.

Materials

Stimulus film

The non-violent stimulus film, presented without audio track, lasted 3:14 min and showed the staged theft of a laptop. Six amateur actors (4 men, 2 women, aged 21 to 36) appeared in the film. The scene was situated at a communal area at a university. One student left his laptop unattended, whereupon the thief, incited by the accomplice, stole it and both left the scene.

Interviews

Self-Administered Interview

German and Dutch translations of the original English version of the SAI (Gabbert et al., 2009; see Hope et al., 2011, for a detailed description) were used (see chapter 4, for previous use of these materials). First, witnesses were asked to mentally

reinstate the context. That is, they were instructed to think back to the witnessed incident and picture what they could see or hear, and what they thought or felt at the time. Hereafter, witnesses described the course of events. In subsequent sections, non-leading cues were used to prompt descriptions of the appearance of the perpetrator(s), and, if applicable, of potential other witnesses or vehicles involved. Witnesses were also asked to draw a sketch of the scene, so as to facilitate recall of positions and directions (i.e., spatial information). Accordingly, the SAI relied on multiple and varied retrieval. In the final sections, details relating to the witnessing conditions (e.g., lighting conditions) were prompted and witnesses had the opportunity to write down any other information that came to mind. In every section, witnesses were reminded to provide the most complete and accurate account of the witnessed incident, but not to guess.

Free recall

Following Gabbert et al. (2009), in the FR form, witnesses were merely requested to provide a description of the sequence of actions and events and a description of all persons involved, including the perpetrator(s) and other witnesses. As in the SAI, participants were reminded to provide the most complete and accurate account possible, but not to guess. In contrast to the SAI, the FR lacked memory-enhancing components (e.g., mental context reinstatement). It entailed only one instead of multiple and varied retrieval attempts and did not provide prompts to cue recall.

Measures of working and source memory

Working memory task: Operation span task

As a complex span task, the Operation span (Ospan) task required participants to pursue a secondary task, while remembering items (Engle, Cantor, & Carullo, 1992; Turner & Engle, 1989). Specifically, participants solved arithmetic problems, while remembering words. They were presented with operation strings, that is, equation-word pairs (e.g., "Is $(10/5)-3=2$? PAINT"). They had to read aloud the equation and subsequently determine by mental arithmetic, whether the suggested solution was correct or incorrect. Hereafter, participants read aloud the to-be-remembered word whereupon the next equation-word pair appeared. After the last operation string, three question marks appeared on screen, which marked the end of a trial and was the prompt for participants to write down the to-be-remembered words from the previous trial. Set size (i.e., number of equation-word pairs) of a given trial varied from two to five. For each set size, there were three trials, yielding 12 trials in total. Set size varied pseudo-randomly. We used the partial-credit unit scoring to quantify performance (Conway et al., 2005). That is, a correctly recalled word was considered a correct response, irrespective of whether the word had

been recalled in the correct order. Subsequently, average accuracy across trials was calculated.

Source monitoring tests

In line with Unsworth and Brewer (2010a), two source monitoring tests were used to measure source memory performance. In the picture source recognition test, participants were shown 30 pictures that appeared for 1 s, one at a time, in one of four quadrants on screen. Participants were instructed to pay attention to both the picture and the quadrant in which the picture appeared. At test, they were presented with 30 old and 30 new pictures, each displayed in the center of the screen. Participants had to indicate, whether a picture was old or new. If considered old, they were asked to specify in which quadrant it had appeared. The pictures were taken from Rossion and Pourtois (2004).

In the gender source recognition test, participants heard 30 English one-syllable nouns, which were spoken by either a female or male voice. They were instructed to pay attention to both the word and the gender of the voice. At test, participants were presented with 30 old and 30 new words that were shown to them on screen. They were instructed to indicate, whether a word was old or new. If considered old, they had to specify whether it had been spoken by the male or the female voice.

For both tests, the order of the stimuli during encoding and at test had initially been randomly determined and the same order was then applied to all participants. No time limits were imposed on the responses. However, after 5 s had elapsed without a response, a warning appeared urging participants to respond faster. For each participant, two scores were calculated, a picture source monitoring score and a gender source monitoring score. As in Unsworth and Brewer (2010a), the score was the proportion of correct responses across all items of each task.¹⁶

¹⁶ We also calculated conditionalized source identification scores (see Unsworth & Brewer, 2009). Unlike the proportion correct score, this score only considers the old items. Specifically, it is a ratio of the number of correct source identifications of all old items correctly identified as old. Previous research has found that both scores are highly correlated and account for the same individual difference variance (Nash Unsworth, personal communication, April 25th, 2014). Indeed, correlations between overall proportion correct and conditionalized scores were very high (picture source: $r[123] = .91$; gender source: $r[123] = .81$, $ps < .001$). Entering the conditionalized scores into the regression analyses yielded practically the same results as the regression equations with the proportion correct scores. Only the interaction between gender source monitoring score and presence of retrieval support on the number of incorrect details changed from significant to marginally significant ($p = .059$). The analyses of proportion correct are reported in the current manuscript as this score has the advantage that it uses information from all test items.

Procedure

Approval for the experiments was obtained by the local ethics committee of the faculty. We made use of a cover story telling participants that the research dealt with social perception. After signing an informed consent and providing demographic information, participants were shown the stimulus film and were asked to watch carefully. The film was presented on a screen with a screen size of 22 in. (55.88 cm) and a resolution of 1920 x 1080 pixels. Hereafter, the Ospan and source monitoring tests were completed, which were computer-based and run with E-Prime 1.1.4.1 and Presentation 14.8, respectively. Participants were then taken to a different room to avoid effects of physical context on recall performance. Here they completed their interview, either an SAI ($n = 63$) or an FR ($n = 62$), depending on the retrieval support condition. After finishing data collection, participants were fully debriefed.

Coding

Participants' statements were transcribed and coded against an elaborate coding scheme. For example, the statement "The thief sat at the rightmost table." would yield four details (for further details see chapters 2 and 4). Subjective responses, such as "He was attractive", were not scored. A detail was coded as correct, if it matched the content of the film, and coded as incorrect, if it did not. Confabulations were defined as incorrect details referring to non-existent details (e.g., the thief attacked the victim; see Dando, Wilcock, & Milne, 2009). To code the accuracy of age, height, and weight estimates, we accepted deviations of 2 years, 4 cm, or 3 kg from the true value (see Fahsing, Ask, & Granhag, 2004). In the SAI, information from the sketch was also coded. As was the case for verbal recall, a-priori coding rules were also specified for non-verbal output. Specifically, objects and their positions as well as the positions and moving directions (in the SAI, witnesses are informed that they may use arrows for indicating directions) of the persons drawn in the sketch were scored. Only objects and persons that were labeled (e.g., thief) in the sketch by the participant were coded.

Inter-coder reliability

Regarding inter-coder reliability, the randomly selected statements of ten Dutch and ten German statements (i.e., 20 statements in total) were independently coded by two coders (see chapter 4). There was high agreement for correct (German: $\kappa = .99$; Dutch: $\kappa = .98$; $ps \leq .001$) and incorrect recall (German: $\kappa = .98$; Dutch: $\kappa = .94$; $ps \leq .001$).

Results

In line with the recommendations provided by Conway et al. (2005) and to ensure that enough attention was paid to the verification element of the Ospan task, participants whose accuracy rate was less than 85% ($n = 5$) in the verification element were excluded from analysis¹⁷. Although the inclusion or exclusion of these participants had no overall effect on results, these participants were excluded from further analyses.

In the Ospan task, the average memory scores obtained were $M = 80.93\%$ ($SD = 8.65$; $range = 48.47\% - 98.33\%$). For the picture and gender source test, the average scores were $M = 82.51\%$ ($SD = 9.20$; $range = 55.00\% - 98.33\%$) and $M = 65.44\%$ ($SD = 10.36$; $range = 40.00\% - 95.00\%$), respectively. Performance in the Ospan task was not correlated with performance in the picture source recognition score, $r(118) = -.09$, $p = .337$, but positively correlated with the gender source recognition score, $r(118) = .23$, $p = .012$. The correlation between picture and gender source monitoring scores was $r(123) = .07$, $p = .430$.

We performed multiple regression analyses (enter method) on the data. As predictors, a dummy variable to code for the presence of retrieval support (retrieval support = 1; FR = 0), the centered Ospan and source monitoring scores, and the interaction terms were entered into the equations (i.e., seven predictors in total). An a-priori power analysis with G*Power 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) yielded a required total sample size of $N = 103$ participants, given $\beta = .80$, $\alpha = .05$, and a medium effect size. The number of correct and incorrect details, and accuracy (number of correct details divided by all reported details; see Meissner, Sporer, & Susa, 2008) were the dependent variables. When the interaction terms were significant, we analyzed the simple slopes of both interview groups. When the interaction terms were non-significant, they were removed one at a time and the analyses re-run until only significant interaction terms or main effects remained. Table 6.1 shows the means and standard deviations of the recall performance for the retrieval support and the FR group. Table 6.2 shows the results of the final regression equations, after non-significant interactions were removed.

¹⁷ Unfortunately, the number of calculation errors was not available for participants from Experiment 1. Hence, participants from Experiment 1 who failed to pass the 85% accuracy limit in the verification element could not be identified and removed from the sample, as would have been the recommended procedure. Although this is not optimal, we do not believe that this is a threat to the validity of our results for the following reasons. First, even though the cases from Experiment 2 that were removed scored rather low in the Ospan task (percentile ranks ranging from 2.5%- 39.5% of the Experiment 2 sample), they were by no means outliers. As the sample of Experiment 1 was in many ways comparable to the sample of Experiment 2, there is no reason to believe that patterns would have been different in Experiment 1. Second, usually a very small fraction of participants is affected by an excessive number of errors (Kane & Engle, 2000; Turner & Engle, 1989). Third, and most importantly, previous research found that the pattern of correlations of the Ospan score with other working memory measures remained unaffected after participants with an accuracy rate of less than 85% were excluded from analysis (Unsworth, Heitz, Schrock, & Engle, 2005).

Table 6.1

Means and Standard Deviations for Number of Correct and Incorrect Details, and Accuracy as a Function of Presence of Retrieval Support

	FR (<i>n</i> = 62)		Retrieval support (<i>n</i> = 63)		Across interviews	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Correct details	85.89	24.52	117.16	28.20	101.65	30.66
Incorrect details	10.06	5.16	14.43	5.72	12.26	5.85
Accuracy (%)	89.55	4.59	88.83	4.37	89.19	4.48

Note. The means and standard deviations reported in this table result from a combination of the recall data of Experiment 1 and Experiment 2 reported in chapter 4.

Table 6.2

Regression of Recall Performance on Ospan and Picture and Gender Source Monitoring Score, and Presence of Retrieval Support

Variable	<i>B</i>	95% <i>CI</i> (<i>B</i>)	<i>SE</i>	β	<i>t</i>	<i>p</i>	<i>R</i> ²	95% <i>CI</i> (<i>R</i> ²)	<i>F</i>
Correct details							.30	[.17, .43]	12.07**
D1	31.01	[21.39, 40.62]	4.85	.50	6.39	< .001			
Ospan	30.97	[-26.56, 88.49]	29.04	.09	1.07	.289			
Picture source	64.42	[11.58, 117.26]	26.68	.19	2.42	.017			
Gender source	-11.71	[-59.01, 35.59]	23.88	-.04	-0.49	.625			
Incorrect details							.23	[.10, .36]	6.87**
D1	4.66	[2.72, 6.60]	0.98	.39	4.76	< .001			
Ospan	6.15	[-5.45, 17.75]	5.86	.09	1.05	.296			
Picture source	-4.83	[-15.48, 5.83]	5.38	-.07	-0.90	.372			
Gender source	10.32	[-3.20, 23.84]	6.82	.18	1.51	.133			
Gender source x D1	-29.99	[-48.51, -11.47]	9.35	-.38	-3.21	.002			
Accuracy							.13	[.02, .24]	3.45*
D1	-0.01	[-0.03, 0.01]	0.01	-.10	-1.18	.241			
Ospan	-0.02	[-0.11, 0.08]	0.05	-.03	-0.37	.711			
Picture source	0.09	[< 0.01, 0.17]	0.04	.18	2.02	.046			
Gender source	-0.14	[-0.25, -0.03]	0.06	-.31	-2.45	.016			
Gender source x D1	0.26	[0.11, 0.41]	0.08	.44	3.47	.001			

Note. D1 = dummy variable to code for the presence of retrieval support. ** *p* < .001, * *p* < .01.

Ospan Score

In contrast to our hypotheses, all interactions between presence of retrieval support and Ospan score were non-significant, $|B|s \leq 27.05$, $|b|s \leq .06$, $ps \geq .650$. Nei-

ther were there significant main effects of the Ospan score after removing the interaction terms, $p_s \leq .289$ (see Table 6.2).

Picture and Gender Source Monitoring Score

Number of correct details

In contrast to our hypotheses, the interactions of presence of retrieval support with both picture source monitoring score, $B = 66.27$, 95% CI [-39.50, 172.04], $SE = 53.39$, $\beta = .13$, $p = .217$, and gender source monitoring score, $B = 21.83$, 95% CI [-70.25, 113.92], $SE = 46.48$, $\beta = .05$, $p = .639$, were non-significant. As can be derived from the positive regression coefficient in Table 6.2, irrespective of the presence of retrieval support, a higher picture source monitoring score was associated with the recall of more correct details ($p = .017$). The main effect of the gender source monitoring score was non-significant ($p = .625$).

Number of incorrect details

Unexpectedly, the interaction between presence of retrieval support and picture source monitoring score was not significant, $B = -8.13$, 95% CI [-29.58, 13.32], $SE = 10.83$, $\beta = -.08$, $p = .454$. The same held for the main effect of the picture source monitoring score, $p = .372$ (see Table 6.2).

The interaction between presence of retrieval support and gender source monitoring score was significant, $B = -29.99$, 95% CI [-48.51, -11.47], $SE = 9.35$, $\beta = -.38$, $p = .002$. Contrary to our expectations, the analyses of the simple slopes yielded a non-significant effect for the FR group, $B = 10.32$, 95% CI [-3.20, 23.84], $SE = 6.82$, $\beta = .18$, $p = .133$. For the retrieval support group, however, a higher gender source recognition score was associated with fewer incorrect details, $B = -19.67$, 95% CI [-32.75, -6.60], $SE = 6.60$, $\beta = -.35$, $p = .004$. Figure 6.1 displays the results of this simple slope analysis for the number of incorrect details.

Accuracy

As with correct and incorrect details, there was no significant interaction between the picture source monitoring score and presence of retrieval support, $B = 0.15$, 95% CI [-0.02, 0.33], $SE = 0.09$, $\beta = .21$, $p = .080$. However, a higher picture source monitoring score was associated with higher recall accuracy, $p = .046$ (see Table 6.2).

The interaction between presence of retrieval support and gender source monitoring score was significant, $B = 0.26$, 95% CI [0.11, 0.41], $SE = 0.08$, $\beta = .44$, $p = .001$. Hence, we analyzed the simple slopes of the two interview groups. Contrary to expectations, for the FR group, a lower gender source monitoring score was

associated with a higher accuracy, $B = -0.14$, 95% CI [-0.25, -0.03], $SE = 0.06$, $\beta = -.31$, $p = .016$. In contrast, for the retrieval support group, a lower gender source monitoring score was associated with lower accuracy, $B = 0.13$, 95% CI [0.02, 0.23], $SE = 0.05$, $\beta = .29$, $p = .019$. Figure 6.2 displays the results of this simple slope analysis for recall accuracy.

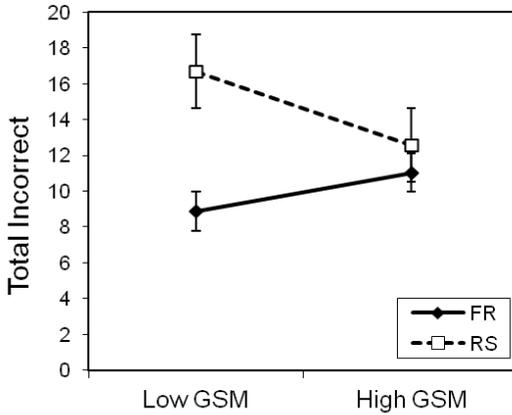


Figure 6.1.

Simple slopes of the interaction between gender source monitoring score and presence of retrieval support for the number of incorrect details. Low and high source monitoring is defined as 1 *SD* below or above the mean score. RS = Retrieval support; FR = Free recall; GSM = Gender source monitoring.

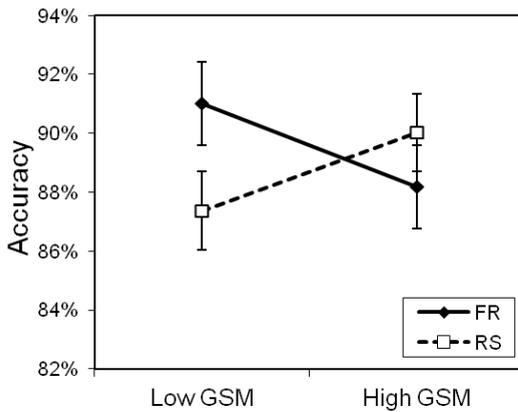


Figure 6.2.

Simple slopes of the interaction between gender source monitoring score and presence of retrieval support for accuracy. Low and high source monitoring is defined as 1 *SD* below or above the mean score. RS = Retrieval support; FR = Free recall; GSM = Gender source monitoring.

Discussion

In the present study, we sought to extend the existing research on the relationship between working memory and source monitoring and recall performance within the context of eyewitness testimony. More specifically, the aim of the present study was to examine the effect of retrieval support in witnesses with different levels of WMC and source monitoring abilities on recall performance. The recently developed SAI (Gabbert et al., 2009) was selected as a proxy tool for providing witnesses with retrieval support. We hypothesized that providing retrieval support would be more beneficial than an FR for individuals with lower WMC or source monitoring abilities.

In contrast to our hypotheses, there were no significant interactions between presence of retrieval support and WMC to predict recall performance. This suggests that the greater structure provided by retrieval support may not be a suitable means to compensate deficits in executive attention found in individuals with reduced WMC (Engle et al., 1999; Engle & Kane, 2004). Given that the SAI already entails a considerable amount of retrieval support, it is unlikely that a further increase in retrieval support would bring about a significant change of the pattern of results. We had reasoned that the greater structure provided in the SAI because of its different recall sections would help witnesses focus on the recall task at hand and suppress irrelevant information. As described in the introduction, this cognitive inhibition (attending to relevant information and suppressing irrelevant information) is but one task working memory is involved in (Engle et al., 1999; Engle & Kane, 2004). Although speculative, it may be the case that retrieving information and holding the retrieved information active require so much executive attention that the retrieval support that helps focus on the recall task at hand does not provide sufficient compensation.

The question could be raised whether retrieval support with the SAI may actually increase cognitive load. One argument in favor of this could be that the more comprehensive recall instructions used in the SAI may be more difficult to memorize than those used in the FR. However, we do not think that retrieval support with the SAI increases cognitive load. If anything, the SAI reduces the necessity to memorize instructions. As it is a *written* interview, the instructions are always visible to the witnesses and can be reread if need be. Furthermore, the SAI is completed in the absence of an interviewer. The presence of others has been found to increase cognitive load and to impair performance in tasks involving frontal activity (Wagstaff et al., 2008). Retrieval from memory is such a task (Kane & Engle, 2002). Note that both issues (i.e., interviewer presence and possibility to review instructions) are different in oral interviews, such as the Cognitive Interview, in which an interviewer is present and witnesses usually hear the instructions only once. Therefore, it is unlikely that the use of the Cognitive Interview as a proxy tool for high retrieval support would have yielded more positive results. Furthermore,

the notion that the SAI induces higher cognitive load is not supported by the data. If retrieval support with the SAI had increased cognitive load, a significant interaction opposite to the hypothesized pattern (i.e., a stronger positive relationship between WMC and recall performance in the retrieval support than in the FR group) would have emerged. However, no significant interactions were obtained. Since WMC did not predict recall performance in the present study at all (i.e., there were no significant main effects, either), our findings should, perhaps, be considered with caution. There is no obvious reason for these non-significant results. Previous studies also involving homogenous samples of undergraduates have identified associations between WMC as measured with the Ospan task and memory performance (Gerrie & Garry, 2007; Jaschinski & Wentura, 2002; but see Peters, Jelacic, et al., 2007, who did not find a relation between memory performance and Ospan score within an undergraduate sample). In light of the majority of previous research it seems unlikely that range restriction due to a homogeneous sample was responsible for the present results.

With respect to source monitoring performance, the hypotheses could not be confirmed, either. Specifically, for the gender source monitoring test, we found results opposite to the expected pattern. That is, in the retrieval support group, a higher gender source recognition score was associated with fewer recall errors and more accurate reports. In contrast, in the FR group, a lower score was associated with more accurate reports. Although there were no interactions between presence of retrieval support and the picture source monitoring score, we obtained the usual pattern that better source monitoring was associated with better recall performance. Unlike previous findings (Unsworth & Brewer, 2010a), this pattern was not due to fewer errors, but due to recall of more correct details.

Contrary to previous assumptions (Memon et al., 2010), it thus seems that providing witnesses with retrieval support does not facilitate source monitoring. Importantly, our results are consistent with findings obtained in a sample with autism spectrum disorders (ASD), a condition known to be associated with reduced source monitoring performance (Bowler, Gardiner, & Berthollier, 2004). Specifically, in two studies, Maras and Bowler (2010, 2012) found that individuals with ASD did not benefit from recall with the CI and the use of mental context reinstatement in particular. When interviewed with a structured interview that lacks mental context reinstatement, no differences emerged between participants with ASD and participants without ASD. Source monitoring is dependent not only on the quality of the retrieved information and of the judgment processes, but also on the quality of the encoded information (Johnson et al., 1993). To the extent that the latter of these three factors is decisive for successful source monitoring, providing retrieval support that is targeted on the retrieval phase is unlikely to enhance remembering. Indeed, problems with source monitoring have been related to binding deficits during encoding (Chalfonte & Johnson, 1996). These refer to problems in binding details of the target event with feature information necessary to specify

the source, such as spatial or temporal (i.e., context) information (see also the Constructive Memory Framework; Schacter et al., 1998). For successful mental context reinstatement, however, intact connections between context cues and details of the target event are essential. If the context is not connected to the event, the context cues cannot provide access to the memory of the event (see Maras & Bowler, 2012). Failure to establish the links during encoding could be the reason why mental context reinstatement may be ineffective for participants with lower source monitoring abilities. Therefore, retrieval support with the SAI that (unlike the FR) makes use of context reinstatement may not have had beneficial effects for the participants with lower source monitoring abilities. Note, however, that the advantage of retrieval support for individuals with high source monitoring abilities was observed for only the gender source monitoring task. As for the picture source monitoring task and recall performance, the interview type was irrelevant. Therefore, the interpretation of the source memory results should be considered with caution and we strongly encourage future research to explore these issues further. Yet, the results are supported by the findings of a very recent study that was published, while this dissertation was near completion (Maras, Mulcahy, Memon, Picariello, & Bowler, 2014). Specifically, Maras et al. (2014) found that the SAI elicited less accurate reports in individuals with ASD than a control interview, whereas no differences emerged for the participants without ASD.

Overall, the results of the current study appear to suggest that providing retrieval support is not an effective means to help witnesses with reduced WMC or source monitoring abilities remember. While retrieval support seems to be equally effective for witnesses with high or low WMC, source monitoring ability appears to be critical for completing an interview with retrieval support that comprises mental context reinstatement. As such, it may be prudent to advise caution when administering such interviews to witnesses with reduced source monitoring abilities. However, it is important to note that this is the first study to examine the interaction of retrieval support and WMC and source monitoring performance to predict eyewitness recall performance. Hence, replications are critical, especially with a more diverse sample than the one used here.

This notion leads over to the limitation of this study, to wit that the sample predominantly consisted of undergraduates. Indeed, it could be criticized that we did not test participants who are characterized by reduced WMC and source monitoring abilities (e.g., individuals with low intelligence). Moreover, testing undergraduates who cognitively function at a high level and are unlikely to differ much regarding cognitive tasks (see Peters, Jelicic, et al., 2007) made the sample rather homogenous. Therefore, if anything, our findings probably underestimated the associations between WMC and source monitoring and recall performance. However, previous research found significant associations between cognitive measures and recall performance even in undergraduate samples (e.g., Jaszchinski & Wentura, 2002; Peters, Jelicic, et al., 2007; Unsworth & Brewer, 2010a, 2010b). Nevertheless,

our findings require replication in mixed samples of people with low WMC and source memory (e.g., patient groups) and healthy participants.

Apart from examining the relationship between retrieval support and WMC and source memory in a more diverse sample, there is a second research line that could be addressed in future research. Given that retrieval support was not effective, it should be the aim of future research to investigate whether there are other means to help witnesses with reduced WMC and source monitoring abilities remember and obtain highly accurate and comprehensive accounts. The importance of such an endeavor is underlined by the fact that many victims of crimes are psychiatric patients (Walsh et al., 2003) who often have poor executive functioning (e.g., Abraham et al., 2007). We therefore urge further research on this important topic.

CHAPTER 7 **General Discussion**

Aims of the Present Dissertation

Following the approach of a positive legal psychology (Meissner, Hartwig, & Russano, 2010; Sauerland & Krix, 2011), the aim of this dissertation was to examine the effects of retrieval support on recall performance in eyewitness interviews. Throughout, the SAI (Gabbert, Hope, & Fisher, 2009) served as a proxy tool for providing witnesses with ample retrieval support. Given that the SAI is a written interview, we first examined whether either written or spoken interviews are more suitable to facilitate eyewitness recall (chapter 2). In chapters 3 and 4, the effects of retrieval support following different suboptimal witnessing conditions emerging in the encoding phase (i.e., stress and divided attention) were examined. Chapters 4 and 5 focused on retrieval support in repeated interviews. Finally, in chapter 6, the effects of retrieval support given a suboptimal condition during retrieval (poor working and source memory) were examined. In the following, the main findings of the studies will be discussed and implications for interviewing practice as well as suggestions for future research will be provided.

Summary and Discussion of the Research Findings

Modality of Eyewitness Accounts

Examining the effect of the modality of eyewitness accounts (chapter 2), we found no evidence for a spoken superiority effect. If anything, there was some indication of a written superiority effect for the quantity, not accuracy, of the event descriptions. Yet, none of the comparisons between the written and the classic spoken condition (i.e., spoken-videotaped with an interviewer present) reached significance. Additionally, differences observed during free report were compensated when followed up with cued questions. The combination of free report and cued question is the usual procedure during police interviews (Wells, Memon, & Penrod, 2006). That is, police officers already make use of the measure that can be employed to compensate any possible disadvantages of spoken statements.

Neither level of executive functioning nor instruction comprehensiveness interacted with the modality of the interview. The data, however, once more showed the importance of using supporting retrieval instructions (Fisher & Geiselman, 1992), as the comprehensive recall instructions elicited more detailed accounts than the scarce instructions.

From the results of the two experiments reported in chapter 2, we can draw the conclusion that the modality may not be decisive when it comes to the quantity and quality of eyewitness accounts. Since the tested samples consisted of students, this conclusion necessarily needs to be confined to highly-educated samples. We encourage replication of the findings, because they contradict a previous study that

used a similar methodology (Sauerland & Sporer, 2011). For the time being, it can be concluded that the police practice of rather arbitrarily using written or spoken interviews seems unproblematic. Importantly, the results also suggest that it does not seem to be problematic that the SAI is a written interview. This may be different, of course, for witnesses experiencing literacy difficulties or witnesses who do not feel confident expressing themselves in writing. Indeed, Hope, Gabbert, and Fisher (2011) advise to be careful regarding the use of the SAI in these cases, although, of course, this remains an empirical question to be tested.

Retrieval Support and Suboptimal Conditions Emerging during Encoding

The results of the experiments reported in chapters 3 and 4 that focused on the effectiveness of retrieval support with the SAI following suboptimal conditions that emerge during encoding (i.e., exposure to stress and distraction) were mixed. While they were more positive for retrieval support when the witnesses experienced stress, they were less positive when witnesses were distracted while observing a crime. In the following, the findings will be discussed in more detail.

Retrieval support with the SAI and stress

In the experiment described in chapter 3, half of the participants were exposed to stress. Subsequently, all participants witnessed a live staged crime and completed either an SAI or a free recall (FR). While the stress manipulation was successful as indicated by the significant increase in cortisol levels and negative affect in the stress group, there was no effect of stress on memory performance. As previously found (e.g., Gabbert et al., 2009), the SAI elicited more comprehensive accounts (corresponding to a large effect), albeit not more accurate accounts than the FR. Since there were no interactions between level of stress and interview type, the police can safely use the SAI, irrespective of the stress levels experienced by the witnesses. This finding is important, because the SAI is particularly administered in cases with many eyewitnesses (Hope et al., 2011) who are likely to vary in the level of stress they experience.

To account for the non-significant effects of stress on memory performance, we referred to neurobiological studies that have pointed to the importance of the timing of the stressor. While stress during encoding has been found to facilitate memory performance (Cahill, Gorski, & Le, 2003; Smeets, Otgaar, Candel, & Wolf, 2008), stress during retrieval seems to compromise memory performance (de Quervain et al., 2000; Kuhlmann, Piel, & Wolf, 2005). Unfortunately, this issue has largely been neglected in eyewitness research. As testing occurred in a single session in our study, the stress reaction was present both during encoding and retrieval. Hence, the memory-enhancing and memory-impairing effects of stress may have canceled each other out. The SAI, however, was developed to be completed

immediately after the crime (Gabbert et al., 2009) when witnesses may still experience distress. From a theoretical perspective, it would be interesting to examine what the results would be if the SAI was completed after a short delay when participants are not stressed any more. Drawing from the results obtained in the neurobiological field, memory performance should be enhanced in the stress group in this situation.

These considerations demonstrate the importance of an integrative approach when studying the effects of stress on eyewitness performance. Ideally, it should combine the expertise of eyewitness researchers with the expertise of stress researchers who have a neurobiological background. The former have profound knowledge regarding what interviews to use and what coding procedures to apply to eyewitness statements. The latter have in-depth knowledge regarding the nature of stress reactions and how to induce and measure them. Such cooperation may have the potential to significantly take forward the research on stress and eyewitness memory.

Retrieval support with the SAI and divided attention

In Experiment 1 of chapter 4, we explored the question, whether retrieval support with the SAI helps witnesses provide comprehensive accounts when they were distracted during the crime. In line with the outshining hypothesis (Smith, 1988, 1994), we expected an interaction. Specifically, while retrieval support with the SAI should always lead to better recall performance than FR, this difference should be more pronounced when participants had been distracted. Yet, this was not the case. Retrieval support with the SAI only elicited more correct details in the full attention condition. In the divided attention condition, no such differences emerged. Although the SAI led to more incorrect details than FR, there were no effects on accuracy. In sum, retrieval support with the SAI did not seem suitable to compensate the detrimental effects of divided attention during encoding. However, because the SAI did not elicit less accurate reports than FR, distraction is not a contraindication of using the SAI. Rather, the police should be aware that retrieval support loses its advantage relative to an interview without retrieval support.

A closer look at the interpretation of the findings of chapter 6 may provide an account for these unexpected results. In chapter 6, we argued that the effectiveness of mental context reinstatement depends on the presence of links between target and context information. Otherwise, the context cues are unable to provide access pathways to the recollections. The lack of a link between target and context information can be caused by disruptions during encoding (Johnson, Hashtroudi, & Lindsay, 1993). Indeed, divided attention has clearly been found to disrupt encoding processes (N. D. Anderson, Craik, & Naveh-Benjamin, 1998; Craik, Govoni, Naveh-Benjamin, & Anderson, 1996) and may therefore impede link formation. This may be the reason why retrieval support with the SAI that relies on mental

context reinstatement did not boost recall performance relative to FR that does not use context reinstatement, when participants had been distracted during encoding.

Retrieval Support and Suboptimal Conditions during Retrieval

The study described in chapter 6 was an overarching analysis of the two experiments reported in chapter 4 and focused on working memory capacity (WMC) and source monitoring abilities. Low WMC and poor source memory are associated with poor memory performance (e.g., McCabe, Roediger, McDaniel, Balota, & Hambrick, 2010; Rosen & Engle, 1997; Unsworth & Brewer, 2010a, 2010b). In fact, a considerable proportion of potential witnesses show reduced levels of WMC and source monitoring. These include the elderly (Johnson et al., 1993; Salthouse, 1994) and individuals with lower intelligence (Engle, Tuholski, Laughlin, & Conway, 1999) or mental disorders (e.g., Abraham, Windmann, McKenna, & Güntürkün, 2007; Marchetta, Hurks, Krabbendam, & Jolles, 2008).

We expected that retrieval support would be especially beneficial for witnesses with reduced WMC and source monitoring abilities, because retrieval support provides a greater structure for recall and may also facilitate source monitoring. Although the hypotheses were not confirmed, an interesting pattern of results emerged. If anything, retrieval support seemed to be less efficient for witnesses with reduced source memory. Specifically, within the retrieval support group, low source monitoring scores (in one of the two source monitoring tasks) were associated with more recall errors and less accurate reports. In contrast, low source monitoring scores were associated with more accurate reports in the FR group. There were no significant associations with WMC.

The results of the study are in line with the finding that the Cognitive Interview and context reinstatement in particular are ineffective for individuals with autism spectrum disorders (ASD; Maras & Bowler, 2010, 2012) who are known to have difficulties with source monitoring (Bowler, Gardiner, & Berthollier, 2004). If reduced source monitoring is due to difficulties in binding the target information with feature context during encoding (Chalfonte & Johnson, 1996), context reinstatement that is an essential part of the SAI cannot facilitate retrieval due to the missing link between context and target information. This indicates that even though reduced source monitoring is a condition that manifests itself in the retrieval phase, the origin of low source memory may go back to the encoding phase (see Johnson et al., 1993). Since retrieval support is targeted at the retrieval phase, it may not be suitable to tackle deficits that start in the encoding phase.

The police should be cautious to administer retrieval-support interviews that make use of mental context reinstatement, such as the SAI, with witnesses who exhibit source monitoring deficits. This conclusion is backed by a recent study that was published at the same time as the completion of this thesis. This study found that the SAI was indeed ineffective for witnesses with ASD (Maras, Mulcahy,

Memon, Picariello, & Bowler, 2014). We are aware of potential practical difficulties in identifying witnesses with source monitoring deficits beforehand, since it is not feasible, of course, to administer the corresponding tests prior to an interview. If, however, there are indications for the presence of a condition associated with source monitoring problems (e.g., old age; Johnson et al., 1993), the police may use the more careful strategy and refrain from using context reinstatement.

Retrieval Support through Repeated Interviews: Recall Consistency

Chapter 5 dealt with the provision of retrieval support through repeated interviews. The first question raised in the chapter, to wit, whether repeated interviews facilitate retrieval, can be affirmed. Every single participant recalled at least one reminiscent item. Relative to the first interview, conducting a second interview elicited on average 30% more correct information. This is quite a large amount of information that would have been lost if only one interview had been conducted. Although reminiscent details ($M = 85.53\%$) were not as accurate as consistent details ($M = 91.96\%$), the accuracy rate was still rather high and significantly higher than the accuracy of contradictory information ($M = 39.87\%$). Interestingly, for the accuracy of the reminiscent details, it was not decisive how much retrieval support the first interview offered. In sum, the results are well in line with previous research investigating the effect of repeated eyewitness interviews (Brock, Fisher, & Cutler, 1999; Gilbert & Fisher, 2006; Hope, Gabbert, Fisher, & Jamieson, 2014; Odnot, Wolters, & van Giezen, 2013).

The second question we wished to address was whether experienced police detectives see the merit of repeated eyewitness interviews, as previous research had shown that police officers regard inconsistencies, such as reminiscence, as an indicator of overall inaccuracy (e.g., Brewer, Potter, Fisher, Bond, & Luszcz, 1999). Indeed, reminiscence apparently violates the notion of fading memories over time, although it can easily be explained with a change of retrieval cues (Fisher, Brewer, & Mitchell, 2009). In short, the answer to the question whether police detectives see the merit of repeated interviews is no. The detectives exhibited beliefs that correspond to those summarized in the so-called courtroom approach of memory (Fisher et al., 2009). That is, the police detectives expected an immense drop of accuracy from the first to the second recall attempt while actual accuracy remained stable. Moreover, they expressed a profound distrust in reminiscence: First, the police detectives presumed that the accuracy of reminiscent information would be as low as the accuracy of contradictions. Second, they voiced the concern that reminiscence was due to detrimental external influences (e.g., co-witness misinformation or media coverage).

Given that a similar pattern of results was obtained in a student sample (Oeberst, 2012), experience with (interviewing) witnesses does not seem to be a decisive factor when it comes to estimating witnesses' recall performance. Interesting-

ly, estimating recall performance is not the only area in which experience with witnesses does not seem to make a difference. Specifically, it reminds one of findings from other research areas, that is, the detection of false confessions and the detection of deception. Experienced police officers are not better at detecting false confessions than college students, yet the former are more confident in their decisions than the latter (Kassin, Meissner, & Norwick, 2005). Moreover, police officers and laypersons alike overestimate their ability to detect deception (Schell & Kassin, 2009, 2012). The results clearly show the need to educate police personnel about the structure and functioning of human memory. This issue will be addressed in more detail in the sections on practical implications and future directions further below.

Self-Administered Interview vs. Free Recall: A Meta-Analytical Synthesis

The SAI was a main focus in this dissertation, as it was deemed the proxy tool for providing witnesses with ample retrieval support. Since it is a relatively new tool, there has been no meta-analysis summarizing its effect. To determine what the effects of SAI completion on recall performance *across* the studies presented in this dissertation were, a meta-analysis was conducted, using an unpublished spreadsheet by Siegfried L. Sporer (personal communication, May 26th, 2014). For this purpose, the data of the four experiments reported in chapters 3-5 were entered, comprising $N = 369$ participants. Regarding the data of chapter 3, we only considered the perpetrator and verifiable event details, because correct and incorrect details could not be determined for the unverifiable event details. Moreover, perpetrator and verifiable event details of this study were collapsed. The independent variable was the type of interview at the first recall attempt (SAI vs. FR). The number of correct and incorrect details and accuracy at the first and the second recall attempt were the dependent variables. Accordingly, not only immediate effects of SAI completion, but also carry-over effects of initially completing an SAI could be examined. When discussing the meta-analytical results of the second recall attempt, we will also address the role of the subsequent interview in the memory-preserving effect of initial SAI completion, which was a focus of chapter 4. The results of the meta-analysis are presented in Table 7.1. The effect sizes used are Cohen's d for independent comparisons (Cohen, 1988; Sporer & Cohn, 2011). The mean weighted effect sizes were calculated by weighting the individual effect sizes with their inverse variances (Lipsey & Wilson, 2001).

Table 7.1

Mean Weighted Effect Sizes for the Type of Initial Interview (SAI vs. Free Recall) for Number of Correct and Incorrect Details and Accuracy at the First (Time 1) and Second Recall Attempt (Time 2)

Dependent variable	<i>k</i>	<i>N</i>	<i>d</i> ¹	95% CI		<i>Z</i>	<i>p</i> _z
				<i>LL</i>	<i>UL</i>		
Correct details Time 1	4	369	1.16	0.94	1.39	10.21	< .001
Incorrect details Time 1	4	369	0.74	0.53	0.95	6.88	< .001
Accuracy Time 1	4	369	-0.15	-0.35	0.05	1.43	.152
Correct details Time 2	3	241	-0.18	-0.43	0.06	1.46	.144
Incorrect details Time 2	3	241	-0.24	-0.49	0.01	1.92	.054
Accuracy Time 2	3	241	0.10	-0.15	0.34	0.76	.446

Note. A fixed effects model was applied. CI = confidence interval; LL = lower limit; UL = upper limit.

¹Positive values indicate increases in the respective dependent variable in the SAI relative to the FR group.

At the first recall attempt, large effect sizes were found for both correct and incorrect details, while the effect size for accuracy was small and non-significant. This means that immediately after the crime, the SAI leads to more comprehensive accounts (involving an increase in both correct and incorrect details) than the FR, while accuracy rates remain unaffected. Similar meta-analytical results were obtained with the related Cognitive Interview (Köhnken, Milne, Memon, & Bull, 1999; Memon, Meissner, & Fraser, 2010). Specifically, these meta-analyses found large effects for correct recall (Köhnken et al., 1999: $d = 0.87$; Memon et al., 2010: $d = 1.20$), with the Cognitive Interview leading to more correct details than control interviews. The magnitude of the effect size obtained by Memon et al. is comparable to the effect size found here. For incorrect recall, the Cognitive Interview meta-analyses yielded small, yet significant effects (Köhnken et al., 1999: $d = 0.28$; Memon et al., 2010: $d = 0.24$), such that the Cognitive Interview also elicited more incorrect details. This means that for incorrect recall, the effect size of the present SAI studies is considerably larger. Unfortunately, effect sizes for accuracy were not reported in the Cognitive Interview meta-analyses so that we cannot compare these here. The increase in incorrect details can be problematic, because for the police it is unknown whether a given detail is correct or incorrect. However, it should be noted that the increase in incorrect details was balanced by an (even larger) increase in correct details. Consequently, accuracy remained unaffected and was at a high level (88% in the SAI groups across the four studies).

The question arises why the SAI elicits more incorrect details relative to a control interview than the Cognitive Interview does. Although speculative, this may be because the SAI only shares the cognitive memory-enhancing components (i.e., mental context reinstatement, “report-everything” component, and multiple and varied retrieval) with the Cognitive Interview (Gabbert et al., 2009). Additionally, the Cognitive Interview features social components, such as rapport-building,

which have been found to enhance recall performance (e.g., Vallano & Schreiber Compo, 2011). Of course, the database of the present meta-analysis is rather small, consisting of only four studies. Thus, it remains to be seen whether this pattern prevails if more SAI studies are meta-analytically summarized.

Turning to the second recall attempt, small and non-significant effect sizes were obtained for correct recall and accuracy. The effect size for incorrect recall was also small, but marginally significant, indicating that initially completing an SAI relative to an FR tended to reduce the number of incorrect information reported in the subsequent interview. In sum, we could not detect a reliable memory-preserving effect of the SAI across all studies. This may not be surprising given that the meta-analysis of the second recall attempt comprised only three experiments, one of them being Experiment 1 of chapter 4 that failed to detect carry-over effects. Still, the finding is noteworthy and deserves further discussion, because one of the reasons why the SAI was developed was that it should preserve memory for a subsequent personal interview (Hope et al., 2011). Moreover, in some countries, such as the United Kingdom, only the testimony obtained in the personal interview, but not the statement elicited with the SAI can serve as evidence in court. Yet, even in countries where statements obtained with the SAI are admissible, such as in Germany or, with some limitations, the Netherlands, it can be assumed that the courts will, and should, insist on the implementation of additional personal interviews, at least in high-stakes cases.

The absence of a memory-preserving effect contradicts network models of memory (e.g., J. R. Anderson, 1983) according to which a more effortful recall attempt, such as is obtained with the SAI, strengthens the memory network more than a less effortful recall attempt with the FR. As a consequence, memory traces that have been strengthened to a higher degree should be less prone to forgetting. It is, however, well in line with some Cognitive Interview studies that did not find carry-over effects, either (e.g., Brock, Fisher, & Cutler, 1999; Memon, Wark, Bull, & Köhnken, 1997).

In chapter 4, the results suggested a reporting issue instead of a memory issue (i.e., failure to overtly report details from memory, not forgetting; Koriat & Goldsmith, 1996) to account for the failure of the memory-preserving effect to occur. Accordingly, the memory-preserving effect of the SAI seems to surface only with subsequent interviews fostering comprehensive statements at the second recall attempt. However, this notion is difficult to reconcile with the data from chapter 5. Here, after completing an SAI or an FR as the first interview, all participants from the memory group received an SAI at the second recall attempt. However, no memory-preserving effect was observed.

In the study by Hope et al. (2014), in which a memory-preserving effect of the SAI was found, the Cognitive Interview was used as the subsequent interview. The question arises how the discrepant findings between our studies and Hope et al.'s may be explained. The Cognitive Interview and the SAI, which were used as T2-

interviews, are closely related, although the former is an oral interview and the latter a written one. Based on the results of chapter 2, it can be concluded that the modality of the interviews is not the decisive aspect to account for the differences. It rather seems that the explanation provided in chapter 4 regarding the emergence of a memory-preserving effect may need to be extended. Specifically, the participants who had to fill in the SAI both at the first and the second recall attempt may have wondered why they should complete exactly the same interview twice. This may have resulted in a motivation loss which led them to not writing down everything they could recall. This would again imply the presence of a reporting issue (Koriat & Goldsmith, 1996). It may also be conceivable that these participants, assuming that they still somewhat remembered the instructions from the week before, did not read and follow the instructions (e.g., with respect to mental context reinstatement) as carefully as they had done the first time. As a result, they may have underperformed in the second interview. It may hence not only be essential to use an interview during the second recall attempt that fosters comprehensive accounts, but to also use an interview that differs from the first one in order to keep witnesses motivated. Whether this explanation is applicable remains to be empirically tested.

Evaluation of the Self-Administered Interview

Given the present (meta-analytical) findings, how should the SAI be evaluated? On the one hand and similar to the Cognitive Interview (Memon et al., 2010), the SAI significantly increased incorrect recall relative to FR in all experiments of this dissertation. This finding certainly requires a critical remark. Strikingly, previous research somehow seemed to avoid confrontation with this topic, by either not reporting the means of incorrect recall at all (e.g., Gawrylowicz, Memon, & Scoboria, 2014) or by reporting the means, but not the corresponding inferential statistics (e.g., Gabbert et al., 2009). In real cases this increase in errors can be problematic, because, unlike in controlled laboratory settings in which the ground truth is known, the police do not know whether a given detail is correct or incorrect and every incorrect detail can potentially have dire consequences.

On the other hand, it should not be disregarded that the increase in incorrect details was balanced by an even larger increase in correct details, which is also found in Cognitive Interview research (Memon et al., 2010). As a result, accuracy did not decline but was the same as obtained with FR. The gain of correct details relative to FR may well yield the decisive lead for the police investigations. It is also important to note that the SAI does not produce poor-quality recall. Specifically, accuracy rates obtained with the SAI were high, amounting to 88% across the four experiments. So, in spite of an increase in errors, the participants who received an SAI on average still recalled 9 out of 10 details correctly. This means that

the SAI leads to more comprehensive accounts, but does not compromise accuracy. This is not trivial, because increases in quantity can be accompanied by decreases of accuracy, as shown in previous research (Evans & Fisher, 2011; Koriat & Goldsmith, 1996). As such, the SAI, which is based on cognitive theory of memory (e.g., J. R. Anderson, 1983; Tulving & Thomson, 1973; Tulving & Watkins, 1975), must be differentiated from other interviewing techniques, such as forensic hypnosis or the Amytal interview. In the Amytal or barbiturate-induced interview, the witness is administered a barbiturate (i.e., a hypnotic that leads to drowsiness, but also to increased talkativeness) prior to the interview (Piper, 1993). Both hypnosis and Amytal interview have been found to elicit excessive recall errors and to increase suggestibility, making the obtained statements unreliable (Piper, 1993; Sanders & Simmons, 1983). Moreover, forensic hypnosis is associated with the development of pseudomemories and increased confidence in incorrectly recalled details (Stebly & Bothwell, 1994). Perhaps unsurprisingly then, reviewing the literature on forensic hypnosis and the Cognitive Interview, Kebbell and Wagstaff (1998) come to the conclusion that the Cognitive Interview is to be preferred to hypnosis. Similar to the Cognitive Interview, the SAI does not compromise recall accuracy and even seems to reduce suggestibility, as it protects against the reporting of misinformation (Gabbert, Hope, Fisher, & Jamieson, 2012; Paterson, Eijkemans, & Kemp, 2014).

When evaluating the SAI, it should also be taken into consideration for which purpose the tool was devised. Specifically, the SAI was devised to give witnesses the opportunity to provide an early statement in cases when the personal interview will be delayed. Research has shown that writing down the recollections with the SAI is to be preferred to no immediate recall opportunity. That is, relative to a no-initial-interview control group, the SAI group provided more accurate accounts in a subsequent personal interview (Gabbert et al., 2009, Experiment 2; Hope et al., 2014).

In sum, it is concluded that under “normal” conditions, the SAI seems to be an effective interview tool to enhance eyewitnesses’ statements. In the present dissertation, we sought to test whether this also holds for suboptimal conditions. For these situations, a somewhat different picture emerged. Clearly, the research described in this dissertation identified two situations in which the SAI did not enhance recall performance, to wit, when witnesses are distracted (chapter 4) and for witnesses with poor source memory (chapter 6). In the latter case, the use of the SAI may even be advised against and instead FR may be more suitable. This is important to consider when providing recommendations for the use of the SAI.

Irrespective of how the effectiveness of the SAI is evaluated, it is important to think of ways how the impact of errors can be reduced. One strategy could be to regard “one SAI as no SAI”. That is, the police could make use of the fact that the SAI is usually applied when multiple witnesses observed the crime. Hence, one option could be to compare the different accounts and only consider those details

for investigation purposes that were consistently mentioned by different witnesses (see Waubert de Puiseau, Assfalg, Erdfelder, & Bernstein, 2012). Such an approach of obtaining an aggregate statement may increase the accuracy further.

As indicated above and due to the inconsistent research findings, no final evaluation can be made regarding the memory-preserving effect of the SAI. If future research failed to find a reliable effect, this would considerably limit the applicability of the tool, as it was developed to preserve memory for a subsequent personal interview. This limitation would especially hold for countries, such as the United Kingdom, where the statement elicited with the SAI does not count as evidence.

Methodological Remarks

Just as previously published studies on the SAI (Gabbert et al., 2009; Gabbert et al., 2012; Gawrylowicz et al., 2014; Hope et al., 2014), the research presented in this dissertation almost exclusively relied on stimulus films. Even though the latter yield a higher level of standardization than live staged events, they may be criticized for a lack of realism. We tried to accommodate for this potential shortcoming by employing a live staged scenario (i.e., the theft of a cell phone) in the study described in chapter 3. To our knowledge, this was the first experimental study employing the SAI that made use of such an ecologically valid scenario. As this experiment found comparable results to the published SAI studies and the other studies presented in this dissertation that made use of films, it can be concluded that the medium of event presentation (film vs. staged event) does not seem to be decisive. This speaks to the robustness of the finding that the SAI elicits more comprehensive, albeit not more accurate accounts than FR (see also the meta-analysis section above for an in-depth discussion of this finding). The lack of influence of presentation medium observed here is consistent with the finding that it is not a moderator of recall performance for the closely related Cognitive Interview, either (Memon et al., 2010).

The second methodological remark refers to the sample composition of the present research. The samples tested in this dissertation largely consisted of students. Indeed, the high costs that are associated with testing participants from the general public did not allow exclusively testing community samples. Moreover, finding interested and motivated members of the general public willing to participate was rather difficult. However, there is good reason to believe that the results obtained with student and community samples are quite similar. Specifically, Gawrylowicz et al. (2014) who tested the SAI in a community sample found comparable effect sizes to the ones identified in the meta-analysis reported above (i.e., large positive effects for correct recall and a non-significant effect size for accura-

cy). Hence, we do not believe that the use of student samples undermined the validity of the results obtained in the present dissertation.

Practical Implications

The research findings presented in this dissertation have important implications for interviewing practice. Across four experiments, it was shown that, like the Cognitive Interview (Fisher & Geiselman, 1992), retrieval support with the SAI led to more comprehensive accounts, while accuracy remained unaffected (see the meta-analytical findings). Similar to the eye-closure technique (see Vredeveltdt, Baddeley, & Hitch, 2014; Vredeveltdt, Hitch, & Baddeley, 2011), the SAI also has the advantage that it is resource-saving. That is, it is easy and cost-effective to conduct and does not require extensive special training or the presence of an interviewing police officer. Therefore, it can be concluded that the SAI is a useful tool to support the police in obtaining comprehensive statements from eyewitnesses in the immediate aftermath of a crime.

In Germany, especially after low-stakes incidents like car accidents, witnesses often only provide a written statement and are not invited to a personal interview by the police at all. For this purpose, so called witness questionnaires (“Zeugenfragebogen”) are used (Nikola Hahn, personal communication, February 12th, 2014). Even though the title may suggest otherwise, this questionnaire is even scarcer than the FR used in the studies of this dissertation and frequently does not elicit particularly useful statements. Administration of the SAI would most likely improve this situation. In fact, the situation may change soon, since field trials of the SAI are currently underway in one German police force, the Hessian police, to examine the usefulness of the SAI in Germany.

As pointed out in chapter 1, the SAI is not meant to be a replacement of a personal interview and there are countries, such as the United Kingdom, in which the statement obtained with the SAI is not admissible as evidence (Hope et al., 2011). Hence, knowledge as to when the memory-preserving effect of initial retrieval support with the SAI surfaces is very important. While this issue has unfortunately not fully been solved yet, there are indications that the type of the subsequent interview plays an important role (see chapter 4 and the meta-analysis section in this chapter). Therefore, it is imperative that the police apply best practice during the subsequent interview, irrespective of whether witnesses received retrieval support with the SAI or not prior to the personal interview. It should be noted that the question of a memory-preserving effect of the SAI may be less pressing in countries where the statements obtained with the SAI are admissible as evidence in court, such as Germany. In these countries it would be possible to combine the accounts obtained in the SAI and the subsequent interview.

While the present dissertation has found beneficial effects of retrieval support with the SAI, it is important to realize that the instrument is not a “silver bullet”. That is, we identified situations in which the ample retrieval support provided in the SAI may not be more beneficial than an ordinary FR, such as when the witness was distracted (chapter 4), or even less beneficial, such as for witnesses with reduced source memory (chapter 6). Therefore, the police would be well advised to enquire about the witnessing conditions and characteristics of the witness to be able to assess the effectiveness of retrieval support with the SAI for the case at hand.

The current findings also underscore the importance of a differentiated view on eyewitness recall performance. Indeed, overgeneralization in terms of a general mistrust in eyewitnesses is inappropriate, as it may lead to unjustified discrediting of witnesses as was shown in chapter 5 regarding the recall of reminiscent information. Unfortunately, such beliefs do not only exist in the field of repeated eyewitness interviews, but also in the field of child eyewitnesses. Specifically, there is the belief that child eyewitnesses are generally less credible than adult eyewitnesses. Yet, recent research has shown that in some cases, children’s recollections may be *less* vulnerable to suggestive influence than adults’ recollections (Brainerd, 2013; Otgaar, Howe, Peters, Sauerland, & Raymaekers, 2013). These observations clearly call for a more profound education of legal professionals, such as police officers, lawyers, and judges, in terms of the functioning of human memory.

Finally, in order to continuously enhance the methods and techniques in the field of interviewing eyewitnesses, a close cooperation between researchers and practitioners (i.e., the police) is essential, which is also a core element of a positive legal psychology (Meissner et al., 2010). This means that, on the one hand, research questions should be geared more strongly to the problems and needs of interviewers in the field. Such cooperation can be fruitful as can be seen in the cases of the development of both the Cognitive Interview (Fisher & Geiselman, 1992) and the SAI (Gabbert et al., 2009). On the other hand, empirical findings need to be systematically put into practice by practitioners (Fisher, 2010). To this end, the police should make use of evidence-based interviewing techniques only. Unfortunately, this is not always the case. For instance, in Germany, forensic hypnosis is still administered as an interview technique (Beetz & von Delhaes, 2011) in spite of empirical evidence showing that it is ineffective at best (Erdelyi, 1994) and increases suggestibility at worst (Sanders & Simmons, 1983). This issue does not speak in favor of a close enough cooperation. To be sure, clear evidence-based guidelines for interviewing eyewitnesses as well as a regular exchange between practitioners and scientific experts in the field of interviewing eyewitnesses are necessary to improve interviewing practice.

Future Directions

While the present dissertation yielded new insights into the field of retrieval support for eyewitnesses, it also identified avenues for future research. Moreover, given that the SAI is a rather new tool, there are inevitably research questions that remain unanswered. In this respect, similar to the approach taken in the present dissertation, future research could continue investigating the SAI as a proxy tool for providing witnesses with retrieval support.

The first research line that should be pursued further pertains to the memory-preserving effect of initial retrieval support. In chapter 4, we narrowed down the conditions regarding when providing initial retrieval support with the SAI preserves memory for a subsequent interview. Yet, as the lack of a memory-preserving effect in the memory group of chapter 5 and the results of the meta-analysis show, the question as to when the memory-preserving effect of initial retrieval support surfaces in the subsequent interview has not ultimately been solved. Taking into account the considerations regarding the necessary features of the subsequent interview that were described in the meta-analysis section, future studies should continue researching this question. As mentioned previously, this is an important endeavor, because in some jurisdictions the initial statement made with the SAI does not count as evidence (Hope et al., 2011). Regarding the subsequent personal interview, another aspect would be worth examining. Specifically, it would be interesting to know whether it helps witnesses to read their statements provided in the SAI before the start of the interview. Rereading the statement may pose a source of context reinstatement (Sauerland, Holub, & Sporer, 2008; Sporer, 2007), which may additionally be beneficial for retrieval.

A second future research line could focus on variables inherent in the witness. The present dissertation has taken a first step into this direction by investigating the relationship between retrieval support and working and source memory. Other sources of individual differences, such as educational background or personality traits, and their relationship with retrieval support could be studied. Indeed, research in the field of eyewitness testimony has recently started examining the role of personality traits regarding recall performance (e.g., Pozzulo, Coplan, & Wilson, 2005; Zhu et al., 2010b). One question that could be addressed is whether retrieval support with the SAI can help anxious witnesses who may have difficulties testifying when interviewed by a police officer achieve comprehensive accounts. Another personality trait that could be investigated in the framework of SAI research is the need for cognitive closure (NfCC; Webster & Kruglanski, 1994). Individuals with a high NfCC are characterized by desire for predictability and structure, as well as by aversion to ambiguity. The results of a recent eyewitness study suggest that witnesses high in NfCC are more prone to misinformation effects than witnesses low in NfCC (Pica, Pierro, Bélanger, & Kruglanski, 2014). More specifically, Pica et al. found that heightened NfCC led to increased retrieval-induced forgetting (i.e., for-

getting of non-retrieved related information following partial retrieval; M. C. Anderson, Bjork, & Bjork, 1994). This was probably due to increased inhibition that resulted from the tendency to reduce ambiguity. Increased retrieval-induced forgetting in turn made these witnesses more prone to misinformation effects. The SAI provides structure for recall, which could accommodate witnesses with high NfCC, and protects against reporting of misinformation (Gabbert et al., 2012). Moreover, because the SAI elicits comprehensive accounts, there is less room for retrieval-induced forgetting. In combination, it could be argued that the SAI may be more suitable for witnesses with high NfCC compared to FR.

It would certainly also be worthwhile to test the SAI in more heterogeneous witness populations, such as adolescents or the elderly. As in an ageing population the elderly are more likely to become witnesses, the question whether retrieval support with the SAI helps them remember at all is an important one, especially since ageing is associated with reduced source memory (Johnson et al., 1993). Recall that in chapter 6 reduced source memory was associated with less accurate reports when participants had been provided with retrieval support with the SAI.

The third potential future research line pertains to the translations and the format of the SAI. To reach as many potential witnesses as possible, it is clear that the SAI should be translated into more languages than the currently available translations (e.g., Dutch, English, German, and Norwegian). Therefore, eyewitness researchers whose native language is one of the “missing” languages are called upon to prepare further translations. Furthermore, at present, the SAI is only available as a booklet that has to be completed handwritten. It is conceivable that a spoken version is devised should future research show that a written SAI is not suitable for witnesses who have difficulties expressing themselves in writing (Hope et al., 2011). A next step could also be to make use of the new technical capabilities, by devising a digital version of the SAI. Such an “eyewitness mobile app” would allow witnesses to provide an early account at the crime scene using their smartphones or tablet computers. After completion, the form could directly be sent electronically to the police station. This could probably also maximize the return rate of SAI forms.

Interestingly, the Los Angeles County Sheriff’s Department has recently launched a software program called Large Emergency Event Digital Information Repository (LEEDIR) for which a website and a mobile app exist.¹⁸ LEEDIR was specifically designed for the use by eyewitnesses of major crimes (e.g., mass shootings) or catastrophes to upload photos and videos of the incident that were taken with the witnesses’ smartphones. The police can subsequently access this footage for investigation purposes. This path was embarked upon after the experience of the police in the aftermath of the Boston Marathon bombings. Specifically, after calling upon witnesses to send in photos and videos of the incident, the police were

¹⁸ <http://www.scrp.org/news/2014/04/10/43430/la-county-sheriff-launches-app-for-eyewitnesses-to/>. Retrieved on August 27th, 2014.

overwhelmed with the vast number of pictures they received and confronted with collapsing systems that were not designed for such data volume. As LEEDIR is a new software, it has not been evaluated and it remains to be seen whether it has the potential to support the police during the investigations. For example, the question arises whether there are actually sufficient resources to screen all uploaded footage. It would also be interesting to test whether LEEDIR could be used to upload verbal witness reports recorded with the microphones of the smartphones together with the photos. Finally, privacy issues may also need to be addressed given that the provider is a private company.

A final future research line could be concerned with training programs for police officers. That is, in chapter 5, police detectives were identified to have significant knowledge gaps concerning the structure and functioning of eyewitnesses' memories in spite of long lasting experience with interviewing eyewitnesses. To prevent unjustified discrediting of witnesses, it would be important to develop and subsequently evaluate respective training programs for police officers. Such trainings should increase knowledge on the principles of human memory, including grain size theory (Goldsmith, Koriat, & Weinberg-Eliezer, 2002), network models of memory (J. R. Anderson, 1983), and the principle of multiple and varied retrieval (Tulving & Watkins, 1975). The trainings should also reduce the bias against reminiscence by demonstrating that it is a normal phenomenon that occurs at the same time as forgetting (Erdelyi, 2010). This could be achieved by incorporating comprehensive practical elements. For instance, training participants could interview each other on several occasions about an incident and in the process directly experience reminiscence themselves. Such practice interviews would also serve the purpose of practicing and strengthening the police officers' interview skills.

Concluding Remarks

The present dissertation dealt with the effectiveness of retrieval support for eyewitnesses. For this purpose, the SAI served as a proxy tool for providing eyewitnesses with retrieval support. The tool contains several memory-enhancing components and was designed to help witnesses achieve early comprehensive statements (Gabbert et al., 2009). As was shown across several experiments reported in this dissertation, retrieval support with the SAI has the potential to significantly support the police during the investigations and to elicit more comprehensive accounts than FR. On the other hand, the findings showed that the tool is not a "silver bullet" (i.e., it does not enhance memory when the witness was distracted or displays poor source memory) and does not exempt the police from applying best practice during the subsequent interview. As one of the first countries worldwide (next to the United Kingdom and Norway), the SAI is already used in the

Netherlands. It has recently been approved for nationwide use by the Dutch police and will become mandatory from 2015 onwards.

Even in the era of DNA testing, eyewitness testimony remains an important piece of evidence for solving cases (de Poot, Bokhorst, van Koppen, & Muller, 2004; Fisher, 2010). Still, many factors can cause faulty witness accounts and ultimately lead to wrongful convictions, as was illustrated in the first case in chapter 1. If, however, eyewitness evidence is collected properly, by using the SAI or other evidence-based techniques, such as those described in chapter 1 that are in line with a positive legal psychology (Meissner et al., 2010), chances are high to receive accurate testimonies. Therefore, eyewitness accounts should of course continue to be admissible as evidence in court, although nobody should be sentenced solely on the grounds of eyewitness evidence. To detect possible sources of error during the interview, it is of paramount importance to audio- or videotape all interviews in full length (c.f. Kassin, 1998; Sporer, 1992, 1993). In addition, the police should inquire about potential detrimental influences during both the encoding stage (e.g., divided attention) and the retention interval (e.g., exposure to misinformation).

References

- Abraham, A., Windmann, S., McKenna, P., & Güntürkün, O. (2007). Creative thinking in schizophrenia: The role of executive dysfunction and symptom severity. *Cognitive Neuropsychiatry*, *12*, 235-258. doi:10.1080/13546800601046714
- Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of Verbal Learning and Verbal Behavior*, *22*, 261-295. doi:10.1016/S0022-5371(83)90201-3
- Anderson, M. C., Bjork, R. A., & Bjork, E. L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term-memory. *Journal of Experimental Psychology: Learning Memory and Cognition*, *20*, 1063-1087. doi:10.1037/0278-7393.20.5.1063
- Anderson, N. D., Craik, F. I. M., & Naveh-Benjamin, M. (1998). The attentional demands of encoding and retrieval in younger and older adults: 1. Evidence from divided attention costs. *Psychology and Aging*, *13*, 405-423. doi:10.1037/0882-7974.13.3.405
- Anderson, R. C., & Pichert, J. W. (1978). Recall of previously unrecalled information following a shift in perspective. *Journal of Verbal Learning and Verbal Behavior*, *17*, 1-12. doi:10.1016/S0022-5371(78)90485-1
- Ayers, M. S., & Reder, L. M. (1998). A theoretical review of the misinformation effect: Predictions from an activation-based memory model. *Psychonomic Bulletin & Review*, *5*, 1-21. doi:10.3758/BF03209454
- Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences*, *4*, 417-423. doi:10.1016/S1364-6613(00)01538-2
- Baddeley, A. (2010). Working memory. *Current Biology*, *20*, R136-R140. doi:10.1016/j.cub.2009.12.014
- Baddeley, A., Emslie, H., Kolodny, J., & Duncan, J. (1998). Random generation and the executive control of working memory. *The Quarterly Journal of Experimental Psychology*, *51A*, 819-852. doi:10.1080/027249898391413
- Baddeley, A. D., & Hitch, G. (1974). Working memory. In G.A. Bower (Ed.), *The Psychology of Learning and Motivation* (pp. 47-89). New York: Academic Press.
- Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- Beetz, A. M., & von Delhaes, A. (2011). Forensische Hypnose: Der Einsatz von Hypnose als erinnerungsunterstützendes Verfahren im Rahmen polizeilicher Ermittlungen [Forensic hypnosis: Using hypnosis as a recall-facilitating tool during police investigations]. *Hypnose-ZHH*, *6*, 165-187.
- Bekerian, D. A., & Dennett, J. L. (1990). Spoken and written recall of visual narratives. *Applied Cognitive Psychology*, *4*, 175-187. doi:10.1002/acp.2350040303
- Bensi, L., Nori, R., Gambetti, E., & Giusberti, F. (2011). The enhanced Cognitive Interview: A study on the efficacy of shortened variants and single techniques. *Journal of Cognitive Psychology*, *23*, 311-321. doi:10.1080/20445911.2011.497485
- Bergmann, M. M., Jacobs, E. J., Hoffmann, K., & Boeing, H. (2004). Agreement of self-reported medical history: Comparison of an in-person interview with a self-administered questionnaire. *European Journal of Epidemiology*, *19*, 411-416. doi:10.1023/B:EJEP.0000027350.85974.47
- Bond, C. F. & Titus, L. J. (1983). Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin*, *94*, 265-292. doi:10.1037/0033-2909.94.2.265
- Boon, J., & Noon, E. (1994). Changing perspectives in cognitive interviewing. *Psychology, Crime & Law*, *1*, 59-69. doi:10.1080/10683169408411936
- Bornstein, B. H., Liebel, L. M., & Scarberry, N. C. (1998). Repeated testing in eyewitness memory: A means to improve recall of a negative emotional event. *Applied Cognitive Psychology*, *12*, 119-131. doi:10.1002/(sici)1099-0720(199804)12:2<119::aid-acp500>3.0.co;2-4
- Bowler, D. M., Gardiner, J. M., & Berthollier, N. (2004). Source memory in adolescents and adults with Asperger's syndrome. *Journal of Autism and Developmental Disorders*, *34*, 533-542. doi:10.1007/s10803-004-2548-7

REFERENCES

- Boyce, M. A., Lindsay, D. S., & Brimacombe, C. A. E. (2008). Investigating investigators: Examining the impact of eyewitness identification evidence on student-investigators. *Law and Human Behavior, 32*, 439-453. doi:10.1007/s10979-007-9125-5
- Brainerd, C. J. (2013). Developmental reversals in false memory: A new look at the reliability of children's evidence. *Current Directions in Psychological Science, 22*, 335-341. doi:10.1177/0963721413484468
- 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. doi:10.1002/(sici)1099-0720(199908)13:4<297::aid-acp578>3.3.co;2-j
- Brock, P., Fisher, R. P., & Cutler, B. L. (1999). Examining the cognitive interview in a double-test paradigm. *Psychology, Crime & Law, 5*, 29-45. doi:10.1080/10683169908414992
- Buchanan, T. W., Tranel, D., & Adolphs, R. (2006). Impaired memory retrieval correlates with individual differences in cortisol response but not autonomic response. *Learning & Memory, 13*, 382-387. doi:10.1101/lm.206306
- Busey, T. A., & Loftus, G. R. (2007). Cognitive science and the law. *Trends in Cognitive Sciences, 11*, 111-117. doi:10.1016/j.tics.2006.12.004
- Cahill, L., Gorski, L., & Le, K. (2003). Enhanced human memory consolidation with post-learning stress: Interaction with the degree of arousal at encoding. *Learning & Memory, 10*, 270-274. doi:10.1101/lm.62403
- Campos, L., & Alonso-Quecuty, M. L. (2008). Language crimes and the cognitive interview: Testing its efficacy in retrieving a conversational event. *Applied Cognitive Psychology, 22*, 1211-1227. doi:10.1002/acp.1430
- Castel, A. D., & Craik, F. I. M. (2003). The effects of aging and divided attention on memory for item and associative information. *Psychology and Aging, 18*, 873-885. doi:10.1037/0882-7974.18.4.873
- Chalfonte, B. L., & Johnson, M. K. (1996). Feature memory and binding in young and older adults. *Memory & Cognition, 24*, 403-416. doi:10.3758/bf03200930
- Christianson, S.-Å. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin, 112*, 284-309. doi:10.1037/0033-2909.112.2.284
- Christianson, S.-Å., & Loftus, E. F. (1987). Memory for traumatic events. *Applied Cognitive Psychology, 1*, 225-239. doi:10.1002/acp.2350010402
- Christianson, S.-Å., & Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Cognition and Emotion, 5*, 81-108. doi:10.1080/02699939108411027
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Collins, R., Lincoln, R., & Frank, M. G. (2002). The effect of rapport in forensic interviewing. *Psychiatry, Psychology and Law, 9*, 69-78. doi:10.1375/pplt.2002.9.1.69
- Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12*, 769-786. doi:10.3758/BF03196772
- Craik, F. I. M., Govoni, R., Naveh-Benjamin, M., & Anderson, N. D. (1996). The effects of divided attention on encoding and retrieval processes in human memory. *Journal of Experimental Psychology: General, 125*, 159-180. doi:10.1037/0096-3445.125.2.159
- Dando, C., Wilcock, R., & Milne, R. (2009). The Cognitive Interview: The efficacy of a modified mental reinstatement of context procedure for frontline police investigators. *Applied Cognitive Psychology, 23*, 138-147. doi:10.1002/acp.1451
- Dando, C., Wilcock, R., Milne, R., & Henry, L. (2009). A modified cognitive interview procedure for frontline police investigators. *Applied Cognitive Psychology, 23*, 698-716. doi:10.1002/acp.1501
- Davis, M. R., McMahon, M., & Greenwood, K. M. (2005). The efficacy of mnemonic components of the cognitive interview: Towards a shortened variant for time-critical investigations. *Applied Cognitive Psychology, 19*, 75-93. doi:10.1002/acp.1048
- de Poot, C. J., Bokhorst, R. J., van Koppen, P. J., & Muller, E. R. (2004). Rechercheportret: Over dilemma's in de opsparing [Detectives' portrait: On dilemmas in police investigations]. Alphen aan den Rijn: Kluwer.

- de Quervain, D. J. F., Roozendaal, B., Nitsch, R. M., McGaugh, J. L., & Hock, C. (2000). Acute cortisone administration impairs retrieval of long-term declarative memory in humans. *Nature Neuroscience*, *3*, 313-314. doi:10.1038/73873
- Deffenbacher, K. A. (1994). Effects of arousal on everyday memory. *Human Performance*, *7*, 141-161. doi:10.1207/s15327043hup0702_3
- Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior*, *28*, 687-706. doi:10.1007/s10979-004-0565-x
- Dickerson, S. S., & Kemeny, M. E. (2004). Acute stressors and cortisol responses: A theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, *130*, 355-391. doi:10.1037/0033-2909.130.3.355
- Dror, I. E., & Hampikian, G. (2011). Subjectivity and bias in forensic DNA mixture interpretation. *Science and Justice*, *51*, 204-208. doi:10.1016/j.scijus.2011.08.004
- 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. doi:10.1037/0021-9010.83.5.745
- Elliott, R. (2003). Executive functions and their disorders. *British Medical Bulletin*, *65*, 49-59. doi:10.1093/bmb/65.1.49
- Engle, R. W., Cantor, J., & Carullo, J. J. (1992). Individual differences in working memory and comprehension: A test of four hypotheses. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *18*, 972-992. doi:10.1037/0278-7393.18.5.972
- Engle, R. W., & Kane, M. J. (2004). Executive attention, working memory capacity, and a two-factor theory of cognitive control. In B. Ross (Ed.), *The psychology of learning and motivation* (Vol. 44, pp. 145-199). New York: Elsevier.
- Engle, R. W., Tuholski, S. W., Laughlin, J. E., & Conway, A. R. A. (1999). Working memory, short-term memory, and general fluid intelligence: A latent-variable approach. *Journal of Experimental Psychology: General*, *128*, 309-331. doi:10.1037/0096-3445.128.3.309
- Erdelyi, M. H. (1994). Hypnotic hypermnesia: The empty set of hypermnesia. *International Journal of Clinical and Experimental Hypnosis*, *42*, 379-390. doi:10.1080/00207149408409366
- Erdelyi, M. H. (2010). The ups and downs of memory. *American Psychologist*, *65*, 623-633. doi:10.1037/a0020440
- Evans, J. R., & Fisher, R. P. (2011). Eyewitness memory: Balancing the accuracy, precision and quantity of information through metacognitive monitoring and control. *Applied Cognitive Psychology*, *25*, 501-508. doi:10.1002/acp.1722
- Fahsing, I. A., Ask, K., & Granhag, P. A. (2004). The man behind the mask: Accuracy and predictors of eyewitness offender descriptions. *Journal of Applied Psychology*, *89*, 722-729. doi:10.1037/0021-9010.89.4.722
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*, 175-191. doi:10.3758/bf03193146
- Fawcett, J. M., Russell, E. J., Peace, K. A., & Christie, J. (2013). Of guns and geese: A meta-analytic review of the 'weapon focus' literature. *Psychology, Crime & Law*, *19*, 35-66. doi:10.1080/1068316x.2011.599325
- Fernandez, A., & Glenberg, A. M. (1985). Changing environmental context does not reliably affect memory. *Memory & Cognition*, *13*, 333-345. doi:10.3758/BF03202501
- Fisher, R. P. (2010). Interviewing cooperative witnesses. *Legal and Criminological Psychology*, *15*, 25-38. doi:10.1348/135532509X441891
- Fisher, R. P., Brewer, N., & Mitchell, G. (2009). The relation between consistency and accuracy of eyewitness testimony: Legal versus cognitive explanations. In R. Bull, T. Valentine, & T. Williamson (Eds.), *Handbook of psychology of investigative interviewing: Current developments and future directions* (pp. 121-136). Chichester, UK: John Wiley.

REFERENCES

- Fisher, R. P., & Chandler, C. C. (1991). Independence between recalling interevent relations and specific events. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *17*, 722-733. doi:10.1037/0278-7393.17.4.722
- Fisher, R. P., & Geiselman, R. E. (1992). Memory enhancing techniques for investigative interviewing: The cognitive interview. Springfield, IL: Charles C. Thomas.
- Fisher, R. P., & Schreiber, N. (2007). Interviewing protocols to improve eyewitness memory. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *Handbook of eyewitness psychology: Vol. 1. Memory for events* (pp. 53-80). Mahwah, NJ: Lawrence Erlbaum Associates.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*, 378-382. doi:10.1037/h0031619
- Foos, P. W., & Fisher, R. P. (1988). Using tests as learning opportunities. *Journal of Educational Psychology*, *80*, 179-183. doi:10.1037//0022-0663.80.2.179
- Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness evidence: Examining the efficacy of a self-administered interview tool. *Law and Human Behavior*, *33*, 298-307. doi:10.1007/s10979-008-9146-8
- Gabbert, F., Hope, L., Fisher, R. P., & Jamieson, K. (2012). Protecting against misleading post-event information with a self-administered interview. *Applied Cognitive Psychology*, *26*, 568-575. doi:10.1002/acp.2828
- Gabbert, F., Memon, A., Allan, K., & Wright, D. B. (2004). Say it to my face: Examining the effects of socially encountered misinformation. *Legal and Criminological Psychology*, *9*, 215-227. doi:10.1348/1355325041719428
- Gawrylowicz, J., Memon, A., & Scoboria, A. (2014). Equipping witnesses with transferable skills: The Self-Administered Interview©. *Psychology, Crime & Law*, *20*, 315-325. doi:10.1080/1068316x.2013.777961
- Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1985). Eyewitness memory enhancement in the police interview: Cognitive retrieval mnemonics versus hypnosis. *Journal of Applied Psychology*, *70*, 401-412. doi:10.1037/0021-9010.70.2.401
- Gerrie, M. P., & Garry, M. (2007). Individual differences in working memory capacity affect false memories for missing aspects of events. *Memory*, *15*, 561-571. doi:10.1080/09658210701391634
- Gilbert, J. A. E., & Fisher, R. P. (2006). The effects of varied retrieval cues on reminiscence in eyewitness memory. *Applied Cognitive Psychology*, *20*, 723-739. doi:10.1002/acp.1232
- Ginet, M., & Verkamp, F. (2007). The cognitive interview: Is its benefit affected by the level of witness emotion? *Memory*, *15*, 450-464. doi:10.1080/09658210601092670
- Ginsburg, N., & Karpik, P. (1994). Random number generation: Analysis of the responses. *Perceptual and Motor Skills*, *79*, 1059-1067. doi:10.2466/pms.1994.79.3.1059
- Ginsburg, N., & Karpik, P. (1995). Simulation of human performance on a random generation task. *Perceptual and Motor Skills*, *81*, 1183-1186. doi:10.2466/pms.1995.81.3f.1183
- Goldsmith, M., Koriat, A., & Pansky, A. (2005). Strategic regulation of grain size in memory reporting over time. *Journal of Memory and Language*, *52*, 505-525. doi:10.1016/j.jml.2005.01.010
- Goldsmith, M., Koriat, A., & Weinberg-Eliezer, A. (2002). Strategic regulation of grain size in memory reporting. *Journal of Experimental Psychology: General*, *131*, 73-95. doi:10.1037//0096-3445.131.1.73
- Gould, J. D. (1978). How experts dictate. *Journal of Experimental Psychology: Human Perception and Performance*, *4*, 648-661. doi:10.1037/0096-1523.4.4.648
- Gould, J. D., & Boies, S. J. (1978). How authors think about their writing, dictating, and speaking. *Human Factors*, *20*, 495-505. doi:10.1177/001872087802000415
- Grabowski, J. (2007). The writing superiority effect in the verbal recall of knowledge: Sources and determinants. In G. Rijlaarsdam (Series Ed.) and M. Torrance, L. van Waes, & D. Galbraith (Volume Eds.), *Writing and cognition: Research and applications* (Studies in writing, Vol. 20, pp. 165-179). Amsterdam: Elsevier.
- Hellhammer, J., & Schubert, M. (2012). The physiological response to Trier Social Stress Test relates to subjective measures of stress during but not before or after the test. *Psychoneuroendocrinology*, *37*, 119-124. doi:10.1016/j.psyneuen.2011.05.012

- Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & Cognition*, *18*, 496-506. doi:10.3758/BF03198482
- Hope, L., Gabbert, F., & Fisher, R. P. (2011). From laboratory to the street: Capturing witness memory using the self-administered interview. *Legal and Criminological Psychology*, *16*, 211-226. doi:10.1111/j.2044-8333.2011.02015.x
- Hope, L., Gabbert, F., Fisher, R. P., & Jamieson, K. (2014). Protecting and enhancing eyewitness memory: The impact of an initial recall attempt on performance in an investigative interview. *Applied Cognitive Psychology*, *28*, 298-313. doi:10.1002/acp.2984
- Hope, L., Mullis, R., & Gabbert, F. (2013). Who? What? When? Using a timeline technique to facilitate recall of a complex event. *Journal of Applied Research in Memory and Cognition*, *2*, 20-24. doi:10.1016/j.jarmac.2013.01.002
- Horowitz, M. W., & Newman, J. B. (1964). Spoken and written expression: An experimental analysis. *Journal of Abnormal and Social Psychology*, *68*, 640-647. doi:10.1037/h0048589
- Houston, K. A., Clifford, B. R., Phillips, L. H., & Memon, A. (2013). The emotional eyewitness: The effects of emotion on specific aspects of eyewitness recall and recognition performance. *Emotion*, *13*, 118-128. doi:10.1037/a0029220
- Hulse, L. M., & Memon, A. (2006). Fatal impact? The effects of emotional arousal and weapon presence on police officers' memories for a simulated crime. *Legal and Criminological Psychology*, *11*, 313-325. doi:10.1348/135532505x58062
- Ibabe, I., & Sporer, S. L. (2004). How you ask is what you get: On the influence of question form on accuracy and confidence. *Applied Cognitive Psychology*, *18*, 711-726. doi:10.1002/acp.1025
- Jahanshahi, M., Saleem, T., Ho, A. K., Dirnberger, G., & Fuller, R. (2006). Random number generation as an index of controlled processing. *Neuropsychology*, *20*, 391-399. doi:10.1037/0894-4105.20.4.391
- Jamieson, K., Gabbert, F., Allan, K., & Carson, D. (2009, July). *The effects of actual versus perceived memory quality on our susceptibility to memory conformity*. Paper presented at the meeting of the Society for Applied Research in Memory and Cognition, Kyoto, Japan.
- Jaschinski, U., & Wentura, D. (2002). Misleading postevent information and working memory capacity: An individual differences approach to eyewitness memory. *Applied Cognitive Psychology*, *16*, 223-231. doi:10.1002/acp.783
- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, *114*, 3-28. doi:10.1037/0033-2909.114.1.3
- Kane, M. J., & Engle, R. W. (2000). Working-memory capacity, proactive interference, and divided attention: Limits on long-term memory retrieval. *Journal of Experimental Psychology: Learning Memory, and Cognition*, *26*, 336-358. doi:10.1037//0278-7393.26.2.336
- Kane, M. J., & Engle, R. W. (2002). The role of prefrontal cortex in working-memory capacity, executive attention, and general fluid intelligence: An individual-differences perspective. *Psychonomic Bulletin & Review*, *9*, 637-671. doi:10.3758/bf03196323
- Kassin, S. M. (1998). Eyewitness identification procedures: The fifth rule. *Law and Human Behavior*, *22*, 649-653. doi:10.1023/A:1025702722645
- Kassin, S. M., Meissner, C. A., & Norwick, R. J. (2005). "I'd know a false confession if I saw one": A comparative study of college students and police investigators. *Law and Human Behavior*, *29*, 211-227. doi:10.1007/s10979-005-2416-9
- Kebbell, M. R., & Milne, R. (1998). Police officers' perceptions of eyewitness performance in forensic investigations. *The Journal of Social Psychology*, *138*, 323-330. doi:10.1080/00224549809600384
- Kebbell, M. R., Milne, R., & Wagstaff, G. F. (1999). The cognitive interview: A survey of its forensic effectiveness. *Psychology, Crime & Law*, *5*, 101-115. doi:10.1080/10683169908414996
- Kebbell, M. R., & Wagstaff, G. F. (1998). Hypnotic interviewing: The best way to interview eyewitnesses? *Behavioral Sciences & the Law*, *16*, 115-129. doi:10.1002/(SICI)1099-0798(199824)16:1<115::AID-BSL296>3.0.CO;2-I
- Kellogg, R. T. (2007). Are written and spoken recall of text equivalent? *American Journal of Psychology*, *120*, 415-428. doi:10.2307/20445412

REFERENCES

- Kirschbaum, C., Kudielka, B. M., Gaab, J., Schommer, N. C., & Hellhammer, D. H. (1999). Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. *Psychosomatic Medicine*, *61*, 154-162. doi:10.1097/00006842-199903000-00006
- Köhnken, G., Milne, R., Memon, A., & Bull, R. (1999). The cognitive interview: A meta-analysis. *Psychology, Crime & Law*, *5*, 3-27. doi:10.1080/10683169908414991
- Koriat, A., & Goldsmith, M. (1994). Memory in naturalistic and laboratory contexts: Distinguishing the accuracy-oriented and quantity-oriented approaches to memory assessment. *Journal of Experimental Psychology: General*, *123*, 297-315. doi:10.1037//0096-3445.123.3.297
- Koriat, A., & Goldsmith, M. (1996). Monitoring and control processes in the strategic regulation of memory accuracy. *Psychological Review*, *103*, 490-517. doi:10.1037/0033-295x.103.3.490
- Kuhlmann, S., Piel, M., & Wolf, O. T. (2005). Impaired memory retrieval after psychosocial stress in healthy young men. *Journal of Neuroscience*, *25*, 2977-2982. doi:10.1523/jneurosci.5139-04.2005
- La Rooy, D., Katz, C., Malloy, L. C., & Lamb, M. E. (2010). Do we need to rethink guidance on repeated interviews? *Psychology, Public Policy, and Law*, *16*, 373-392. doi:10.1037/a0019909
- Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159-174. doi:10.2307/2529310
- Lane, S. M. (2006). Dividing attention during a witnessed event increases eyewitness suggestibility. *Applied Cognitive Psychology*, *20*, 199-212. doi:10.1002/acp.1177
- Lehrer, J. (2010, Dec 13). The truth wears off. *The New Yorker*. Retrieved from http://www.newyorker.com/reporting/2010/12/13/101213fa_fact_lehrer
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.
- Lipton, J. P. (1977). On the psychology of eyewitness testimony. *Journal of Applied Psychology*, *62*, 90-95. doi:10.1037/0021-9010.62.1.90
- Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory*, *12*, 361-366. doi:10.1101/lm.94705
- Loftus, E. F., & Palmer, J. C. (1974). Reconstruction of automobile destruction: An example of the interaction between language and memory. *Journal of Verbal Learning and Verbal Behavior*, *13*, 585-589. doi:10.1016/S0022-5371(74)80011-3
- Logie, R. H., Gilhooly, K. J., & Wynn, V. (1994). Counting on working-memory in arithmetic problem-solving. *Memory & Cognition*, *22*, 395-410. doi:10.3758/bf03200866
- Logie, R. H., Maylor, E. A., Della Sala, S., & Smith, G. (2004). Working memory in event- and time-based prospective memory tasks: Effects of secondary demand and age. *European Journal of Cognitive Psychology*, *16*, 441-456. doi:10.1080/09541440340000114
- Maras, K. L., & Bowler, D. M. (2010). The cognitive interview for eyewitnesses with autism spectrum disorder. *Journal of Autism and Developmental Disorders*, *40*, 1350-1360. doi:10.1007/s10803-010-0997-8
- Maras, K. L., & Bowler, D. M. (2012). Context reinstatement effects on eyewitness memory in autism spectrum disorder. *British Journal of Psychology*, *103*, 330-342. doi:10.1111/j.2044-8295.2011.02077.x
- Maras, K. L., Mulcahy, S., Memon, A., Picariello, F., & Bowler, D. M. (2014). Evaluating the effectiveness of the Self-Administered Interview© for witnesses with autism spectrum disorder. *Applied Cognitive Psychology*, *28*, 693-701. doi:10.1002/acp.3055
- Marchetta, N. D. J., Hurks, P. P. M., Krabbendam, L., & Jolles, J. (2008). Interference control, working memory, concept shifting, and verbal fluency in adults with attention-deficit/hyperactivity disorder (ADHD). *Neuropsychology*, *22*, 74-84. doi:10.1037/0894-4105.22.1.74
- Marsh, E. J., Tversky, B., & Hutson, M. (2005). How eyewitnesses talk about events: Implications for memory. *Applied Cognitive Psychology*, *19*, 531-544. doi:10.1002/acp.1095
- McCabe, D. P., Roediger, H. L., McDaniel, M. A., Balota, D. A., & Hambrick, D. Z. (2010). The relationship between working memory capacity and executive functioning: Evidence for a common executive attention construct. *Neuropsychology*, *24*, 222-243. doi:10.1037/a0017619
- McCauley, M. R., & Fisher, R. P. (1995). Facilitating children's eyewitness recall with the revised cognitive interview. *Journal of Applied Psychology*, *80*, 510-516. doi:10.1037//0021-9010.80.4.510

- McDaniel, M. A., & Fisher, R. P. (1991). Tests and test feedback as learning sources. *Contemporary Educational Psychology, 16*, 192-201. doi:10.1016/0361-476X(91)90037-L
- Meissner, C. A., Hartwig, M., & Russano, M. B. (2010). The need for a positive psychological approach and collaborative effort for improving practice in the interrogation room. *Law and Human Behavior, 34*, 43-45. doi:10.1007/s10979-009-9205-9
- Meissner, C. A., Sporer, S. L., & Susa, K. J. (2008). A theoretical review and meta-analysis of the description-identification relationship in memory for faces. *European Journal of Cognitive Psychology, 20*, 414-455. doi:10.1080/09541440701728581
- Memon, A., Meissner, C. A., & Fraser, J. (2010). The cognitive interview: A meta-analytic review and study space analysis of the past 25 years. *Psychology, Public Policy, and Law, 16*, 340-372. doi:10.1037/a0020518
- Memon, A., Wark, L., Bull, R., & Köhnken, G. (1997). Isolating the effects of the cognitive interview techniques. *British Journal of Psychology, 88*, 179-197. doi:10.1111/j.2044-8295.1997.tb02629.x
- Memon, A., Zaragoza, M., Clifford, B., & Kidd, L. (2010). Inoculation or antidote? The effects of cognitive interview timing on false memory for forcibly fabricated events. *Law and Human Behavior, 34*, 105-117. doi:10.1007/s10979-008-9172-6
- Meyer, T., Smeets, T., Giesbrecht, T., Quaedflieg, C. W. E. M., & Merckelbach, H. (2013). Acute stress differentially affects spatial configuration learning in high and low cortisol-responding healthy adults. *European Journal of Psychotraumatology, 4*, 19854. doi:10.3402/ejpt.v4i0.19854
- Mitchell, K. J., Johnson, M. K., Raye, C. L., & Greene, E. J. (2004). Prefrontal cortex activity associated with source monitoring in a working memory task. *Journal of Cognitive Neuroscience, 16*, 921-934. doi:10.1162/0898929041502724
- Naveh-Benjamin, M., Guez, J., & Marom, M. (2003). The effects of divided attention at encoding on item and associative memory. *Memory & Cognition, 31*, 1021-1035. doi:10.3758/bf03196123
- Neisser, U. (1996). Remembering as doing. *Behavioral and Brain Sciences, 19*, 203. doi:10.1017/S0140525X00042308
- Nicolson, N. A. (2008). Measurement of cortisol. In L. J. Luecken, & L. C. Gallo (Eds.), *Handbook of physiological research methods in health psychology* (pp. 37-74). Los Angeles, CA: Sage.
- Odinot, G., Memon, A., La Rooy, D., & Millen, A. (2013). Are two interviews better than one? Eyewitness memory across repeated cognitive interviews. *PLoS ONE, 8*, e76305. doi:10.1371/journal.pone.0076305
- Odinot, G., Wolters, G., & van Giezen, A. (2013). Accuracy, confidence and consistency in repeated recall of events. *Psychology, Crime & Law, 19*, 629-642. doi:10.1080/1068316x.2012.660152
- Oeberst, A. (2012). If anything else comes to mind . . . better keep it to yourself? Delayed recall is discrediting: Unjustifiably. *Law and Human Behavior, 36*, 266-274. doi:10.1037/h0093966
- Otgaar, H., Howe, M. L., Peters, M., Sauerland, M., & Raymaekers, L. (2013). Developmental trends in different types of spontaneous false memories: Implications for the legal field. *Behavioral Sciences & the Law, 31*, 666-682. doi:10.1002/bsl.2076
- Paterson, H. M., Eijkemans, H., & Kemp, R. I. (2014). Investigating the impact of delayed administration on the efficacy of the Self-Administered Interview. *Psychiatry, Psychology and Law*. Advance online publication. doi:10.1080/13218719.2014.947670
- Perfect, T. J., Wagstaff, G. F., Moore, D., Andrews, B., Cleveland, V., Newcombe, S., Brisbane, K.-A., & Brown, L. (2008). How can we help witnesses to remember more? It's an (eyes) open and shut case. *Law and Human Behavior, 32*, 314-324. doi:10.1007/s10979-007-9109-5
- Peters, M., Giesbrecht, T., Jellic, M., & Merckelbach, H. (2007). The random number generation task: Psychometric properties and normative data of an executive function task in a mixed sample. *Journal of the International Neuropsychological Society, 13*, 626-634. doi:10.1017/s1355617707070786
- Peters, M., Jellic, M., Haas, N., & Merckelbach, H. (2006). Mild executive dysfunctions in undergraduates are related to recollecting words never presented. *International Journal of Neuroscience, 116*, 1065-1077. doi:10.1080/00207450600808768

REFERENCES

- Peters, M. J. V., Jelicic, M., Verbeek, H., & Merckelbach, H. (2007). Poor working memory predicts false memories. *European Journal of Cognitive Psychology, 19*, 213-232. doi:10.1080/09541440600760396.
- Pica, G., Pierro, A., Bélanger, J. J., & Kruglanski, A. W. (2014). The role of need for cognitive closure in retrieval-induced forgetting and misinformation effects in eyewitness memory. *Social Cognition, 32*, 337-359.
- Piper, A. (1993). "Truth serum" and "recovered memories" of sexual abuse: A review of the evidence. *The Journal of Psychiatry & Law, 21*, 447-471.
- Potter, R., & Brewer, N. (1999). Perceptions of witness behaviour-accuracy relationships held by police, lawyers and mock-jurors. *Psychiatry, Psychology and Law, 6*, 97-103. doi:10.1080/13218719909524952
- Powell, M. B., Fisher, R. P., & Wright, R. (2005). Investigative interviewing. In N. Brewer, & K. D. Williams (Eds.). *Psychology and law: An empirical perspective* (pp. 11-42). New York: Guilford.
- Pozzulo, J. D., Coplan, R. J., & Wilson, J. (2005). Not too 'shy' to help catch a thief: Recall memory of shy versus non-shy witnesses. *Personality and Individual Differences, 39*, 1361-1372. doi:10.1016/j.paid.2005.06.010
- Quaedflieg, C. W. E. M., Meyer, T., Schwabe, L., & Smeets, T. (2013). Time dependent effects of stress prior to encoding on processing efficiency and 24h delayed retrieval. *Psychoneuroendocrinology, 38*, 3057-3069. doi:10.1016/j.psyneuen.2013.09.002
- Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*, 249-255. doi:10.1111/j.1467-9280.2006.01693.x
- Roosendaal, B. (2002). Stress and memory: Opposing effects of glucocorticoids on memory consolidation and memory retrieval. *Neurobiology of Learning and Memory, 78*, 578-595. doi:10.1006/nlme.2002.4080
- Roosendaal, B., & McGaugh, J. L. (2011). Memory modulation. *Behavioral Neuroscience, 125*, 797-824. doi:10.1037/a0026187
- Rosen, V. M., & Engle, R. W. (1997). The role of working memory capacity in retrieval. *Journal of Experimental Psychology: General, 126*, 211-227. doi:10.1037//0096-3445.126.3.211
- Rosenthal, R. (2002). Covert communication in classrooms, clinics, courtrooms, and cubicles. *American Psychologist, 57*, 839-849. doi:10.1037/0003-066X.57.11.839
- Rossion, B., & Pourtois, G. (2004). Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception, 33*, 217-236. doi:10.1068/p5117
- Rubin, D. C., & Wenzel, A. E. (1996). One hundred years of forgetting: A quantitative description of retention. *Psychological Review, 103*, 734-760. doi:10.1037/0033-295x.103.4.734
- Saks, M. J., & Koehler, J. J. (2005). The coming paradigm shift in forensic identification science. *Science, 309*, 892-895. doi:10.1126/science.1111565
- Salthouse, T. A. (1994). The aging of working memory. *Neuropsychology, 8*, 535-543. doi:10.1037/0894-4105.8.4.535
- Sanders, G. S., & Simmons, W. L. (1983). Use of hypnosis to enhance eyewitness accuracy: Does it work? *Journal of Applied Psychology, 68*, 70-77. doi:10.1037/0021-9010.68.1.70
- Sauerland, M., Holub, F. E., & Sporer, S. L. (2008). Person descriptions and person identifications: Verbal overshadowing or recognition criterion shift? *European Journal of Cognitive Psychology, 20*, 497-528. doi:10.1080/09541440701728417
- Sauerland, M., & Krix, A. C. (2011). Vorbereitung und Durchführung von Gegenüberstellungen und Lichtbildvorlagen vor dem Hintergrund einer positiven Rechtspsychologie [Lineup preparation and administration in the context of a positive legal psychology]. *Praxis der Rechtspsychologie, 21*, 119-132.
- Sauerland, M., Raymaekers, L. H. C., Otgaar, H., Memon, A., Waltjen, T. T., Nivo, M., Slegers, C., Broers, N. J., & Smeets, T. (2014). *Stress, stress-induced cortisol responses, and eyewitness identification performance*. Manuscript submitted for publication.
- Sauerland, M., & Sporer, S. L. (2011). Written vs. spoken eyewitness accounts: Does modality of testing matter? *Behavioral Sciences & the Law, 29*, 846-857. doi:10.1002/bsl.1013

- Schacter, D. L., Norman, K. A., & Koutstaal, W. (1998). The cognitive neuroscience of constructive memory. *Annual Review of Psychology, 49*, 289-318. doi:10.1146/annurev.psych.49.1.289
- Scheck, B., Neufeld, P., & Dwyer, J. (2003). *Actual innocence: When justice goes wrong and how to make it right*. New York: New American Library.
- Schell, J. M., & Kassir, S. M. (2009, September). *Welche Auswirkungen können Geständnisse haben? Laien-Ansichten über Ermittlungsverfahren und deren Risiken [What impacts can confessions have? Laypersons' beliefs about police investigations and their risks]*. Paper presented at the 13th biannual meeting of the German Association of Psychology and Law, Gießen, Germany.
- Schell, J. M., & Kassir, S. M. (2012, March). *Police practices and beliefs: A survey of European investigators*. Paper presented at the annual meeting of the American Psychology-Law Society, San Juan, Puerto Rico.
- Schooler, J. W. (2011). Unpublished results hide the decline effect. *Nature, 470*, 437. doi:10.1038/470437a
- Schwabe, L., Joëls, M., Roozendaal, B., & Oitzl, M. S. (2012). Stress effects on memory: An update and integration. *Neuroscience & Biobehavioral Reviews, 36*, 1740-1749. doi:10.1016/j.neubiorev.2011.07.002
- Scrivner, E., & Safer, M. A. (1988). Eyewitnesses show hypermnnesia for details about a violent event. *Journal of Applied Psychology, 73*, 371-377. doi:10.1037//0021-9010.73.3.371
- Shaffer, J. P. (1986). Modified sequentially rejective multiple test procedures. *Journal of the American Statistical Association, 81*, 826-831. doi:10.1080/01621459.1986.10478341
- Sixth Circuit Criminal Pattern Jury Instructions. (2013). Retrieved April 24th, 2014, from http://www.ca6.uscourts.gov/internet/crim_jury_insts.htm
- Smeets, T., Candel, I., & Merckelbach, H. (2004). Accuracy, completeness, and consistency of emotional memories. *American Journal of Psychology, 117*, 595-609. doi:10.2307/4148994
- Smeets, T., Cornelisse, S., Quaedflieg, C. W. E. M., Meyer, T., Jelicic, M., & Merckelbach, H. (2012). Introducing the Maastricht Acute Stress Test (MAST): A quick and non-invasive approach to elicit robust autonomic and glucocorticoid stress responses. *Psychoneuroendocrinology, 37*, 1998-2008. doi:10.1016/j.psyneuen.2012.04.012
- Smeets, T., Otgaar, H., Candel, I., & Wolf, O. T. (2008). True or false? Memory is differentially affected by stress-induced cortisol elevations and sympathetic activity at consolidation and retrieval. *Psychoneuroendocrinology, 33*, 1378-1386. doi:10.1016/j.psyneuen.2008.07.009
- Smith, S. M. (1988). Environmental context dependent memory. In G. M. Davies, & D. M. Thomson (Eds.), *Memory in context: Context in memory* (pp. 13-33). New York: Wiley.
- Smith, S. M. (1994). Theoretical principles of context-dependent memory. In M. Gruneberg, & P. E. Morris (Eds.), *Theoretical aspects of memory* (pp. 168-195). London: Routledge.
- Smith, S. M., Glenberg, A. G., & Bjork, R. A. (1978). Environmental context and human memory. *Memory & Cognition, 6*, 342-353. doi:10.3758/BF03197465
- Smith, S. M., & Vela, E. (2001). Environmental context-dependent memory: A review and meta-analysis. *Psychonomic Bulletin & Review, 8*, 203-220. doi:10.3758/bf03197114
- Sporer, S. L. (1992). Post-dicting eyewitness identification accuracy: Confidence, decision-times and person descriptions among choosers and non-choosers. *European Journal of Social Psychology, 20*, 157-180. doi:10.1002/ejsp.2420220205
- Sporer, S. L. (1993). Eyewitness identification accuracy, confidence and decision-times in simultaneous and sequential lineups. *Journal of Applied Psychology, 78*, 22-33. doi:10.1037/0021-9010.78.1.22
- Sporer, S. L. (2007). Person descriptions as retrieval cues: Do they really help? *Psychology, Crime & Law, 13*, 591-609. doi:10.1080/10683160701253986
- Sporer, S. L., & Cohn, L. D. (2011). Meta-analysis. In B. D. Rosenfeld, & S. D. Penrod (Eds.), *Research methods in forensic psychology* (pp. 43-62). New York: Wiley.
- Stebly, N. M., & Bothwell, R. K. (1994). Evidence for hypnotically refreshed testimony: The view from the laboratory. *Law and Human Behavior, 18*, 635-651. doi:10.1007/BF01499329
- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. New York: Pearson.

REFERENCES

- Tuckey, M. R., & Brewer, N. (2003). The influence of schemas, stimulus ambiguity, and interview schedule on eyewitness memory over time. *Journal of Experimental Psychology: Applied*, *9*, 101-118. doi:10.1037/1076-898X.9.2.101
- Tulving, E., & Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, *80*, 352-373. doi:10.1037/h0020071
- Tulving, E., & Watkins, M. J. (1975). Structure of memory traces. *Psychological Review*, *82*, 261-275. doi:10.1037/h0076782
- Turner, M. L., & Engle, R. W. (1989). Is working memory capacity task dependent? *Journal of Memory and Language*, *28*, 127-154. doi:10.1016/0749-596X(89)90040-5
- Tversky, B., & Marsh, E. J. (2000). Biased retellings of events yield biased memories. *Cognitive Psychology*, *40*, 1-38. doi:10.1006/cogp.1999.0720
- Unsworth, N. (2008). Exploring the retrieval dynamics of delayed and final free recall: Further evidence for temporal-contextual search. *Journal of Memory and Language*, *59*, 223-236. doi:10.1016/j.jml.2008.04.002
- Unsworth, N., & Brewer, G. A. (2009). Examining the relationships among item recognition, source recognition, and recall from an individual differences perspective. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *35*, 1578-1585. doi:10.1037/a0017255
- Unsworth, N., & Brewer, G. A. (2010a). Individual differences in false recall: A latent variable analysis. *Journal of Memory and Language*, *62*, 19-34. doi:10.1016/j.jml.2009.08.002
- Unsworth, N., & Brewer, G. A. (2010b). Variation in working memory capacity and intrusions: Differences in generation or editing? *European Journal of Cognitive Psychology*, *22*, 990-1000. doi:10.1080/09541440903175086
- Unsworth, N., & Engle, R. W. (2007). The nature of individual differences in working memory capacity: Active maintenance in primary memory and controlled search from secondary memory. *Psychological Review*, *114*, 104-132. doi:10.1037/0033-295x.114.1.104
- Unsworth, N., Heitz, R. R., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods*, *37*, 498-505. doi:10.3758/bf03192720
- Uviller, H. R. (1993). Credence, character, and the rules of evidence: Seeing through the liar's tale. *Duke Law Journal*, *42*, 776-832. doi:10.2307/1372715
- Vallano, J. P., & Schreiber Compo, N. (2011). A comfortable witness is a good witness: Rapport-building and susceptibility to misinformation in an investigative mock-crime interview. *Applied Cognitive Psychology*, *25*, 960-970. doi:10.1002/acp.1789
- Vredeveltdt, A., Baddeley, A. D., & Hitch, G. J. (2014). The effectiveness of eye-closure in repeated interviews. *Legal and Criminological Psychology*, *19*, 282-295. doi:10.1111/lcrp.12013
- Vredeveltdt, A., Hitch, G. J., & Baddeley, A. D. (2011). Eye-closure helps memory by reducing cognitive load and enhancing visualisation. *Memory & Cognition*, *39*, 1253-1263. doi:10.3758/s13421-011-0098-8
- Vredeveltdt, A., & Penrod, S. D. (2013). Eye-closure improves memory for a witnessed event under naturalistic conditions. *Psychology, Crime & Law*, *19*, 893-905. doi:10.1080/1068316x.2012.700313
- Wagenaar, W. A., & van der Schrier, J. H. (1996). Face recognition as a function of distance and illumination: A practical tool for use in the courtroom. *Psychology, Crime & Law*, *2*, 321-332. doi:10.1080/10683169608409787
- Wagstaff, G. F., Wheatcroft, J., Cole, J. C., Brunas-Wagstaff, J., Blackmore, V., & Pilkington, A. (2008). Some cognitive and neuropsychological aspects of social inhibition and facilitation. *European Journal of Cognitive Psychology*, *20*, 828-846. doi:10.1080/09541440701469749
- Walsh, E., Moran, P., Scott, C., McKenzie, K., Burns, T., Creed, F., Tyrer, P., Murray, R. M., & Fahy, T. (2003). Prevalence of violent victimisation in severe mental illness. *The British Journal of Psychiatry*, *183*, 233-238. doi:10.1192/bjp.183.3.233
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063-1070. doi:10.1037/0022-3514.54.6.1063

- Waubert de Puiseau, B., Assfalg, A., Erdfelder, E., & Bernstein, D. M. (2012). Extracting the truth from conflicting eyewitness reports: A formal modeling approach. *Journal of Experimental Psychology: Applied*, *18*, 390-403. doi:10.1037/a0029801
- Weber, N., & Brewer, N. (2008). Eyewitness recall: Regulation of grain size and the role of confidence. *Journal of Experimental Psychology: Applied*, *14*, 50-60. doi:10.1037/1076-898x.14.1.50
- Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*, 1049-1062. doi:10.1037//0022-3514.67.6.1049
- Wechsler, D. (1997). *Administration and scoring manual for the Wechsler Adult Intelligence Scale* (3rd ed.). San Antonio, TX: Psychological Corporation.
- Wells, G. L. (1978). Applied eyewitness-testimony research: System and estimator variables. *Journal of Personality and Social Psychology*, *36*, 1546-1557. doi:10.1037//0022-3514.36.12.1546
- Wells, G. L., Memon, A., & Penrod, S. D. (2006). Eyewitness evidence: Improving its probative value. *Psychological Science in the Public Interest*, *7*, 45-75. doi:10.1111/j.1529-1006.2006.00027.x
- Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, *22*, 603-647. doi:10.1023/A:1025750605807
- Wessel, I., van der Kooy, P., & Merckelbach, H. (2000). Differential recall of central and peripheral details of emotional slides is not a stable phenomenon. *Memory*, *8*, 97-111. doi:10.1080/096582100387641
- Wright, A. M., & Holliday, R. E. (2007). Enhancing the recall of young, young-old and old-old adults with cognitive interviews. *Applied Cognitive Psychology*, *21*, 19-43. doi:10.1002/acp.1260
- Zajonc, R. B. (1965). Social facilitation. *Science*, *149*, 269-274. doi:10.1126/science.149.3681.269
- Zhu, B., Chen, C. S., Loftus, E. F., Lin, C. D., He, Q. H., Chen, C. H., Li, H., Xue, G., Lu, Z. L., & Dong, Q. (2010a). Individual differences in false memory from misinformation: Cognitive factors. *Memory*, *18*, 543-555. doi:10.1080/09658211.2010.487051
- Zhu, B., Chen, C., Loftus, E. F., Lin, C., He, Q., Chen, C., . . . Dong, Q. (2010b). Individual differences in false memory from misinformation: Personality characteristics and their interactions with cognitive abilities. *Personality and Individual Differences*, *48*, 889-894. doi:10.1016/j.paid.2010.02.016

Summary

English

Eyewitness testimony is an important piece of evidence that often provides the critical leads for the investigations. In many cases it is the only available evidence (de Poot, Bokhorst, van Koppen, & Muller 2004; Fisher, 2010). On the other hand, incorrect eyewitness testimony is the major cause for wrongful convictions (Wells et al., 1998). Therefore, it is essential to study techniques that improve witness accounts and support witnesses in providing complete and accurate statements. In line with such a positive legal psychology that aims to increase the reliability of eyewitness accounts (Meissner, Hartwig, & Russano, 2010), the present dissertation sought to examine the beneficial effects of retrieval support in eyewitnesses. For this purpose, the Self-Administered Interview© (SAI; Gabbert, Hope, & Fisher, 2009) served as a proxy tool for providing eyewitnesses with retrieval support. The SAI, which contains several memory-enhancing components, is independently completed by the witnesses at the crime scene to ensure an early comprehensive statement.

In *chapter 1* two case vignettes are presented. These cases illustrate how, on the one hand, eyewitness evidence can lead to wrongful convictions and how, on the other hand, it can be crucial for apprehending the true perpetrator. Building on this, the concept and the importance of a positive psychological approach, which goes beyond documenting problems and shortcomings of eyewitness statements, are presented. Subsequently, an overview of recent research in the field of interviewing eyewitnesses, including the SAI, that is in line with a positive legal psychology is given. Additionally, the outline of this dissertation is presented and the following research questions to be addressed in the following chapters are formulated: Can retrieval support with the SAI help witnesses achieve comprehensive and accurate accounts, even when the witnessing conditions during encoding or retrieval were suboptimal? Are repeated interviews a means to provide retrieval support and do police detectives see the merit of repeated interviews?

The SAI is a written interview, while police interviews are usually conducted in spoken. Before conducting in-depth research on the SAI, the question arises whether the modality of the interview (written vs. spoken) influences recall performance. *Chapter 2* describes two experiments ($N = 135$ and $N = 124$) that examined this question. Across both experiments, there was no support for a modality effect. Moreover, level of executive functioning and instruction comprehensiveness did not interact with the modality to influence recall performance. It thus seems that the modality is not a decisive factor when it comes to the quantity and quality of eyewitness testimony, at least for highly educated witnesses. Therefore, the fact that the SAI is a written interview may be considered unproblematic. The experiment reported in *chapter 3* ($N = 127$) focused on the suboptimal condition of stress during encoding. Prior to witnessing a live staged crime, participants either underwent a stress task or a control task. Participants then completed an

SAI or a free recall (FR). While the stress manipulation was successful, as indicated by increased cortisol levels and increased negative affect in the stress group, stress had no impact on memory performance. As expected, retrieval support with the SAI elicited more comprehensive statements than FR. Importantly, there were no interactions between stress and interview type. Hence, it can be concluded that the SAI can be safely administered under different levels of stress.

Chapter 4 comprises two experiments. Experiment 1 ($N = 88$) dealt with the suboptimal situation of divided attention during encoding. It addressed the question whether retrieval support with the SAI can help witnesses provide more comprehensive and more accurate statements than FR even when the witness was distracted during the crime. In contrast to what we expected, retrieval support with the SAI led to more correct details than an FR only under full attention conditions. If the witness had been distracted during encoding, the advantage of retrieval support disappeared. Previous research has shown that witnesses can achieve more accurate reports in a second delayed interview if they completed the SAI relative to an FR immediately after the crime (Hope, Gabbert, Fisher, & Jamieson, 2014). According to Hope et al., this is due to interviews with ample retrieval support, such as the SAI, strengthening the memories more than less effortful retrieval attempts, such as FR. However, in Experiment 1, no such memory-preserving effect was detected. Therefore, Experiment 2 ($N = 81$) was conducted to address the conflicting findings. The results suggested that providing retrieval support during the first interview does not automatically preserve memory in the subsequent interview. Rather, the memory-preserving effect of initial retrieval support only seems to surface if an interview is used at the second recall attempt that fosters comprehensive recall.

Chapter 5 also focused on repeated interviews. More specifically, the effect of retrieval support through repeated interviews was examined, especially in terms of reminiscence, the recall of previously unrecalled information. Reminiscence is a normal memory phenomenon that occurs parallel to forgetting (Erdelyi, 2010). Yet, in court, recall of reminiscent information is often used to discredit the witness' credibility (Potter & Brewer, 1999). Participants of the memory group ($N = 84$) underwent two recall attempts about the same incident with a one week interval between the first and second interview. The results confirmed that repeated interviews support retrieval by promoting reminiscence. Specifically, every participant experienced reminiscence and the accuracy of reminiscent information was high. The latter was not influenced by the amount of retrieval support (SAI vs. FR) in the first interview. To examine whether experienced police detectives see the merit of repeated interviews, the recall performance of the memory group was compared with the police detectives' estimations of the performance. This was not the case, as this estimation group ($N = 81$) expressed a profound mistrust in reminiscence by severely underestimating the accuracy of reminiscence. Moreover, they articulated the concern that reminiscence was the result of negative external

SUMMARY

influences (e.g., co-witness misinformation). The results demonstrate the need to revise interview trainings for members of the legal system.

In *chapter 6* an overarching analysis ($N = 125$) of the experiments reported in *chapter 4* are described. This analysis focused on a suboptimal condition eminent during retrieval, that is, reduced working memory and source memory (i.e., memory for the origin of a recollection). Both have been associated with recall of fewer correct information and more recall errors (Unsworth & Brewer, 2010a, 2010b). We examined whether retrieval support with the SAI (relative to FR) could compensate for this deficit by providing more structure for retrieval and facilitating source monitoring. For this purpose, participants completed working and source memory tasks and subsequently completed either an SAI or an FR. In contrast to our expectations, the presence of retrieval support did not moderate the relationship between working memory capacity and recall performance. In one source memory task, better source memory was associated with more accurate accounts in the retrieval support group, whereas in the FR group, lower source memory was associated with higher accuracy. This suggests that high retrieval-support interviews may not be suitable for witnesses with reduced source memory.

Chapter 7 contains a general discussion of the main findings. Moreover, a meta-analytical synthesis of the four experiments in which retrieval support with the SAI was compared to FR in terms of recall performance is presented. This meta-analysis found that at the first recall attempt the SAI led to more correct and more incorrect details than FR, while accuracy was unaffected. At the second recall attempt, no significant effect sizes were obtained, indicating that across the experiments of the present dissertation no memory-preserving effect of initial SAI completion was found. Finally, an evaluation of the effectiveness of the SAI, as well as practical implications and suggestions for future research are presented.

German

Zeugenaussagen sind wichtige Beweismittel, die oftmals die entscheidenden Hinweise für die polizeiliche Ermittlungsarbeit erbringen. In vielen Fällen stellen sie sogar das einzig verfügbare Beweismittel dar (de Poot, Bokhorst, van Koppen, & Muller 2004; Fisher, 2010). Andererseits gelten fehlerhafte Zeugenaussagen als Hauptgrund für Fehltritte (Wells et al., 1998). Daher ist es von wesentlicher Bedeutung, Verfahren zu erforschen, die Zeugenaussagen verbessern und Zeugen dabei unterstützen, möglichst umfangreiche und richtige Aussagen abzugeben. Im Einklang mit einer solchen positiven Rechtspsychologie, die beabsichtigt, die Zuverlässigkeit von Zeugenaussagen zu erhöhen (Meissner, Hartwig, & Russano, 2010), war es das Ziel der vorliegenden Dissertation, die positiven Effekte von Abrufunterstützung bei Augenzeugen zu untersuchen. Zu diesem Zweck diente das Eigenständige Vernehmungsprotokoll für Augenzeugen (EVA; Englisch: Self-Administered Interview©, SAI; Gabbert, Hope, & Fisher, 2009) stellvertretend für Verfahren, die ein hohes Maß an Abrufunterstützung bieten. Das EVA, das verschiedene Bestandteile enthält, welche die Erinnerungsleistung verbessern, wird von Zeugen völlig eigenständig noch am Tatort ausgefüllt, um den Zeugen eine frühzeitige umfangreiche erste Aussage zu ermöglichen.

In *Kapitel 1* werden zunächst zwei Fallvignetten beschrieben. Diese Fälle sollen zeigen, wie Zeugenaussagen einerseits zu Fehltritten führen können, wie sie andererseits jedoch auch von wesentlicher Bedeutung für die Ergreifung des wahren Täters sein können. Darauf aufbauend werden das Konzept und die Bedeutung eines positiven rechtspsychologischen Ansatzes erläutert, der über das ledigliche Aufzeigen von Problemen und Defiziten in Zeugenaussagen hinausgeht. Hiernach erfolgt ein Überblick über neuere Forschungsarbeiten aus dem Bereich der Zeugenvernehmung, die im Zeichen einer positiven Rechtspsychologie stehen, darunter das EVA. Schließlich wird die Gliederung der vorliegenden Dissertation vorgestellt und es werden die folgenden Forschungsfragen formuliert, auf die in den nachfolgenden Kapiteln eingegangen werden soll: Kann Abrufunterstützung mit dem EVA Zeugen auch dann zu umfangreichen und richtigen Aussagen verhelfen, wenn die Zeugenbedingungen während Enkodierung und Abruf nicht optimal waren? Sind wiederholte Vernehmungen ein Mittel zur Abrufunterstützung und erkennen erfahrene Polizeibeamte den Nutzen wiederholter Vernehmungen?

Während es sich bei dem EVA um ein schriftliches Vernehmungsverfahren handelt, werden polizeiliche Vernehmungen normalerweise mündlich durchgeführt. Daher sollte zunächst die Frage geklärt werden, ob die Modalität der Vernehmung (schriftlich vs. mündlich) die Abrufleistung beeinflusst. *Kapitel 2* befasst sich mit zwei Experimenten ($N = 135$ und $N = 124$), die genau dieser Frage nachgegangen sind. In beiden Experimenten fand sich kein Beleg für einen Modalitätseffekt. Auch zeigten sich keine signifikanten Interaktionen zwischen Modalität und Exekutivfunktionen bzw. Umfang der Abrufinstruktionen auf die Erinnerungsleistung.

tung. Damit scheint die Modalität der Vernehmung kein entscheidender Einflussfaktor hinsichtlich Umfang und Richtigkeit von Zeugenaussagen zu sein, zumindest bei Zeugen mit einem hohen Bildungsgrad. Daher erscheint die Tatsache, dass das EVA ein schriftliches Vernehmungsverfahren ist, unproblematisch.

Das Experiment ($N = 127$), das in *Kapitel 3* beschrieben wird, befasste sich mit der suboptimalen Zeugenbedingung von erhöhtem Stress in der Enkodierungsphase. Die Versuchsteilnehmer durchliefen entweder eine Stress auslösende Aufgabe oder eine Kontrollaufgabe und wurden danach Zeuge eines nachgestellten Diebstahls. Anschließend bearbeiteten die Versuchspersonen entweder ein EVA oder einen freien Abruf. Obwohl die Stressmanipulation erfolgreich war, was sich an einem Anstieg der Kortisolwerte und einem Anstieg von negativen Emotionen innerhalb der Stressgruppe zeigte, hatte der Stress keinen Einfluss auf die Erinnerungsleistung. Erwartungsgemäß erzielten die Versuchspersonen, die Abrufunterstützung mit dem EVA erhielten, umfangreichere Aussagen als die Versuchspersonen mit einem freien Abruf. Da es keine signifikanten Interaktionen zwischen Stress und Art des Vernehmungsverfahrens gab, lässt sich schließen, dass das EVA ohne weiteres eingesetzt werden kann, unabhängig davon, welches Ausmaß an Stress ein Zeuge während der Tat erlebt hat.

Kapitel 4 umfasst zwei Experimente. Dabei beschäftigte sich Experiment 1 ($N = 88$) mit einer weiteren suboptimalen Bedingung in der Enkodierungsphase, nämlich mit geteilter Aufmerksamkeit. Das Experiment ging der Frage nach, ob Abrufunterstützung mit dem EVA Zeugen auch dann zu umfangreichen und richtigen Aussagen verhilft, wenn der Zeuge während der Tat abgelenkt war. Entgegen unserer Erwartungen führte Abrufunterstützung mit dem EVA nur dann zu mehr richtig erinnerten Details als ein freier Abruf, wenn die Versuchspersonen den Stimulusfilm mit voller Aufmerksamkeit gesehen hatten. Waren die Zeugen jedoch abgelenkt, gab es keinen Unterschied mehr bezüglich der Anzahl der richtigen Details in Abhängigkeit vom eingesetzten Vernehmungsverfahren. Frühere Forschung konnte zeigen, dass Versuchspersonen in einer späteren zweiten Vernehmung eine höhere Richtigkeit erzielten, wenn sie unmittelbar nach dem beobachteten Ereignis zunächst ein EVA (anstelle eines freien Abrufs) ausgefüllt hatten (Hope, Gabbert, Fisher, & Jamieson, 2014). Laut Hope et al. ließ sich dieses Ergebnis damit erklären, dass Vernehmungsmethoden, die, wie das EVA, eine umfangreiche Abrufunterstützung bieten, die Gedächtnisinhalte in einem höheren Ausmaß stärken als weniger umfassende Vernehmungsmethoden, wie der freie Abruf. Jedoch zeigte sich im Experiment 1 von Kapitel 4 kein solcher gedächtniskonservierender Effekt. Aus diesem Grund wurde ein weiteres Experiment durchgeführt (Experiment 2, $N = 81$), um den widersprüchlichen Ergebnissen nachzugehen. Die Ergebnisse dieses Experiments legten nahe, dass Abrufunterstützung während der ersten Vernehmung nicht automatisch zu einem gedächtniskonservierenden Effekt führt. Stattdessen scheint der Effekt nur dann aufzutreten, wenn zum zweiten Ab-

rufzeitpunkt ein Interview verwendet wird, das einen umfangreichen Abruf fördert.

In *Kapitel 5* ging es erneut um wiederholte Vernehmungen. Dabei wurde der Effekt von Abrufunterstützung mittels wiederholter Vernehmungen speziell im Hinblick auf Reminiszenz untersucht. Reminiszenz bezeichnet die Erinnerung von Informationen, die in einer vorherigen Vernehmung noch nicht erinnert wurden. Reminiszenz ist ein normales gedächtnispsychologisches Phänomen, das gleichzeitig mit dem Vergessen von Informationen auftritt (Erdelyi, 2010). Dagegen wird die Erinnerung von reminiszenten Informationen vor Gericht oftmals als Zeichen gewertet, dass die Aussage des Zeugen nicht glaubwürdig ist (Potter & Brewer, 1999). Im vorliegenden Experiment gaben die Versuchsteilnehmer aus der Gedächtnisgruppe ($N = 84$) im Abstand von einer Woche zwei Aussagen über einen Vorfall ab. Wie erwartet zeigten die Ergebnisse, dass wiederholte Vernehmungen die Abrufleistung mittels Reminiszenz befördern. So trat bei jedem Versuchsteilnehmer Reminiszenz auf. Außerdem wiesen reminiszente Details eine hohe Richtigkeit auf. Diese wurde nicht vom Ausmaß der Abrufunterstützung (EVA vs. freier Abruf) in der ersten Vernehmung beeinflusst. Um zu untersuchen, ob erfahrene Polizeibeamte den Nutzen von wiederholten Vernehmungen erkennen, wurde die Abrufleistung der Gedächtnisgruppe mit den Schätzungen der Polizeibeamten verglichen. Tatsächlich sah diese Schätzgruppe ($N = 81$) keinen Nutzen in wiederholten Vernehmungen. Stattdessen brachten die Polizeibeamten ein großes Misstrauen gegenüber Reminiszenz zum Ausdruck, indem sie die Richtigkeit reminiszenter Informationen stark unterschätzten und angaben, dass Reminiszenz ausschließlich das Ergebnis von negativen äußeren Einflüssen sei (z. B. Falschinformation von anderen Augenzeugen oder aus Medienberichten). Diese Ergebnisse verdeutlichen die Notwendigkeit von Anpassungen von Schulungen auf dem Gebiet der Zeugenvernehmung für Polizisten.

Kapitel 6 beinhaltet eine übergreifende Analyse ($N = 125$) der beiden Experimente aus Kapitel 4. Dabei ging es um eine suboptimale Situation in der Abrufphase, nämlich geringe Kapazitäten des Arbeits- und Quellengedächtnisses (d. h. Gedächtnis für den Ursprung einer Erinnerung). In der Vergangenheit konnte gezeigt werden, dass geringere Kapazitäten des Arbeits- und Quellengedächtnisses zum Abruf von weniger richtigen und mehr falschen Informationen führten (Unsworth & Brewer, 2010a, 2010b). Ziel der vorliegenden Analyse war es herauszufinden, ob Abrufunterstützung mit dem EVA (im Vergleich zu einem freien Abruf) diese Defizite ausgleichen kann, indem sie eine höhere Struktur für den Abruf bietet sowie die Quellenüberwachung erleichtert. Hierzu bearbeiteten die Versuchspersonen einige Arbeits- und Quellengedächtnisaufgaben und füllten anschließend entweder ein EVA oder einen freien Abruf aus. Anders als erwartet war das Vorliegen von Abrufunterstützung kein Moderator des Zusammenhangs zwischen Arbeitsgedächtniskapazität und Abrufleistung. Bei einer (von zwei) Quellengedächtnisaufgabe ergab sich folgendes Ergebnismuster: Für die Versuchspersonen, die Abruf-

SUMMARY

unterstützung erhielten, galt: je besser das Quellengedächtnis, desto höher die Richtigkeit der Aussage. Für die Versuchspersonen, die einen freien Abruf bearbeiteten, galt dagegen: je schlechter das Quellengedächtnis, desto höher die Richtigkeit der Aussage. Dies legt nahe, dass Interviews mit Abrufunterstützung für Zeugen mit geringem Quellengedächtnis womöglich nicht sonderlich geeignet sind.

Die allgemeine Diskussion der Hauptergebnisse ist in *Kapitel 7* nachzulesen. In diesem Kapitel erfolgt auch eine metaanalytische Zusammenfassung der vier Experimente dieser Dissertation, in denen Abrufunterstützung mit dem EVA im Hinblick auf die Erinnerungsleistung mit dem freien Abruf verglichen wurde. Die Metaanalyse konnte zeigen, dass zum ersten Abrufzeitpunkt das EVA zu mehr richtigen und mehr falschen Details führte als der freie Abruf. Die Richtigkeit der Aussagen wurde hingegen nicht beeinflusst. Beim zweiten Abrufzeitpunkt ergaben sich keinerlei signifikante Effektstärken. Dies bedeutet, dass über alle Experimente hinweg kein gedächtniskonservierender Effekt des EVA auf die nachfolgende Aussage gefunden wurde. Abschließend wird in diesem Kapitel eine Bewertung der Effektivität des EVA abgegeben sowie auf praktische Auswirkungen und Ideen für zukünftige Forschungsvorhaben eingegangen.

Valorization Addendum

Relevance

As mentioned in chapter 1, even in the era of DNA testing, eyewitness evidence is the only available evidence in many criminal cases and often provides the critical leads for the investigation (de Poot, Bokhorst, van Koppen, & Muller, 2004; Kebbell & Milne, 1998). On the other hand, errors in eyewitness accounts are the primary cause for wrongful convictions (Saks & Koehler, 2005). Therefore, the societal relevance of research that is aimed at improving the reliability of eyewitness testimony, such as was presented in this dissertation, and that can ultimately reduce the risk of wrongful convictions is self-evident. Miscarriages of justice are not only a tragedy for the wrongfully convicted individual. They also undermine people's trust in the judicial system and as such affect society as a whole. Moreover, if the wrong person is apprehended and sentenced, – this is something that is easily overlooked – the real perpetrator goes free and has the opportunity to reoffend, which, of course, comes at a cost to society. But even if we do not assume the worst case (i.e., a wrongful conviction), the problem of the perpetrator going free persists, if a crime cannot be solved because of scarce eyewitness evidence that, for example, did not allow for a reconstruction of the crime. Hence, it is important to obtain comprehensive and highly accurate accounts. Furthermore, obtaining such testimony may expedite trials at court, which saves resources. The experiments described in this dissertation showed that the Self-Administered Interview© (SAI) has the potential to elicit both comprehensive and accurate accounts in the immediate aftermath of a crime and has the potential to support the police.

Target groups

First and foremost, law enforcement agencies are the obvious target group of the research presented here, as the SAI, which had a central role in the described experiments, is an interview tool that is distributed by the police at the crime scene (Gabbert, Hope, & Fisher, 2009). It was developed to support the police immediately after the crime, especially when there are multiple witnesses. As mentioned in chapter 1, the police often do not have the time and resources to conduct timely interviews and interviews are delayed as a result (Gabbert et al., 2009). Delaying the interview, however, is problematic because witnesses are likely to forget details and/or to be exposed to misinformation (Ayers & Reder, 1998; Rubin & Wenzel, 1996). When a crime occurs, the police have to deal with multiple tasks. In addition to obtaining eyewitness evidence, they have to secure the crime scene and block it from gapers and other possible trespassers, collect physical evidence, or initiate a search for the perpetrator, just to name a few. Use of the SAI is likely to come as a relief to the police, as it does not produce extra workload, but saves resources for other urgent matters that need to be taken care of directly after the crime. Likewise, the results may be of interest to public prosecutors and judges

who are, of course, also interested in obtaining and preserving reliable eyewitness accounts. Although the police may be the primary target of this research, we believe that it should be of interest to every member of the society who can potentially become a witness of a crime. More specifically, witnesses and victims may feel that they are not taken seriously, if their interviews are delayed and they are not given the opportunity to make an early statement. Providing them with an SAI in cases in which no early personal interview is possible may prevent the development of resentment towards the police and maintain witnesses' willingness to cooperate with the police.

Activities and Products

Since an early stage of the PhD project, I have been in contact and cooperating with the primary target group of my research, that is, the police, both in the Netherlands and in Germany. This cooperation has resulted in concrete output. Specifically, as mentioned in chapter 7, the Dutch version of the SAI is already in use by the Dutch police and has been approved for nationwide use by the Expertgroep Algemeen Rechercheren (EGAR; English: expert group on general investigations). Moreover, the tool will become mandatory from 2015 onwards. In Germany, field trials of the German translation have recently started with the police of Hesse as the cooperating police force. In this field trial, not only the statements provided by the witnesses will be evaluated (e.g., in terms of quantity and level of specificity and whether the instructions were correctly understood), but also police officers' satisfaction with the tool. It is possible that the findings of the field trials will result in the publication of an article in a journal that is specifically targeted to a law enforcement audience. I have been pursuing such a strategy before, by publishing a paper that presented the SAI in detail (Krix & Sauerland, 2011). This paper was the key to establishing contact with the German police.

Taking into consideration the results of chapter 5, it is conceivable to devise and conduct training programs for members of the police to improve teaching of the structure of human memory and eventually reduce the bias against recall of reminiscent information. Conducting the training should, of course, be accompanied by a sound evaluation. Again, the results of such an evaluation could be published in relevant journals.

Innovation

By definition, PhD theses and their results are innovative, because one requirement of a PhD project is that it should yield new insights and ideas. Incidentally, this criterion is part of the PhD regulations of Maastricht University and should be applied by the PhD supervisor when assessing the candidate's thesis. That is why it does not go without a certain oddity to emphasize the innovations in this adden-

dum. Because the new insights and implications have been stated in detail in chapter 7, I will repeat them here only briefly and would like to refer the interested reader to chapter 7. The project is innovative, because it was conducted in close cooperation with the primary target group, that is, the police. Moreover, in line with a positive legal psychology (Meissner, Hartwig, & Russano, 2010), the research was directly targeted to supporting the police, instead of pointing to the error-proneness of eyewitness accounts. The project has yielded two translated versions of the SAI, a German and a Dutch one. Finally, the experiments have increased the knowledge as to when the SAI can preserve memory for a subsequent interview and in what situations it is more or less advisable to administer the SAI.

Schedule and Implementation

As mentioned in the previous sections, the “dissemination and marketing” (as, by the way, I would have preferred to term this addendum instead of valorization) is already quite advanced. Specifically, the SAI has been implemented in all provinces of the Netherlands. In Germany, field trials are underway in one Bundesland (Hesse). The schedule for the field trials has been set to six months, but as agreed with the police it can be extended if need be (e.g., if there are not enough cases). Depending on the success of the trials, further Bundesländer could be approached. Implementation of the SAI does not require a huge financial investment. In fact, given the current difficult budget situation in Germany, one of the requirements of the Hessian police before the onset of the field trial was that it should be relatively cost-neutral. Implementation will cause some costs for printing the forms, as well as personnel costs for publicizing and explaining the tool and its application range among the police forces, and being available for questions. However, since the tool is rather self-explanatory and does not require, for example, extensive training of interview skills, it is cost-saving. Also, it is possible to use the already existing infrastructure of the police for implementation purposes. For instance, the police academies have their own printing offices that can be used for printing the forms. Moreover, they have mailing lists at their disposal that can be used to distribute information material on the SAI. Finally, the police academies conduct trainings (e.g., interview trainings) on a regular basis that could be used for introducing the SAI. Generally, the risks of implementing the SAI are low, for example because of the low financial costs and because the existing infrastructure can be used (i.e., a reorganization is not necessary).

Acknowledgments

After four years in Maastricht, my PhD project has been completed. It has been a very instructive time and great experience that I am sure I will always fondly remember. So let me take this opportunity to thank all people who have supported me and contributed to this thesis in many different yet significant ways.

First of all, I wish to thank my supervisors Prof. Dr. Harald Merckelbach and Dr. Melanie Sauerland. Supportive and inspiring supervisors are an important pillar of a successful PhD project and I am glad that I had two such supervisors. Harald and Melanie, you taught me a lot during my PhD years and it was a great privilege working with you!

Harald, I thank you for your trust in me when you accepted me as a PhD candidate and welcomed me with open arms to the Forensic Psychology Section. Your enthusiasm for research and science is palpable every day and it is contagious. I am also grateful for your fast and instructive feedback. Your creative ideas allowed me to view my project from different angles and strengthened my research.

Melanie, es ist nun fast zehn Jahre her, dass ich zum ersten Mal an deine Bürotür geklopft und meine Semesterarbeit bei dir geschrieben habe. Schon damals habe ich immer sehr gerne mit dir zusammengearbeitet. Als ich gehört habe, dass du an die Uni zurückkehrst, und sich die Möglichkeit eröffnete, bei dir zu promovieren, stand für mich fest, diese Chance zu ergreifen. Eine bessere tägliche Promotionsbetreuerin als dich hätte ich mir nicht vorstellen können. Im Laufe des Projekts hast du mir immer wieder Chancen aufgezeigt, etwa indem du Anregungen zu Publikationen (z. B. in *Polizei & Wissenschaft*) gegeben hast. Das war großartig. Vielen Dank für deine außerordentliche Unterstützung, deine guten Ideen und dein superschnelles Feedback. Auf dich konnte ich mich immer verlassen. Ich hoffe, dass wir auch in Zukunft zusammenarbeiten können.

I also especially want to thank my (former) colleagues from the Forensic Psychology Section: Alfons, Anna, Conny, Corine, Dalena, David, Elly, Ewout, Glynis, Harald, Henry, Jenny, Jill, Kim, Lieke, Linsey, Maarten, Maartje, Manuela, Marko, Melanie, Nathalie, Saskia, Thomas, and Tom. I enjoyed working with you very much. No matter what question on research issues occurred to me there sure was one expert among you who knew the answer. You are a very special and inspiring group that may be unparalleled. Of course, I should also mention the innumerable lunches, "borrels", and retreats that were great fun.

Maartje, Manuela, and Lieke, it was a pleasure sharing a room with you for almost two years. You were great roommates. Het was heel gezellig.

ACKNOWLEDGMENTS

My thanks are also due to all other colleagues from CPS who have made my time at Maastricht University enjoyable. Truus, Jessie, and Lindy, thank you for your organizational support and patiently answering my questions on such things as promotiebrieven, VVV-bonnen, archiving data, etc.

I also wish to thank my cooperation partners. First, I should mention the inventors of the SAI, Fiona Gabbert and Lorraine Hope. Fiona and Lorraine, thank you for sharing the SAI and cooperating on this research project.

Imke Rispens en Roel Boon, bedankt voor de goede samenwerking omtrent de ZeG. Imke, jij ook bedankt voor het testen van de RIMOZ-rechercheurs.

Clemens Lorei möchte ich dafür danken, dass er mir die Testung an der Hessischen Hochschule für Polizei und Verwaltung ermöglicht hat.

Nikola Hahn, Ihnen möchte ich ganz herzlich für Ihr Interesse an EVA, Ihre Ausdauer bei der Genehmigung des Feldversuchs der hessischen Polizei und allgemein für Ihre Offenheit für rechtspsychologische Forschung danken.

Quite a few students were involved in my experiments and I would like to thank all of them for their contributions.

Maartje and Jessica, I am so glad that you agreed to be my paranimfen and that you will support me during the defense.

Maartje, toen ik net in Maastricht begon, gaf je mij meteen het gevoel om erbij te horen. Dit vond ik ontzettend lief van je. Bedankt voor de dagelijkse gezellige koffiepauzes en de vele leuke gesprekken. Ik heb er altijd erg van genoten. Bovendien was je een super taallerares. Met jou heb ik graag Nederlands geleerd.

Jessica, du warst im letzten Jahr in Maastricht ein wunderbares „kamergeenootje“ und bist mir inzwischen eine liebe Freundin geworden. Ich vermisse das gemeinsame Teewasserholen (inklusive der Gespräche dabei) sowie das Schimpfen über die Zeitschriftengutachter. Vielen Dank für deine Unterstützung und die schöne Zeit! Auch möchte ich mich bedanken, dass du dich um die kleinen Dinge, wie meine Post, gekümmert hast, als ich schon wieder zurück in Deutschland war.

I would also like to thank two very special people who have influenced my life in important ways but who are sadly no longer among us: Elsie and Karl Heinz, I wish I could share this moment with you. I am sure you would be mighty proud now.

Mama, vielen Dank für deine liebevolle und unglaublich vielseitige Unterstützung mein Leben lang. Deine Unterstützung ist mit Worten nicht zu beschreiben und gar nicht hoch genug einzuschätzen. Du hast mir erst das Psychologiestudium ermöglicht und mir beigebracht, dass man mit Fleiß und harter Arbeit sehr viel im Leben erreichen kann, auch wenn die Umstände nicht immer für einen sprechen. Danke für alles!

Frank, thank you for introducing me to Texas and for welcoming Timm and me so warmly to your home. Spending time at the ranch really allowed me to relax and recover from stressful times.

Timm, hab vielen Dank für deine unendliche Geduld, dein Verständnis und Interesse. Ich bin dir sehr dankbar, dass du die verrückte Zeit des Pendelns zwischen Maastricht, Gießen und Stuttgart mitgemacht hast. Das ist alles andere als selbstverständlich und ich weiß es sehr zu schätzen. Du hast mich während der Zeit in Maastricht nicht nur emotional unterstützt, sondern mir auch konkret mit etlichen Makros und MacGyver-Einsätzen in der Wohnung das Leben erleichtert. Dafür danke ich dir sehr. Du bist ganz außergewöhnlich. Ich liebe dich.

Curriculum Vitae

Alana C. Krix was born on October 8th, 1984 in Wetzlar, Germany. Finishing secondary school, she received her Abitur from Goetheschule Wetzlar in 2004. That same year, she started studying psychology at the University of Giessen, Germany, and graduated with the degree of Diplom-Psychologin in 2010. From 2010 to 2014, she was a PhD candidate at the Forensic Psychology Section/Department of Clinical Psychological Science of Maastricht University. For her PhD project, she received a scholarship from Studienstiftung des deutschen Volkes (German National Academic Foundation). She currently works as a freelance lecturer at the University of Giessen and the University of Applied Sciences for Police and Administration of Hesse, Germany.

Publications

International journal articles

Krix, A. C., Sauerland, M., Gabbert, F., & Hope, L. (2014). Providing eyewitnesses with initial retrieval support: What works at immediate and subsequent recall? *Psychology, Crime and Law, 20*, 1005-1027. doi:10.1080/1068316X.2014.902456

Sauerland, M., & Krix, A. C. (2012). Would you recognize the perpetrator? What do you need to know when you have to make an identification from a lineup? *In-Mind Magazine, 2/2012*. Retrieved from <http://www.in-mind.org/article/would-you-recognize-the-perpetrator-what-do-you-need-to-know-when-you-have-to-make-an>

Sauerland, M., Krix, A. C., van Kan, N., Glunz, S., & Sak, A. (2014). Speaking is silver, writing is golden? The role of cognitive and social factors in written versus spoken witness accounts. *Memory & Cognition, 42*, 978-992. doi:10.3758/s13421-014-0401-6

German and Dutch journal articles

Krix, A. C., & Sauerland, M. (2011). Das Eigenständige Vernehmungsprotokoll für Augenzeugen: Ein neues Verfahren, mit dem Zeugen sich selbst vernehmen können [The Self-Administered Interview: A new tool which witnesses can use to interview themselves]. *Polizei und Wissenschaft, 4/2011*, 30-35.

Krix, A. C., & Sauerland, M. (2011). Im Zweifel für den Angeklagten: Über Fehlerquellen von Zeugenaussagen [Giving the defendant the benefit of the doubt: Sources of errors in eyewitness accounts]. *In-Mind Magazine, 1/2011*. Retrieved from <http://de.in-mind.org/article/im-zweifel-fuer-den-angeklagten-ueber-fehlerquellen-von-zeugenaussagen>

Krix, A. C., & Sauerland, M. (2013). Wie lassen sich Zeugenaussagen verbessern? – Neue Entwicklungen und Methoden [How can eyewitness accounts be improved? – New developments and methods]. *Praxis der Rechtspsychologie, 23*, 136-150.

Krix, A. C., Sauerland, M., & Schreuder, M. J. (2013). Hoe getuigen zichzelf kunnen verhoren: de Zelfrapportage voor getuigen [How witnesses can interview themselves: The Self-Administered Interview]. *Expertise en Recht, 5/6/2013*, 180-184.

Sauerland, M., & Krix, A. C. (2011). Vorbereitung und Durchführung von Gegenüberstellungen und Lichtbildvorlagen vor dem Hintergrund einer positiven Rechtspsychologie [Preparing and administering lineups in the context of a positive legal psychology]. *Praxis der Rechtspsychologie*, 21, 119-132.

Submitted manuscripts

Krix, A. C., Sauerland, M., Lorei, C., & Rispens, I. (2014). *Consistency across repeated eyewitness interviews: Contrasting police detectives' beliefs with actual eyewitness performance*. Manuscript submitted for publication.

Krix, A. C., Sauerland, M., Merckelbach, H., Gabbert, F., & Hope, L. (2014). *How effective is retrieval support for witnesses with different levels of working and source memory?* Manuscript submitted for publication.

Krix, A. C., Sauerland, M., Raymaekers, L. H. C., Memon, A., & Smeets, T. (2014). *Eyewitness evidence obtained with the Self-Administered Interview© is unaffected by stress*. Manuscript submitted for publication.

Sauerland, M., Mehlkopf, S., Krix, A. C., & Sagana, A. (2014). *Deceiving suspects about the content of their alibis: Consequences for truthful and untruthful suspects*. Manuscript submitted for publication.

Conference presentations

Sauerland, M., Krix, A. C., Raymaekers, L. H. C., Otgaar, H., Memon, A., Nivo, M., Slegers, C., Waltjen, T. T., & Smeets, T. (2014, June). *Stress, stress-induced cortisol responses, and eyewitness performance*. Paper presented at the annual meeting of the International Investigative Interviewing Research Group, Lausanne, Switzerland.

Krix, A. C., Sauerland, M., Merckelbach, H., Hope, L., Gabbert, F. (2013, September). *Interviewing witnesses with a Self-Administered Interview: The role of executive functioning*. Paper presented at the annual meeting of the European Association of Psychology and Law, Coventry, England.

Krix, A. C., Sauerland, M., Merckelbach, H., Hope, L., Gabbert, F. (2013, September). *Vernehmung mit dem Eigenständigen Vernehmungsprotokoll für Augenzeugen: Die Bedeutung der Exekutivfunktionen [Interviewing witnesses with the Self-Administered Interview: The role of executive functioning]*. Paper presented at the 15th biannual meeting of the German Association of Psychology and Law, Bonn, Germany.

Krix, A. C., Sauerland, M., Hope, L., Gabbert, F., & Merckelbach, H. (2012, April). *Exploring the memory-preserving effect of the Self-Administered Interview*. Paper presented at the annual meeting of the European Association of Psychology and Law, Nicosia, Cyprus.

Krix, A. C., Sauerland, M., Gabbert, F., Hope, L., & Merckelbach, H. (2012, March). *The effectiveness of the Self-Administered Interview when the witness was distracted during the crime*. Paper presented at the annual meeting of the American Psychology-Law Society, San Juan, Puerto Rico.

Krix, A. C., Sauerland, M., Merckelbach, H., Gabbert, F., & Hope, L. (2011, September). *Wie Zeugen sich selbst vernehmen können: Erste Ergebnisse zur Deutschen und Niederländischen Form des Self-Administered Interviews [How witnesses can interview themselves: First results on the German and Dutch version of the Self-Administered Interview]*. Paper presented at the 14th biannual meeting of the German Association of Psychology and Law, Muenster, Germany.

Krix, A. C., Sauerland, M., Merckelbach, H., Gabbert, F., & Hope, L. (2011, July). *How witnesses can interview themselves: The Self-Administered Interview*. Paper presented at the 1st meeting Forensische Psychologie Update of the Forensic Psychology Section of Maastricht University, Maastricht, The Netherlands.