The Effectiveness of the Self-Administered Interview©

- A Meta-Analytic Review and

Empirical Study with Older Adult Witnesses -

Katrin Pfeil

INSTITUTE OF CRIMINOLOGY

UNIVERSITY OF CAMBRIDGE

This dissertation is submitted for the degree of

Doctor of Philosophy

Selwyn College March 2017

DECLARATION

This dissertation is the result of my own work and includes nothing that is the outcome of work done in collaboration except where specifically indicated in the text and acknowledgements.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other university or similar institution except where specifically indicated in the text and acknowledgements. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other university or similar institution except where specifically indicated in the text and acknowledgements.

This dissertation, including footnotes, does not exceed the permitted length of 80.000 words.

Katrin Pfeil

March 2017

ACKNOWLEDGEMENTS

Undertaking this PhD has been a truly life-changing experience for me and it would not have been possible to do without the support and guidance that I received from many people.

I would like to first say a very big thank you to my supervisor Dr Katrin Mueller-Johnson for all the support and encouragement she gave me, both during the journey that was my PhD and during the detour I took in South Africa to pursue a small dream that will hopefully turn into a research career. Without her guidance and feedback this PhD would not have been achievable. Danke!

I gratefully acknowledge the funding received towards my PhD from Gates Cambridge. The scholarship has not only supported me throughout my PhD, but also during maternity leave, which is far from self-evident, even in 2017. Thank you for supporting women who pursue both a career and a family.

A very special thanks goes to my fellow Gates Scholars. Being part of this community has inspired me in unimaginable ways. Thank you all for showing me that you should always follow your heart and that you can achieve anything if you just set your mind to it. Without your inspiration I would not have volunteered in protected areas in South Africa and thus found a way to combine my passion and my work. Thank you also for fighting battles with me and help dust off some of the patriarchal structures. And finally, thank you for walking with me along the way. Anna, Andrew, Collin, Devani, Hanna, Hanns Hagen, Halliki, Linda, Lindsey, Luis, Rebekah, Tory, Vai, Zoë, and so many more – Thank you!

My appreciation goes to our two student research fellows, Valerie Erkens and Eva Klein, who supported me throughout data collection. Recruiting, scheduling and testing nearly 150 participants, in almost 300 sessions, was a big undertaking and I am grateful for their help. I also deeply appreciate the help of all my participants, who volunteered their time and cognitive capacity in return for some chocolate (and to further my research of course!). I would like to especially thank the University of the Third Age who helped me advertise my study (again!). A big thank you goes to Kirstine Szifris, her family and friends for being my partners in crime and helping me

to shoot the video material used for this dissertation. And finally, thank you to my fellow PhD students and friends at the Institute of Criminology, especially Pascale, Kritika, Sara and Cristobal, for making life in the office and beyond fun and for contributing to a wonderful Cambridge experience.

I am particularly grateful to my parents and parents-in-law for their continued support, encouragement, and for giving me the feeling that I can achieve anything. Vielen Dank für alles!

'Last but not least' is a common phrase to end a list of acknowledgements. However, it does not do justice to who I am about to thank. My deepest and most heartfelt gratitude goes to my husband Andy and our two sons, Leonas and Bennett. You are my life and love and I would not be where I am today without you.

Thank you. I love you.

"Yesterday, my life was headed in one direction. Today, it is headed in another.
Yesterday, I believed that I would never have done what I did today."

- David Mitchell, CLOUD ATLAS -

ABSTRACT

Name: Katrin Pfeil

Title: The Effectiveness of the Self-Administered Interview – A Meta-Analytic Review and Empirical Study with Older Adult Witnesses

The Self-Administered Interview© (SAI©) is an eyewitness interviewing tool designed to help protect eyewitness memory and elicit a comprehensive initial statement (Gabbert, Hope, & Fisher, 2009). Early research shows promising results in general adult samples. Whether the SAI© is an effective tool for older adult witnesses has not yet been fully addressed. Older adults will become increasingly important as a witness population in the future, yet perform worse compared to young adults. Some attempts have been made to aid older adult witnesses, but an easy-to-apply and effective method is yet to be introduced. This dissertation presents an overview of current knowledge on eyewitnesses and provides a theoretical basis for the empirical chapters. It further presents results of a systematic review and several meta-analyses on the effectiveness of the SAI© as a means to enhance eyewitness testimony. The meta-analyses cover 38 experimental comparisons from 22 empirical studies representing 1712 interviewees. Results indicate a strong benefit of the SAI both immediately after the witnessed crime ($d = 1.20$) and in a delayed recall ($d = 0.92$ compared to no initial recall) after one to three weeks. The third large chapter of this dissertation presents the results of an experiment that investigated the effectiveness of the SAI© for older witnesses' testimony, suggestibility and lineup performance. 144 participants, half of which were 60 years or older and half aged 18-30 years, took part in two sessions. In the first session, they were shown a film of a staged crime and either filled in the SAI©, gave a written free recall or no initial recall. In the second session after one week they were then asked to give a free recall of what they remembered, answer questions including suggestive questions, and also to identify the perpetrator from the film from a 6-person simultaneous photo lineup. Results confirm the classic SAI© effect for young adults, show a small beneficial effect for older adults and also indicate a beneficial effect for lineup performance for the first time.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

In recent years, the likelihood of terrorist attacks in western countries seems an ever-present threat. When we consider attacks such as the ones on the airport in Brussels, Belgium (22nd March 2016), on the promenade in Nice, France (14th July 2016) or on the Christmas market in Berlin, Germany (19th December 2016), we have to try to comprehend the sheer number of witnesses that were present. Hundreds of people were wounded and may have critical information about the attack, in addition to an unknown number of bystanders on the scene who may also hold valuable information. Gathering evidence from hundreds of witnesses with potentially case-breaking information, while an urgent response is understandably needed at the scene, quickly amounts to an unsolvable task for police forces. Valuable information may be lost in hectic scenes, if witnesses think they are less important or if officers fail to identify them as significant.

The fact is that any of those witnesses may hold valuable or even critical information. For police officers to determine on-site who to question more thoroughly is nearly impossible, which only leaves the possibility to arrange future interviews for each witness. This practice often leads to a delay between witnessing a crime and being interviewed (Kebbell, Milne, & Wagstaff, 1999) of days or even weeks. These delays create two grave disadvantages: First, the longer the interval, the more witnesses will forget (Wixted & Ebbesen, 1991), making their testimony less complete. Second, a longer delay also presents more opportunities for memory contamination, e.g. through post-event misinformation (Zaragoza & Lane, 1994), thus making their testimony less accurate.

"The truth is that nearly everybody is right about some things
and wrong about most things;
and if a man's testimony is not to be taken
until he is right on every subject,
witnesses will be extremely scarce."

Robert Green Ingersoll
(American lawyer and political leader, 1833 – 1899)

This quote nicely illustrates the three main issues that come to mind when talking about eyewitnesses: a) Memory is fallible and witnesses make mistakes; b) Yet they are an essential part of the Criminal Justice System; and c) Effective tools and protocols are needed to aid them. A number of tools have been put forward in the past 40 decades, and yet there is much room for further improvement and the development of more effective tools to elicit the most complete and most accurate witness testimony, as well as aid witnesses in making the correct choice when presented with a lineup. How this outcome can be achieved is the subject matter of this doctoral thesis. With the Self-Administered Interview[1] (Gabbert et al., 2009), a simple tool has been introduced that may have the potential to revolutionize investigative interviewing. It aims to elicit a comprehensive initial witness statement immediately after an incident, thus minimising delay and memory contamination, while also opening up the possibility to elicit statements from several witnesses at the same time without having to increase police resources.

This thesis will provide a comprehensive analysis of the effectiveness of the Self-Administered Interview in three main chapters. Following this introduction, the *first chapter* will provide the theoretical basis for this dissertation. It comprises an overview of current knowledge on eyewitnesses and their performance in testimony and lineup tasks. It will further focus on older adults as witnesses, underlying cognitive characteristics and discuss attempts that have been made so far to improve their performance. *Chapter two* will focus on the Self-Administered Interview and present a systematic review of the SAI literature and several meta-analyses on its effectiveness. The results, limitations and implications will be discussed in the end. The *third chapter* will present a comprehensive experimental study that examined the effectiveness of the SAI for older adult witness in comparison to young ones. It will provide analyses on the impact of the SAI on older adults' testimony, susceptibility to suggestions and on their lineup performance. The results, limitations and implications of the empirical findings will be thoroughly discussed. Concluding remarks will close this thesis.

---

[1] The Self-Administered Interview, or short SAI, is copyrighted (Copyright © 2006, Gabbert, Hope & Fisher). For readability however, the copyright symbol was omitted throughout this thesis.

1. EYEWITNESSES

1.1. A Criminological Framework

In court, eyewitness evidence is regarded to be among the most incriminating types of evidence, second only after confession evidence itself (Kassin & Neumann, 1997). However, eyewitness evidence is not always accurate and can lead to miscarriages of justice. To date, over 350 wrongfully convicted people have been exonerated in the U.S. by DNA testing, including 20 persons who had been sentenced to death and were awaiting enforcement of the judgement (Innocence Project, 2016). On average, these wrongfully convicted people had served 14 years each in prison before they were exonerated and released. Sources of wrongful convictions include e.g. forensic errors, false confessions, perjured testimony, and as indicated above, eyewitness misidentification. In the U.S. over 75% of known wrongful convictions, many of them in rape cases, are at least in part due to mistaken eyewitness identification (Gould & Leo, 2010), making it the single greatest cause of wrongful convictions (Innocence Project, 2013). Of course, not every witness misremembers important details of a crime or fails to correctly identify the perpetrator, and data from exoneration cases is trivially small compared to the number of convictions based on eyewitness evidence (Boyce, Beaudry, & Lindsay, 2007).

There are several reasons leading to mistakes in eyewitness performance, the most apparent being the memory itself. It does not work like a tape recorder, neither when memorizing an event, nor when attempting to recall it (Boyce et al., 2007). Concerns about memory become even more apparent when older adult witnesses are involved, bearing in mind their declining sensory and memory systems. And although older adults make up only a relatively small number of all victims of crime (e.g. Lanier & Dietz, 2009), they constitute a sizeable number of affected individuals. Moreover, recent years have shown an increased awareness of crimes affecting this age group, such as physical, sexual, and financial abuse of older vulnerable adults; or distraction burglary (e.g. Bachman & Meloy, 2008; McCabe & Gregory, 1998). Furthermore, the global population is rapidly aging, and by 2050 the number of older adults will exceed the number of young persons for the first time in history (United Nations Population Division, 2008). Globally, the number of persons aged 60 or above is expected to more than double by 2050, rising from 962 million in 2017 to 2.1

billion in 2050. The number of persons aged 80 or over is projected to even triple in that period. In Europe, 25 per cent of the population is already aged 60 years or over and that proportion is projected to reach 35 per cent in 2050 (United Nations Population Division, 2017). With more older adults present, and furthermore them remaining fit and active up to a high age (e.g. Memon, Gabbert, & Hope, 2004), they are more likely to witness a crime and be involved in the Criminal Justice System. However, before specifically looking at older adult witnesses, a brief overview on eyewitness performance and person identification research in general will be given.

1.2. Eyewitness Performance

Witness testimony and identification has long fascinated practice and research, for obvious reasons outlined above: It is well known that witnesses make mistakes, however witness evidence is still very much needed in court. There are various reasons why these mistakes can occur and research has, using both experimental designs and archival studies, examined, challenged and changed the ways in which witness evidence is gathered. An important differentiation in witness research was pointed out by Wells (1978), who distinguished between so-called estimator and system variables. Estimator variables cannot be influenced by the legal system as they constitute situational or environmental factors, such as poor visibility of the perpetrator due to poor light conditions, as well as individual differences in the witness, such as the personality or *age of the witness*. System variables on the other hand can be influenced by the legal system, such as the *interviewing technique* or the structure of the line-up. This study focuses on older witnesses as a variable whose impact has to be estimated by the legal system, and also provides a possible system variable, the Self-Administered Interview, as a means to improve identification performance. But first, a brief overview on testimony and suggestibility will be provided.

1.2.2. Testimony and Suggestibility

When assessing the quality of an eyewitness recollection, two different properties need to be addressed: the quantity and the accuracy of information (Pansky, Koriat, & Goldsmith, 2005). In general memory research, the focus has traditionally been on the quantity of retrieval, and memory was treated as a 'storehouse' (Pansky, et al., 2005) with percent recall being the standard measure of memory quantity. This has been useful for examining forgetting curves, and the impact of study time, divided attention and level of processing. Whereas it is evident that one cannot expect an eyewitness to remember every detail about a crime, one would like to be able to rely on the accuracy of the information provided. Accuracy of information however becomes more important in eyewitness settings and reflects the likelihood that each reported items is correct and thus evaluates the dependability of memory. Whereas the percent recall measure assesses the person for what she or he fails to report, the accuracy measure only assesses the person for what she or he does report (Pansky, et

al., 2005). In other words, a witness might only remember 30% of what happened, but might be 100% accurate in what they remember.

There are many variables that influence the quantity and accuracy of witness accounts. The distinction between system and estimator variables mentioned earlier can be even further separated out in a temporal manner - into witness characteristics (which are set before the event), characteristics of the event and post-event influences (Memon, 2008). Figure 1.1. gives an overview of possible variables that can influence witness testimony at different stages in memory processing.

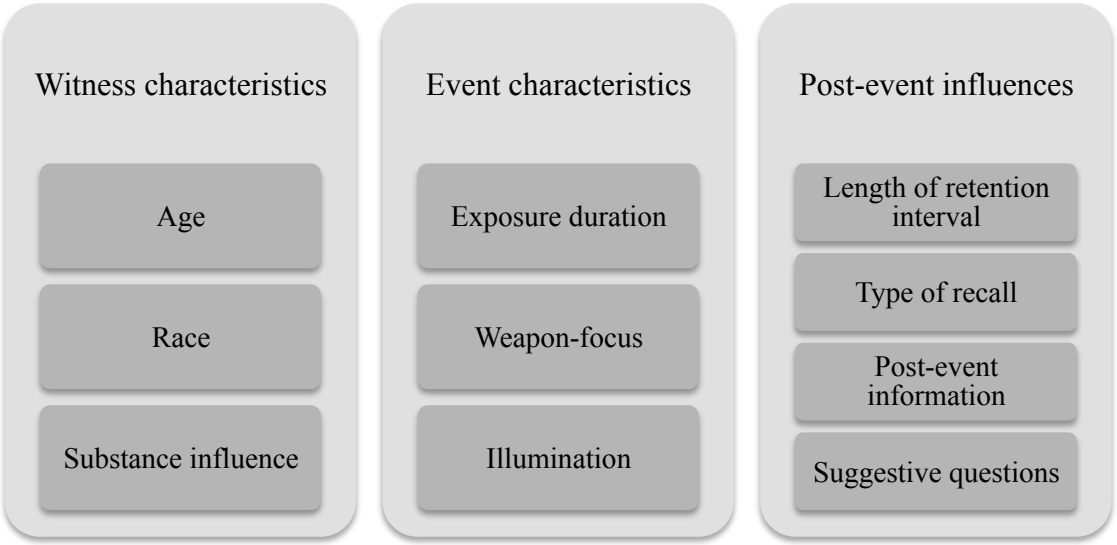


Figure 1.1. Selection of variables that can impact witness testimony at different stages in memory processing.

Witness characteristics include for example the age and race of the witness, as well as substance influence. As will be discussed in more detail later on, perception and memory functioning declines with age resulting in older adults remembering fewer details and making more mistakes in recall and recognition tasks (e.g. Mueller-Johnson & Ceci, 2004; Wilcock, Bull, & Vrij, 2007). Regarding the ethnicity of witnesses, it is generally found that cross-racial identifications are more difficult and that testimony is less reliable when the race of the witness and suspect are not the same (see Meissner & Brigham, 2001b for a meta-analytic review). People encode more qualitative information about own-race faces (Memon, 2008), and the own-race bias is influenced by familiarity and thus can be decreased with frequency and quality of contact (Chiroro & Valentine, 1995). As for the impact of substances on memory,

there is overwhelming evidence that alcohol reduces witness accuracy, affecting both the encoding of information and retrieval (e.g. Yuille & Tollestrup, 1990).

Event characteristics on the other hand include the duration and illumination of the event, as well as whether weapons are present, amongst others. While it is common sense that the longer a witness can observe an event and perpetrator, the more details they might remember, research findings on the duration-accuracy relationship vary. Whereas e.g. Clifford and Richards (1977) found better recall after 30 seconds than 15 seconds exposure in an experiment, archival studies of real-life crimes have not found such a relationship (Fahsing, Ask, & Granhag, 2004). Regarding the presence of a weapon, there are also mixed research results. Laboratory studies usually support a so-called weapon-focus, which suggests that witnesses focus on the weapon rather than the offender and can therefore report fewer offender-related details, whereas archival studies do not support a weapon-focus (see Steblay, 1992 for a meta-analytic review). An undoubted factor that impacts witness testimony is illumination – the amount of light at the crime scene. Witnesses typically remember less about an event that took place at twilight rather than during the day, and the accuracy of details and recognition of people is also better in daylight (e.g. Yarmey, 1986).

Post-event influences include for example the length of the retention interval, the type of recall, post-event information, and leading questions. As for the interval between witnessing an event and being interviewed by the police, it is well established that the sooner the recall is made, the more details are given and the better is the accuracy of the account. Regarding the type of recall, witnesses tend to give more information in an interrogative recall, i.e. answers to specific questions, compared to a free recall, i.e. telling everything they can remember in their own words at their own pace (e.g. Sternberg, Lamb, Orbach, Esplin, & Mitchell, 2001). However, they also make more mistakes in the interrogative recall compared to free recall and practitioners should be aware of this trade-off. Lastly, among the most widely studied post-event characteristics are post-event misinformation and leading questions. Both may alter a witnesses' memory of the event and lead to misremembering of information, although it is still unclear whether this effect is irrevocable or whether the original memory can still be retrieved under the right conditions (Loftus, 2005; McCloskey & Zaragoza, 1985).

Suggestibility generally describes the susceptibility to accept misleading information from others. This process happens unnoticed e.g. in reaction to misleading post-event information, being pressurized during an interview and being asked leading questions (Gudjonsson & Clark, 1986). Planting misinformation into a witness' memory in experimental settings has lead them to remember a car being a different colour, seeing broken glass where there was none and even whole buildings that were not there (Williams, Loftus, & Deffenbacher, 1992). Leading questions are a very common way of contaminating a witness' memory by introducing new details that were not present in the event (Loftus, 2005). They can be used intentionally, e.g. during trial by a barrister ("You do agree with this, don't you?") and unintentionally, e.g. during a police interview when the interviewer unintentionally provides information gathered from a previous witness. This should of course be avoided in police interviewing (Gudjonsson & Clark, 1986); however, it does happen and can have disastrous effects on not only the witness memory but ultimately the outcome of a trial. The theoretical mechanisms behind suggestibility are still unclear and being discussed, such as the 'trace-alteration account' (the original memory trace is altered) (Loftus, 2005), the original memory inhibition during recalling misinformation (Saunders & MacLeod, 2002) and memory coexistence, i.e. the original memory stays unaltered and can be correctly recalled under the right conditions (McCloskey & Zaragoza, 1985).

Recent studies have highlighted ways to reduce susceptibility to suggestions. Saywitz, Wells, Larson, and Hobbs (2016) have conducted a meta-analysis on the effects of interviewer support on children's suggestibility and found that children were more resistant and less acquiescent to suggestive questions when interviewers were supportive as compared to non-supportive or neutral. Szpitalak and Polczyk (2016) found that reinforced self-affirmation in the form of positive feedback about one's memory skills and reflecting on one's achievements can reduce vulnerability to interrogative suggestibility. The test group who had performed reinforced self-affirmation before completing a memory test showed significantly lower scores for all measures of interrogative suggestibility, while there were no differences in memory skills between the groups. Moreover, Huff, Weinsheimer, and Bodner (2016) shed further light on the effect of initial retrieval to reduce a person's susceptibility to misinformation. They found that a protective effect of testing emerged on a final free recall test following a delay and on a final source-memory test regardless of delay.

In sum, in criminal investigations, eyewitness testimony often is crucial evidence, but it is not only important to get an account that is as complete as possible, but rather the most accurate account possible. Witness memory is influenced by a number of variables, such as the age of the witness, the illumination during the event and the time span between the event and the recall. Furthermore, post-event misinformation and leading questions during the interview present common threats to accuracy. Next an overview on person identification and police practice will be given.

1.2.3. Person Identification

When witnesses are asked to view a lineup, they not only have to recall information about the perpetrator's appearance, but ultimately make a decision about which person shown committed the crime (in case of the perpetrator being present) or decide that the perpetrator is not amongst them. It has been suggested that recognition involves two different processes or judgement strategies: one based on recalling exact details ('remember') and the other one based on assessment of familiarity ('know', see Wilcock, Bull, & Milne, 2008 for an overview). The latter refers to a situation in which a witness immediately recognizes a face, but is unable to place a context or name the person. However, this automatic recognition has been found to be associated with accuracy in lineup decisions: the face just 'popped out' (Dunning & Perretta, 2002). Conversely, the strategy to look at faces and actively retrieve details from memory to recognize a person has been found to be associated with false identifications (Dunning & Stern, 1994), suggesting that the witness may pick the person who is most similar to their memory of the perpetrator relative to the other lineup members.

Another distinction in judgement strategies can be found when looking at different lineup types. In simultaneous lineups, i.e. when several photos are presented at the same time, witnesses may be more likely to examine each face and compare them with one another. This may again lead to the witness falsely identifying the person that is most similar to the perpetrator ('relative decision strategy', Lindsay & Wells, 1985). In sequential lineups however, when the photographs are presented one at a time, witnesses are more likely to compare each photograph with their memory image of the perpetrator, known as 'absolute decision strategy' (Lindsay & Wells, 1985). In a meta-analytic comparison between these two lineup presentation methods,

Steblay, Dysart, Solomon & Lindsay (2001) found that correct rejection rates, indicating that the perpetrator was not present, were indeed more frequent in sequential lineups. However, correct identifications were more frequent in simultaneous lineups.

In a real-life situation, police does not know whether their suspect is in fact the real perpetrator and thus, if the perpetrator is amongst the persons in the lineup. In research it is therefore important to distinguish between these so-called target-absent (TA) and target present (TP) lineups. False identifications, i.e. falsely identifying an innocent person or a foil from a lineup as being the perpetrator, are more likely to occur from TA lineups than from TP lineups (Wells, 1993). It was also found that the behaviour of the lineup administrator has an impact on the witness (e.g. Phillips, McAuliff, Kovera, & Cutler, 1999), e.g. intentional and unintentional hinting towards the suspect or accuracy of the identification. To minimize the impact of suggestion and feedback (Gould & Leo, 2010), identification procedures should be administered double-blind, meaning that neither the witness nor the police officer or researcher administering the lineup knows the identity of the suspect. This procedure has been adopted in several states (e.g. New Jersey, North Carolina) and the UK (Greathouse & Kovera, 2009; Hutton, Johnston, & Sampson, 2005). Also, in the UK the standard identification procedure is a video identification, meaning that short videos of the suspect and foils are presented sequentially to the witness (Hutton et al., 2005).

In order to gain a better understanding of how people recognize faces, it is important to distinguish between the recognition of known or familiar faces, and that of unknown or unfamiliar faces (Wilcock et al., 2008). Generally, people tend to be very good at recognizing familiar faces, and at the same time very poor with unfamiliar ones (Bruce, Burton, & Hancock, 2007). Unfortunately, as witnesses to a crime, especially when it comes to lineup procedures, it is mostly the second type of task that is demanded: recognizing an unfamiliar face. In this situation, perception and memory is known to be image-specific, and the cognitive capacity to generalize from one image to another is quite limited (Bruce et al., 2007). This means for the eyewitness situation that witnesses keep a specific image of the perpetrator's face in mind, and experience difficulty when confronted with altered characteristics of this image, e.g. with a different hair cut or even at a different angle. Impressions of a face are vulnerable to various influences, such as viewpoint and lighting, and people are

poor in recognizing the same face or even matching pictures of the same face, when these variables change (Hancock, Bruce, & Burton, 2000).

Changes in the picture of a suspect, i.e. the angle or expression in the face or both, were for instance shown to impair recognition of unfamiliar faces in a study by Bruce (1982). Here, compared to an accuracy of 90% when being presented the same picture, the ability to recognize the face dropped to 76% for a change in either viewpoint or expression, and to 61% when both were changed. It was especially concerning that there was also a higher rate of false identifications, i.e. when viewpoint and expression changed, 12% of participants falsely identified a new face as a previously seen one. A change in viewpoint and expression most closely represents the forensic setting. It is very probable that the picture of a perpetrator provided to a witness will show this person from a different viewpoint and with a different facial expression compared to when the perpetrator was seen by the witness while committing the crime. This suggests that even when the initial picture of an unfamiliar face is clear, even small changes regarding the picture affect recognition accuracy (Wilcock et al., 2008). Applied to an eyewitness setting, this has e.g. implications for CCTV footage: even in cases in which a witness has had a good look of a suspect, recognizing this person from a poor CCTV image is much more difficult.

There are a number of theoretical models that provide a framework for understanding how faces are recognized, most of which derived from empirical findings in laboratory studies and clinical cases. Bruce and Young (1986) were the first to propose a very influential model of face recognition. They suggested that there are seven different types of information, or so-called codes, deriving from faces: pictorial, structural, visually derived semantic, identity specific semantic, name, expression and facial speech. The probability to positively identify a face increases with the number of codes that are available to the person (Wilcock et al., 2008). For unfamiliar faces, only some codes are likely to be available, such as pictorial, structural, expression and facial speech. When looking at familiar faces however, name codes and identity specific semantic codes can additionally be involved (Wilcock et al., 2008). This means that with unfamiliar faces, recognition stays at a somewhat basic level, whereas with familiar faces, person identity information may help to positively identify a face (Bruce et al., 2007). Bruce et al. (2007) further argue that the expertise for familiar faces arises through accumulation and averaging of countless different images of a person' face rather than quality of processing. Support

for this theoretical position is given by Troje & Kersten (1999), who showed that participants were good at recognizing full-face views and profile views of familiar faces and also full-face views of their own face (like in a mirror), but poor at recognizing profile-views of their own face (which they are unlikely to see often).

Another model of face recognition was put forward by Valentine (1991). The Multidimensional Space Framework (or short Face-Space) suggests that faces are encoded within a multidimensional space and defined by a number of dimensions to distinguish between different faces, e.g. eye colour or face shape. Typical looking faces, which mean faces which people are more likely to see in their environment, accumulate around the so-called central tendency of the dimensions. Around this centre are a lot of typical faces arranged. More unusual faces on the other hand will be more distant from the central tendency, where there are fewer faces (Wilcock et al., 2008). The model also proposes an explanation for the difficulty to distinguish between faces of other ethnic backgrounds ('own-race-bias', see Meissner & Brigham, 2001b for a meta-analytic review). It is suggested that dimensions used to encode own-ethnicity faces may be inappropriate for other race faces, and when there are limited or no appropriate dimensions available, faces are densely clustered (Wilcock et al., 2008). This results in the heightened difficulty to distinguish between them. Practical applications for Face-Space in person identification include its use for computer-generated facial composites and the design of fair lineups to identify suspects with distinguishing features (Valentine, Lewis, & Hills, 2016).

In summary, in face recognition it is important to distinguish between familiar faces, a task that people accomplish with expertise on a daily basis, and unfamiliar faces, which can pose a great difficulty. Two models regarding face recognition were introduced, Bruce & Young's (1986) model of facial codes and Valentine's (1991) Face-Space, providing an understanding of how faces are encoded and recognized. However, these are general models of face recognition, developed for the general population and it is unclear if they generalize to older adults. Bearing in mind that witnesses make mistakes in recalling details of an event as well as in lineup situations, it is important to elicit information on how and why these mistakes occur. With regards to older adults, memory decline plays a major role in witness testimony and face recognition. Theories put forward to account for age-related differences in retrieval and recognition focus on both the encoding and the retrieval stage of memory processing, and will be introduced in the next section.

1.3. Older Adults' Cognition

Two aspects are important when examining older adults' eyewitness performance: perceiving and memorizing the event including the characteristics of the perpetrator, and the ability to retrieve this information, both in terms of giving a verbal account and also when it comes to recognizing whether the perpetrator is among the people in a lineup. Age differences can occur at any of these stages of cognitive performance and will be discussed in the following sections, based on theoretical models and empirical findings.

At the initial stage of memorizing information, that is perceiving and paying attention to the event, older adults are already at disadvantage to younger ones. Losses in the sensory system increase with age, such as diminished vision and hearing (e.g. Congdon et al., 2004; Lin, Niparko, & Ferrucci, 2011). Changes in the structure of the eye result in less efficient processing of visual stimuli and half of the adults aged 75 to 79 suffer from measurable hearing loss (Schneider & Pichora-Fuller, 2000). For being able to perceive an event, witnesses must also selectively pay attention to it and ignore other irrelevant information, an ability that also declines with age (Van Gerven & Guerreiro, 2016) and leads to less efficient encoding and processing of information.

When coming to memory processes in specific, it is well known that memory performance declines with age, but also, that not all aspects of memory are impaired. However, the type of memory relied upon most in eyewitness situations, the episodic long-term memory, is particularly affected by old age (Schacter, Koutstaal, & Norman, 1997; Souchay, Isingrini, & Espagnet, 2000). In contrary to implicit memory, which holds information such as how to ride a bike, explicit memory serves intentional retrieval, e.g. "What did the perpetrator look like?". The latter can be further refined in semantic memory, which contains general knowledge of facts and words, and episodic memory, that is to remember personally experienced events in a particular setting at a particular time. In fact, older adults are said to have a much larger disruption in episodic memory than in semantic memory (Nilsson, 2003; Souchay et al., 2000), which is exactly the kind of memory that is needed for remembering the details of a crime. Moreover, the ability to remember details over a long period of time (long-term memory), as opposed to shortly maintain small amounts of information in immediate awareness (short-term memory) underlies major changes with age (Nilsson, 2003).

Thus, deficits in accurately remembering the details of a crime can occur at three distinct stages: when perceiving the event and encoding the information, when retaining i.e. maintaining the information until an account is made, and lastly when retrieving the stored information, e.g. in a police interview (Mueller-Johnson & Ceci, 2007). Though it was shown that older adults are at disadvantage both in the first (perception) and the second (retention) stage, the greatest differences between older adults and young ones were found in the last stage, the retrieval of stored memory. Especially when asked to freely recall information, as compared to recognition, older adults were especially impaired (Craik & McDowd, 1987). In addition, older adults' source memory gets worse (Memon, Bartlett, Rose, & Gray, 2003; Multhaup, de Leonardis, & Johnson, 1999). Source-monitoring refers to the ability to identify where information was learnt, e.g. whether the person has experienced an event personally or rather heard about it from another person or read about it in the newspaper. This difficulty in source memory can make older adults particularly vulnerable to misleading post-event information, such as suggestive questions, and can lead to inaccurate testimony as they have more difficulty to place an event, or a person's face, and thus rely more on a general feeling of familiarity rather than explicit recollection (Wilcock et al., 2008).

Different theoretical explanations have been advanced by cognitive aging research to account for these age-related changes. The processing speed theory for instance assumes a general slowing in cognitive processing in age (Cerella, 1985). More recent studies have shown that processing speed has a mediator function between age and cognitive functions, such as memory processes (Salthouse, 1996). Balota et al. (2000) conclude that age and memory performance are only indirectly related, and thus that age differences in memory reflect in fact age differences in the speed of processing. Another approach is that of attentional capacity: according to this framework, cognitive resources are limited for any given cognitive task (see Balota et al., 2000). It has been found that when the cognitive task gets more difficult and thus requires more attentional capacity, older adults encode information more general instead of encoding specific details (Rabinowitz, Craik, & Ackerman, 1982), which leads to a poorer memory performance. Similar, but based on Baddeley's (1995) working memory model, is the reduced working memory theory. According to this theory, storage capacity and manipulation of information in working memory are

limited with increasing age and thus older adults' performance declines in complex mental operations (e.g. Salthouse, Mitchell, Skovronek, & Babcock, 1989).

There are also neurological changes that underlie the memory deficits and are said to play an important role in the reduced attentional capacity and processing speed, such as reductions in cerebral volume, decreased brain metabolism, reduced blood flow, and altered neurochemical systems (N. D. Anderson & Craik, 2000). Furthermore, they argue that these age-related reductions in attentional capacity and processing speed both reduce cognitive control, i.e. the ability to manage one's thoughts, recollections, and actions towards a task-related goal, and that this in turn leads to impaired inhibition and reduced conscious recollection. With relation to source monitoring, Glisky, Rubin, and Davidson (2001) found that especially older adults with below average frontal function show deficits in source memory. They stress the importance of frontal function during the encoding of source and suggest that older adults with reduced frontal lobe function fail to initiate the processes required to integrate contextual information with focal content during study. Interestingly, they also found that these deficits could be eliminated when older adults were required to consider the relation between an items and its context during studying. Cabeza, Anderson, Locantore, and McIntosh (2002) looked at low- and high-performing older adults and compared their prefrontal cortex (PFC) activity during recall and source memory. Results suggested that low-performing older adults recruited similar prefrontal cortex regions as young adults (right PFC), but used those ineffectively, whereas high-performing older adults engaged the PFC bilaterally. This suggests that high-performing older adults counteracted age-related neural decline through plastic reorganization of frontal networks. Finally, Sala-Llonch et al. (2014) conducted a large-scale resting-functional magnetic resonance imaging study with healthy older adults to analyse the impact of age on functional brain connectivity and related differences in memory performance. They found that ageing was associated with less connectivity of long-range connections (fronto-pariental and fronto-occipital) and with less integrated and more segregated global networks. This was further related to lower performance in verbal and visual memory functions.

Other theories provide the basis to decide what interventions could be helpful in improving memory account, e.g. the environmental support theory (Craik, 1986) and inhibition theory (Hasher & Zacks, 1988). The former theory describes how the influence of context supports the memory retrieval in older adults. Specifically, if the

task provides environmental support, as in a recognition task compared to a free recall, age-related memory differences diminish (Craik & McDowd, 1987). The latter theory relates to the inhibitory function that is necessary to suppress irrelevant information and activate information that is relevant to the actual task. Especially for the inhibitory functioning theory, there is some support that older adults are more distracted by irrelevant information (e.g. environmental details) and thus have greater difficulties in memory tasks (e.g. Hartman & Hasher, 1991; Hasher & Zacks, 1988).

In summary, older adult's sensory systems diminish, making it more difficult to perceive relevant information. Furthermore, the episodic long-term memory declines with age (Balota et al., 2000), which is exactly the memory system that is needed to remember the details of a crime. There are several theories that attempt to explain the observed memory deficits, e.g. the inhibition theory (Hasher & Zacks, 1988). Furthermore, several neurological changes are said to underlie these memory deficits by indirectly impairing cognitive requirements such as inhibitory functioning (N. D. Anderson & Craik, 2000). Although research on older adults' general memory deficits is well established over the past decades, relatively little research has been done in the specific field of eyewitness psychology, where most literature is based on young adults and children (Bartlett & Memon, 2007; Memon et al., 2004). Thus the next section will present what is known about older adult witnesses, regarding their testimony and susceptibility to suggestions, and their performance in person identification tasks.

1.4. The Ageing Eyewitness

When examining older adults' performance as eyewitnesses, there are three major aspects to address: the amount and accuracy of the witness account, the susceptibility to suggestions and the accuracy in person identification. Although research mostly focusses on either the verbal account (and possibly its susceptibility to suggestions) or person identification (and possibly suggestive influences), in real-life situations witnesses are often confronted with both tasks. Thus in this study, both aspects were integrated in the experimental part, and both aspects will be discussed in the following section.

1.4.1. Testimony and Suggestibility

Focusing on testimony, that is the produced verbal account of an event, it is generally found that older adults produce less accurate person details, action details and setting details than young adults (Yarmey, 2001), both in free recall and in answers to questions (Yarmey, Jones, & Rashid, 1984; Yarmey & Kent, 1980). Coxon and Valentine (1997) moreover found, that older adults not only gave fewer correct, but also more incorrect answers to questions than young adults. Additionally, older adults are particularly vulnerable to misleading post-event information, which means information that is presented between the witnessed event and the interview (Bornstein, 1995). This post-event information can influence the witnesses' testimony in the way that they accept suggested information, e.g. from a discussion about the incident with other witnesses, as actually having seen or heard it themselves. Thus eyewitness testimony often reflects not only the actually witnessed scene, but also information that was obtained later on (Moulin, Thompson, Wright, & Conway, 2007). Findings about suggestibility of older adults however vary across studies: While some researchers found a higher degree of susceptibility to suggestions in older adults as compared to young ones (e.g. Karpel, Hoyer, & Toglia, 2001; Mitchell, Johnson, & Mather, 2003), others did not find reliable age differences (e.g. Coxon & Valentine, 1997; Gabbert, Memon, & Allan, 2003), and some studies even found younger adults to be more suggestible than older ones (Huff & Umanath, 2017; Marche, Jordan, & Owre, 2002).

Although being able to remember the details of a crime and giving a comprehensive verbal account is without doubt very important, another and quite

distinctive task that is often asked of eyewitnesses is to identify the perpetrator. Findings about older adults' person identification performance and, related to this, their confidence for those decisions, will be presented next.

1.4.2. Person Identification

With regards to performance on a lineup, it is a consistent finding that older adults perform more poorly, both in TP and in TA lineups compared to young adults (see Erickson, Lampinen, & Moore, 2016; Fitzgerald & Price, 2015 for two separate meta-analyses). This means that regardless of lineup type (i.e. whether the perpetrator was present or not) and perpetrator age, older adults were found to be reliably worse at making correct lineup decisions compared to young adults (Erickson et al., 2016).

Bartlett & Memon (2007) presented averaged data over 10 studies comparing young and older adult witnesses. They reported an average drop of the proportion of correct rejections in TA lineups of .22, namely from .53 for young adults to .31 for older ones, meaning that, across these studies, only one-third of older adults correctly indicated that the perpetrator was not in the lineup. Furthermore, Memon, Bartlett, Rose & Gray (2003), examined the impact of the retention period on older adults, which in this case means the delay between witnessing a crime and viewing a lineup. They found that when there was only a short delay of about 35 minutes, older and young adults performed similarly on the lineup. However, when the lineup was viewed after 1 week, older witnesses were significantly less accurate than young ones. Another possible impact on the lineup performance is the so-called 'own-age bias'. Most research on person identification uses young adult's faces as targets, i.e. for the lineup photographs (Anastasi & Rhodes, 2005) and it was suggested that young witnesses outperforming older ones might be due to them being of the same age as the lineup faces (e.g. List, 1986). However, support for an own-age bias in older adults is weak - some studies did not find an own-age bias in older adults (Havard & Memon, 2009; Wilcock et al., 2007), others concluded that there is an OAB in older adults, but it is weaker than for young ones (see Rhodes & Anastasi, 2012 for a meta-analysis).

A further important issue in the study of person identification is the relationship between confidence and accuracy of witness' identification decision. In court the confidence with which a witness makes an identification is very persuasive to jurors: it was shown that they were more likely to believe evidence given by a

confident compared to a less confident witness (Brewer & Burke, 2002; Cutler, Berman, Penrod, & Fisher, 1994; Wells, Lindsay, & Ferguson, 1979). However, vast amount of research exists showing the fallibility of this assumption and discussing possible influences and methodological issues of studies (see e.g. Deffenbacher, 1980; Leippe & Eisenstadt, 2007), which show that there is, in fact, no simple relationship between confidence and accuracy.

On way to disentangle the relation was presented in a meta-analysis by Sporer, Penrod, Read & Cutler (1995). Although the overall confidence-accuracy correlation was weak, they found a difference when including choice as a moderator variable. For people who identified someone from a lineup (choosers) the relationship was reliably higher (r = .41) than for people who did not identify someone (non-choosers, r = .12), suggesting that once witnesses have identified someone from a lineup, their confidence may be a stronger predictor of their accuracy. In a recent synthesis, Wixted and Wells (2017) even argue that confidence and accuracy are, in fact, strongly related under *pristine testing conditions*. This means that if memory is uncontaminated, the lineup is fair and administered double-blind with unbiased lineup instructions, and the confidence statement is taken immediately after viewing the lineup, high-confidence suspect identifications are remarkably accurate. Conversely, this also implies that under non-pristine testing conditions, e.g. unfair lineups are used or confidence in the courtroom is taken, the accuracy of even highly confident identifications is compromised. Data from the US indicates that some of these conditions are commonly met in the field, whereas others are frequently not (Loftus & Greenspan, 2017). For example, witness confidence was only documented 76% of the time for identifications, and 44% of the time for non-identifications. Moreover, lineups were only administered double-blind in 31% of cases.

Specifically regarding older adults, Scogin, Calhoon & D'Errico (1994) presented a photo lineup after having shown a film of a crime to young (18-35 years), young-old (59-74 years) and old-old (75-94 year). They found no significant correlation between line-up accuracy and the self-rating of confidence of the participants in any of these age-groups. Even if an older witness was very confident to have chosen the correct person this did not mean they actually were correct (Wilcock et al., 2008). This finding was replicated in several other studies (Memon et al., 2003; Memon, Hope, Bartlett, & Bull, 2002; Wilcock et al., 2007). Regarding older adults' confidence compared to young adults results are mixed: whilst some research found

older adults to be less confident than young ones (e.g. Memon et al., 2003; Memon et al., 2002), others did not find a difference between age-groups (e.g. Searcy, Bartlett, & Memon, 1999).

Confidence in person identification decisions is very malleable and easily altered by suggestive influences, which was e.g. shown by Wells & Bradfield (1998). In this study, participants who had seen a TA lineup and had made a false identification received a confirming feedback saying they had identified the correct person. Participants who received this feedback were significantly more certain about their choice than participants who had not received feedback. Neuschatz et al. (2005) investigated the post-identification feedback effect in older adults and replicated the finding for this age-group. Older adults were as susceptible to the effects of feedback from a lineup administrator as young adults and it is therefore important to obtain information about a witnesses' confidence immediately after they make their choice and before they learn new information.

In summary, the confidence-accuracy relationship is a frequently discussed issue in person identification research, the main reason being the persuasive power of confident witnesses in court, when in fact confidence is a fallible indicator for accuracy. But under certain conditions, such as pristine testing conditions or in cases in which the witness has made an identification from the lineup, the correlation is higher. Generally, meta-analyses have shown that older adults perform poorer in lineups than young adults – they make fewer correct identifications in TP and more false identifications in TA lineups. To counter this trend, some attempts have been made to aid witnesses in general and older adults in specific. The most important advances will be presented next.

1.5. Attempts to Improve Eyewitness Performance

Some approaches have been made to improve the quality of eyewitness testimony, and the Cognitive Interview (CI) devised by Fisher and Geiselman (see 1992 for an overview) is among the most rigorously tested and most widely accepted methods (A. M. Wright & Holliday, 2007a). The original version consists of four basic memory aids: the instruction to report everything, even when it is considered trivial; context reinstatement, i.e. the mental recreation of the original environmental, cognitive, physiological, and affective states; recalling the event in a different temporal sequence, e.g. backwards; and recalling the event from several different perspectives (Köhnken, Schimossek, Aschermann, & Höfer, 1995; A. M. Wright & Holliday, 2007a). Fisher and Geiselman (1992) further refined the original version by addressing the social dynamics and communication between the interviewer and the witness and called this the Enhanced Cognitive Interview. The CI was found to enhance memory in written reports (Aschermann, Mantwill, & Köhnken, 1991) as well as in oral interviews (Geiselman, Fisher, MacKinnon, & Holland, 1986). Furthermore it increased the amount of correct details without increasing the number of inaccurate details, for both young adults and child witnesses (Köhnken, Milne, Memon, & Bull, 1999). Wright and Holliday (2007a, 2007b) replicated these findings for older adults.

However, the CI has not been found to improve person identification performance from lineups (Clifford & Gwyer, 1999; Gwyer & Clifford, 1997). On the contrary, producing a comprehensive verbal recall of the perpetrator's characteristics, as in the CI, was even found to hinder a subsequent identification ('verbal overshadowing', Searcy, Bartlett, & Memon, 2000; D. B. Wright & Stroud, 2002). In a meta-analysis, Meissner & Brigham (2001a) found an overall small, but significant negative effect, showing that the recall did impact negatively on identification accuracy. Furthermore, learning to deliver the Cognitive Interview involves considerable training (e.g. Memon, Wark, Bull, & Köhnken, 1997) and there is some evidence that it is not always implemented accurately by police officers, e.g. due to time limitations (Kebbell et al., 1999).

More generally than the CI, eyewitness interviews should consist of two parts, a free recall part and question part (Greuel, 2008). This approach is also recommended for older witnesses, as for instance Bornstein (1995) summarizes in his

work on improving elderly eyewitnesses' memory. Specifically he recommends to ask precise questions and to emphasize recognition memory rather than recall. He also underlines that these questions should not be presented in a leading manner. As a helpful component he proposes to reinstate the context by instructing the participant to imagine the scene of the crime and also, to ask questions in the correct chronological sequence rather than switching between times and incidents. Finally, an effective way to improve the memory of older eyewitnesses would be to begin with unstructured, free recall, then move to specific questions and finish with a series of yes/no questions (Bornstein, 1995).

Another attempt to improve especially identification performance focusses on the lineup presentation. In many jurisdictions the standard lineup procedure is to present photographs simultaneously, i.e. several photographs at the same time (Steblay et al., 2001). Different from that are sequential lineups, in which the photographs or videos of suspects are presented one at a time, so that choices have to be made individually for each one. These are usually found to reduce false identifications compared to simultaneous lineups (see Steblay et al., 2001 for a meta-analysis). This effect has been found for young and for older adults (Searcy et al., 2000); however, sequential lineups also seem to reduce correct identifications in TP lineups (Rose, Bull, & Vrij, 2005; Steblay et al., 2001). Regardless, sequential lineups are a requirement for police forces in the UK (Hutton et al., 2005).

A successful approach to reduce false identification rates are non-biased lineup instructions, informing the witness that the perpetrator may or may not be present. Malpass & Devine (1981) found a decrease in false identification rates from 78% without this instruction to 33% with instruction. As a result, police forces in the UK and the US have been required to give witness such non-biased lineup instructions for the past decades (Wells et al., 2000; Zander, 1990). However, evidence suggests that older adults fail to remember instructions given to them prior to the conduction of lineups. When they were asked after the lineup procedure to recall the instructions they had been given, only 46% of older adults were able to correctly recall them, compared to 68% of young adults (Rose et al., 2005). Even more important, failing to remember the instructions was significantly associated with failing in the lineup itself: participants who failed to remember the information made also more false identifications from the lineup. Also, enhanced lineup instructions especially for older adults, e.g. emphasizing the importance of only picking a person

from the lineup when the witness is certain that it was the perpetrator, found no effect on lineup performance (Wilcock, Bull, & Vrij, 2005). However, older adults benefitted from other pre-lineup procedures, such as practice lineups and pre-lineup question (e.g. "Bearing in mind the criminal may not be present in the lineup, how sure are you that you will be able to correctly reject a lineup, that is, say none of them if the photograph of the criminal isn't there?") and were significantly less likely to make a false identification from the lineup (Wilcock & Bull, 2010).

In summary, eyewitness performance is subject to mistakes in general. Older adults are found to perform even worse in eyewitness situations than young adults, both regarding testimony and person identification. Some approaches have been made to aid eyewitness performance in general and specifically for older adults, but they have their disadvantages: the Cognitive Interview is very time-consuming and has no beneficial effects on person identification. Sequential lineups seem to reduce false identifications from TA lineups only at the cost of correct identifications from TP ones. Pre-lineup procedures also only in some cases seem to have an impact. An easy-to-apply and effective method to aid older eyewitnesses is yet to be introduced.

1.5.1. The Self-Administered Interview

The Self-Administered Interview (SAI) is an eyewitness interviewing tool developed and described by Prof. Fiona Gabbert, Goldsmiths University of London, Prof. Lorraine Hope, University of Portsmouth, and Prof. Ronald P. Fisher, Florida International University (one of the creators of the Cognitive Interview). It was designed to elicit a comprehensive initial statement and help protect eyewitness memory for later police interviewing (Gabbert et al., 2009). The booklet format contains open-ended questions and asks the witness to write down as much detailed information about the crime as they can remember (see 3.2.3.2. for further details on the SAI). Due to its self-explaining instructions, the SAI can be filled in without further explanation of police officers, making it especially helpful when a large number of witnesses is present. To date, research addressing the SAI is at an early stage. However, the first two laboratory-based evaluation studies (Gabbert et al., 2009) showed very promising results: In the first study, participants who had been shown a film of a staged crime, were given either a free recall task about the crime or completed the SAI, while a third group was interviewed using the Cognitive

Interview. Results showed that participants who were given the SAI performed better than participants who gave a free recall, and at the same level as participants, who were given the very resource intensive CI. Moreover they found that participants with the SAI were able to recall more correct details, including forensically relevant person details, in a one week delayed recall test than participants without the SAI. Thus, the SAI seems effective in providing an enhanced and high-quality witness statement. Field trials have been conducted in several UK police forces and on Norwegian police force (Hope, Gabbert, & Fisher, 2011). The SAI has further been translated into several foreign languages, e.g. German, Dutch and Norwegian (Krix, Sauerland, Merckelbach, Gabbert, & Hope, September, 2011).

In the years following its development, research on the SAI has slowly picked up, bringing the total number of studies to more than 20 in 2017. However, the evaluation research on the SAI is still in its infancy. In particular, to date research did only marginally address the application to different age groups (Gawrylowicz, Memon, Scoboria, Hope, & Gabbert, 2014) and no research has been done concerning the impact on person identification. A number of studies have included measurements of suggestibility (Gittins, Paterson, & Sharpe, 2015; Mackay & Paterson, 2015; McPhee, Paterson, & Kemp, 2014), either in the form of suggestive questions during recall or in the form of introducing misleading post-event information before the interview. And indeed, there are theoretical reasons to expect positive impacts of the SAI on suggestibility: Giving an immediate statement in form of a standardized protocol naturally minimizes the risk of misleading post-event information and the use of leading questions and pressurizing the witness. These factors known to distort memory derive from the social interaction during an interview, which is not the case when using the SAI. It may also help strengthen episodic memory trace (Gabbert et al., 2009) and thus prevent susceptibility to suggestions (Ceci, Toglia, & Ross, 1988; Pezdek & Roe, 1995) in the later police interview. However, the lack of a social component compared to a face-to-face interview also leads to a potential shortfall of the SAI: It may not be applicable to vulnerable witnesses, such as rape victims, who need social support during interviewing. To gain a better understanding of the research on the SAI that has been done to date, gather evidence as to its effectiveness and to identify possible areas on which research needs to focus further, a systematic review of evaluation studies and meta-analysis of effects has been conducted and will be presented in the next chapter.

2. THE EFFECTIVENESS OF THE SAI - A SYSTEMATIC REVIEW AND META-ANALYSIS

This chapter provides an overview of the current level of knowledge concerning the Self-Administered Interview in the form of a systematic review and meta-analysis. The research goals and justification of chosen method will be outlined, followed by an explanation of study methods and the presentation of both the systematic review and meta-analysis results. In the end, results and implications will be thoroughly discussed.

2.1. Research Goals

The present study intends to produce a systematic analysis of the literature available on the Self-Administered Interview as well as a standardized measure of the association between the SAI and eyewitness testimony.

Systematic and meta-analytic reviews are the most rigorous methods for assessing the effectiveness of interventions (Welsh & Farrington, 2007). Systematic reviews look at the methods and results of studies on a specific topic to reach a research-based consensus, i.e. they locate, appraise and synthesize evidence from previous evaluation studies. This includes the following key features (Welsh & Farrington, 2007):

- Explicit objectives, i.e. a clear rationale for conducting the review.
- Explicit eligibility criteria, i.e. inclusion and exclusion criteria for studies are specified in detail and explicitly stated.
- Study search to reduce bias, i.e. it is explicitly stated how the literature was searched and what measurements were undertaken to locate studies reported outside academic journals, in foreign languages etc.
- Study screening and exclusion justification, i.e. each located study must be screened for eligibility and all excluded studies must be listed with reasons for exclusion.
- Data assembly as complete as possible, i.e. all relevant data must be extracted and coded, and where necessary, original authors contacted to receive more information.

- Quantitative techniques are used when appropriate, i.e. a meta-analysis should be conducted as well when suitable.
- Structured and detailed report, i.e. a clear description of each research phase, decision and conclusions.

Meta-analytic reviews can complement a systematic review and involve the statistical analysis of the results of previous research studies (Lipsey & Wilson, 2001), i.e. they summarize effect size data to determine the average effect of a particular intervention. Hereby, effect sizes of individual studies are weighted according to their sample size, with larger studies thus having more impact on the resulting weighted average effect size (Borenstein, Hedges, Higgins, & Rothstein, 2009). Meta-analyses are very transparent in their explication of chosen methods and included studies, and thus easily replicable by other researchers. They are also able to handle a large number of studies, which would otherwise be too complex to analyze for example in a systematic review alone (Lipsey & Wilson, 2001). Another advantage is that with meta-analysis, researchers are also able to analyze impacts on the summary effect size e.g. in a moderator analysis. The most important and relevant advantage however is the ability of meta-analysis to translate important yet complex information derived from a large body of research into bite-sized summaries that are easy to understand and interpret by laypeople. They thus provide the most reliable and comprehensive analysis of what works and since there is an increasing interest in evidence-based policy and practice, making informed decisions becomes more practical with the help of meta-analytic results.

However, systematic reviews and meta-analyses are not free from disadvantages. They are very time-consuming and require significant effort and resources. Specifically with meta-analysis, there are usually also statistical hurdles that are hard to overcome, such as dealing with different outcomes in primary studies, which ties into the critique that they are sometimes unable to synthesize more complex patterns of some individual studies. Finally, there is the problem of dissemination or making the results not only available but also known to relevant decision-makers. On a more theoretical level, Borenstein et al. (2009) discuss the following criticism on meta-analysis:

- A number cannot summarize a research field, i.e. a summary effect constitutes an insufficient simplification of findings.
- The file drawer problem, i.e. if the included studies are a biased sample of all possible sample, the meta-analysis will reflect this bias (e.g. publication bias).
- Mixing apples and oranges, i.e. combining different kinds of studies in the same analysis and ignoring important differences between them.
- Garbage in, garbage out, i.e. inclusion of many low-quality studies may lead to an erroneous meta-analysis.
- Ignoring important studies, i.e. exclusion of possibly important studies.
- Disagreement with randomized trials, i.e. sometimes large-scale randomized trials yield contrasting results to meta-analyses on the same question.
- Meta-analyses are sometimes performed poorly, i.e. they are so complicated that researchers make mistakes in the application of the method.

In order to overcome possible flaws and avoid most of the aforementioned points, it is most important to adhere to the key features of systematic reviews as mentioned above and discussed in more detail by Welsh and Farrington (2007). To reiterate, this involves having explicit objectives as well as the rigorous and explicit use of clear protocols and transparent methods, such as stating the inclusion and exclusion criteria in detail. Therefore, the present study was guided by the following research questions:

- Is the SAI more effective in eliciting a comprehensive initial statement compared to a simple recall attempt? (Time 1, thereafter abbreviated T1)
- Is the SAI more effective in preserving eyewitness memory for a later police interview compared to no initial recall and also compared to a simple initial recall attempt? (Time 2, thereafter abbreviated T2)

In order to provide a comprehensive analysis of these questions, results will not only be based on the detailed statistical analysis, but also include a review of the literature in more depth. This narrative synthesis provides background information on all included studies and thus will allow for assessing them in more detail, as well as identifying gaps in the current body of research.

2.2. Methods

In this section, first an overview of search strategies and study selection will be given. It will be outlined how studies were coded and what inclusion and exclusion criteria were chosen. Then it will be explained how effect sizes were obtained and computed where necessary.

2.2.1. Study Selection

As a first step, a literature search was conducted using *Google scholar*, searching for "self-administered interview" and also "SAI" both in the title and in text, from 2009 onwards. The cutoff was chosen due to the introductory paper on the SAI being published in 2009 (Gabbert et al., 2009). This was followed by a search of potentially relevant references cited in any of the publications found. In addition, the developers of the SAI and authors of the original paper were contacted and asked for contact information of academic collaborators. Then, several authors of primary articles were contacted and asked for any additional relevant work, published or unpublished, and the names of other relevant researchers.

Inclusion and exclusion criteria were formulated based on a priori considerations and also loosely on the examples of Köhnken et al. (1999) and Memon et al. (2010) and their respective meta-analyses on the effects of the Cognitive Interview. Thus, criteria for *including* studies in the final sample were that (i) the manuscript was written in English or German; (ii) a Self-Administered Interview was conducted in either the original or minimally modified version[2]; (iii) its effects were compared with some form of control measurement either in an initial recall (e.g. SAI vs written free recall [wFR] at T1), as a impact measurement in a delayed interview (e.g. effect SAI vs no initial recall on interview after one week at T2), or both; and (iv) dependent measures of recall were available (correct details and accuracy rate). In contrast to Memon et al. (2010), both published as well as unpublished studies were included in the analysis, so as to not run at risk of publication bias. Table 2.1. gives an overview of included studies with their respective study description.

[2] Minimally modified would for example be if the authors omitted the "sketch the scene" part of the SAI, as this part usually does not get coded or analysed, even if included in the study.

Table 2.1. **Included studies** for the association between the SAI and correctly remembered details and accuracy rate. Each study description includes sample age group, immediate or delayed outcome comparison; kind of control condition(s); free or cued recall at T2; oral or written recall at T2.

1	Af Hjelmsäter, Strömwall & Granhag (2012)
	Children; delayed comparison (2 weeks); SAI vs. no recall vs. FR; T2 free recall; oral.
2	Boessenrodt (2011)
	Adults; delayed comparison (1 week); SAI vs. FR; T2 free recall; oral.
3	Colomb & Gabbert (2013)
	Adults; delayed comparison (1 week); SAI vs. no recall; T2 free recall; oral.
4	Gabbert, Hope & Fisher (2009)
	Study 1: Adults; immediate comparison; SAI vs. FR; written.
	Study 2: Adults; delayed comparison (1 week); SAI vs. no recall; T2 free recall; written.
5	Gabbert, Hope, Fisher & Jamieson (2012)
	Study 1: Adults; delayed comparison (1 week); SAI vs. no recall; T2 free recall; written.
	Study 2: Adults; delayed comparison (3 week); SAI vs. no recall; T2 cued recall; written.
6	Gawrylowicz, Memon & Scoboria (2014a)
	Adults; immediate comparison; SAI vs. FR; written.
7	Gawrylowicz, Memon & Scoboria (2014b)
	Older adults; immediate comparison; SAI vs. FR; written.
8	Gittins, Paterson & Sharpe (2015)
	Adults; delayed comparison (1 week); SAI vs. no recall; T2 cued recall; written.
9	Hope, Gabbert, Fisher & Jamieson (2014)
	Adults; immediate and delayed comparison (1 week); SAI vs. FR vs. no recall; written and oral.
10	Kemp, Paterson & Yu (2016)
	Adults; immediate comparison; SAI vs. FR; written.
11	Krix, Sauerland, Lorei & Rispens (2015)
	Adults; immediate and delayed comparison (1 week); SAI vs. FR; written.
12	Krix, Sauerland, Merckelbach, Gabbert & Hope (2015)
	Adults; immediate comparison; SAI vs. FR; written.
13	Krix, Sauerland, Raymaekers, Memon, Quaedflieg & Smeets (2016)
	Adults; immediate comparison; SAI vs. FR; written.
14	Mackay & Paterson (2015)
	Adults; delayed comparison (1 week); SAI vs. no recall; T2 cued recall; written.
15	Maras, Mulcahy, Memon, Picariello & Bowler (2014)
	Adults; immediate and delayed comparison (1 week); SAI vs. SR; T2 cued recall; written.
16	Mauer (2013)
	Adults; immediate and delayed comparison (1 week); SAI vs. FR; T2 free recall; oral.
17	McPhee, Paterson & Kemp (2014)
	Adults; delayed comparison (1 week); SAI vs. no recall; T2 cued recall; oral.
18	Paterson, Eijkemans & Kemp (2015)
	Adults; delayed comparison (2 weeks); SAI vs. no recall; T2 free recall; written.
19	Pfeil (2016)
	Adults and older adults; immediate and delayed comparison (1 week); SAI vs. FR vs. no recall; T2 free recall; oral.
20	Schoof (2014)
	Adults; immediate and delayed comparison (1 week); SAI vs. FR; T2 free recall; oral.
21	Stephan (2013)
	Adults; immediate and delayed comparison (1 week); SAI vs. FR; T2 free recall; oral.
22	Zeier, Hewig, Kraus & Wagner (2016)
	Adults; immediate and delayed comparison (1 week); SAI vs. SR; T2 cued recall; written.

Note. FR = Free recall. SR = Structured recall.

Studies or subgroups of studies were *excluded* from the final sample if they failed to comply with the above-mentioned inclusion criteria, and in particular if they (i) did not have a non-SAI control group; (ii) used clinical samples; (iii) did not report sufficient data; (iv) used a Cognitive Interview as control group; (v) delayed the application of the SAI to longer than one hour after the event; (vi) measured a transfer effect to a different scenario. See table 2.2. for an overview of excluded studies and subgroups of studies with their respective exclusion criteria.

Table 2.2. **Excluded studies and subgroups** for the association between the SAI and correctly remembered details and accuracy rate with description of exclusion criteria.

Studies	
1	Boon (2012)
	Study did not include non-SAI control group.
2	Curtis (2013)
	Study did not include non-SAI control group.
3	Hope, Gabbert & Fisher (2011)
	Overview paper, no detailed data presented.
4	Krix, Sauerland, Gabbert & Hope (2014)
	Data is re-analysed and extended in Krix et al. (2015).
Subgroups	
1	Gabbert, Hope & Fisher (2009)
	Comparison with CI excluded.
2	Gawrylowicz, Memon & Scoboria (2014a)
	T2 data excluded as used different scenario from T1 (event transfer effect).
3	Gawrylowicz, Memon & Scoboria (2014b)
	T2 data excluded as used different scenario from T1 (event transfer effect).
4	Kemp, Paterson & Yu (2016)
	Comparison with CI excluded.
5	Mackay & Paterson (2015)
	Comparisons with delayed SAI application excluded.
6	Maras, Mulcahy, Memon, Picariello & Bowler (2014)
	Subsample of Autism Spectrum Disorder excluded.
7	Paterson, Eijkemans & Kemp (2015)
	Comparisons with delayed SAI application excluded.

Note. CI = Cognitive Interview.

The final sample consisted of 38 experimental comparisons from 22 empirical studies representing 1712 interviewees. Of these studies, 14 were published manuscripts, 3 unpublished manuscripts (in preparation to be published) and 5 unpublished theses or dissertations. Table 2.3. gives an overview of all 38 experimental comparisons with respective study codes, short subgroup description, means, standard deviations, and group and effect sizes for the association of the SAI and correctly remembered details.

Table 2.3. Means, standard deviations, group and effect sizes for correct details for all experimental comparisons used in the different meta-analyses.

		SAI			Control			
Study/Subgroup code	Subgroup	M	SD	n	M	SD	n	d
Af Hjelmsaeter 2012 (1)	T2 FR, SAI vs. no	21.02	6.58	65	16.98	6.81	64	0.60
Af Hjelmsaeter 2012 (2)	T2 FR, SAI vs. FR	21.02	6.58	65	18.30	7.83	65	0.38
Boessenrodt 2011 (1)	T2 FR, SAI vs. FR	41.35	7.10	20	33.35	9.39	20	0.96
Colomb 2013	T2 FR, SAI vs. no	51.61	12.02	36	35.95	9.20	37	1.46
Gabbert 2009 (1)	T1, SAI vs. FR	70.70	20.46	19	41.50	14.00	18	1.67
Gabbert 2009 (2)	T2 FR, SAI vs. no	62.38	22.04	21	45.90	24.02	21	0.71
Gabbert 2012 (1)	T2 FR, SAI vs. no	86.25	23.25	20	48.10	21.97	20	1.69
Gabbert 2012 (2)	T2 CR, SAI vs. no	25.77	14.49	31	11.09	5.13	31	1.35
Gawrylowicz 2014a	T1, SAI vs. FR	69.07	21.40	42	45.05	18.32	42	1.21
Gawrylowicz 2014b	T1, SAI vs. FR	83.26	24.95	40	66.19	22.50	40	0.72
Gittins 2015	T2 CR, SAI vs. no	46.42	17.73	42	38.43	14.10	38	0.50
Hope 2014 (1)	T1, SAI vs. FR	126.45	35.59	20	67.85	15.54	20	2.13
Hope 2014 (2)	T2 FR, SAI vs. no	176.95	41.33	20	140.15	54.56	20	0.76
Hope 2014 (3)	T2 FR, SAI vs. FR	176.95	41.33	20	126.15	43.45	20	1.20
Kemp 2016	T1, SAI vs. FR	92.68	24.84	28	71.07	23.99	28	0.88
Krix 2015a (1)	T1, SAI vs. FR	112.95	24.01	43	93.10	24.29	41	0.82
Krix 2015a (2)	T2 CR, SAI vs. FR	102.35	26.11	43	109.59	26.14	41	-0.28
Krix 2015b	T1, SAI vs. FR	117.16	28.20	63	85.89	24.52	62	1.18
Krix 2016	T1, SAI vs. FR	21.29	4.94	64	13.34	4.90	63	1.62
Mackay 2015	T2 CR, SAI vs. no	52.60	21.53	27	42.40	20.72	27	0.48
Maras 2014 (1)	T1, SAI vs. SR	44.59	12.08	18	32.22	13.37	17	0.97
Maras 2014 (2)	T2 CR, SAI vs. SR	-	-	14	-	-	16	0.66
Mauer 2013 (1)	T1, SAI vs. FR	33.00	9.74	20	22.55	7.25	20	1.22
Mauer 2013 (2)	T2 FR, SAI vs. FR	26.00	8.45	20	21.55	7.81	20	0.55
McPhee 2014	T2 CR, SAI vs. no	84.48	17.48	21	55.29	18.79	21	1.61
Paterson 2015	T2 FR, SAI vs. no	48.58	17.28	26	38.96	12.49	26	0.64
Pfeil 2016 (1)	T1, SAI vs. FR	112.88	54.39	24	70.13	24.29	24	1.01
Pfeil 2016 (2)	T1, SAI vs. FR	77.92	26.34	24	71.25	18.28	24	0.29
Pfeil 2016 (3)	T2 FR, SAI vs. no	97.21	34.64	24	60.54	20.66	24	1.29
Pfeil 2016 (4)	T2 FR, SAI vs. no	97.17	28.66	24	85.96	35.94	24	0.34
Pfeil 2016 (5)	T2 FR, SAI vs. FR	97.21	34.64	24	78.04	29.26	24	0.60
Pfeil 2016 (6)	T2 FR, SAI vs. FR	97.17	28.66	24	96.71	23.32	24	0.02
Schoof 2014 (1)	T1, SAI vs. FR	40.19	9.69	26	21.46	7.27	28	2.19
Schoof 2014 (2)	T2 FR, SAI vs. FR	26.62	9.88	26	20.32	6.71	28	0.75
Stephan 2013 (1)	T1, SAI vs. FR	55.88	13.90	24	30.83	13.55	23	1.82
Stephan 2013 (2)	T2 FR, SAI vs. FR	38.92	11.39	24	30.91	13.25	23	0.65
Zeier 2016 (1)	T1, SAI vs. SR	44.05	11.16	21	36.48	6.85	21	0.82
Zeier 2016 (2)	T2 CR, SAI vs. SR	10.21	2.44	21	11.70	2.11	21	-0.65

Note. T1 = Initial comparison at time 1. T2 = Impact measurement at time 2. FR = Free recall. CR = Cued recall. SR = Structured recall. No = No initial recall.

Table 2.4. provides the same overview for or the association of the SAI and accuracy rate, respectively. Additionally, the following information was coded from each study or subgroup: (a) year of publication; (b) publication type; (c) laboratory where study was conducted (SAI developers or other labs); (d) sample size; (e) age group of interviewees; (f) type of control condition (no recall or other form of recall); (g) type of comparison (initial at T1 or delayed at T2); (h) recall modality at T2 (oral or written); (i) type of recall at T2 (free or cued). A table containing the detailed moderator variable coding can be found in appendix A. It was expected that the effect sizes for correct details would be greater in an initial recall comparison rather than a delay, and in a comparison with no initial recall rather than another form of initial recall. No specific hypotheses were formulated with regard to the effect sizes for accuracy rate.

Table 2.4. Means, standard deviations, group and effect sizes for the accuracy rate for all experimental comparisons used in the different meta-analyses.

		SAI			Control			
Study/Subgroup code	Subgroup	M	SD	n	M	SD	n	d
Af Hjelmsaeter 2012 (1)	T2 FR, SAI vs. no	0.90	0.08	65	0.91	0.08	64	-0.13
Af Hjelmsaeter 2012 (2)	T2 FR, SAI vs. FR	0.90	0.08	65	0.93	0.08	65	-0.38
Boessenrodt 2011 (1)	T2 FR, SAI vs. FR	0.94	0.03	20	0.94	0.05	20	0.00
Colomb 2013	T2 FR, SAI vs. no	0.91	0.03	36	0.87	0.06	37	0.84
Gabbert 2009 (1)	T1, SAI vs. FR	0.89	0.05	19	0.91	0.03	18	-0.49
Gabbert 2009 (2)	T2 FR, SAI vs. no	0.93	0.05	21	0.88	0.06	21	0.91
Gabbert 2012 (1)	T2 FR, SAI vs. no	0.96	0.03	20	0.95	0.05	20	0.24
Gabbert 2012 (2)	T2 CR, SAI vs. no	0.81	0.13	31	0.73	0.11	31	0.66
Gawrylowicz 2014a	T1, SAI vs. FR	0.96	0.03	42	0.95	0.05	42	0.24
Gawrylowicz 2014b	T1, SAI vs. FR	0.97	0.03	40	0.98	0.02	40	-0.39
Gittins 2015*	T2 CR, SAI vs. no	84.79	6.88	42	80.04	8.49	38	0.61
Hope 2014 (1)	T1, SAI vs. FR	-	-	20	-	-	20	-0.38
Hope 2014 (2)	T2 FR, SAI vs. no	-	-	20	-	-	20	0.61
Hope 2014 (3)	T2 FR, SAI vs. FR	-	-	20	-	-	20	0.57
Kemp 2016*	T1, SAI vs. FR	89.69	3.87	28	90.91	5.15	28	-0.27
Krix 2015a (1)*	T1, SAI vs. FR	86.58	3.66	43	88.42	5.47	41	-0.40
Krix 2015a (2)*	T2 CR, SAI vs. FR	85.51	4.66	43	86.44	4.97	41	-0.19
Krix 2015b*	T1, SAI vs. FR	88.83	4.37	63	89.55	4.59	62	-0.16
Krix 2016*	T1, SAI vs. FR	97.92	3.19	64	98.07	4.14	63	-0.04
Mackay 2015	T2 CR, SAI vs. no	0.88	0.06	27	0.82	0.09	27	0.78
Maras 2014 (1)	T1, SAI vs. SR	0.87	0.04	18	0.87	0.08	17	0.00
Maras 2014 (2)	T2 CR, SAI vs. SR	0.86	0.05	14	0.84	0.11	16	0.23
Mauer 2013 (1)	T1, SAI vs. FR	0.94	0.05	20	0.95	0.60	20	-0.02
Mauer 2013 (2)	T2 FR, SAI vs. FR	0.95	0.06	20	0.95	0.05	20	0.00
McPhee 2014	T2 CR, SAI vs. no	-	-	21	-	-	21	1.62
Paterson 2015	T2 FR, SAI vs. no	0.92	0.12	26	0.90	0.08	26	0.20
Pfeil 2016 (1)	T1, SAI vs. FR	0.93	0.03	24	0.95	0.03	24	-0.67
Pfeil 2016 (2)	T1, SAI vs. FR	0.90	0.04	24	0.90	0.04	24	0.00
Pfeil 2016 (3)	T2 FR, SAI vs. no	0.94	0.03	24	0.92	0.05	24	0.49
Pfeil 2016 (4)	T2 FR, SAI vs. no	0.92	0.03	24	0.87	0.07	24	0.93
Pfeil 2016 (5)	T2 FR, SAI vs. FR	0.94	0.03	24	0.95	0.03	24	-0.33
Pfeil 2016 (6)	T2 FR, SAI vs. FR	0.92	0.03	24	0.89	0.05	24	0.73
Schoof 2014 (1)	T1, SAI vs. FR	0.95	0.04	26	0.98	0.04	28	-0.75
Schoof 2014 (2)	T2 FR, SAI vs. FR	0.96	0.04	26	0.97	0.06	28	-0.20
Stephan 2013 (1)	T1, SAI vs. FR	0.95	0.04	24	0.97	0.03	23	-0.57
Stephan 2013 (2)	T2 FR, SAI vs. FR	0.96	0.04	24	0.96	0.04	23	0.00
Zeier 2016 (1)	T1, SAI vs. SR	0.93	0.04	21	0.94	0.03	21	-0.28
Zeier 2016 (2)	T2 CR, SAI vs. SR	0.86	0.17	21	0.78	0.13	21	0.53

Note. T1 = Initial comparison at time 1. T2 = Impact measurement at time 2. FR = Free recall. CR = Cued recall. SR = Structured recall. No = No initial recall.

*In these studies, means and standard deviations of the accuracy rate were given as a percentage. They were not converted to decimals, as percentage is more precise.

2.2.2. Calculation of Effect Sizes

For this meta-analysis, Cohen's d was used as the effect size measurement. Cohen's d represents the magnitude of an effect and is calculated from the difference between the means of recalled details in SAI and control conditions, divided by the pooled standard deviation (Cohen, 1988). Cohen (1988) considered $d = 0.2$ to represent a small, around $d = 0.5$ a medium-sized and $d = 0.8$ and upwards a large effect. Sawilowsky (2009) further expanded these to include $d = 1.20$ as a very large and $d = 2.0$ as a huge effect. Effect sizes associated with higher means in the SAI or dependent on the SAI in the initial recall were given a positive sign, and those associated with higher means in the control conditions were given a negative sign. Separate effect sizes were computed for correct details and accuracy rate. For most studies means and standard deviations of the dependent variables were available and were thus used for the calculation of the effect sizes. In some studies, d was given and thus input directly, or calculated from *F*-tests and *t*-tests using the online effect-size calculator of the Campbell Collaboration (Wilson, 2001).

Further, many studies in this review used complex data structures in that they did not contribute only one effect size, but several. According to Borenstein et al. (2009), there are different types of complex data structures and different ways of dealing with them in a meta-analysis. The first type is multiple independent subgroups within a study, for example when different effects are reported separately for young and older adults (Pfeil, 2016; see experimental chapter). Since each participant provided unique information and contributed to only one effect size, the subgroups were treated as separate studies and were included in the meta-analysis as such.

The second type is multiple outcomes within a study, where data is reported on several related, but distinct outcomes. In this case, both the correct details and the accuracy rate were used as dependent measures of recall. Within each study, both outcomes are based on the same participants, which would be highly problematic if they were to be combined. However, since the goal was to compute a summary effect for the impact of the SAI on the correct details, and separately for the impact on the accuracy rate, two separate sets of meta-analyses were performed, which results in each participant contributing to only one effect size per analysis. Another form of multiple outcomes within a study were studies that provided multiple measurements for the impact of the SAI on a later interview – namely in the form of one

measurement for a free recall and a second one for a subsequent cued recall. So whereas most studies only looked at the impact of the SAI on a delayed free recall, and some only looked at the impact on a delayed cued recall, a few studies provided information for both types of recall (i.e. first participants were asked to provide a free recall and then afterwards answer a number of questions). If the latter was the case, only data for the free recall was used in the meta-analyses as it provides the first and most unaffected recall attempt available from that specific study[3].

The third and last type of complex data structures, and arguably the most complicated one, is multiple comparisons within a study, i.e. studies that use a single control group and several treatment groups. In this case, it is rather a single treatment group and several control groups, as a number of studies have compared the impact of the SAI to a) no initial recall and b) to another form of initial recall, such as a written free recall. This means that participants in the SAI group contributed information to more than one effect size (SAI vs. no recall and SAI vs. free recall, e.g. Hope et al., 2014), which is problematic for two reasons if these data sets were to be treated as separate studies. The first problem is that this approach would assign more weight to studies with two outcomes than to studies with one outcome (SAI vs. no recall, e.g. Gabbert, Hope & Fisher, 2009, study 2), as in a meta-analysis studies are weighted by sample size. The second problem is that this leads to an improper estimate of the precision of the summary effect (Borenstein et al., 2009), because the SAI scores come from the same set of participants and are therefore not independent of each other. Simply treating these comparisons as separate studies would underestimate the error and overestimate the precision of the summary effect.

One solution to this problem is to perform separate meta-analyses to compute a summary effect for SAI vs. no initial recall and another summary effect for SAI vs. another form of recall. Whereas this results in a more focused set of analyses, it would also be interesting to compute a summary effect for the SAI vs. any control group to be able to draw a broader conclusion on its impact. So in addition to running separate meta-analyses, the effect sizes of both comparisons (SAI vs. no initial recall and SAI vs. another form of recall) within each affected study were pooled and only a single composite effect was used for a broader, more inclusive meta-analysis. The composite

[3]All but one study that have separately used and analyzed a free and cued recall have done so in order to include suggestive questions in the cued recall (Gabbert et al., 2009; Kemp, Paterson, & Yu, 2016).

effect was created by simply taking the mean effect size of the comparisons SAI versus no recall and SAI versus other recall (see equation 1).

$$\overline{ES} = \frac{1}{2}(ES_1 + ES_2) \quad (1)$$

The variance of this composite was computed based on the variance of each single effect size as well as on the correlation between the two effects (see equation 2; Borenstein et al., 2009).

$$var(\overline{ES}) = \frac{1}{4}(var_{ES1} + var_{ES2} + 2r\sqrt{var_{ES1}}\sqrt{var_{ES2}}) \quad (2)$$

Following Borenstein et al.'s (2009) argumentation, the correlation can be estimated accurately based on the number of participants in each comparison group. With equal sample sizes in each group[4], the correlation between the two control groups, *no recall* and *other recall*, is 0, whereas the 'correlation' of the SAI experimental group is 1 (as it is ultimately the same data set), resulting in a combined correlation of 0.50 (Borenstein et al., 2009).

In total, for the association between the SAI and correct details a composite effect size was computed for four studies (Af Hjelmsaeter, 2012 1+2; Hope, 2014 2+3; Pfeil, 2016 3+5; and Pfeil, 2016 4+6; see table 2.5.).

Table 2.5. Individual and composite effect sizes (SAI vs. no recall and SAI vs. other recall) for correct details.

Subgroup			Composite		
Study Code	d	var	Study Code	d	var
Af Hjelmsaeter 2012 (1)	0.60	0.03	Af Hjelmsaeter 2012 (1+2)	0.49	0.02
Af Hjelmsaeter 2012 (2)	0.38	0.03			
Hope 2014 (2)	0.76	0.11	Hope 2014 (2+3)	0.98	0.08
Hope 2014 (3)	1.20	0.12			
Pfeil 2016 (3)	1.29	0.10	Pfeil 2016 (3+5)	0.94	0.07
Pfeil 2016 (5)	0.60	0.09			
Pfeil 2016 (4)	0.35	0.09	Pfeil 2016 (4+6)	0.18	0.06
Pfeil 2016 (6)	0.02	0.08			

[4] Most studies used equal sample sizes for each experimental group. In some studies the sample size varied up to *n* +/- 2 for groups, which is a negligible difference.

The same was done for the association between the SAI and accuracy rate, resulting in a composite effect size for the same four studies (Af Hjelmsaeter, 2012 1+2; Hope, 2014 2+3; Pfeil, 2016 3+5; and Pfeil, 2016 4+6; see table 2.6.).

Table 2.6. Individual and composite effect sizes (SAI vs. no recall and SAI vs. other recall) for accuracy rate.

Subgroup			Composite		
Study Code	d	var	Study Code	d	var
Af Hjelmsaeter 2012 (1)	-0.13	0.03	Af Hjelmsaeter 2012 (1+2)	-0.25	0.02
Af Hjelmsaeter 2012 (2)	-0.38	0.03			
Hope 2014 (2)	0.61	0.11	Hope 2014 (2+3)	0.59	0.08
Hope 2014 (3)	0.57	0.10			
Pfeil 2016 (3)	0.49	0.09	Pfeil 2016 (3+5)	0.08	0.06
Pfeil 2016 (5)	-0.33	0.08			
Pfeil 2016 (4)	0.93	0.09	Pfeil 2016 (4+6)	0.83	0.07
Pfeil 2016 (6)	0.73	0.09			

For all the reasons and proceedings outlined above, no study participants were thus represented more than once in any of the final meta-analyses.

2.3. Results

First, results from the systematic review of methodologies will be presented. Then, effect sizes for correct details and associated study characteristics will be analysed, followed by results from effect sizes for accuracy rate and again the associated study characteristics. A summary completes this section.

2.3.1. Systematic Review of Methodologies

The Self-Administered Interview was developed in the mid 2000s, with the first paper being published in 2009 (Gabbert et al., 2009). After the occasional publication in subsequent years, research has picked up and more studies were generated from 2014 onwards. In the following, the 22 studies that were identified as fitting the inclusion criteria for this meta-analytic review will be systematically reviewed. For ease of understanding, they were categorized by study design: Five studies conducted an initial comparison of the SAI versus another form of recall right after the crime had been witnessed (at time 1); eight studies conducted an impact measurement of the SAI at a delayed interview (at time 2); and nine studies conducted both an initial comparison and an impact measurement (time 1 and 2).

In the first group, studies that looked at a time 1 comparison of the SAI and another form of initial recall, four out of five studies were published in peer-reviewed journals (Gawrylowicz, Memon, & Scoboria, 2014; Gawrylowicz, Memon, Scoboria, et al., 2014; Krix, Sauerland, Merckelbach, Gabbert, & Hope, 2015; Krix et al., 2016), while one study is currently submitted for publication (Kemp et al., 2016). Participants were mainly students and young adults from the general public with mean ages per study of around 22 years. However, one study focused on older adults with an age range of 65-95 years (Gawrylowicz, Memon, Scoboria, et al., 2014). In all five studies, participants were randomly assigned to the recall conditions, which were either to complete the SAI or a written free recall booklet that simply instructed the participants to report everything they can remember in as much detail as possible without guessing. Thus, the type of comparison was consistent across the time 1 studies. However, instead of completing the recall conditions by paper and pen, one study opted for a computer-assisted completion (Kemp et al., 2016). In terms of outcome measures, all papers reported the classic memory recall coding schemes, i.e. correctly and incorrectly recalled details, as well as type-specific details such as

person or action details. In addition, Krix et al. (2016) manipulated the level of stress participants encountered during encoding, while Krix, Sauerland, Merckelbach, et al. (2015) also measured the working memory capacity and source monitoring abilities of participants. Gawrylowicz, Memon, and Scoboria (2014) and Gawrylowicz, Memon, Scoboria, et al. (2014) looked at a transfer effect of the SAI and measured the impact of the initial retrieval on recalling a second, different event after a delay. Lastly, Kemp et al. (2016) compared three recall conditions, the SAI, a written free recall, and the Cognitive Interview.

In the second group, studies that looked at a time 2 impact measurement of the SAI on a delayed interview, six out of eight studies were published in peer-reviewed journals (af Hjelmsäter, Strömwall, & Granhag, 2012; Gabbert, Hope, Fisher, & Jamieson, 2012; Gittins et al., 2015; Mackay & Paterson, 2015; McPhee et al., 2014; Paterson, Eijkemans, & Kemp, 2015), while one study is being prepared for publication (Colomb & Gabbert, 2013) and another study was published internally as a master's thesis (Bössenrodt, Hewig, Kraus, & Paelecke, 2011). Notably, four out of the six journal articles derive from the same working group of Dr Helen Paterson, University of Sydney, Australia. Participants were again mainly students and young adults from the general public with mean ages per study of around 20 years. One study focused on children aged 11-12 years (af Hjelmsäter et al., 2012). In seven out of eight studies, participants were randomly assigned to the recall conditions. Bössenrodt et al. (2011) chose a randomized block design in that they randomised participants, but also controlled for an equal distribution of gender across conditions.

All but one study was designed to compare the impact of the SAI vs. no initial recall on a delayed interview. Bössenrodt et al. (2011) compared the impact of the SAI to an oral free recall, but did not report any data on this initial (T1) comparison. af Hjelmsäter et al. (2012) additionally included an initial written free recall condition, but did not report data on this T1 comparison. Two studies further included conditions on a delayed application of the SAI, e.g. the next day (Mackay & Paterson, 2015; Paterson et al., 2015), which were excluded from any analysis in this study. Delays between application of the SAI and the interview recall ranged from 1 to 3 weeks. However, only two studies looked at a delay of two weeks (af Hjelmsäter et al., 2012; Paterson et al., 2015), and only one study conducted the interview after three weeks (Gabbert et al., 2012, study 2), leaving the main body of evidence based on only a one-week delay. The interviews at time 2 differed in modality and type:

Half of the studies conducted an oral interview (af Hjelmsäter et al., 2012; Bössenrodt et al., 2011; Colomb & Gabbert, 2013; McPhee et al., 2014), while the other half asked participants for a written recall of events (Gabbert et al., 2012; Gittins et al., 2015; Mackay & Paterson, 2015; Paterson et al., 2015); and again half of the studies asked participants for a free (af Hjelmsäter et al., 2012; Bössenrodt et al., 2011; Colomb & Gabbert, 2013; Gabbert et al., 2012, study 1; Paterson et al., 2015) and a cued recall (Gabbert et al., 2012, study 2; Gittins et al., 2015; Mackay & Paterson, 2015; McPhee et al., 2014), respectively.

In terms of outcome measures, all papers reported the classic memory recall coding schemes, such as correctly recalled details and accuracy rate. In addition, seven studies (all but Colomb and Gabbert (2013)) included measurements of suggestibility, either in the form of suggestive questions during a cued recall or in the form of introducing misleading post-event information before the interview and measuring susceptibility to those. Furthermore, Gittins et al. (2015) manipulated the level of stress participants encountered during encoding, while Mackay and Paterson (2015) measured psychological distress across all participants. And lastly, Colomb and Gabbert (2013) compared two interview types at time 2, the Modified Cognitive Interview and a structured interview.

Finally, in the third group, studies that looked at both the initial comparison and the impact of the SAI on a delayed interview, four out of nine studies were published in peer-reviewed journals (Gabbert et al., 2009; Hope, Gabbert, Fisher, & Jamieson, 2014; Krix, Sauerland, Lorei, & Rispens, 2015; Maras, Mulcahy, Memon, Picariello, & Bowler, 2014), while one study is being submitted for publication (Zeier, Hewig, Kraus, & Wagner, 2016), three studies were published internally as master's theses (Mauer, Hewig, & Kraus, 2013; Schoof, Hewig, & Kraus, 2014; Stephan, Hewig, & Kraus, 2013) and one study is being presented as part of this doctoral dissertation (Pfeil & Mueller-Johnson, 2016). Notably, the three master's theses and the study currently under review derive from the same working group of Dr Uta Kraus, University of Wuerzburg, Germany. Participants were mostly a mixture of students and more adults from the general public, resulting in higher mean ages of around 30 years. However, one study also focused on older adults with an age range of 60 years and above (Pfeil & Mueller-Johnson, 2016). In five out of nine studies, participants were randomly assigned to the recall conditions. The remaining four studies chose a randomized block design, i.e. they randomised participants to

recall conditions, but also controlled for either an equal distribution of gender across conditions (Mauer et al., 2013; Stephan et al., 2013), a person's individual suggestibility (Schoof et al., 2014) or for chronological age and verbal IQ (Maras et al., 2014).

Regarding control groups, the study designs varied slightly in the way that at time 1, seven studies compared the SAI with an initial free recall, whereas two studies used an initial cued recall (Maras et al., 2014; Zeier et al., 2016). Furthermore, about half of these control recalls were administered written (Gabbert et al., 2009; Hope et al., 2014; Krix, Sauerland, Lorei, et al., 2015; Maras et al., 2014; Pfeil & Mueller-Johnson, 2016) and orally (Mauer et al., 2013; Schoof et al., 2014; Stephan et al., 2013; Zeier et al., 2016), respectively. Additionally, three studies included a no initial recall control group (Gabbert et al., 2009; Hope et al., 2014; Pfeil & Mueller-Johnson, 2016) and one study further included a Cognitive Interview group (Gabbert et al., 2009) at time 1. All nine studies employed a delay of 1 week between the initial recall and the interview at time 2. These interviews at time 2 also differed in modality and type: For six studies, data was available from a free recall of events (Gabbert et al., 2009; Hope et al., 2014; Mauer et al., 2013; Pfeil & Mueller-Johnson, 2016; Schoof et al., 2014; Stephan et al., 2013)[5]. The remaining three studies either employed a cued recall or a mixture of free and cued recall, specifically a second SAI (Krix, Sauerland, Lorei, et al., 2015), a structured recall similar to the SAI but without the cognitive and memory-enhancing techniques (Maras et al., 2014) or a short structured recall with specific questions (Zeier et al., 2016). Again, half of the studies conducted an oral interview (Hope et al., 2014; Mauer et al., 2013; Pfeil & Mueller-Johnson, 2016; Schoof et al., 2014; Stephan et al., 2013), while the other half asked participants for a written account of events (Gabbert et al., 2009; Krix, Sauerland, Lorei, et al., 2015; Maras et al., 2014; Zeier et al., 2016).

In terms of outcome measures, all papers reported the classic memory recall coding schemes, such as correctly recalled details and accuracy rate. In addition, four studies provided data on suggestive questions during the delayed interview (Mauer et al., 2013; Pfeil & Mueller-Johnson, 2016; Schoof et al., 2014; Stephan et al., 2013). Further, Krix, Sauerland, Lorei, et al. (2015) additionally looked at police detectives

[5] Some of these studies also provided data from a subsequent cued recall. However, this data was excluded from analysis as mentioned previously.

and their beliefs on consistency and reminiscence of witness testimony. Maras et al. (2014) was the only study to include an experimental group of mentally ill people, namely those with autism spectrum disorder. And lastly, the study presented in this dissertation in the only one to also measure the impact of the SAI on witnesses' lineup performance (Pfeil & Mueller-Johnson, 2016).

2.3.2. Meta-analyses

In the following, results for several meta-analyses will be presented. For each analysis, results include the standard effect summary and forest plot, and detailed publication bias analyses as well as moderator analyses.

For the standard effect summary and forest plot, as well as for the publication bias analyses, the software Comprehensive Meta-Analysis (CMA) 2.0 (Borenstein, Hedges, Higgins, & Rothstein, 2005) was used. All summary effects presented below are based on the random effects model rather than the fixed effects model. Whereas a fixed-effect model assumes that there is one true effect size underlying all studies in the analysis, and that all differences in observed effects are due to sampling error alone (Borenstein et al., 2009), a random-effects model allows for variation of the effect size from study to study. These variations could for example be due to differences in age groups, health or education among study samples, resulting in higher or lower effects in children, more impaired or less educated samples. They could also be due to differences in interventions such as using modified versions, or in intervention intervals. Thus, as studies differ in their chosen samples and implementations of interventions, there may be different effect sizes underlying these studies, which represent a random sample of the true effect sizes (Borenstein et al., 2009).

As described above, studies included in the following meta-analyses all comply with very narrow inclusion criteria. And yet, they are not functionally equivalent and there is some considerable variation in their specific study designs, as was discussed in the systematic review. For example, most studies used student participants, but some looked at children and older adults; most studies using an impact measurement had a one week delay, but some used two or even three weeks; most studies asked participants to give a free recall, but some chose to ask more specific questions instead. These variations are likely to reflect different underlying effect sizes. Additionally, the goal of the meta-analyses in this study is to generalize to other populations and inform on a more broad effect, rather than summarize the effect for one specific population. Therefore, the random-effects model was chosen for analyses.

Following the basic analyses of summary effects, possible publication biases were then analysed across all meta-analyses. While a meta-analysis yields a

mathematically accurate synthesis of the included studies, if these studies are a biased sample of all relevant studies, then the mean effect size will reflect this bias (Borenstein et al., 2009). As described in the study selection procedure, every precaution was taken to ensure representation of all eligible studies in the final meta-analyses. This entailed contacting several relevant authors and including published and unpublished studies (Masters and PhD theses, unpublished manuscripts). To provide a statistical analysis, Duval and Tweedie (2000a, 2000b) trim-and-fill procedure was employed. This is an iterative process that produces the best estimate of an unbiased effect size. It removes the most extreme small studies until a funnel plot is symmetrical around an adjusted effect size and then inserts them back into the analysis to correct the variance (Borenstein et al., 2009). The resulting funnel plot then shows imputed studies in black circles (and actual included studies in blank circles) and an adjusted effect size as a black diamond (with the original effect size as a blank diamond).

Additionally, Rosenthal's Fail-Safe N test was used (Rosenthal, 1979). It addresses the concern that studies with small or no effects might be missing from the analysis, which, if included, would reduce the observed effect. Rather than simply speculating about the impact of the missing studies, this method calculates how many studies with no effect would need to be found in order to make the p-value non-significant. If this number is relatively small, then there is indeed cause for concern. However, if fail-safe N is large, it indicates that the observed effect is real, even if possibly inflated by the exclusion of some studies. These measures and precautions taken together will give an indication of any possible publication bias.

Finally, moderator analyses were conducted to assess the relationship between study-level covariates and effect size. For this analysis, IBM SPSS (version 22.0) was used, including a macro for meta-regression (Wilson, 2010). First, heterogeneity measurements will be reported to give an indication of the methodological (e.g. study design) and clinical diversity (e.g. variability in participants) of studies. These include Cochran's Q-statistic with its relevant degrees of freedom and p-value, and the I-squared statistic. The Q-statistic represents an analogue to the chi-squared statistic and is the standard measure of heterogeneity (Higgins & Green, 2011). However, Q has low power to detect inconsistencies of studies' results when the number of studies in the meta-analysis is small, and conversely, it has too much power when the number of studies is large (Higgins, Thompson, Deeks, & Altman, 2003). I-squared describes

the percentage of the variability in effect estimates that is due to heterogeneity rather than sampling error or chance alone (Higgins & Green, 2011). According to the Cochrane Handbook (Higgins & Green, 2011) a rough guide to the interpretation of I^2 is as follows: 0% to 40% does not indicate any important heterogeneity among studies; 30% to 60% indicates moderate heterogeneity; 50% to 90% represents substantial heterogeneity, and 75% to 100% represents considerable heterogeneity. It may be noted that the value of I^2 depends on the magnitude and direction of effects as well as on the strength of evidence for heterogeneity as e.g. indicated by the p-value of Q (Higgins & Green, 2011). As in the primary analyses, moderator analyses are based on the random-effects model, as again, it allows for variation of the effect size from study to study, e.g. due to differences in age groups or intervention intervals.

2.3.2.1. The Association between the SAI and Correct Details

This section contains results for several meta-analyses on the association between the Self-Administered Interview and correctly remembered details in a recall attempt. First, results for an initial comparison at time 1 will be presented, i.e. a direct comparison of details elicited in the SAI and in another form of initial recall, such as a written free recall. This is followed by results from impact measurements at time 2, i.e. the impact the SAI has on a later witness interview compared to either no initial recall, another form of initial recall and a combination of both control groups.

2.3.2.1.1. Initial Comparison at Time 1.

A total of 15 studies presented data on correctly recalled details at time 1 for the comparison of the SAI versus other types of initial recall. The summary effect size across these studies showed a very large and substantial effect, d = 1.20, p = .000 (95% CI: 0.95–1.46; z = 9.40). All but one study yielded a significant effect in favour of the SAI as shown in the forest plot in figure 2.1. This means that immediately after the event, when witnesses filled in the SAI they provided significantly and substantially more correct information that in a control recall task (e.g. a written free recall).

Model	Study name	Subgroup within study	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
	Pfeil 2016 (2)	T1, SAI vs. FR	0.294	0.290	0.084	-0.275	0.863	1.014	0.311
	Gawrylowicz 2014b	T1, SAI vs. FR	0.719	0.231	0.053	0.266	1.171	3.114	0.002
	Zeier 2016 (1)	T1, SAI vs. SR	0.818	0.321	0.103	0.188	1.447	2.545	0.011
	Krix 2015a (1)	T1, SAI vs. FR	0.822	0.227	0.052	0.377	1.268	3.616	0.000
	Kemp 2016	T1, SAI vs. FR	0.885	0.280	0.078	0.336	1.434	3.160	0.002
	Maras 2014 (1)	T1, SAI vs. SR	0.972	0.358	0.128	0.271	1.673	2.719	0.007
	Pfeil 2016 (1)	T1, SAI vs. FR	1.015	0.307	0.094	0.414	1.616	3.309	0.001
	Krix 2015b	T1, SAI vs. FR	1.183	0.194	0.038	0.803	1.563	6.100	0.000
	Gawrylowicz 2014a	T1, SAI vs. FR	1.206	0.237	0.056	0.741	1.671	5.083	0.000
	Mauer 2013 (1)	T1, SAI vs. FR	1.217	0.344	0.119	0.542	1.892	3.535	0.000
	Krix 2016	T1, SAI vs. FR	1.616	0.204	0.042	1.215	2.016	7.905	0.000
	Gabbert 2009 (1)	T1, SAI vs. FR	1.657	0.381	0.145	0.910	2.404	4.347	0.000
	Stephan 2013 (1)	T1, SAI vs. FR	1.824	0.347	0.121	1.144	2.505	5.255	0.000
	Hope 2014 (1)	T1, SAI vs. FR	2.134	0.396	0.157	1.358	2.910	5.387	0.000
	Schoof 2014 (1)	T1, SAI vs. FR	2.198	0.345	0.119	1.523	2.874	6.375	0.000
Fixed			1.169	0.071	0.005	1.029	1.308	16.430	0.000
Random			1.204	0.128	0.016	0.953	1.455	9.400	0.000

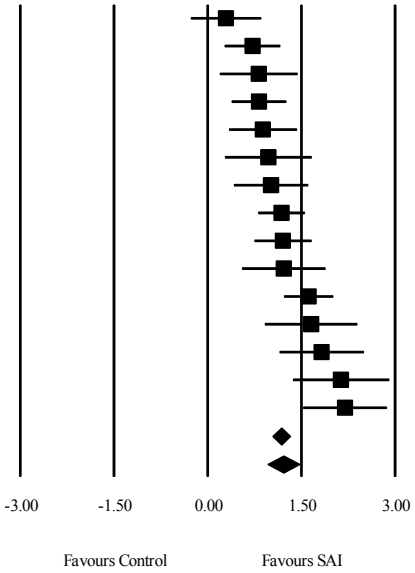


Figure 2.1. Forest plot for SAI vs. other types of initial recall at T1 for correct details.

Publication bias analyses through Duval and Tweedie's trim-and-fill procedure did not reveal any imputed effect sizes in the relevant funnel plot (fig. 2.2.). This means that under the random effects model the point estimate and 95% confidence interval [*d* = 1.20 (95% CI: 0.95–1.46)] remain the same using trim-and-fill. Furthermore, Rosenthal's fail-safe N test for this meta-analysis is 1034, which means that 1034 studies containing an initial comparison of the SAI and another type of recall would need to be located *and* would need to show no effect in order for the 2-tailed *p*-value to exceed 0.05. In other words, there would need to be 69 missing 'null' studies for *every* observed study for this effect to be nullified. It seems extremely unlikely that such a large number of studies have been missed in the literature search. Thus, the overall results from the sensitivity analyses tests strongly suggest the absence of publication bias in this meta-analysis.

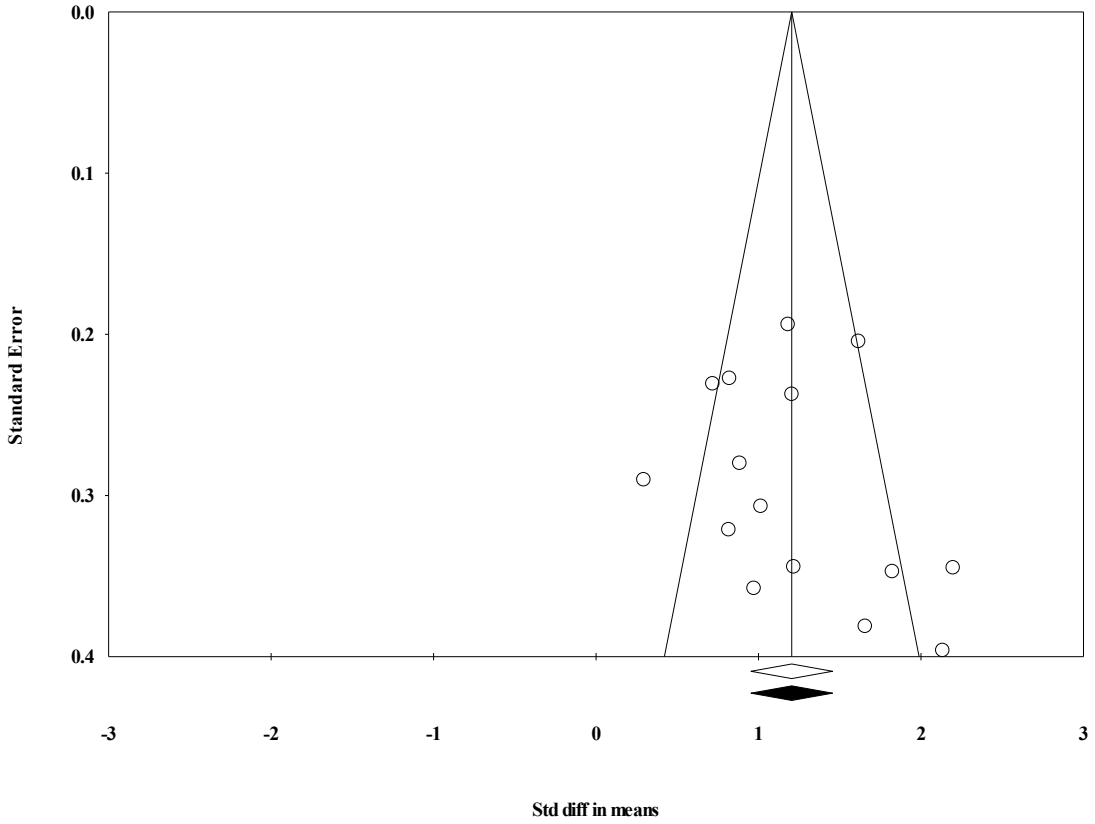


Figure 2.2. Funnel plot for SAI vs. other types of initial recall at T1 for correct details.

Although all but one study yielded an effect size supporting the link between the SAI and an increase in correct details at T1, the magnitude and significance of the

effect varied across these studies. The Q and I^2 statistics confirmed that the there was substantial heterogeneity across the studies ($Q(14) = 42.88$, $p < .001$, $I^2 = 67.35$) and thus moderator analyses were conducted to try to explain this variability. A weighted least squares regression model was conducted to predict the variance in effect sizes across studies based upon the following moderators: age group (adults vs. children and older adults), laboratory (original authors vs. other), publication type (published vs. unpublished manuscript), and recall modality at T1 (written vs. oral). Studies were weighted by the reciprocal of the variance of each effect size, as this procedure gives the most accurate mean effect size (Hedges & Olkin, 1985). All predictors were entered simultaneously into the regression model and results are shown in table 2.7. The overall regression model proved to just be significant ($p < .05$). However, none of the regression coefficients in the model showed a significant impact on the model, which indicates that the given moderators did not impact on the overall effect.

Table 2.7. Results of meta regression for SAI vs. other types of initial recall with effect sizes for correct details at T1 as dependent variable.

Predictor	B	SE B	β	Z-value
Age group	0.54	0.34	.38	1.57
Laboratory	0.36	0.33	.28	1.11
Publication type	0.29	0.31	.28	0.91
T1 recall modality	0.59	0.36	.50	1.63

Note. All p-values are non-significant.

2.3.2.1.2. Impact Measurement at Time 2.

This meta-analysis represents an overall effect of the impact of the SAI on a later interview recall. It contains comparisons of the SAI versus no initial recall as well as the SAI versus another form of initial recall. A total of 19 studies presented data on the impact of the Self-Administered Interview versus any type of control group and correctly recalled details at time 2. The summary effect size across these studies showed a highly significant, medium-sized effect $d = 0.70$, $p = .000$ (95% CI: 0.45–0.95; z = 5.43). Thirteen out the 19 studies yielded a significant effect in favour of the SAI as shown in the forest plot in figure 2.3. This means that witnesses who filled in the SAI after having witnesses a crime provided significantly and substantially more correct information in an interview recall after a delay of 1-3

weeks than those who did not have an initial recall opportunity or who have filled in another form of initial recall (such as a written free recall).

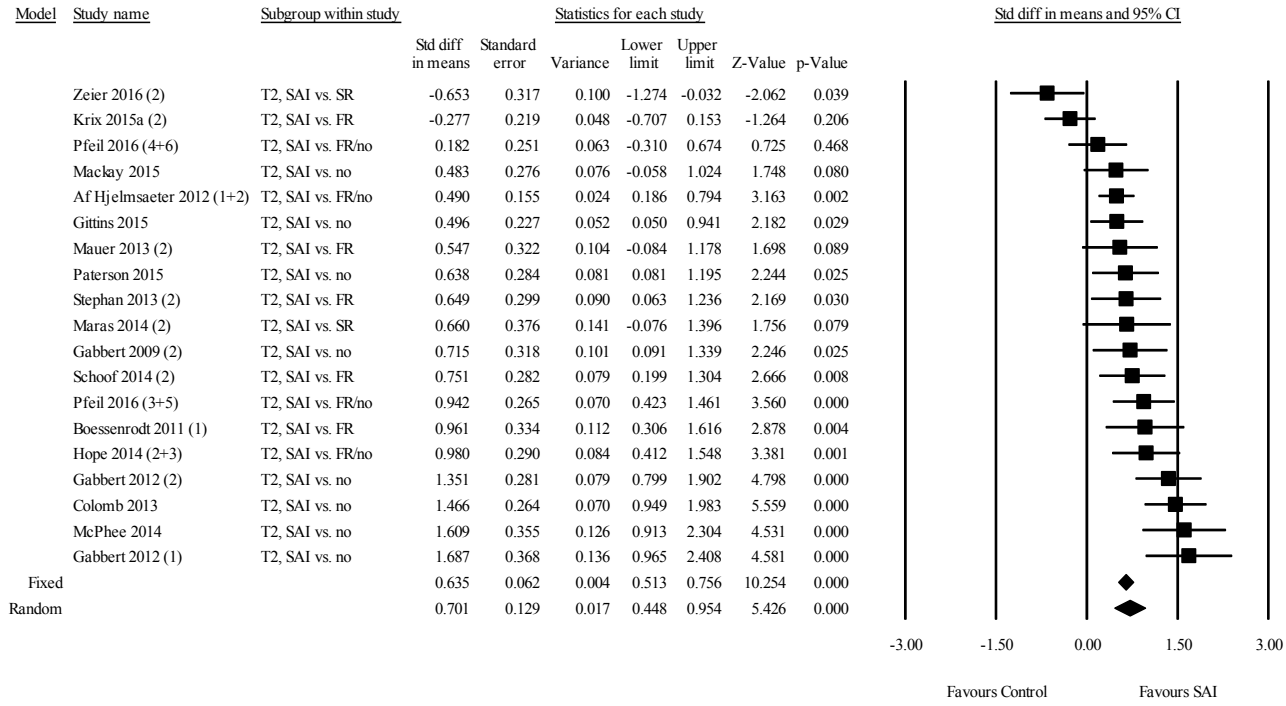


Figure 2.3. Forest plot for SAI vs. any control group at T2 for correct details.

Publication bias analyses through Duval and Tweedie's trim-and-fill procedure again revealed only one imputed effect size in the relevant funnel plot (fig. 2.4.), resulting in a minimal shift of the imputed summary effect size. Under the random effects model the point estimate and 95% confidence interval is d = 0.70 (95% CI: 0.45–0.95). Using trim-and-fill the imputed point estimate is d = 0.65 (95% CI: 0.40–0.91). This suggests an only trivial overestimation of the current summary effect size. Furthermore, Rosenthal's fail-safe N test for this meta-analysis is 546, which means that 546 studies containing an impact measurement of the SAI and either no initial recall or another form of initial recall would need to be located *and* would need to show no effect in order for the 2-tailed p-value to exceed 0.05. In other words, there would need to be 29 missing 'null' studies for *every* observed study for this effect to be nullified. It seems extremely implausible that such a large number of studies have been missed in the literature search. The overall results from the sensitivity analyses tests suggest the absence of publication bias in this meta-analysis as well.

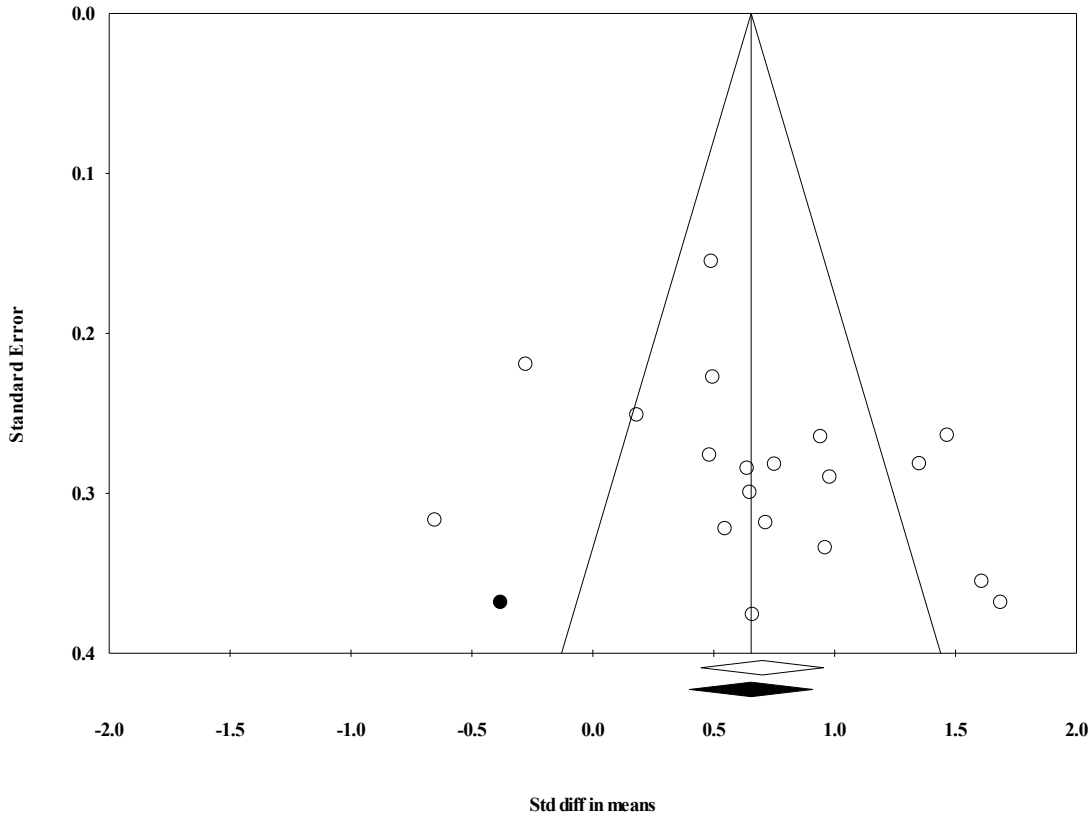


Figure 2.4. Funnel plot for SAI vs. any control group at T2 for correct details.

Although most studies yielded an effect size supporting the link between the SAI and an increase in correct details at T2, the magnitude and significance of the effect varied across these studies. The Q and I^2 statistics confirmed that the there was substantial heterogeneity across the studies ($Q(18) = 74.75$, $p < .001$, $I^2 = 75.92$) and thus moderator analyses were conducted to try to explain this variability. A weighted least squares regression model was conducted to predict the variance in effect sizes across studies based upon the following moderators: age group (adults vs. children and older adults), laboratory (original authors vs. other), publication type (published vs. unpublished manuscript), control condition at T1 (no initial recall vs. other type of initial recall), recall type at T2 (free vs. cued recall), and recall modality at T2 (written vs. oral). Studies were again weighted by the reciprocal of the variance of each effect size. All predictors were entered simultaneously into the regression model and results are shown in table 2.8. The overall regression model proved to be significant ($Q(6) = 46.70$, $p < .001$) with effect sizes for correct details at T2 significantly influenced by the control condition at T1 ($Z = -4.12$, $p < .001$) and the recall modality at T2 ($Z = 4.06$, $p < .001$). This means that it made a difference to the effect size if the SAI

was compared to no initial recall at T1 or to another form of initial recall, and if the recall in the interview at T2 was done orally or written. This will be further expanded below.

Table 2.8. Results of meta regression for SAI vs. any control group with effect sizes for correct details at T2 as dependent variable.

Predictor	B	SE B	β	Z-value	
Age group	0.45	0.25	.32	1.85	
Laboratory	0.43	0.23	.30	1.86	
Publication type	0.21	0.22	.18	0.95	
T1 control condition	-0.75	0.18	-.67	-4.12	***
T2 recall type	-0.06	0.23	-.05	-0.25	
T2 recall modality	0.98	0.24	.89	4.06	***

***$p < .001$.

In addition to the meta-regression, univariate testing for moderators that yielded significant effects was conducted. This is analogous to testing main effects in an ANOVA and compares categorical subgroups with each other. For the moderator *T1 control* condition, eight studies looked at the comparison of the SAI versus no initial recall, whereas eleven studies compared the SAI versus another type of initial recall.[6] A comparison of these two subgroups revealed a significant difference in the mean effect sizes for correct details in both the multivariate and the univariate analysis. The mean effect size was larger when comparing the SAI versus no initial recall ($d = 1.03$, $p < .001$) compared to giving participants a form of type of initial recall ($d = 0.46$, $p < .01$; $Q_b = 5.60$, $p < .05$). For the moderator *T2 recall modality*, nine studies asked participants to give a written account of their memory, whereas ten studies asked for a recall in an oral interview. Whereas the multivariate analysis revealed a highly significant effect, a univariate comparison of these two subgroups showed no significant difference between mean effect sizes for correct details (written: $d = 0.54$, $p < .01$, oral: $d = 0.84$, $p < .001$; $Q_b = 1.34$, *n.s.*).

[6] For this subgroup analysis, the four studies in which composite effect sizes across both control conditions were calculated, were coded as 'another type of recall', as this represents the more conservative approach.

Further to the overall effect of the SAI at time 2, two additional meta-analyses were run to separately analyse and compare summary effects for the two distinct control conditions 'SAI versus no initial recall' and 'SAI versus another form of initial recall'.[7] Whereas the former presents the current practice after a crime or incident and is therefore interesting from a practitioner's point of view, the latter challenges the SAI in that it may show an equal advantage of an even simpler recall instruction such as a written free recall and may thus prove the SAI unnecessary.

In the first meta-analysis, a total of 12 studies presented data on the impact of the Self-Administered Interview versus no initial recall on correctly recalled details at time 2. The summary effect size across these studies showed a large and substantial effect, $d = 0.92$, $p = .000$ (95% CI: 0.66–1.18; $z = 6.84$). All but two individual studies yielded a significant effect in favour of the SAI as shown in the forest plot in figure 2.5. This means that witnesses who filled in the SAI after having witnesses a crime provided significantly and substantially more correct information in an interview recall after a delay of 1-3 weeks than those who did not have any initial recall opportunity.

Model	Study name	Subgroup within study	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
	Pfeil 2016 (4)	T2, SAI vs. no	0.345	0.291	0.085	-0.225	0.915	1.186	0.236
	Mackay 2015	T2, SAI vs. no	0.483	0.276	0.076	-0.058	1.024	1.748	0.080
	Gittins 2015	T2, SAI vs. no	0.496	0.227	0.052	0.050	0.941	2.182	0.029
	Af Hjelmsaeter 2012 (1)	T2, SAI vs. no	0.603	0.180	0.032	0.251	0.956	3.351	0.001
	Paterson 2015	T2, SAI vs. no	0.638	0.284	0.081	0.081	1.195	2.244	0.025
	Gabbert 2009 (2)	T2, SAI vs. no	0.715	0.318	0.101	0.091	1.339	2.246	0.025
	Hope 2014 (2)	T2, SAI vs. no	0.760	0.327	0.107	0.119	1.402	2.322	0.020
	Pfeil 2016 (3)	T2, SAI vs. no	1.286	0.317	0.101	0.664	1.907	4.055	0.000
	Gabbert 2012 (2)	T2, SAI vs. no	1.351	0.281	0.079	0.799	1.902	4.798	0.000
	Colomb 2013	T2, SAI vs. no	1.466	0.264	0.070	0.949	1.983	5.559	0.000
	McPhee 2014	T2, SAI vs. no	1.609	0.355	0.126	0.913	2.304	4.531	0.000
	Gabbert 2012 (1)	T2, SAI vs. no	1.687	0.368	0.136	0.965	2.408	4.581	0.000
Fixed			0.860	0.079	0.006	0.705	1.016	10.849	0.000
Random			0.920	0.135	0.018	0.656	1.184	6.840	0.000

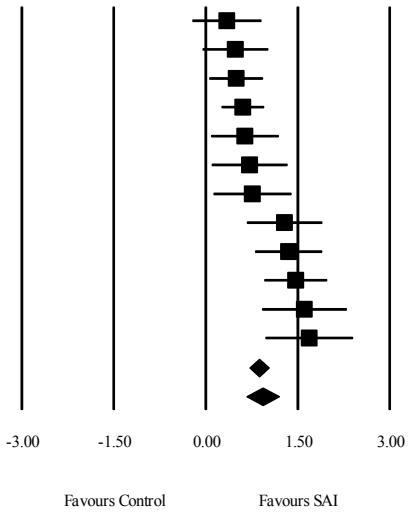


Figure 2.5. Forest plot for SAI vs. no initial recall at T2 for correct details.

[7] In addition to the information gained from the moderator analysis of the overall effect, these separate analyses provide a cleaner way of examining the SAI. They do not rely on calculating composite effects, which is always a statistical compromise. Instead, they only include original effects found in the individual studies.

Publication bias analyses through Duval and Tweedie's trim-and-fill procedure revealed two imputed effect sizes in the relevant funnel plot (fig. 2.6.), resulting in a slight shift of the imputed summary effect size and suggesting a trivial overestimation of the current summary effect size. Under the random effects model the point estimate and 95% confidence interval is d = 0.92 (95% CI: 0.66–1.18). Using trim-and-fill the imputed point estimate is d = 0.81 (95% CI: 0.53–1.08). Furthermore, Rosenthal's fail-safe N test for this meta-analysis is 380 which means that there would need to be 32 missing 'null' studies for *every* observed study for this effect to be nullified. Again, it seems extremely unlikely that such a large number of studies have been missed in the literature search. The overall results from the sensitivity analyses tests suggest the absence of publication bias in this meta-analysis.

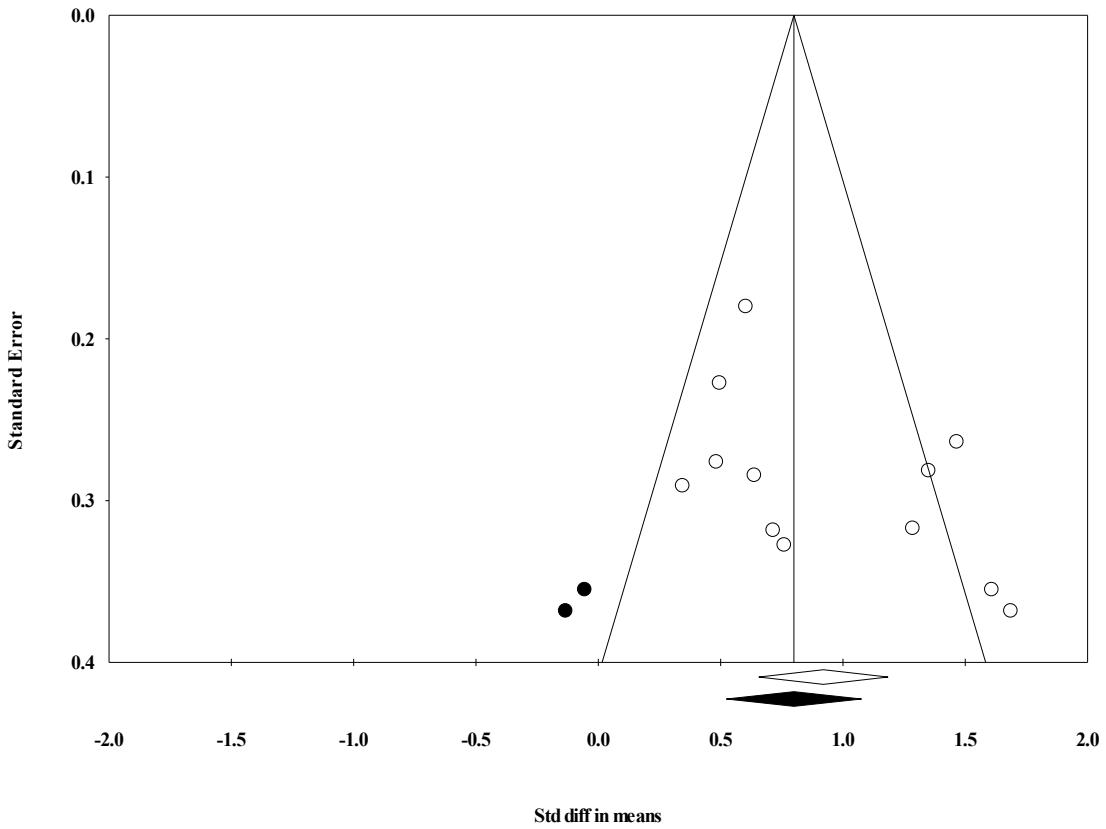


Figure 2.6. Funnel plot for SAI vs. no initial recall at T2 for correct details.

In the second meta-analysis, a total of 11 studies presented data on the impact of the Self-Administered Interview versus another form of initial recall on correctly recalled details at time 2. The summary effect size across these studies showed a

significant, medium-sized effect, $d = 0.42$, $p = .008$ (95% CI: 0.11–0.73; $z = 2.64$). Six out of the 11 individual studies yielded a significant effect in favour of the SAI, four yielded non-significant effects and one even yielded a significant effect that favoured the control recall (see fig. 2.7.). Despite the fact that the summary effect is smaller than in the SAI vs. no recall meta-analysis, this still means that overall, witnesses who filled in the SAI after having witnesses a crime provided significantly more correct information in a delayed interview recall than those who filled in another form of initial recall and thus shows the superiority of the SAI over other forms of initial recall, such as a written free recall instruction.

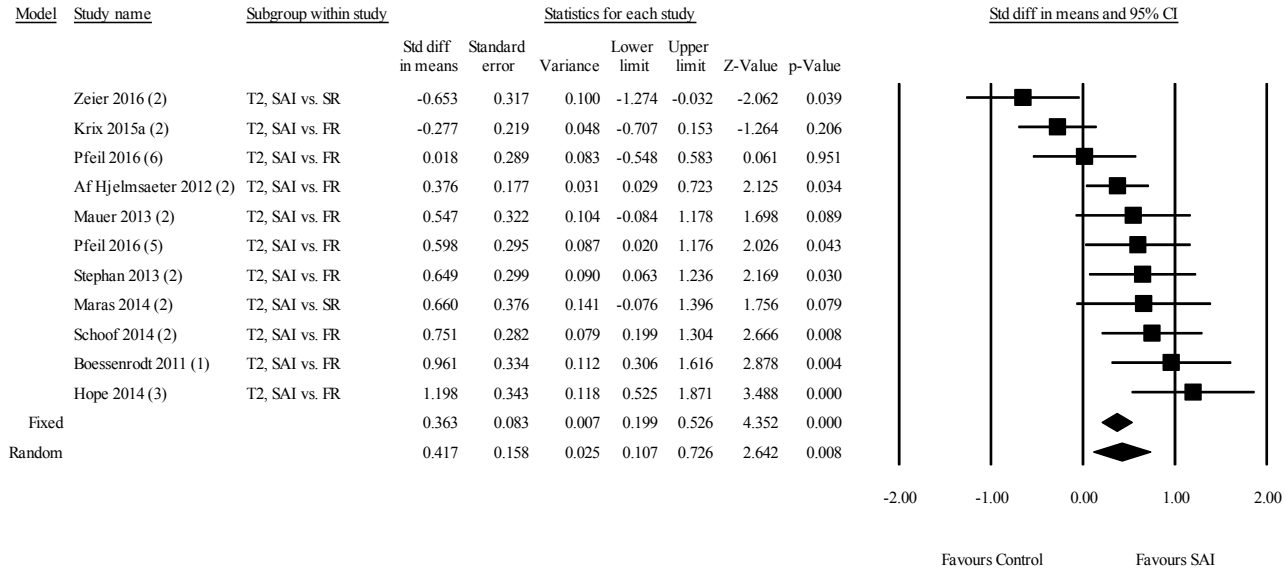


Figure 2.7. Forest plot for SAI vs. other types of initial recall at T2 for correct details.

Publication bias analyses through Duval and Tweedie's trim-and-fill procedure revealed two imputed effect size in the relevant funnel plot (fig. 2.8.), resulting in a slight shift of the imputed summary effect size. Under the random effects model the point estimate and 95% confidence interval is $d = 0.42$ (95% CI: 0.11–0.73). Using trim-and-fill the imputed point estimate is $d = 0.35$ (95% CI: 0.04–0.66), suggesting a trivial overestimation of the current summary effect size. Furthermore, Rosenthal's fail-safe N test for this meta-analysis is 52 which means that 52 studies containing an impact measurement of the SAI versus another form of initial recall would need to be located *and* would need to show no effect in order for the 2-tailed *p*-value to exceed 0.05. Despite the fact that this number is not quite as impressive as in the previous

meta-analyses above, it still seems highly unlikely that this rather large number of studies have been missed in the literature search. Therefore, the overall results from the sensitivity analyses tests still suggest the absence of publication bias in this meta-analysis.

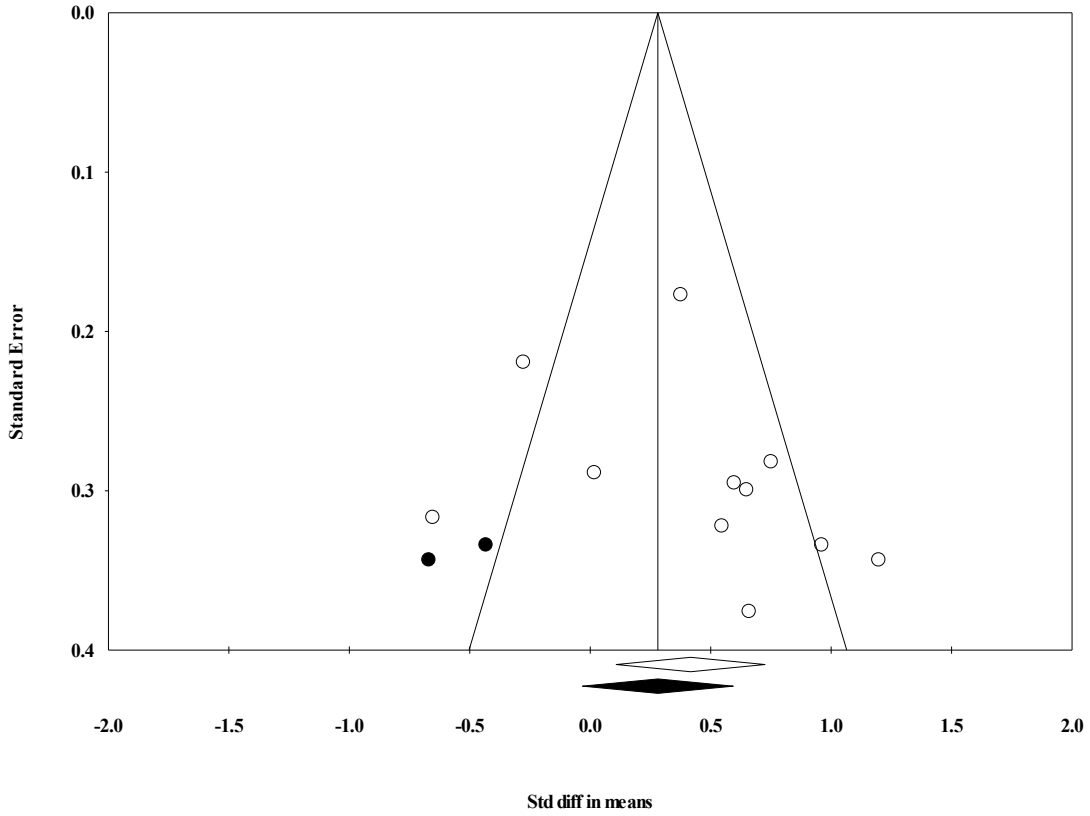


Figure 2.8. Funnel plot for SAI vs. other types of initial recall at T2 for correct details.

2.3.2.2. The Association between the SAI and Accuracy Rate

This section contains results for several meta-analyses on the association between the Self-Administered Interview and accuracy rate in a recall attempt. First, results for an initial comparison at time 1 will be presented, i.e. a direct comparison of details elicited in the SAI and in another form of initial recall, such as a written free recall. This is followed by results from impact measurements at time 2, i.e. the impact the SAI has on a later witness interview compared to either no initial recall, another form of initial recall and a combination of both control groups.

2.3.2.2.1. Initial Comparison at Time 1.

Fifteen studies presented data on the association of the Self-Administered Interview and accuracy rate at time 1. The summary effect size across these studies showed a small, yet significant effect in favour of the *control condition*, d = -0.25, p = .001 (95% CI: -0.39– -0.11; z = -3.43). However, only two out of the 15 individual studies yielded a significant effect, whereas results from the remaining 13 studies were non-significant, as shown in the forest plot in figure 2.9. This means that immediately after the event, when witnesses filled in the SAI their accuracy rates were slightly lower than in a control recall task (e.g. a written free recall), suggesting that in addition to a substantial increase in correct details, the SAI also lead to a slight increase in incorrect details.

Publication bias analyses through Duval and Tweedie's trim-and-fill procedure did not reveal any imputed effect size in the relevant funnel plot (fig. 2.10.). This means that under the random effects model the point estimate and 95% confidence interval [d = -0.25 (95% CI: -0.39– -0.11)] remain the same using trim-and-fill. Furthermore, Rosenthal's fail-safe N test for this meta-analysis is 44, which means that 44 studies containing an initial comparison of the SAI and another type of recall would need to be located *and* would need to show no effect in order for the 2-tailed p-value to exceed 0.05. In other words, there would need to be 3 missing 'null' studies for every observed study for this effect to be nullified. In conclusion, the overall results from the sensitivity analyses tests still suggest the absence of publication bias in this meta-analysis.

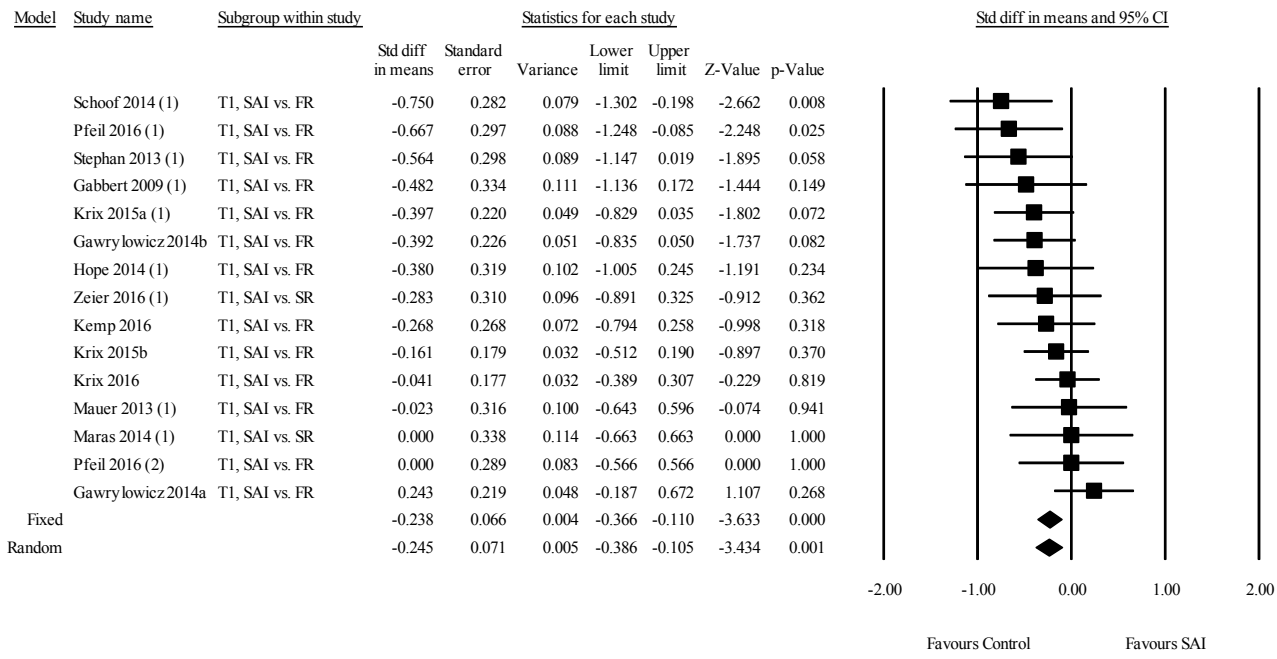


Figure 2.9. Forest plot for SAI vs. other types of initial recall at T1 for accuracy rate.

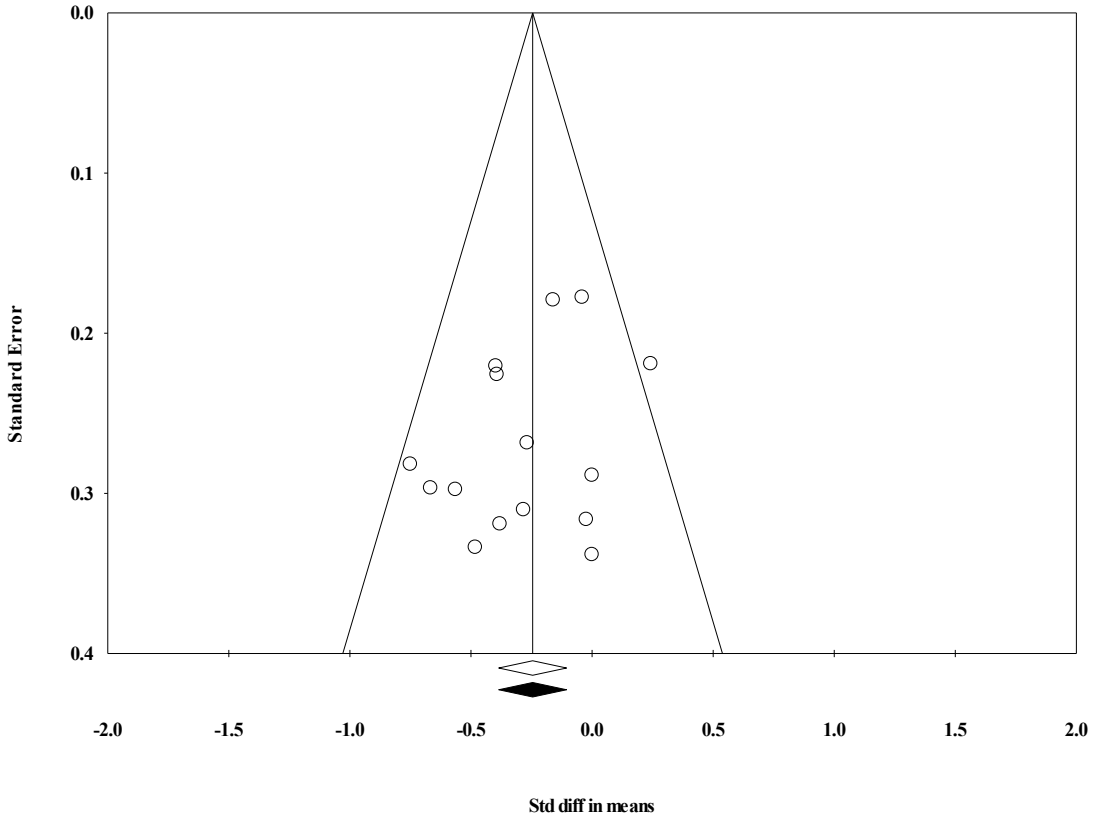


Figure 2.10. Funnel plot for SAI vs. other types of initial recall at T1 for accuracy rate.

The overall majority of studies failed to yield a significant effect linking either the SAI or the control condition to a higher accuracy rate at T1. Heterogeneity analyses underlines this low variability among studies further ($Q(14) = 16.22$, p = n.s., I^2 = 13.67). When running a weighted least squares regression model using the moderators age group, laboratory, publication type, and recall modality at T1, the overall regression model proved to be non-significant (p = n.s.). It can thus be concluded that there is no indication of heterogeneity among studies in this meta-analysis.

2.3.2.2.2. Impact Measurement at Time 2.

The first meta-analysis at time 2 again represents an overall effect of the impact of the SAI on a later interview recall. It contains comparisons of the SAI versus no initial recall as well as the SAI versus another form of initial recall. A total of 19 studies presented data on the impact of the Self-Administered Interview versus any type of control group and accuracy rate at time 2. The summary effect size across these studies showed a medium-sized and significant effect, $d = 0.38$, $p = .001$ (95% CI: 0.15–0.60; z = 3.32). Out of the 19 individual studies, 8 yielded a significant effect in favour of the SAI, whereas results from the remaining 11 were non-significant, as shown in figure 2.11. This means that witnesses who filled in the SAI after having witnesses a crime performed with a higher accuracy rate in an interview recall after a delay of 1-3 weeks than those who did not have an initial recall opportunity or who have filled in another form of initial recall (such as a written free recall).

Publication bias analyses through Duval and Tweedie's trim-and-fill procedure revealed six imputed effect sizes in the relevant funnel plot (fig. 2.12.). This means that under the random effects model the point estimate and 95% confidence interval [d = 0.38 (95% CI: 0.15–0.60) remain the same using trim-and-fill. Furthermore, Rosenthal's fail-safe N test for this meta-analysis is 143, which means that 143 studies containing an impact measurement of the SAI and either no initial recall or another form of initial recall would need to be located *and* would need to show no effect in order for the 2-tailed p-value to exceed 0.05. The overall results from the sensitivity analyses tests suggest the absence of publication bias in this meta-analysis.

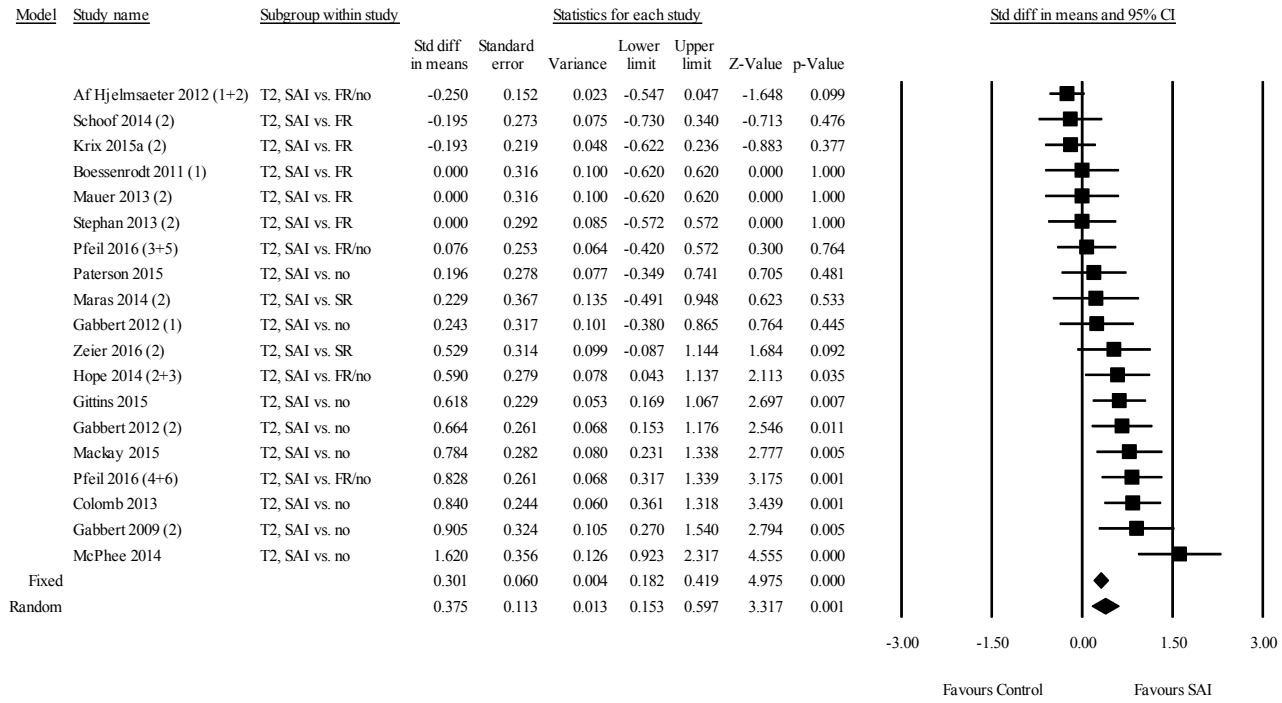


Figure 2.11. Forest plot for SAI vs. any control group at T2 for accuracy rate.

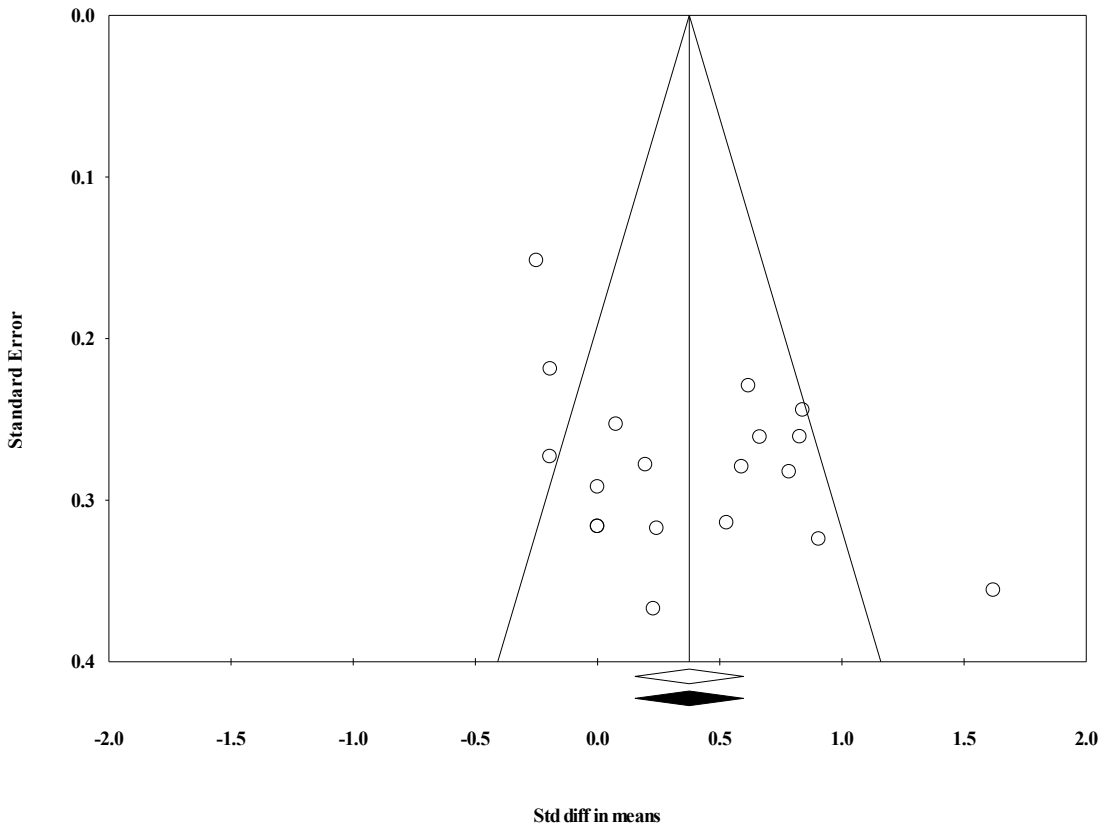


Figure 2.12. Funnel plot for SAI vs. any control group at T2 for accuracy rate.

As less than half of the individual studies yielded a significant effect size supporting the link between the SAI and a higher accuracy rate at T2, and because the magnitude and significance of the effect varied substantially across studies as indicated by the Q and I^2 statistics ($Q(18) = 60.05$, $p < .001$, $I^2 = 70.02$), moderator analyses were conducted to try to explain this variability. A weighted least squares regression model was conducted to predict the variance in effect sizes across studies based upon the following moderators: age group (adults vs. children and older adults), laboratory (original authors vs. other), publication type (published vs. unpublished manuscript), control condition at T1 (no initial recall vs. other type of initial recall), recall type at T2 (free vs. cued recall), and recall modality at T2 (written vs. oral). Studies were again weighted by the reciprocal of the variance of each effect size. All predictors were entered simultaneously into the regression model and results are shown in table 2.9. The overall regression effect proved to be significant ($p < .01$) with effect sizes for accuracy rate at T2 significantly influenced by the control condition at T1 ($Z = -3.89$, $p < .01$) and the recall type at T2 ($Z = 2.73$, $p < .01$). The moderators recall modality at T2 ($Z = 1.88$, $p = .06$) and laboratory ($Z = 1.81$, $p = .07$) just failed to reach the significance level.

Table 2.9. Results of meta regression for SAI vs. any control group with effect sizes for accuracy rate at T2 as dependent variable.

Predictor	B	SE B	β	Z-value	
Age group	-0.27	0.28	-.21	1.53	
Laboratory	0.45	0.25	.38	1.81	
Publication type	-0.36	0.24	-.38	-1.50	
T1 control condition	-0.78	0.20	-.82	-3.89	**
T2 recall type	0.67	0.24	.69	2.73	**
T2 recall modality	0.49	0.26	.53	1.89	

**$p < .01$.

In addition to the meta-regression, univariate testing for moderators that yielded significant effects, and those that were approaching significance was conducted. This is analogous to testing main effects in an ANOVA and compares categorical subgroups with each other. For the moderator *T1 control* condition, eight studies looked at the comparison of the SAI versus no initial recall, whereas eleven

studies compared the SAI versus another type of initial recall.[8] A comparison of these two subgroups revealed a significant difference in the mean effect sizes for accuracy rate in both the multivariate and the univariate analysis. The mean effect size was larger when comparing the SAI versus no initial recall (d = 0.71, p < .001) compared to giving participants another type of initial recall (d = 0.12, p = n.s.; Q_b = 10.72, p < .05). For the moderator *T2 recall type*, twelve studies asked participants to provide free recall, whereas in seven studies participants were asked specific questions (cued recall). Whereas the multivariate analysis revealed a significant effect, a univariate comparison of these two subgroups showed no significant difference between mean effect sizes for accuracy rate (free recall: d = 0.26, p = n.s., cued recall: d = 0.58, p < .05; Q_b = 1.89, *n.s.*). For the moderator *T2 recall modality*, nine studies asked participants to give a written account of their memory, whereas ten studies asked for a recall in an oral interview. Whereas the multivariate analysis revealed an effect that was approaching significance, a univariate comparison of these two subgroups clearly showed no significant difference between mean effect sizes for accuracy rate (written: d = 0.44, p < .01, oral: d = 0.32, p < .05; Q_b = 0.23, *n.s.*). And finally, for the moderator *publication type*, eight studies were unpublished manuscripts and dissertations, whereas eleven studies have been published in academic journals. Whereas the multivariate analysis revealed an effect that was approaching significance, a univariate comparison of these two subgroups clearly showed no significant difference between mean effect sizes for accuracy rate (published: d = 0.45, p < .01, unpublished: d = 0.27 p = n.s.; Q_b = 0.60, *n.s.*).

As in the previous section, further to the overall effect of the SAI at time 2, two additional meta-analyses were run to separately analyse and compare summary effects for the two distinct control conditions 'SAI versus no initial recall' and 'SAI versus another form of initial recall' for the outcome accuracy rate.[9] Whereas the

[8] For this subgroup analysis, the four studies in which composite effect sizes across both control conditions were calculated, were coded as 'another type of recall', as this represents the more conservative approach.

[9] In addition to the information gained from the moderator analysis of the overall effect, these separate analyses provide a cleaner way of examining the SAI. They do not rely on calculating composite

former presents the current practice after a crime or incident and is therefore interesting from a practitioner's point of view, the latter challenges the SAI in that it may show an equal advantage of an even simpler recall instruction such as a written free recall and may thus prove the SAI unnecessary.

In the first meta-analysis, a total of 12 studies presented data on the impact of the Self-Administered Interview versus no initial recall on the accuracy rate at time 2. The summary effect size across these studies showed a highly significant, medium-sized effect, $d = 0.62$, $p = .000$ (95% CI: 0.35–0.88; $z = 4.60$). Seven out of 12 individual studies yielded a significant effect in favour of the SAI with the remaining five failing to present significant results, as shown in the forest plot in figure 2.13. This means that witnesses who filled in the SAI after having witnesses a crime produced testimony with a higher accuracy rate in an interview recall after a delay of 1-3 weeks compared to those who did not have any initial recall opportunity.

Model	Study name	Subgroup within study	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
	Af Hjelmsaeter 2012 (1)	T2, SAI vs. no	-0.125	0.176	0.031	-0.470	0.220	-0.709	0.478
	Paterson 2015	T2, SAI vs. no	0.196	0.278	0.077	-0.349	0.741	0.705	0.481
	Gabbert 2012 (1)	T2, SAI vs. no	0.243	0.317	0.101	-0.380	0.865	0.764	0.445
	Pfeil 2016 (3)	T2, SAI vs. no	0.485	0.293	0.086	-0.089	1.059	1.656	0.098
	Hope 2014 (2)	T2, SAI vs. no	0.610	0.323	0.105	-0.024	1.244	1.886	0.059
	Gittins 2015	T2, SAI vs. no	0.618	0.229	0.053	0.169	1.067	2.697	0.007
	Gabbert 2012 (2)	T2, SAI vs. no	0.664	0.261	0.068	0.153	1.176	2.546	0.011
	Mackay 2015	T2, SAI vs. no	0.784	0.282	0.080	0.231	1.338	2.777	0.005
	Colomb 2013	T2, SAI vs. no	0.840	0.244	0.060	0.361	1.318	3.439	0.001
	Gabbert 2009 (2)	T2, SAI vs. no	0.905	0.324	0.105	0.270	1.540	2.794	0.005
	Pfeil 2016 (4)	T2, SAI vs. no	0.928	0.304	0.092	0.333	1.524	3.056	0.002
	McPhee 2014	T2, SAI vs. no	1.620	0.356	0.126	0.923	2.317	4.555	0.000
Fixed			0.533	0.077	0.006	0.382	0.685	6.902	0.000
Random			0.615	0.134	0.018	0.353	0.877	4.603	0.000

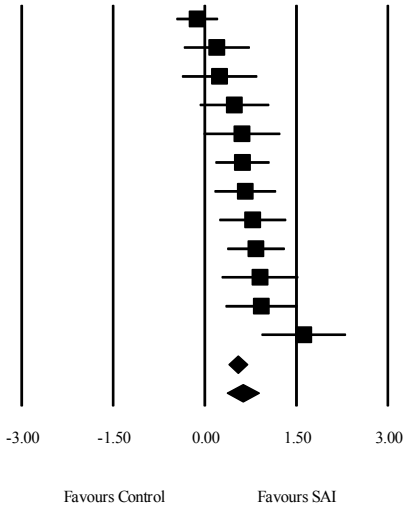


Figure 2.13. Forest plot for SAI vs. no initial recall at T2 for accuracy rate.

Publication bias analyses through Duval and Tweedie's trim-and-fill procedure revealed three imputed effect sizes in the relevant funnel plot (fig. 2.14.), resulting in a shift of the imputed summary effect size. Under the random effects model the point

---

effects, which is always a statistical compromise. Instead, they only include original effects found in the individual studies.

estimate and 95% confidence interval is $d = 0.62$ (95% CI: 0.35–0.88). Using trim-and-fill the imputed point estimate is $d = 0.47$ (95% CI: 0.21–0.73), suggesting a small overestimation of the current summary effect size. Furthermore, Rosenthal's fail-safe N test for this meta-analysis is 167 which means that there would need to be 14 missing 'null' studies for *every* observed study for this effect to be nullified. Again, it seems highly implausible that such a large number of studies have been missed in the literature search. Despite the small shift in the point estimate suggested by the trim-and-sill procedure, the overall results from the sensitivity analyses tests still suggest the absence of publication bias in this meta-analysis.

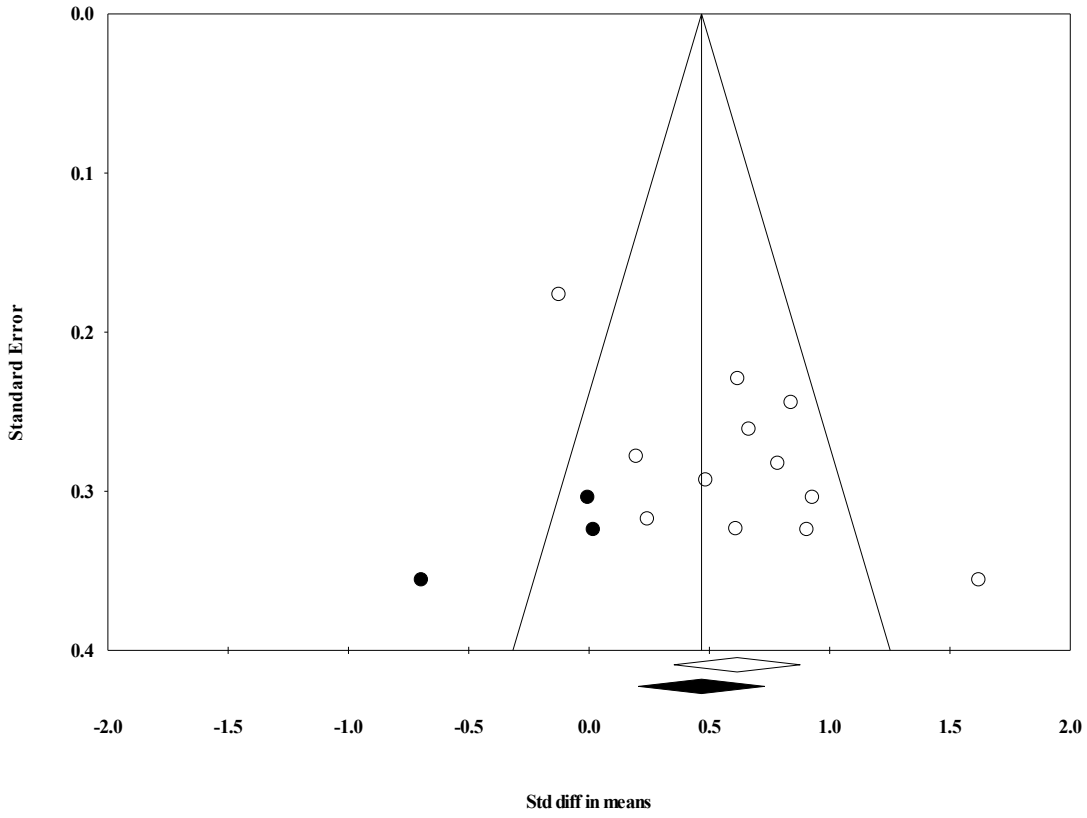
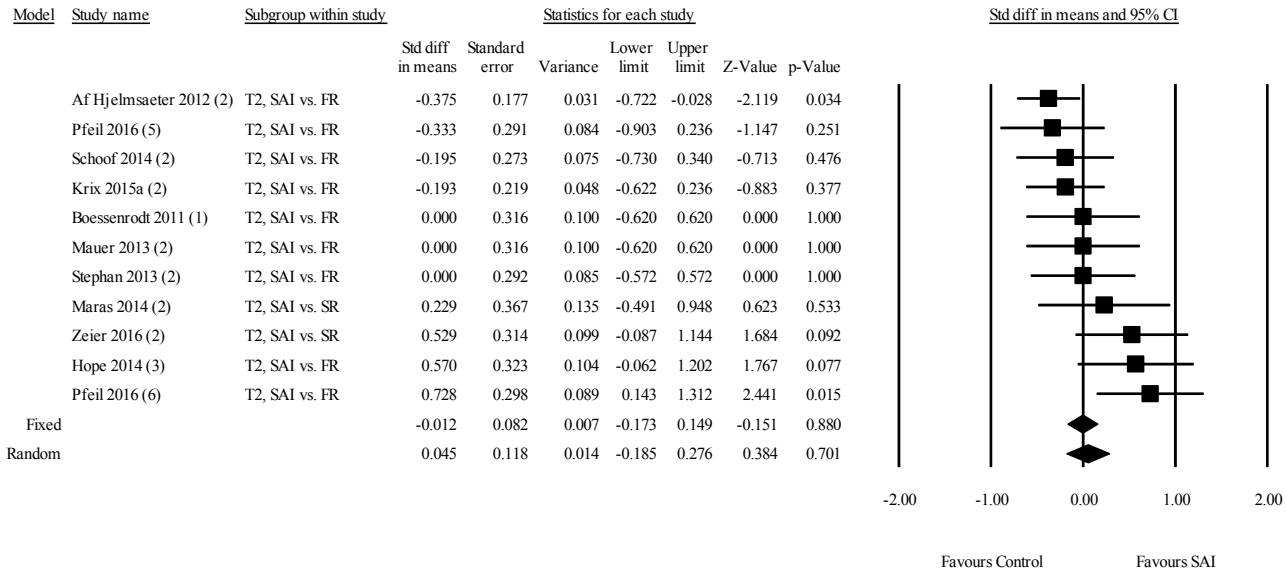


Figure 2.14. Funnel plot for SAI vs. no initial recall at T2 for accuracy rate.

In the second meta-analysis, a total of 11 studies presented data on the impact of the Self-Administered Interview versus another form of initial recall on the accuracy rate at time 2. The summary effect size across these studies did not show a significant effect, $d = 0.05$, p = .701 (95% CI: -0.19–0.28; z = 0.38). Whereas 9 out of the 11 individual studies yielded non-significant results, one study yielded a significant effect in favour of the control recall and one a significant effect in favour

of the SAI (see fig. 2.15.). So compared to having filled in another form of initial recall, witnesses who filled in the SAI, produced testimony at a similar accuracy rate in a delayed interview recall.



Model	Study name	Subgroup within study	Std diff in means	Standard error	Variance	Lower limit	Upper limit	Z-Value	p-Value
	Af Hjelmsaeter 2012 (2)	T2, SAI vs. FR	-0.375	0.177	0.031	-0.722	-0.028	-2.119	0.034
	Pfeil 2016 (5)	T2, SAI vs. FR	-0.333	0.291	0.084	-0.903	0.236	-1.147	0.251
	Schoof 2014 (2)	T2, SAI vs. FR	-0.195	0.273	0.075	-0.730	0.340	-0.713	0.476
	Krix 2015a (2)	T2, SAI vs. FR	-0.193	0.219	0.048	-0.622	0.236	-0.883	0.377
	Boessenrodt 2011 (1)	T2, SAI vs. FR	0.000	0.316	0.100	-0.620	0.620	0.000	1.000
	Mauer 2013 (2)	T2, SAI vs. FR	0.000	0.316	0.100	-0.620	0.620	0.000	1.000
	Stephan 2013 (2)	T2, SAI vs. FR	0.000	0.292	0.085	-0.572	0.572	0.000	1.000
	Maras 2014 (2)	T2, SAI vs. SR	0.229	0.367	0.135	-0.491	0.948	0.623	0.533
	Zeier 2016 (2)	T2, SAI vs. SR	0.529	0.314	0.099	-0.087	1.144	1.684	0.092
	Hope 2014 (3)	T2, SAI vs. FR	0.570	0.323	0.104	-0.062	1.202	1.767	0.077
	Pfeil 2016 (6)	T2, SAI vs. FR	0.728	0.298	0.089	0.143	1.312	2.441	0.015
Fixed			-0.012	0.082	0.007	-0.173	0.149	-0.151	0.880
Random			0.045	0.118	0.014	-0.185	0.276	0.384	0.701

Figure 2.15. Forest plot for SAI vs. other types of initial recall at T2 for accuracy rate.

Publication bias analyses through Duval and Tweedie's trim-and-fill procedure did not reveal any imputed effect size in the relevant funnel plot (fig. 2.16.). This means that under the random effects model the point estimate and 95% confidence interval [$d = 0.05$ (95% CI: -0.19–0.28)] remain the same using trim-and-fill[10], thus suggesting the absence of publication bias in this meta-analysis.

[10] Note that as there is no detected effect in this meta-analysis, Rosenthal's fail-safe N test becomes redundant as a test of publication bias.

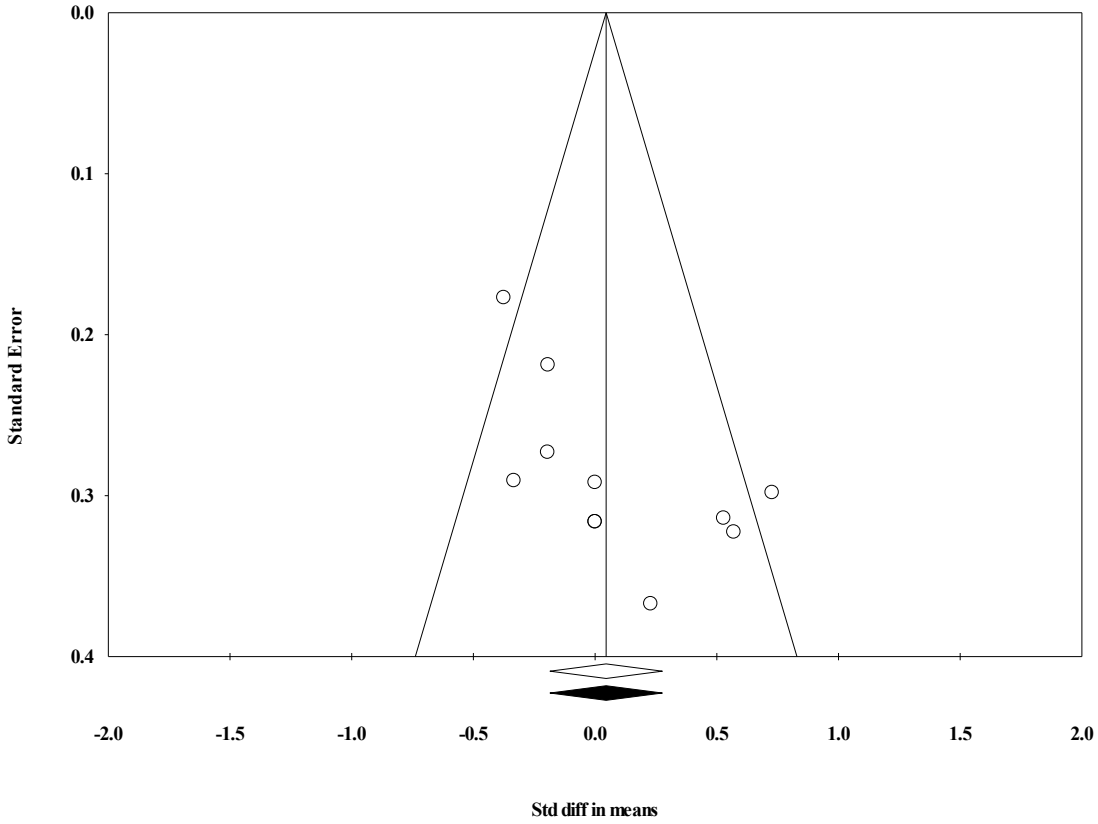


Figure 2.16. Funnel plot for SAI vs. other types of initial recall at T2 for accuracy rate.

2.3.3. Summary

In this results section, first, a narrative review of methodologies of studies included in the review on the effectiveness of the Self-Administered Interview was presented. In total, 22 studies matched the inclusion criteria, i.e. they were written in English or German, conducted a Self-Administered Interview, measured its effects compared with some form of control measurement either in an initial recall, as a impact measurement in a delayed interview, or both, and reported correct details and accuracy rate as dependent measures of recall. Both published and unpublished studies were included in the analysis to minimize publication bias. Studies or subgroups of studies were excluded from the final sample if they did not have a non-SAI control group, used clinical samples, did not report sufficient data, used a Cognitive Interview as control group, delayed the application of the SAI to longer than one hour after the event, or measured a transfer effect to a different scenario.

The final sample consisted of 38 experimental comparisons from 22 empirical studies representing 1712 interviewees. Of these studies, 14 were published manuscripts, 3 in preparation to be published and 5 unpublished theses or dissertations. Five studies compared the quality and quantity of information elicited in the SAI to that elicited in another form of initial recall (T1). Eight studies measured the impact of having an initial recall opportunity with the SAI on the quality and quantity of information given in a delayed interview (T2) and a further nine studies reported data on both an initial comparison and an impact measurement (T1 and T2). Participants were overwhelmingly students and young adults from the local community with only two studies having older adult samples and only one study on children. In 17 studies, participants were randomly assigned to recall conditions, with the remaining five studies having chosen a randomised block design and additionally controlling for gender or other criteria.

The chosen control conditions varied slightly in modality and type. Out of the total 14 studies that reported results of an initial comparison at T1, 12 compared the SAI to a free recall, whereas two chose a cued recall; and ten studies administered this control recall in written form, whereas four interviewed their participants orally. Out of the total 18 studies that reported results on an impact measurement at T2, 11 asked participants for a free recall during the delayed interview, and 7 employed a cued

recall. Exactly half of the studies administered the delayed interview in oral and in written form, respectively. Ten studies measured the impact of the SAI while comparing it to no initial recall and ten studies did so while comparing the SAI to another form of initial recall.[11] With fifteen out of 18 studies having a 1-week delay to the interview and only two and one studies having a 2-week and 3-week delay, respectively, the main body of evidence is based on a rather short retention phase compared to real-life scenarios. In addition to the standard memory outcome measurements, such as correct details and accuracy rate, eleven studies included measurements of suggestibility, either in the form of suggestive questions during a cued recall or in the form of introducing misleading post-event information before the interview and measuring susceptibility to those.

Subsequently to the narrative review, results on several meta-analyses were presented. It was chosen to compute two separate sets of meta-analyses for the two outcome measures, correctly recalled details and accuracy rate. Further, analyses were separated by T1 and T2 results, and within T2, by the kind of control group that was chosen, i.e. the SAI versus no initial recall and the SAI versus another form of initial recall. An overview of all meta-analyses results can be found in table 2.10. With regards to an increase in correct details, results indicate a strong benefit of the SAI both immediately after the witnessed crime and in a delayed recall after one to three weeks. In fact, both the immediate comparison and the impact measurement comparing the SAI to no initial recall (and thus representing current practice) yielded large summary effects (d = 1.20 and d = 0.92, respectively) comparable to that found for the Cognitive Interview (Memon, Meissner, & Fraser, 2010). Generally, as expected summary effects decreased from the initial recall at T1 (d = 1.20) to the delayed recall at T2 (d = 0.70 for SAI vs. any control condition) and also within T2 when comparing the SAI vs. no recall (d = 0.92) to SAI vs. another form of initial recall (d = 0.42). Nevertheless, these all represent substantial and significant increases in correctly recalled details with the Self-Administered Interview.

With regards to the accuracy rate, results were less straightforward. The immediate comparison at T1 actually yielded a small effect (d = -0.25) in favour of the *control condition*, suggesting that the SAI may lead to an increase in immediately recalled incorrect details compared to a simple initial recall task, whereas for the

[11] Note that one study could compare both, the SAI vs. no recall and the SAI vs. other recall.

impact measurement comparing the SAI to any type of control group at T2 there was still a medium-sized summary effect (d = 0.38) favouring the SAI. When looking closer at the impact measurement after a delay and comparing the impact of the SAI to no initial recall at T2 there was a substantial, medium-sized summary effect in favour of the SAI (d = 0.62). Again, this result shows a benefit over the current police practice of not having any formal initial recall for witnesses. However, there was no difference when comparing the SAI to another form of initial recall (d = 0.05) regarding the accuracy rate.

Table 2.10. Mean weighted effect sizes for measures of correct details and accuracy rate at T1 and T2.

Recall Measure Meta-Analytic Comparison	k	N	Weighted Mean *d*	*p*-value	95% CI	N_{FS}
Correct Details						
T1, SAI vs. other recall	15	947	1.20	< .001	(0.95 , 1.46)	1034
T2, SAI vs. any control	19	1048	0.70	< .001	(0.45 , 0.95)	546
T2, SAI vs. no recall	12	710	0.92	< .001	(0.66 , 1.18)	380
T2, SAI vs. other recall	11	603	0.42	< .01	(0.11 , 0.73)	52
Accuracy Rate						
T1, SAI vs. other recall	15	947	-0.25	< .01	(-0.39 , -0.11)	44
T2, SAI vs. any control	19	1048	0.38	< .01	(0.15 , 0.60)	143
T2, SAI vs. no recall	12	710	0.62	< .001	(0.35 , 0.88)	167
T2, SAI vs. other recall	11	603	0.05	n.s.	(-0.19 , 0.28)	n.a.

Note. Effect size data based on random effects model. 'Any control' includes a combination of data on SAI vs. no recall and SAI vs. other recall.

2.3.4. Discussion

In order to evaluate the impact of the Self-Administered Interview on eyewitness accounts, a systematic review of the literature was conducted, as well as several meta-analyses. With the SAI still being a relatively newly developed tool, no such systematic review or meta-analysis of studies exists as of yet. The results of the meta-analyses indicate a very large and significant increase in correct details in an initial recall, and still a medium-sized and substantial increase in correct details after a delay of 1 to 3 weeks with the SAI compared to any control group. Results for the accuracy rate were mixed, indicating a slight increase in incorrect details as well as correct details. The moderator analyses and more detailed, additional meta-analyses indicate a larger summary effect when comparing the SAI to no initial recall tool, as it is current police practice, rather than to a different kind of recall tool. Beyond this moderating variable, the summary effect was relatively unaffected by all other moderators (age group, laboratory, publication type, recall modality and recall type at T2).

However, there is a lack of sufficient data concerning several key areas. One is the applicability to vulnerable witnesses, including different age groups such as children and older adults, as well as witnesses with mental health issues. So far, only one study each explored the potential benefit of the SAI for children and mentally ill people, and only two studies looked at older adult witnesses. Current literature on the SAI also falls short on evidence regarding longer delays. In most studies, interviews followed a delay of one week, whereas only two studies imposed a two-weeks delay and only one study looked at a delay of three weeks. Finally, the experimental study that is being presented in this doctoral dissertation is the first one to explore the impact of the SAI on person identification performance. These issues will be discussed in more detail below, followed by an outline of policy implications.

2.3.4.1. Substantial Increase in Correct Details

The present meta-analysis suggests a substantial increase in correct details with the Self-Administered Interview compared to other recall tools, such as a written free recall, immediately after witnessing an event. The very large summary effect size of $d = 1.20$ is comparable to the benefit found for the Cognitive Interview (see

Memon et al., 2010). More important still, the increase transfers to a later witness interview and, though slightly smaller (d = 0.92 compared to not using any recall tool), is still substantial after a 2 to 3 week delay. A slight decrease of the effect with time is known and typical for memory recall and longer retention intervals.

Current police practice does not involve giving witnesses an early recall tool after the incident – apart from answering a few questions their first comprehensive recall attempt will be giving testimony at the police station after some time. So whereas giving witnesses the SAI to safeguard their memory yields a strong benefit over any comparison group at a later interview, it is noteworthy that this effect is even stronger when looking at the current police practice: Compared to no initial recall, the SAI increased the amount of correct details substantially in the witness interview. This is a promising result in light of the applicability of the SAI and calls for a broader adoption in everyday investigative interviewing.

2.3.4.2. Mixed Results for Accuracy Rate

Whereas the effect sizes for correct details were strong and remarkably consistent, results for accuracy rate have been mixed. For the immediate recall they suggest that witnesses not only produce more correct details in the Self-Administered Interview, but also more incorrect details. The respective meta-analysis even found a small effect favouring other types of initial recall (such as a simple free recall) over the SAI. However, it is important to note that this effect does *not* transfer to a later witness interview. More so, the effect even reverses and when interviewing witnesses after a 1 to 3 weeks delay, results show a medium-sized *increase* in the witness' accuracy rate. So whereas the SAI itself seems to elicit more incorrect details than other initial recall tools, witnesses do produce a more accurate testimony later on (d = 0.38 compared to any control group). This constitutes a huge benefit of the SAI.

When looking at the current police practice and comparing the SAI to no initial recall, there was a substantial positive effect for accuracy rate (d = 0.62). This means that with the SAI, witnesses not only remember more correct details later on, but they also provide a more accurate testimony. This constitutes the largest of all effects for accuracy rate and provides an easy way of enhancing recall without increasing workload for police officers. When comparing the SAI to other simple recall tools, there was no effect for accuracy rate. However, as the SAI was advantageous for correct details, it should still be favoured over other tools.

2.3.4.3. Generalizability to Vulnerable Witnesses

In this meta-analytic review, very few studies included samples from older adults and children or even mentally ill adults. This is probably due to the relative novelty of the SAI, and hinders any conclusions to be drawn regarding vulnerable witnesses as of yet. The self-administered nature of the SAI constitutes its probably biggest advantage regarding applicability within the general population. Witnesses can fill it in themselves and do not need further explanation or guidance. However, the lacking social interaction with a trained interviewer and support from them can be of disadvantage for vulnerable witnesses, such as children, older adults, traumatized victims or people with mental illnesses.

It has been established in the literature and interview guidelines, that rapport building is one of the key factors of successful interviewing (Heubrock & Palkies, 2008). It aims not only to explain the ground rules and expectations of the interview, but also to create an atmosphere of trust and confidence (St-Yves, 2013). This is particularly important when questioning witnesses about traumatic and sensitive information (Powell, Fisher, Wright, Brewer, & Williams, 2005). Furthermore, it has been shown that a supportive interviewing style, such as active listening (Shepherd & Griffiths, 2013) and showing empathy towards the witness (Fisher & Geiselman, 1992), results in a more correct free recall (Bull & Corran, 2002). Support and empathy were also found to lessen children's anxiety and enhance their general wellbeing (Davis & Bottoms, 2002), and to influence rape victims' decisions to go to court and face challenging criminal proceedings (Maddox, Lee, & Barker, 2011).

This social component is missing with the Self-Administered Interview. However, the SAI does not aim to replace a traditional investigative interview, it can rather be seen as an additional tool or step in the investigative process. The actual investigative interview follows later on, as it would in any investigation. Whereas the SAI itself cannot offer emotional support like an interviewer can, it remains to be seen whether it can still be useful for vulnerable witnesses and in which cases it would still be acceptable to distribute it. More research is needed to evaluate the usefulness and acceptance of the SAI within an older adult population and older children (who are able to read and write), and with mentally ill adults. The applicability may reach its limits with traumatized victims, who are most vulnerable and in need for emotional support during interviewing, and further, with anyone who cannot read or write.

2.3.4.4. The SAI in a Real-World Context

It is well established that recall diminishes over time and thus, that investigative interviews should be administered as soon as possible after the event. In the real world, interviews can however follow delays of several weeks or even months. This may be due to shortage of police staff, new developments in the investigative process or unforeseen circumstances. Literature on the Self-Administered Interview so far included only relatively short retention intervals between witnessing the event and remembering details about it in the investigative interview. Most studies used a 1-week delay, with only three studies using 2 or 3 weeks. The substantial benefits of the SAI were retained over these retention intervals. However, sufficient data on witnesses who are interviewed following more lengthy delays is currently lacking.

Furthermore, all but one study on the Self-Administered Interview examined its impact on eyewitness testimony. The study presented in the following chapter of this thesis is the first one to also examine its impact on person identification performance. Witnesses are oftentimes asked not only to remember the details of a crime, but also to identify the perpetrator from a lineup. Other investigative interviewing tools to aid witnesses in remembering more details, such as the Cognitive Interview, were shown to hinder subsequent lineup identifications. It was discussed that this was due to the so-called verbal overshadowing effect, meaning that producing a comprehensive verbal recall of the perpetrator's characteristics impacts negatively on identification accuracy (see Meissner & Brigham, 2001a for a meta-analysis). However, it was also shown that this effect can be overcome by leaving a delay between the interview and the lineup procedure, which was termed 'release from verbal overshadowing' (Finger & Pezdek, 1999). This resembles a real world scenario, as usually a witness is called in to view a lineup some time after they have given their testimony. A large-scale multi-lab study of the verbal overshadowing effect (Alogna et al., 2014) was able to replicate both findings, i.e. a robust verbal overshadowing effect, and also that the effect of providing a verbal description is reduced with a delay between the description and the identification task. With regards to the application of the SAI, a delay of at least several days is to be anticipated between filling in the SAI and being called in for a police lineup procedure. Verbal overshadowing can therefore be of minor concern, if at all, when examining the

impact of the SAI on person identification. In order to be able to further examine any potential benefit of the SAI on person identification performance, more studies need to address this important eyewitness task.

2.3.4.5. Limitations

In general, meta-analyses are known to be prone to a number of difficulties in their application, such as the 'file drawer problem'. As mentioned in the beginning of this chapter, the file drawer problem means that the outcome of a systematic review and meta-analysis can be biased if the original literature is contaminated by publication bias. For example, it has been found that studies with significant results are three times more likely to be published than those with non-significant findings (Dickersin, Chan, Chalmersx, Sacks, & Smith, 1987). Thus, if a meta-analysis only includes published studies it is difficult to account for publication bias and will likely reflect the bias of the original literature (and likely overestimate the summary effect). To address this problem, the present meta-analysis included not only published, but also a number of unpublished studies (e.g. Colomb & Gabbert, 2013; Schoof et al., 2014). Furthermore, a number of analyses were conducted to check for a bias, including funnel plots, Rosenthal's fail-safe N and the trim-and-fill procedure. The results did not indicate a publication bias for any of the meta-analyses in this chapter.

Another common limitation of meta-analyses can be the 'garbage in, garbage out' thesis. This refers to the problem that the inclusion of many low-quality studies may lead to an erroneous meta-analysis. The best way to address this problem is to have inclusion criteria that are based on the quality of studies. For this set of meta-analyses, only studies with adequate control groups were included, whereas studies without non-SAI control groups were excluded (e.g. Boon, 2012; Curtis, 2013).

A limitation more specific to the present meta-analyses could be the inclusion of studies in English and German only, which leaves the risk of several missed studies that were potentially published in other languages. However, the Self-Administered Interview is copyrighted and a proper name and should thus be mentioned in text in any language. Since 'SAI' or 'Self-Administered Interview' was used as the search criterion, and the searched databases also contain journal articles in languages other than English and German, search results should have listed studies published in other languages as well. No such studies were found. Moreover, in email conversations

with the team who developed the SAI, no major research groups working in other languages could be identified that they were aware of.

2.3.4.6. Policy Implications and Way Forward

One of the main goals of compiling research literature is to inform policy, and eventually improve the application of methods and knowledge. While it is often challenging to reach practitioners and change methods and structures that have long been existing, it is critical to identify ways of improving investigative interviewing and thereby reduce miscarriages of justice. Policy recommendations can be based on what Malpass et al. (2008) refer to as the Best Practices (BP) model or the Well Established Knowledge (WEK) model. According to the BP model, conclusions are based on the best evidence available at the time. The downside of this model is that it does not contain any criteria for assessing the strength of the empirical base. This may lead to rapidly changing recommendations as new work is being published. It may also be more difficult to revise and apply best practices if scientists are not routinely involved in implementing new procedures in a specific application environment. The WEK model on the other hand assumes that a) the issue of interest has been extensively studied, b) studies forming the research base are of high scientific standard, and c) findings are well established. This model is more rigorous, requiring that policy formation wait until the research base is well established.

Given that the Self-Administered Interview is a relatively novel development, recommendations deriving from this meta-analytic review can only be based on best practice. In order to advance from best practice to well-established knowledge, a broader and more extensive literature base is needed. Nevertheless, this meta-analytic review presents a first and crucial step in gathering and evaluating the current research on the Self-Administered Interview. The positive effects of the SAI have been well replicated and are robust. There is also general agreement in the scientific community thus far as to the effectiveness of the SAI. The evidence-base around the SAI is growing and future studies should focus on real world application.

In the field, the SAI is easy-to-apply for police officers. Compared to the Cognitive Interview, it does not require any special training, nor does it take more manpower or time than would be required in any case. This suggests that the SAI may be used more readily and incorporated in everyday investigative interviewing than the

CI. Despite the clear benefits of applying a CI, it is not being used at all in the vast majority (83%) of British interviews (Malpass et al., 2008). Subsequent research showed that police officers often find it too complex or feel that the additional time and resources it requires do not warrant application to most of their interviews, which relate to less serious crimes (Clarke & Milne, 2001). It was also highlighted that police officers find the CI demanding, as it does not only take longer to administer, but also involves instructing witnesses in the use of several sophisticated techniques (Dando, Wilcock, & Milne, 2009). Moreover, there are no pre-determined questions, but the police officer has to actively listen and subsequently base their questions on what the witness has provided. This requires social skills in communicating effectively; even more so when vulnerable witnesses are involved.

So far, only two studies provide a direct comparison of the SAI and the CI, with mixed results indicating no clear advantage of using one or the other. Whereas Gabbert et al. (2009) did not find a significant difference in the number of accurate details provided in the SAI and the CI, Kemp et al. (2016) found participants in the CI to remember more accurate details in total than in the SAI. On the other hand, Kemp et al. (2016) found a higher accuracy rate for the SAI, whereas Gabbert et al. (2009) found a lower accuracy rate in the SAI, compared to the CI. Ultimately, studies are needed that combine the use of the SAI with a CI to see whether the application of both is superior to that of either single interview. Given this range of hurdles and difficulties in applying the Cognitive Interview in the field and on the other hand, the similar benefit of applying a Self-Administered Interview, it seems only logical to a) advance research on the SAI and b) promote its application throughout police forces.

Finally, as an impulse for future research, it would be interesting to meta-analyse the impact of the SAI on witness suggestibility. Several studies gathered in the present meta-analytic review have also included a form of suggestibility measurement in addition to the standard memory outcomes. With more and more studies forthcoming on the SAI, it will be worthwhile to conduct a meta-analysis specifically answering the questions of whether or not the SAI can not only preserve witness memory, but also protect witnesses from suggestive influences in the process.

3. THE SAI FOR OLDER WITNESSES – AN EMPIRICAL STUDY

In previous chapters, literature on eyewitness performance in general and on older witnesses in particular was reviewed. Then a closer look at the Self-Administered Interview as a means to enhance performance was taken and all studies known to the author reviewed and meta-analysed. Based on the evidence reviewed, the now following experimental study aims to investigate the effectiveness of the SAI for older adults through the following hypotheses and experimental design.

3.1. Hypotheses

The *testimony part* of this study aims to examine the extent to which older and young adults perform better (i.e. give more detailed and accurate accounts) when first being presented with the SAI vs. a written free recall at time 1 and then interviewed at time 2 as opposed to not having had an initial recall option. It also aims to examine the extent to which older and young adults are less susceptible to misinformation when first being presented the SAI vs. a written free recall (time 1) and then interviewed with leading questions (time 2) as opposed to not having had an initial recall option. In addition, age group differences between older adults and young adults will be investigated.

It is well established, that a detailed and good quality retrieval of details from episodic memory increases the likelihood of recalling these details in subsequent attempts (Marsh, Tversky, & Hutson, 2005; Pansky & Nemets, 2012; Shaw, Bjork, & Handal, 1995). According to the spreading activation theory of memory (also called associative networks of memory, J. R. Anderson, 1983), memory is seen as a network consisting of nodes that represent concepts and share associative links. The quality of the initial coding determines the strength of the associative link, and subsequent retrieval then further strengthens these links across episodic memory. A more extensive retrieval attempt, as it is achieved with the Self-Administered Interview compared to a written free recall, leads to increased activation levels of the encoded details and also the associations between details, and thus supports subsequent retrieval of details. Based on this theory, the SAI with its specific instructions and

prompts will facilitate recall immediately after the event due to an increase in activation levels of the encoded details and their associations within the associative memory network. A simple written free recall task will not lead to increased activation levels and therefore not achieve accounts that are as detailed as with the SAI.

Based on what is known from research on perception, information processing and memory decline in older adults (Balota et al., 2000) as well as from findings specifically on eyewitness accounts (Coxon & Valentine, 1997; Yarmey, 2001), it is hypothesized that younger adults will outperform older ones with regards to the quantity and quality of witness accounts. Moreover, it was established by Craik (1986) and Craik and McDowd (1987) that older adults benefit from more environmental support in a memory task compared to young adults, i.e. they perform better on a recognition compared to a recall task. In other words, older adults seem to be more dependent on external support and will therefore benefit more from a structured tool such as the SAI compared to young adults.

Previous studies on the SAI have further found a reduced susceptibility to suggestions in a later investigative interview (Gabbert et al., 2012). Moreover, in studies by Gittins et al. (2015) and McPhee et al. (2014) SAI participants showed greater rejection of misinformation compared to no-recall participants. According to Loftus (2005), having a strong original memory helps witnesses to detect and ultimately reject discrepant or contradictory information. It is therefore hypothesized that having an initial and comprehensive recall opportunity with the SAI will strengthen the original memory trace and help to inoculate against suggestions.

Hypothesis 1.1: *Initial recall hypothesis*

Participants will give more detailed and accurate accounts in the Self-Administered Interview compared to the written free recall at time 1.

Hypothesis 1.2: *Interview transfer hypothesis*

Participants will give more detailed and accurate accounts in an eyewitness interview at time 2 when having first completed the Self-Administered Interview at time 1 compared to a written free recall or no initial recall at time 1.

Hypothesis 1.3: *Suggestibility hypothesis*

Participants are less susceptible to suggestions at time 2 when first being presented the Self-Administered Interview at time 1 than when being presented with a written free recall or no initial recall at time 1.

Hypothesis 1.4: *Age group hypothesis*

Older adults give less detailed and less accurate accounts both in the initial recall options (SAI and wFR) at time 1 and in the subsequent eyewitness interview at time 2 compared to young adults. Older adults are more susceptible to suggestions in the eyewitness interview compared to young adults.

Hypothesis 1.5: *Interaction hypothesis*

The benefit of the SAI for older adults exceeds that of young adults.

The *person identification part* of this study is designed to explore the influence of the SAI on person identification performance, i.e. it examines the extent to which older and young adults perform better in a lineup at time 2 (i.e. more correct identifications in a target present lineup and more correct rejections in a target absent lineup) when first being presented with the SAI vs. a written free recall at time 1 as opposed to not having had an initial recall option. In addition, age group differences between older adults and young adults will be investigated, as well as the confidence-accuracy relationship.

In addition to the considerations above, literature has consistently shown that older adult witnesses generally perform more poorly compared to young adults in person identification tasks (Havard & Memon, 2009; Searcy et al., 1999; Wilcock et al., 2007). To account for these age-related differences, Searcy et al. (1999) put forward the context recollection theory. It accounts for the problem of recognizing that a face is familiar without being able to identify it. This is due to familiarity being based solely on the characteristics of the face, but not the context in which it was encountered. As such, filler faces in lineups are based on perceived familiarity with the perpetrator mugshot. Identification on the other hand relies on recollection of context, which requires detailed information on the face and the relationship between face and context. According to the context recollection theory, older adults have greater problems with recollecting contextual information required for identification, but not with the perceptual processes required for familiarity. As a result, they rely on familiarity in lineup procedures, i.e. a face standing out as being familiar to them,

more so than young adults, and therefore make more mistakes in choosing a filler face from a lineup. Memon et al. (2002) were able to support this theory in their study.

Based on this theory, the SAI with its specific instructions and prompts will strengthen the original memory and thus help put the face of the perpetrator in the right context for later retrieval in a lineup task. Furthermore, if an early and high-quality recall as it is facilitated in the SAI can strengthen the original memory and thus provide a stronger context for the source of where a face was encoded, it should decrease false identifications for older adults.

With regard to post-identification confidence, research thus far has concluded that there is no simple confidence-accuracy relationship. A meta-analysis by Sporer et al. (1995) found an overall low confidence-accuracy correlation. However, for choosers the confidence–accuracy correlation was reliably and consistently higher than for non-choosers. Specifically regarding older adults, Scogin, Calhoon & D'Errico (1994) found no significant correlation between line-up accuracy and the confidence of participants. This finding was replicated in several other studies (Memon et al., 2003; Memon et al., 2002; Wilcock et al., 2007).

Hypothesis 2.1: *Identification hypothesis*

> Participants perform better in the person identification task at time 2 when first being presented the Self-Administered Interview at time 1 compared to a written free recall or no initial recall at time 1.

Hypothesis 2.2: *Age-group hypothesis*

> Older adults perform worse in the person identification task compared to young adults.

Hypothesis 2.3: *Interaction hypothesis*

> The benefit of the SAI for older adults exceeds that of young adults.

Hypothesis 2.4: *Confidence-accuracy hypothesis*

> Post-identification confidence is not related to accuracy of identification.

Hypothesis 2.5: *SAI confidence-accuracy hypothesis*

> The Self-Administered Interview does not impact the confidence-accuracy relationship.

3.2. Methodology

In order to investigate the effectiveness of the SAI on testimony, suggestibility and person identification performance in older adults and thus to test the above hypotheses, the following experiment was designed. In the methods section, first the participant recruitment will be described, followed by a description of the proposed design and the used apparatus and materials. Finally, the procedure of the experiment and coding of participants' accounts will be presented.

3.2.1. Participants

As the first step power analyses were conducted to estimate the necessary sample size. Previous studies on the effect of the SAI on eyewitness testimony showed large effect sizes (η^2 = .49 for the number of correct details compared to a written free recall in Gabbert et al., 2009, and d = 1.69 for the number of correct details in a delayed recall SAI vs. no initial recall in Gabbert et al., 2012, each study having group sizes of 18-21 participants). Therefore this experiment was also designed to detect large effect sizes. Two power analyses were conducted with the software G*Power (see Faul, Erdfelder, Lang, & Buchner, 2007), one for the testimony part and one for the person identification part. For the *testimony part,* specifications were a) F-test, b) large effect size f = 0.4, c) α = 0.05, d) power of 0.95, e) df = 5, f) number of groups = 6. The analysis revealed a minimum sample size of 130 participants. For the *person identification part*, specifications were a) χ^2-test b) medium to large effect size w = 0.4, c) α = 0.05, d) power of 0.95, e) df = 6. The analysis revealed a minimum sample size of 131 participants.

In accordance with results of the power analyses, a total of 144 participants[12], 72 of which aged 60 years and over (M = 69.00, SD = 5.73) and 72 aged 18-30 (M = 24.26, SD = 2.68), were recruited on a voluntary basis. The age range was chosen according to literature, where an age of 60 years is commonly defined as the cut-off age for "older adults" (see e.g. Bartlett & Memon, 2007). Among older adults, 24 were male and 48 female participants. Similarly, of the 72 young adults, 28 were male and 44 female. Older adults were cognitively normally functioning members of the

[12] A post-hoc power analysis with G*Power to compute the actual achieved power with N = 144 participants revealed a power of 0.97 for both types of analyses.

community and recruited through advertisements and announcements in local community groups, e.g. advertisement in the weekly bulletin of the University of the Third Age, Cambridge, word-of-mouth through participants, and by contacting participants who had taken part in previous eyewitness studies at the Institute of Criminology. All participants had normal or corrected-to-normal vision. Furthermore, older adult participants were screened for cognitive functioning with a short version of the Minimental State Examination (MMSE, Folstein, Folstein, & McHugh, 1975), but no participant had to be excluded based on this result.

3.2.2. Design

For the *testimony and suggestibility part* of this study, a number of variables such as total details, total correct/incorrect/confabulated details, total person/action/object/setting details, susceptibility to suggestibility questions etc. were the dependent variables. Independent variables were recall condition in the first session (SAI vs. written free recall (wFR) vs. no recall) and age-group (young vs. older adults). This resulted in a between-subjects 3 (recall condition) x 2 (age-group) factorial design with 6 different groups (SAI_older, SAI_young, wFR_older, wFR_young, no_older and no_young) and 24 participants in each group. Several ANOVAs, each for the different dependant variables, were conducted.

For the *person identification part* of this study, lineup performance (correct vs. incorrect) was the dependent variable, while recall condition in the first session (SAI vs. written free recall vs. no recall), age-group (young vs. older adults) and target presence in the second session (target present vs. target absent) were independent variables. This resulted in a between-subjects 3 (recall condition) x 2 (age-group) x 2 (target presence) factorial design with 12 groups and 12 participants in each of these groups. As these are four categorical variables in total, a loglinear analysis was conducted, followed up with chi-square analyses for TP and TA lineups, which is the standard statistical procedure for this type of research (see Field, 2009).[13]

[13] It is important to note that all analyses in the person identification part were based on $k = 6$ groups, resulting in the smallest cell size being $n = 24$ (instead of $n = 12$). This means that analyses focused on either target presence differences (combining younger and older adults) or on age differences (combining target absent and present lineups).

Participants were randomly assigned the experimental conditions in order to maximize internal validity and overcome most threats to it (Field & Hole, 2008) and thus to make sure that the variation in the measured outcome (e.g. the amount of correct details) is due to the variation of the intervention (e.g. having filled in the SAI, see Farrington & Welsh, 2005). This means it was assigned randomly which recall condition (SAI vs. wFR vs. no recall) and which lineup (TP or TA lineup) would be presented to participants by using random sampling in Microsoft Office Excel 2007.

3.2.3. Apparatus and Materials

A standard laptop (13" MacBook) was used to show the film and to present the lineup. The film was presented with the same standard media player (Apple QuickTime player) throughout data collection at maximum volume. The lineup was presented using PsychoPy2 (Peirce, 2007), a free software used to design and present experiments. The screen was adjustable for the participants to exclude any light reflection and the approximate distance from the screen was 60cm.

3.2.3.1. Film of a Staged Crime

The video for this study was shot in a small town in England, all actors were taking part voluntarily and consenting to the use of the film for research purposes. The scene depicts an attempted mugging of an older lady, followed by a successful mugging of a younger lady. There is no violence in the film. The whole event lasted two minute and was filmed in digital format. Participants viewed the perpetrator for approximately 30 seconds altogether, 10 seconds of which were a close-up of his face. Some verbal exchange between the perpetrator and the second victim was audible and there were a few other people in the background.

3.2.3.2. Self-Administered Interview

The SAI tool (appendix B) is presented in booklet-format and comprises five sections. Each section contains information and instructions to aid and prompt witnesses and thus elicit both recall and answers to specific questions for a witnessed event, irrespective of the kind of crime. The tool has been repeatedly piloted for clarity, ease of understanding and simplicity. (Gabbert et al., 2009)

Section 1 contains general information about the SAI, such as to follow the instructions and to complete it in the given sequential order. Section 2 comprises a context reinstatement and free recall part, asking the witness to report everything they can remember without guessing. The third section focuses on person description, such as detailed information about the perpetrator's appearance. In the fourth section, witnesses are asked to generate a sketch of the scene including their own position in relation to others. Finally, section 5 focuses on information that witnesses may have not found relevant so far, such as viewing conditions or descriptions of other witnesses. (Gabbert et al., 2009)

3.2.3.3. Written Free Recall Form

The Written Free Recall Form (wFR, appendix C) was designed to allow comparisons of any added value that the SAI might have over a simple recall instruction in the same recall modality. This approach has been used in other studies on the SAI (e.g. Gabbert et al., 2009). Participants were instructed to write down everything they can remember in as much detail as possible, such as details about persons and actions, however unimportant it may seem. They were also instructed to not guess about details they cannot remember and that they are free to use bullet points or full sentences.

3.2.3.4. Eyewitness Interview Form

The Eyewitness Interview Form (appendix D) consists of two parts, a free recall part and a question part, which is generally recommended for interviewing witnesses of all ages (Greuel, 2008) and especially for older adults (Bornstein, 1995). In accordance with Bornstein's recommendations, in the free recall the interviewer asked the participants to think back to the film and tell everything about it they can remember in as much detail as possible. After they provided this free recall account, the interviewer prompted if they could remember anything else or would like to add any detail and then closed the free recall part. The question part of the interview contained different types of questions: specific questions asking for person details, action details, object details and surroundings details, as well as yes/no questions. The question part was designed similar to the structure of the SAI, such as that it contains three major parts – starting with the people directly involved in the crime, first eliciting a description of these people and then clarifying and asking specific

questions about them and closing with yes/no questions; then moving on to potential witnesses, again first person description, then specific questions and yes/no questions; and in the end asking about any other people in the scene in the same manner. The interview closed with a question about the length of the observed film. In total, the interview form consisted of 46 questions (including detailed person description tables). In order to investigate the impact of the SAI on interrogative *suggestibility,* the question part was interspersed with 18 leading questions, both in the form of specific questions and yes/no questions and referring to central (victims and perpetrator) as well as peripheral details (bystanders). Examples of leading questions were "How did the second victim react after having fallen down to the ground?" (when in fact she did not fall), "What was the first thing the policeman did when he arrived?" (when in fact there was no policeman) and "What colour was the child's bike?" (when in fact the child did not have a bike).

3.2.3.5. Lineup

The digital photos for the lineup were obtained from friends and acquaintances, all of whom were voluntarily giving their photos and consenting to its use for research purposes. In total, 15 photographs were obtained. The selection of the photo material was loosely based on the approach adopted by Wilcock and Bull (2010). Specifically, the 15 photographs of possible foils were presented to twenty middle-aged adults (M = 38 years), without them having seen the film beforehand. They were presented with two photos at a time, i.e. the perpetrator on one side and a comparison photo next to it. They were then asked to rate these photographs regarding their similarity to the perpetrator on a seven-level Likert scale ranging from 'not at all similar' (1) to 'highly similar' (7). The six photos with the best fit were used in the lineup, with the target replacement being the foil that was rated most similar to the perpetrator. Sessions were done individually and no time limits were imposed.

Participants were shown either a TP simultaneous lineup or a TA simultaneous lineup. The TP lineup contained the perpetrator and five foils, whereas the TA lineup contained the target replacement and five foils, which is a common approach (e.g. Wilcock & Bull, 2010). The lineup photos comprised six 7cm x 9cm coloured head shots of the face arranged in a 3 x 2 array. Target and foil positions followed a Latin square design, i.e. each item is arranged so that it occurs only once per row and once

per column (Field & Hole, 2008). Thus the arrangement and randomization of photo order was constrained by the requirement that each photo was presented in each of the six positions, comparable to the way the numbers are arranged in a "SUDOKU" puzzle. This resulted in six different arrangements for the TP lineup and six further arrangements for the TA lineup. Twelve participants each were exposed to the same arrangement, taking into account that arrangements were balanced across study conditions.

Testing lineup-fairness: Photographs. To test that the lineup was not biased towards the perpetrator, the proportions technique originally described by Doob & Kirshenbaum (1973) was employed. This is considered the most widely used measure of lineup bias (Brigham & Pfeiffer, 1994) and compares the frequency with which mock witnesses select the perpetrator with the frequency of choices expected by chance alone. For a six-photo lineup, as in the present study, each photo would have a pure chance probability of being picked of .17. If the perpetrator is selected with greater than chance frequency, which is measured by a significant *z* test for proportions, the lineup is biased towards him. In order to test this, 24 young adults, who were not shown the film, but given a written description of the perpetrator, viewed a lineup (twelve participants viewed a TP and twelve a TA lineup) and were asked to choose who they thought the perpetrator was. For the TP lineup, two out of twelve mock witnesses (.17) identified the perpetrator, which indicated no significant bias towards the perpetrator, *z* for proportions = 0.03, $p > .05$. In addition all of the foils were chosen between one and three times. For the TA lineup the target replacement was chosen three times (.25) and all of the foils were chosen between one and three times, which again indicated no significant bias towards the target replacement, *z* for proportions = 0.74, $p > .05$. After the data was collected, incorrect identifications from TA lineups were further examined, showing 52.9% of participants incorrectly identified a foil, whereas 47.1% falsely identified the target replacement. There was no significant effect of optimality on the proportion of participants making foil or target replacement identifications χ^2 (1, n = 17) = 0.14, $p > .05$. These efforts indicated that the resulting choice of photos led to an unbiased lineup in terms of photos.

Testing lineup-fairness: Order effects. After the data was collected, statistical analyses were conducted to check for possible order effects in the lineup performance. Two separate chi-square analyses were employed for TP and TA lineups,

respectively. These revealed no significant effects, $\chi^2(5) = 5.46$, $p > .05$ for TP lineups and $\chi^2(5) = 10.45$, $p > .05$ for TA lineups, which means that participants did not perform significantly better or worse in a specific lineup arrangement compared to the other ones.

3.2.3.6. Lineup Answer Sheet

The Lineup Answer Sheet (appendix E) contained the instructions given to participants, boxes they should tick to identify the perpetrator or indicate he is not present, and a scale to indicate their post-identification confidence. In the instruction for the first task, participants were asked to look at the photographs on the laptop in front of them and told that they find boxes representing each of these photos on the sheet. They were then asked to press the spacebar as soon as they had made their decision and then to tick the box which corresponded to the photo showing the perpetrator from the film. Participants were also provided with the option to not choose any of the photographs in the lineup and instructed that, as with a real police lineup, the perpetrator may or may not be present and that there was a box which said 'none of them'. The boxes on the answer sheet were arranged in the same 3 x 2 array as the lineup on the laptop, and were numbered according to the photographs. Next to the lower row of photographs was the additional box labelled 'none of them'. Participants were instructed that there were no time limits imposed. In the instruction for the second task, participants were asked to indicate on a 7-level Likert scale ranging from 'not at all confident' (1) to 'highly confident' (7) how confident they were about the choice they have made.

3.2.3.7. Minimental State Exam

Older adults were screened for dementia, using a short version of the MMSE (Folstein et al., 1975). Designed to grade the cognitive state of a clinical patient, the full version comprises two sections, the first one to be completed by verbally answering questions, and the second section to be completed by following verbal and written commands. The maximum sore is 30; the test is not timed and usually takes 5-10 minutes to administer. The psychometric characteristics of the MMSE are satisfactory, with an internal consistency, measured by Cronbach's Alpha, of .64 in a sample without cognitive impairments and .81 in a Alzheimer's disease sample (Tombaugh, McDowell, Kristjansson, & Hubley, 1996). The differential sensitivity,

i.e. the ability to discriminate between individuals with and without Alzheimer's disease is .91 (Tombaugh et al., 1996).

In this study a short version (appendix F) was used, comprising a score of 16 out of 30 and employing tasks in orientation (5 points), registration (3 points), attention and calculation (5 points) and recall (3 points). Short forms of the MMSE can be as accurate as the original version (Schultz-Larsen, Lomholt, & Kreiner, 2007), and are especially attractive for research purposes. Given that participants in this study had to remember the correct date and time, read the consent form, write down their name and the date, and fill in several questionnaires, an abbreviated version of the MMSE was considered sufficient for this purpose.

3.2.3.8. Demographic Questionnaire

In the demographic questionnaire (appendix G) demographic as well as background information about each participant were recorded. This included gender, age, ethnic group, general health (ranging from excellent to poor), reported alertness at the time of testing (ranging from completely awake to very tired), time of getting up, frequency of sports/activities (ranging from every day to not regularly) and other possible influencing factors on the alertness, such as naps during the day, caffeinated drinks during the last hours before testing and medication that may have an enhancing or impairing side effect. Furthermore occupation before retirement and highest educational degree were recorded. This information gives an overview of factors that could interfere with the performance in the study, as for instance shown for physical activity (Colcombe & Kramer, 2003; Menec, 2003; Weuve et al., 2004) and caffeinated drinks (Ryan, Hatfield, & Hofstetter, 2002).

3.2.3.9. Filler Tasks

The following questionnaires were used as filler tasks in this study, i.e. they were used to create intervals between watching the film of a staged crime and filling in the SAI in the first session, or between the eyewitness interview and the lineup task in the second session.

3.2.3.9.1. Positive and Negative Effect Schedule.

The PANAS (appendix H) is a 20-item self-report measure of positive and negative affect developed by Watson, Clark & Tellegen (1988). Negative affect (NA)

and positive affect (PA) reflect dispositional dimensions, with high-NA reflecting subjective distress and unpleasurable engagement, and low NA the absence of these feelings. In contrary, PA represents the extent to which an individual experiences pleasurable engagement with the environment. Thus, emotions such as enthusiasm and alertness are indicative of high PA, whilst lethargy and sadness characterize low PA Crawford and Henry (Crawford & Henry, 2004). The reliabilities (internal consistencies) of the PANAS scales were estimated using Crobach's Alpha and showed good results of .89 for the PA scale, and .85 for the NA scale. Moreover, the influence of demographic variables such as gender, occupation and age on the PANAS scores was shown to be neglectable and need not be taken into consideration when interpreting an individual's scores (Crawford & Henry, 2004).

3.2.3.9.2. Center for Epidemiologic Studies Depression Scale Revised.

The CESD-R (appendix I) is a 20-item self-report measurement of depressive symptoms developed by Radloff (1977) and revised by Eaton, Muntaner, Smith, Tien and Ybarra (2004). It is one of the most widely used instruments in psychiatric epidemiology and measures symptoms of depression in nine different groups: Sadness, loss of interest, appetite, sleep, thinking/concentration, guilt, fatigue, movement and suicidal ideation. Possible scores range from 0 (for those who say 'not at all or less than one day' to all 20 questions) and 60 (for those who say '5-7 days' or 'nearly every day for 2 weeks' for all 20 questions). A possible major depressive episode is indicated by a score of at least 16 plus symptoms in at least 2 additional DSM symptom groups (Eaton et al., 2004). In this study, the CESD-R was used as a filler task and to exclude participants with high scores and low retrieval performance, as depression is known to have detrimental effects on memory performance (Burt, Zembar, & Niederehe, 1995). No participants had to be excluded on the basis of their CESD-R score.

3.2.3.9.3. Morningess-Eveningness Questionnaire.

The MEQ (appendix J) by Horne and Östberg (1976) is a further development of the Swedish language Morningness-Eveningness Questionnaire by Östberg (1973) and was modified for the British context by including additional questions and the omission of others (Horne & Östberg, 1976). It is designed to elicit a person's optimal time of day and allocate them to "morning", "evening" or "intermediate"

types, for which it was also used in this study. The English language version consists of 19 items, 14 of which with a four choice selection of answers, indicating definite morning type, moderate morning type, moderate evening type and definite evening type, and five using a time scale. Possible scores range from 16 to 86, with 16-30 indicating the definite evening type, 31-41 the moderate evening type, 42-58 the intermediate type, 59-69 the moderate morning type and 70-86 the definite morning type. The psychometric characteristics of the MEQ show a good reliability (Buela Casal, Caballo, & García Cueto, 1990; Smith, Reilly, & Midkiff, 1989), specifically the internal consistency, measured by Cronbach's Alpha, is .82. The questionnaire has also been shown to correlate with circadian variations in oral temperature, sleep-wake behaviour and periods of perceived alertness and performance (Buela Casal et al., 1990; Horne & Östberg, 1977; Mecacci & Zani, 1983; Smith et al., 1989). Interesting to note is that normative studies have shown the majority of young adults to be evening types (ca. 45%) or intermediate types (ca. 50%), whereas older adults are almost exclusively morning types (ca. 75%) or intermediate types (ca. 25%) according to the MEQ (May, Hasher, & Stoltzfus, 1993).

3.2.3.9.4. SUDOKU.

A choice of four SUDOKU puzzles, two each of low and medium difficulty, was used in both sessions of this study as a filler task (appendix K). This was done to ensure that all participants, regardless of their speed in filling in the other questionnaires, spend the same amount of time before moving on to the final task of the respective session. SUDOKU was chosen as literature shows that there is no detrimental effect on subsequent face recognition compared to cryptic crosswords for instance (Lewis, 2006). Participants were assured that the SUDOKU would merely be used as a filler task and not be analysed. They were also asked whether they had completed a SUDOKU before and to indicate on a seven-level Likert scale how difficult this task was for them.

3.2.4. Procedure

At recruitment people were told that, if participating, they would take part in an experimental eyewitness study. A brief overview of the study was given, including that participants would have to watch a short film of a staged crime, fill in

questionnaires and then come back after one week to report everything they can remember, answer questions about the film and also be shown pictures to identify the perpetrator from the film.

All participants were tested individually. For the young adults, sessions took place in the Institute of Criminology, University of Cambridge. Older adults were tested either in the above or in their own homes. The possibility of home visits was introduced due to the limited mobility of some of the older adults. It was ensured that sessions were free of disturbances and that visual conditions were as good and as constant as possible. Out of 72 older adults, six were visited at their homes.

In the first session, all participants were first given the consent form to read and sign. Afterwards, older adults completed the short version MMSE to ensure normal cognitive functioning, before watching the film of the staged crime. Young adults watched the film right away. It was emphasized that the crime in the film was not real and that all persons in it were volunteer actors. All participants then filled in the first session's filler tasks, i.e. the demographic questionnaire, PANAS, CESD-R and SUDOKU. Participants did not have to complete the SUDOKU, it was rather used as a means to ensure each participant spends the same amount of time with filler tasks (15 minutes). Depending on which study condition the participant was in, they then either could go home ("no" condition), or were asked to fill in either the written free recall form ("wFR" condition) or the Self-Administered Interview ("SAI" condition) and then could go home. They were reminded about the follow-up session one week later at the end of the first session and then again the day before the second appointment via email.

In the second session after a one-week delay, all participants were asked to think back to the film they watched last week and report everything they can remember in as much detail as possible. Subsequently, they were also asked a series of specific questions about the film, including a number of suggestive questions. Both these parts were audio-taped. No time limits were imposed. All participants were then given the second session's filler tasks, i.e. the MEQ, PANAS and SUDOKU (again to ensure an equal interval of 15 minutes for each participant). Then the participants were given the answer sheet for the lineup task to read the instruction, and when ready, presented with either a target present or target absent simultaneous photo lineup on the laptop. They were instructed that the perpetrator may or may not be present in the lineup and that they could take as much time as they wanted to.

Furthermore, the experimenter was sitting opposite the participant and laptop and was not able to see the participant's choice, thereby ensuring that any possible nonverbal cueing behaviour of the researcher did not influence the participant. Finally, the participants received a small reward (chocolate) for completing the study and were thoroughly debriefed, especially on the overall purpose of the study, the necessity of the leading questions during the interview and on the difficulty of the lineup task.

3.2.5. Coding

The coding instructions (appendix L) were constructed in accordance with Wright and Holliday's (2007a) coding instructions, and further refined and adapted for the purpose of this study.

3.2.5.1. SAI, Written Free Recall and Interview Free Recall Part

For the coding of all recall transcripts, i.e. the Self-Administered Interview, the written Free Recall and the Free Recall during the interview, the same scheme and instructions were used, following Wright and Holliday's (2007a) example. They classified details into distinct types, as there were: Person details (any information about an actors' appearance and clothing), action details (any information about what someone was doing), object details (any information about objects) and setting details (any information about the surrounding or setting of the film). Furthermore details were categorized by their accuracy, resulting in correct details (i.e. details that are present in the film), incorrect details (i.e. details that are discrepant from the film) and confabulated details (i.e. details that were not present in the film or that did not happen). In addition, the accuracy rate was calculated (correct details divided by total details), thus providing eight detail scores in total. Examining the recall transcripts on such qualitative level provides a more precise account of the individual performance than a single global score and allows to isolate any strengths and weaknesses for different types of information (A. M. Wright & Holliday, 2007a).

Any information or details that were either too vague (e.g. "he was average height"), subjective (e.g. "he was good-looking") or attributed/assumed (e.g. "perhaps he wanted to phone the police") were not scored. Uncertain responses were also not scored if the participant was really not sure about his statement, rather than just saying 'I think'. If the participant confused details about different persons, the option that

gave the most correct details was chosen. However, exact repetitions of the same information were not scored again.

A second rater coded a random sample of 30 free recall accounts (from 15 young adults and 15 older adults). Inter-rater reliability yielded a significant agreement for the overall amount of details, $r = .97$, $p < .001$, total correct, $r = .98$, $p < .001$, incorrect, $r = .93$, $p < .01$ and confabulated details, $r = .89$, $p < .01$, as well as with regards to the different detail categories total person, $r = .99$, $p < .001$, action, $r = .91$, $p < .001$, object, $r = .95$, $p < .001$, and setting details $r = .90$, $p < .01$.

3.2.5.2. Interview Question Part

Since the question part consisted of different types of questions, a multi-variable coding scheme was developed. The person descriptions and specific questions were coded in the same categorized details as the recall part. The yes/no questions were scored as correct or incorrect and added up, and the suggestive questions were scored referring to their degree of assent und then added up. If the answer to one suggestibility question indicated a rejection ("He didn't have a weapon") the score for this one was 0, if the participant said that he did not know or was unsure about the answer, this was coded 1 and if the participant clearly assented to the suggestion ("He had a knife") this was coded 2. Added up, each participant was given a suggestibility score with higher scores indicating greater susceptibility to suggestions. Finally, the duration of the film was recorded in minutes and was averaged if the participant gave a time span.

3.2.6. Ethical Considerations

An ethics approval was obtained from the Institute of Criminology's ethics committee prior to data collection. Before study sessions started, every participant was given a thorough information sheet about the purpose and nature of the study to give their written informed consent (appendix M). The content was in addition verbally explained to make sure that every participant did fully understand the purpose of the study, the tasks that they would be asked to complete should they take part, and what would happen to their data. It was ensured that every participant understood that the crime in the video was staged by volunteer actors and that none of it was real. It was also emphasized that participation is voluntary and participants

could terminate the session at any time of the session without having to give a reason. Furthermore, they were told that data would be analysed in group form only so that it would not be possible to trace anybody's individual answers. All information given by participants was anonymised and handled confidentially. Participants were given a copy of the consent form with contact details of the experimenter should they have any questions later on.

3.2. Results

In this results chapter, first demographic information will be reported. Then, the witness testimony results will be presented, differentiated by details produced in the first session (SAI and written free recall) and the second session (interview free recall and question part), followed by analyses on interrogative suggestibility. Finally, results of the person identification task will be reported, including analyses on the confidence-accuracy relationship. A brief summary and discussion on each of these subsections complete this chapter.

3.2.1. Demographic Data

First, demographic and background information were analysed by age group and by recall condition (SAI, wFR, no) to examine data for pre-existing differences between groups, such as differences in educational background or perceived health and alertness.

Table 3.1. contains the demographic variables and background characteristics analysed by age group. Group differences were investigated using the appropriate statistical significance testing (independent t-test for metric data, Mann-Whitney test for ordinal data and Chi-square test for nominal data, see Field, 2009). Young adults were an ethnically more diverse group (80% described themselves as being White, 10% as being Asian, 3% Black, and 7% other) compared to older adults (100% White), $\chi^2 = 14.29$, $p < .001$. All young adults attended university, whereas among older adults, 72% reported to hold a university degree, $\chi^2 = 23.23$, $p < .001$. Older adults rated their alertness at the time of witnessing the staged crime as significantly more awake (*Mdn* = 5.00, completely awake) than young adults (*Mdn* = 4.00, fairly awake), $U = 1683.00$, $p < .001$. When looking at the MEQ, older adults on average were moderate morning types ($M = 59.14$, $SD = 8.46$), whereas young adults tended to be intermediate circadian rhythm types ($M = 48.85$, $SD = 9.66$), $t = 6.80$, $p < .001$.[14] When asked about symptoms indicating depression on the CESD-R, young adults scored significantly higher ($M = 9.97$, $SD = 7.44$) than older ones ($M = 6.07$,

[14] This finding is in accordance with literature, showing that an individual's circadian rhythm tends to shift towards morningness with age (May et al., 1993).

SD = 4.91), $t = -3.72$, $p < .001$.[15] Young adults found solving the SUDOKU filler task on average easier (*Mdn* = 5.00, slightly difficult) than older adults (*Mdn* = 7.00, highly difficult), $U = 1866.00$, $p < .01$.

Table 3.1. Demographic variables and background characteristics by age group.

Variable	Age group		Significance test	
	Older adults	Young adults		
N male/female[b]	24/48	28/44	$\chi^2 = 0.48$	
Ethnicity (% white)[b]	100	82	$\chi^2 = 14.29$	***
N university yes/no[b]	52/20	72/0	$\chi^2 = 23.23$	***
Health rating	*Mdn* = 3.00	*Mdn* = 3.50	$U = 2386.00$	
N sports yes/no[b]	68/4	65/7	$\chi^2 = 0.89$	
Sports frequency	*Mdn* = 6.00	*Mdn* = 6.00	$U = 2072.00$	
Alertness rating	*Mdn* = 5.00	*Mdn* = 4.00	$U = 1683.00$	***
MEQ[a]	*M* = 59.14 (*SD* = 8.46)	*M* = 48.85 (*SD* = 9.66)	$t = 6.80$	***
CESD-R[a]	*M* = 6.07 (*SD* = 4.91)	*M* = 9.97 (*SD* = 7.44)	$t = -3.72$	***
SUDOKU	*Mdn* = 7.00	*Mdn* = 5.00	$U = 1866.00$	**

**$p < .01$. ***$p < .001$.
[a]$df = 142$. [b]$df = 1$.

Demographic variables and background characteristics analysed by recall condition are shown in table 3.2. Group differences between the three groups were again investigated using the appropriate statistical significance testing (one-way ANOVA for metric data, Kruskal-Wallis test for ordinal data and Chi-square test for nominal data, see Field, 2009). None of the variables tested revealed significant differences between participants who were assigned to the SAI, the written free recall or no initial recall group and thus showed that randomization appears to have been successful in establishing equal groups.

[15] However, both mean scores are well below the 16-points cut-off that would indicate a depression.

Table 3.2. Demographic variables and background characteristics by recall condition.

Variable	Recall condition			Significance test
	SAI	wFR	no	
Age	*M* = 47.35 (*SD* = 23.18)	*M* = 47.21 (*SD* = 23.64)	*M* = 45.33 (*SD* = 22.31)	*F* = 0.12, df = 2, 141
N male/female	17/31	19/29	16/32	χ^2 = 0.42, df = 2
Ethnicity (% white)	93.8	87.5	91.7	χ^2 = 1.18, df = 2
N home visit yes/no	1/47	4/44	1/47	χ^2 = 3.13, df = 2
N university yes/no	40/8	45/3	39/9	χ^2 = 3.60, df = 2
Health rating	*Mdn* = 3.00	*Mdn* = 3.00	*Mdn* = 3.00	*H* = 0.17, df = 2
N sports yes/no	43/5	45/3	45/3	χ^2 = 0.79, df = 2
Sports frequency	*Mdn* = 6.00	*Mdn* = 6.00	*Mdn* = 6.00	*H* = 2.94, df = 2
Alertness rating	*Mdn* = 5.00	*Mdn* = 5.00	*Mdn* = 4.00	*H* = 1.78, df = 2
MMSE[a]	*M* = 15.50 (*SD* = 0.78)	*M* = 15.50 (*SD* = 0.66)	*M* = 15.38 (*SD* = 0.65)	*F* = 0.26, df = 2, 69
MEQ	*M* = 52.56 (*SD* = 11.36)	*M* = 55.27 (*SD* = 9.22)	*M* = 54.15 (*SD* = 10.60)	*F* = 0.82, df = 2, 141
CESD-R	*M* = 9.31 (*SD* = 7.74)	*M* = 6.81 (*SD* = 4.37)	*M* = 7.94 (*SD* = 7.04)	*F* = 1.76, df = 2, 141
SUDOKU	*Mdn* = 5.00	*Mdn* = 6.00	*Mdn* = 5.00	*H* = 0.93, df = 2

Note. All significance test values are non-significant.

[a]older adults only.

3.2.2. Testimony

This first section begins by presenting the results of the first session, i.e. a comparison of details produced in the Self-Administered Interview and in the written free recall. This is followed by analyses of the second session, i.e. the testimony given in the witness interview, separated by free recall, question part and answers to the specific questions including suggestive questions. In the end, some further analyses will be illustrated and a brief summary completes this section.

On several occasions throughout the testimony results section, a number of multiple comparison tests were conducted, i.e. t-Tests and ANOVAs on a number of recall variables such as total details, person details, accurate details and so forth. Whenever multiple comparisons are conducted, this can increase the likelihood of false positive errors, which means finding a significant difference when in fact, there is none (McDonald, 2009). The heightened risk of false positive results can be corrected for statistically, for example with the Bonferroni correction or Benjamini-Hochberg procedure (Armstrong, 2014; Benjamini & Hochberg, 1995). However, the Bonferroni procedure is conservative and may lead to a very high number of false negatives (McDonald, 2009) and therefore missed significant results.

According to Armstrong (2014) Bonferroni corrections should therefore only be used if it is imperative to avoid a false positive error (e.g. in medical studies) or if a large number of tests are carried out without pre-planned hypotheses. Neither is the case in this study. Furthermore, an underlying assumption for both procedures is that the comparisons are independent of each other. Typical for witness memory studies is however, that they compare multiple variables between groups, and that those variables are likely to be correlated with each other within groups. For example, the variable *total person details* is a compound variable consisting of *correct person details*, *incorrect person details*, and *confabulated person details*, meaning that they are correlated with each other. For all the reasons outlined above, no Bonferroni or Benjamini-Hochberg corrections were conducted. It may also be noted that such corrections are not usually applied in this type of research (Gabbert et al., 2009; Hope et al., 2014; Krix et al., 2016; A. M. Wright & Holliday, 2007a).

3.2.2.1. Initial Recall: Self-Administered Interview and Written Free Recall

In the following, analyses for the first session are presented, this means a comparison of details produced in the Self-Administered Interview and in the written free recall tool. To reiterate, the hypotheses that are relevant to this part of the analyses are:

Initial recall hypothesis

Participants will give more detailed and accurate accounts in the Self-Administered Interview compared to the written free recall at time 1.

Age group hypothesis

Older adults give less detailed and less accurate accounts in the initial recall options (SAI and wFR) at time 1 compared to young adults.

Interaction hypothesis

The benefit of the SAI for older adults exceeds that of young adults.

Table 3.3. comprises means, standard deviations, t-tests and corresponding effect sizes for the details produced in the SAI and the written free recall (wFR) in the first session. Comparisons are shown for the total sample as well as separated for the older and young adults. Starting with the total sample, participants remembered significantly more *correct details* in the SAI ($M = 95.40$, $SD = 45.82$) compared to the wFR ($M = 70.69$, $SD = 21.27$), $t = 3.39$, $p < .001$, which represented a medium-sized effect $d = 0.69$. However, the amount of *incorrect details* also increased in the SAI ($M = 7.75$, $SD = 6.33$) compared to the wFR ($M = 4.98$, $SD = 2.81$), $t = 2.77$, $p < .01$, which again represented a medium-sized effect $d = 0.57$. Further, participants produced significantly more *person details* in the SAI ($M = 53.58$, $SD = 24.56$) compared to the wFR ($M = 30.65$, $SD = 9.26$), $t = 6.05$, $p < .001$, which represented a large effect $d = 1.24$.

When separating the sample by age-group, older adults gave significantly more *person details* in the SAI ($M = 44.04$, $SD = 15.73$) than in the wFR ($M = 30.25$, $SD = 8.53$), $t = 3.78$, $p < .001$, which constitutes a large effect $d = 1.09$. None of the other variables differed significantly between the SAI and the wFR. In the young adults subsample, participants remembered significantly more *correct details* in the SAI ($M = 112.88$, $SD = 54.39$) compared to the wFR ($M = 70.13$, $SD = 24.29$),

$t = 3.52$, $p < .001$, which represented a large effect $d = 1.01$. However, the amount of *incorrect details* also increased in the SAI ($M = 8.08$, $SD = 8.08$) compared to the wFR ($M = 3.63$, $SD = 2.32$), $t = 2.06$, $p < .05$, representing a medium-sized effect $d = 0.75$. Young adults also produced significantly more *person details* ($M = 63.13$, $SD = 28.21$) and *action details* ($M = 28.96$, $SD = 16.61$) in the SAI compared to those who gave a written free recall ($M = 31.04$, $SD = 10.11$ and $M = 19.08$, $SD = 7.55$), $t = 5.25$, $p < .001$ and $t = 2.65$, $p < .05$, respectively. This represented a large effect for the person-related details $d = 1.51$ and a medium-sized effect for the difference in action-related details $d = 0.77$.

Table 3.3. Group differences at T1 by recall condition, for total sample and subsamples

Group	SAI		wFR				
Variable	M	SD	M	SD	t		d
Total sample (df = 94)							
Accuracy rate	0.92	0.04	0.92	0.04	-0.57		0.00
Correct details	95.40	45.82	70.69	21.27	3.39	***	0.69
Incorrect details	7.75	6.33	4.98	2.81	2.77	**	0.57
Confabulated details	0.85	1.47	0.90	1.43	-0.14		0.03
Person details	53.58	24.56	30.65	9.26	6.05	***	1.24
Action details	24.21	13.64	19.98	6.67	1.93		0.39
Object details	8.85	5.50	8.19	3.76	0.69		0.14
Setting details	17.35	11.17	17.75	7.62	-0.20		0.04
Older adults (df = 46)							
Accuracy rate	0.90	0.04	0.90	0.04	0.13		0.00
Correct details	77.92	26.34	71.25	18.28	1.02		0.29
Incorrect details	7.42	4.04	6.33	2.63	1.10		0.32
Confabulated details	1.08	1.59	1.46	1.59	-0.82		0.24
Person details	44.04	15.73	30.25	8.53	3.78	***	1.09
Action details	19.46	7.58	20.88	5.67	-0.73		0.21
Object details	7.29	3.09	8.21	3.53	-0.96		0.28
Setting details	15.63	8.45	19.71	6.48	-1.88		0.54
Young adults (df = 46)							
Accuracy rate	0.93	0.03	0.95	0.03	-1.20		0.67
Correct details	112.88	54.39	70.13	24.29	3.52	***	1.01
Incorrect details	8.08	8.08	3.63	2.32	2.06	*	0.75
Confabulated details	0.63	1.34	0.33	1.01	0.85		0.25
Person details	63.13	28.21	31.04	10.11	5.25	***	1.51
Action details	28.96	16.61	19.08	7.55	2.65	*	0.77
Object details	10.42	6.88	8.17	4.06	1.38		0.40
Setting details	19.08	13.32	15.79	8.28	1.03		0.30

*$p < .05$, **$p < .01$, ***$p < .001$.

To further test for age group effects and interaction effects of age group and initial recall, several factorial 2 × 2 between-subjects ANOVAs were conducted with initial recall condition (SAI and wFR only) and age group as fixed factors and each of the nine total variables as dependent variable, respectively. In order to explore the assumptions of an ANOVA, the Kolmogorov-Smirnov test was performed to test for normal distribution of the total variables and Levene's test was done to test for homogeneity of variances. The Kolmogorov-Smirnov test was significant for most variables, indicating they were not normally distributed. Levene's test was significant for some variables, indicating that variances were significantly different in these cases. Since data was positively skewed a logarithm transformation was implemented. However, this transformation did not improve data in terms of normality. As Field (2009) points out, transforming data is not always useful since it alters the relationship between the original variables in the model and thus limits the interpretation of the data and in some cases even hinders the accuracy of the F-statistic. Furthermore, when group sizes are equal, as in this study, the F-statistic is said to be fairly robust to violations of normality as well as homogeneity of variances and can still perform accurately (Field, 2009). Consideration was also given to excluding outliers using the $M \pm 2SD$ rule (Schendera, 2007). However, this would have unequal group sizes as a result, which in turn is detrimental to the robustness of the F-statistic. Furthermore, in this study the data points and thus any outliers can be assumed to be *legitimate* in the sense that they derive from random sampling of the intended population rather than e.g. data entry errors or intentional mis-reporting (Osborne & Overbay, 2004). Therefore, data is more likely to be representative of that population as a whole if outliers are not removed (Osborne & Overbay, 2004). Due to these considerations, analyses used the original, non-transformed data. The results are shown in table 3.4.

Table 3.4. Factorial ANOVAs at T1.

	Recall condition (R)			Age group (A)			Interaction R x A		
Variable	F[a]		Part. η^2	F[a]		Part. η^2	F[a]		Part. η^2
Accuracy rate	0.41		.00	25.86	***	.22	0.72		.01
Correct details	12.81	***	.12	6.00	*	.06	6.83	**	.07
Incorrect details	7.84	**	.08	1.06		.01	2.91		.03
Confab. details	0.02		.00	7.66	**	.08	1.36		.02
Person details	41.47	***	.31	7.78	**	.08	6.59	*	.07
Action details	4.06	*	.04	3.38		.04	7.24	**	.07
Object details	0.50		.01	2.66		.03	2.81		.03
Setting details	0.04		.00	0.01		.00	3.63		.04

[a]*df* = 1, 92.
*$p < .05$. **$p < .01$. ***$p < .001$.

Main effects for recall condition were found on the variables *correct details*, F (1, 92) = 12.81, $p < .001$, *person details*, F (1, 92) = 41,47, $p < .001$, and *action details* details, F (1, 92) = 4.06, $p < .05$, all of which were in the predicted direction with more details produced in the SAI (M = 95.40, SD = 45.82; M = 53.58, SD = 24.56; M = 24.21, SD = 13.64, respectively) compared to the wFR (M = 70.69, SD = 21.27; M = 30.65, SD = 9.26; M = 19.98, SD = 6.67, respectively). However, the amount of *incorrect details* also increased in the SAI (M = 7.75, SD = 6.33) compared to the wFR (M = 4.98, SD = 2.81), F (1, 92) = 7.84, $p < .01$, which is contrary to the prediction. Effect sizes are indicated by partial eta squared and are between $\eta^2_{partial} = .04$ and $\eta^2_{partial} = .31$ for significant recall condition effects, which represent small to large effects (Cohen, 1988).

Main effects for age group were found on the variables *accuracy rate*, F (1, 92) = 25.86, $p < .001$, *correct details*, F (1, 92) = 6.00, $p < .05$, *confabulated details, F* (1, 92) = 7.66, $p < .01$, and *person details*, F (1, 92) = 7.78, $p < .01$. In line with the hypothesis, young adults provided more *correct* (M = 91.50, SD = 6.94) and *person details* (M = 47.08, SD = 26.50), furthermore had a higher *accuracy rate* (M = 0.94, SD = 0.03) and *confabulated* less details (M = 0.48, SD = 1.19) than older adults (M = 74.58, SD = 22.68; M = 37.15, SD = 14.33; M = 0.90, SD = 0.04; M = 1.27, SD = 1.58, respectively). Effect sizes range between $\eta^2_{partial} = .06$ and $\eta^2_{partial} = .22$ for significant age group effects, which again represent small to large effects (Cohen, 1988).

There were interaction effects between the recall condition and age of the participants on the number of *correct*, F (1, 92) = 6.83, $p < .01$, *person*

F (1, 92) = 6.59, $p < .05$, and *action details F* (1, 92) = 7.24, $p < .01$. This indicates that young and older adults were affected differently by the recall condition. Specifically, for the number of *correct details*, older and young adults performed similarly on the written free recall (*M* = 71.25, *SD* = 18.28 and *M* = 70.13, *SD* = 24.29, respectively); whereas young adults provided significantly more correct details (*M* = 112.87, *SD* = 54.39) in the SAI compared to older adults (*M* = 77.92, *SD* = 26.34; see figure 3.1.), which again represents a medium-sized effect, $\eta^2_{partial} = .07$.

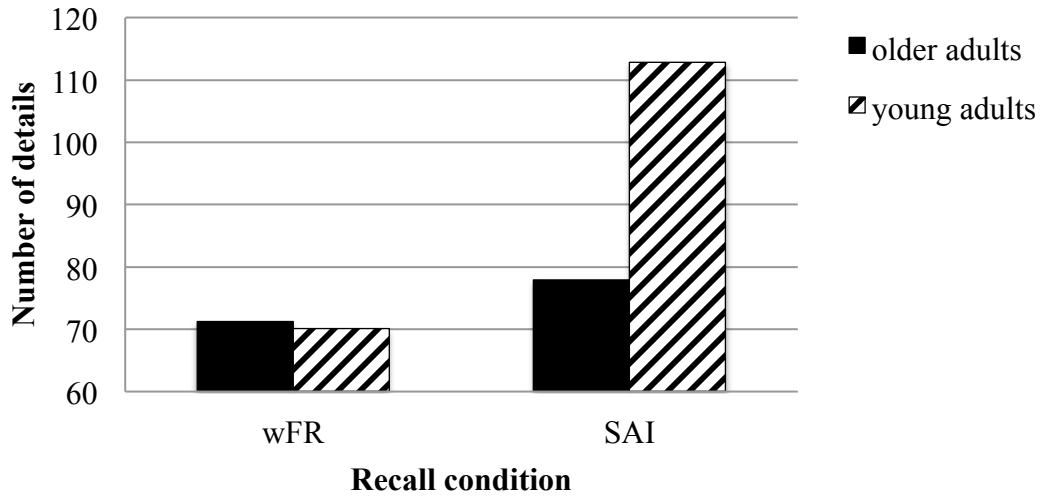


Figure 3.1. Number of correct details as a function of recall condition and age group at T1.

The same pattern repeats again for the variables *person details* (wFR older adults *M* = 30.25, *SD* = 8.53; young adults *M* = 31.04, *SD* = 10.11; SAI older adults *M* = 44.04, *SD* = 15.73; young adults *M* = 63.12, *SD* = 28.21) and *action details* (wFR older adults *M* = 20.88, *SD* = 5.67; young adults *M* = 19.08, *SD* = 7.55; SAI older adults *M* = 19.46, *SD* = 7.58; young adults *M* = 28.96, *SD* = 16.61), which is displayed in figures 3.2. and 3.3., respectively. Both represent medium-sized effects, $\eta^2_{partial} = .07$.

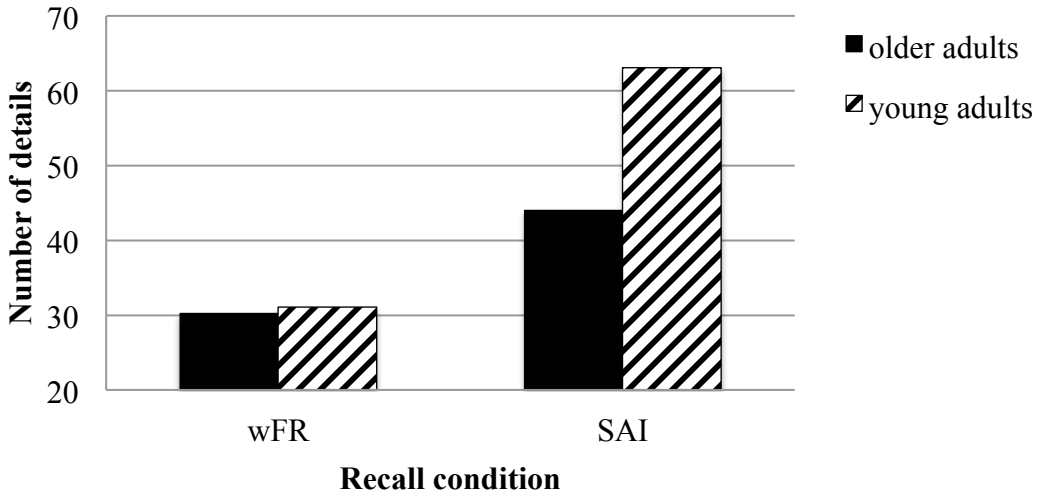


Figure 3.2. Number of person details as a function of recall condition and age group at T1.

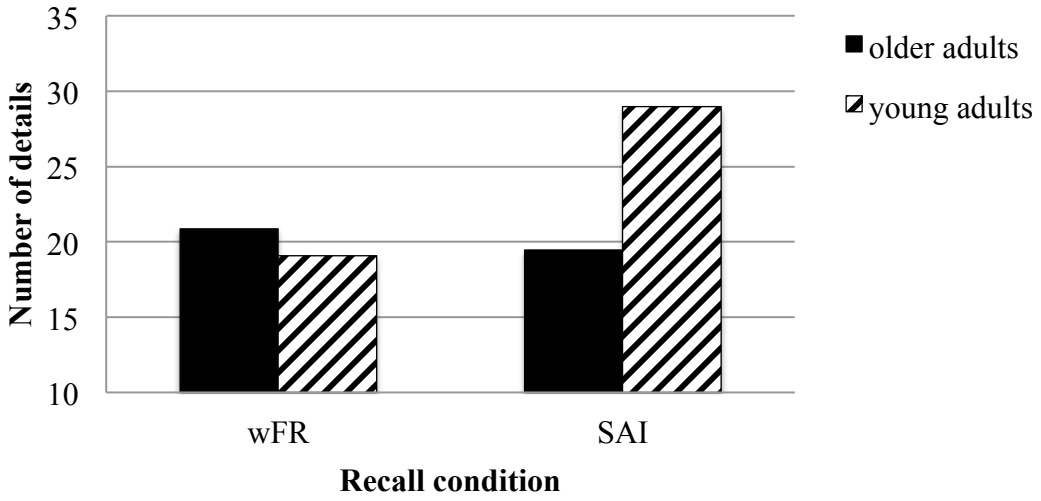


Figure 3.3. Number of action details as a function of recall condition and age group at T1.

Thus, the *initial recall hypothesis* was mainly supported. Participants of either age group produced a more detailed and accurate account in the Self-Administered Interview compared to the simple written free recall at time 1. However, completing the SAI also increased the number of incorrect details compared to the wFR. The *age group hypothesis* was also supported, as older adults gave less detailed and less accurate accounts in the SAI and the wFR at time 1 compared to young adults. The *interaction hypothesis* was not supported – results did not show a greater benefit of the SAI for older adults than for young adults.

3.2.2.2. Interview Free Recall Part

In this section, analyses for the free recall part of the second session are presented, this means a comparison of details produced in the free recall of the witness interview after one week, depending on recall condition in the first session. A complete table of means and standard deviations across all conditions and variables for the free recall part can be found in appendix N. To reiterate, the hypotheses that are relevant to this part of the analyses are:

Interview transfer hypothesis

Participants will give more detailed and accurate accounts in an eyewitness interview at time 2 when having first completed the Self-Administered Interview at time 1 compared to a written free recall or no initial recall at time 1.

Age group hypothesis

Older adults give less detailed and less accurate accounts in the eyewitness interview at time 2 compared to young adults. Older adults are more susceptible to suggestions in the eyewitness interview compared to young adults.

Interaction hypothesis

The benefit of the SAI for older adults exceeds that of young adults.

From a practitioner's point of view it would be interesting to see a comparison of participants who had completed the SAI ("SAI" condition) and participants who did not have an initial recall ("no" condition, which is the current standard in police investigation). Thus, to start off with, t-tests were conducted for the details produced in the interview free recall for the SAI vs. the no condition. Table 3.5. comprises means, standard deviations, t-tests and corresponding effect sizes for the total details. Results are shown for the total sample and for the older and young adults separately.

Starting with the total sample, participants who had completed the SAI a week earlier produced significantly more *correct details* ($M = 97.19$, $SD = 31.45$), as well as more *person* ($M = 47.60$, $SD = 19.01$), *action* ($M = 27.63$, $SD = 7.79$) and *setting details* ($M = 22.19$, $SD = 9.13$) in the delayed free recall compared to those who had not had an initial recall opportunity ($M = 73.25$, $SD = 31.72$; $M = 35.27$, $SD = 16.54$; $M = 22.81$, $SD = 9.86$; and $M = 16.60$, $SD = 9.65$), $t = 3.71$, $p < .001$; $t = 3.39$, $p < .001$; $t = 2.65$, $p < .01$; and $t = 2.91$, $p < .01$, respectively. Effect sizes indicated by

Cohens' d range between $d = 0.54$ and $d = 0.76$ and thus represent medium-sized effects. The SAI group also had a significantly higher *accuracy rate* ($M = 0.93$, $SD = 0.04$) compared to the no group ($M = 0.90$, $SD = 0.06$), $t = 3.01$, $p < .01$, which also represents a medium-sized effect. It may thus be noted, that participants overall produced significantly more correct details without increasing the number of incorrect or confabulated details in the SAI versus the no condition.

When looking at the older adults, the only significant difference was found for the *accuracy rate*. Older adults who had completed the SAI a week earlier reached a significantly higher accuracy rate ($M = 0.92$, $SD = 0.03$) than those who had not had any initial recall ($M = 0.87$, $SD = 0.07$), $t = 3.30$, $p < .001$. This represented a large effect $d = 0.95$. Among the young adults subsample, participants who had completed the SAI a week earlier produced significantly more *correct* ($M = 97.21$, $SD = 34.64$), *person* ($M = 48.21$, $SD = 21.70$), *action* ($M = 27.88$, $SD = 8.39$) and *setting details* ($M = 20.79$, $SD = 8.00$) in the delayed free recall compared to those who had not had an initial recall ($M = 20.66$, $SD = 4.45$; $M = 8.24$, $SD = 4.15$; $M = 6.33$, $SD = 4.58$; and $M = 8.97$, $SD = 3.06$), $t = 4.45$, $p < .001$; $t = 4.15$, $p < .001$; $t = 4.58$, $p < .001$; and $t = 3.06$, $p < .01$, respectively. Effect sizes indicated by *Cohens' d* range between $d = 0.88$ and $d = 1.32$ and thus represent large effects throughout these significant differences in the young adults subsample.

Table 3.5. Group differences at T2 in the interview free recall by recall condition (SAI vs. no), for total sample and subsamples

Group	SAI		no				
Variable	M	SD	M	SD	t		d
Total sample (df = 94)							
Accuracy rate	0.93	0.04	0.90	0.06	3.01	**	0.61
Correct details	97.19	31.45	73.25	31.72	3.71	***	0.76
Incorrect details	6.94	4.35	7.44	6.59	-0.44		0.09
Confabulated details	0.71	1.30	0.98	1.51	-0.94		0.19
Person details	47.60	19.01	35.27	16.54	3.39	***	0.69
Action details	27.63	7.79	22.81	9.86	2.65	**	0.54
Object details	7.42	3.43	6.98	3.21	0.65		0.13
Setting details	22.19	9.13	16.60	9.65	2.91	**	0.59
Older adults (df = 46)							
Accuracy rate	0.92	0.03	0.87	0.07	3.30	**	0.95
Correct details	97.17	28.66	85.96	35.94	1.19		0.34
Incorrect details	7.88	3.54	11.00	7.41	-1.87		0.54
Confabulated details	0.58	1.44	1.04	1.76	-0.99		0.29
Person details	47.00	16.33	42.00	19.92	0.95		0.27
Action details	27.38	7.32	27.58	10.54	-0.08		0.02
Object details	7.67	3.36	8.50	3.27	-0.87		0.25
Setting details	23.58	10.11	19.92	9.32	1.31		0.38
Young adults (df = 46)							
Accuracy rate	0.94	0.03	0.92	0.05	1.05		0.30
Correct details	97.21	34.64	60.54	20.66	4.45	***	1.29
Incorrect details	6.00	4.93	3.88	2.74	1.85		0.53
Confabulated details	0.83	1.17	0.92	1.25	-0.24		0.07
Person details	48.21	21.70	28.54	8.24	4.15	***	1.20
Action details	27.88	8.39	18.04	6.33	4.58	***	1.32
Object details	7.17	3.56	5.46	2.36	1.96		0.57
Setting details	20.79	8.00	13.29	8.97	3.06	**	0.88

**$p < .01$, ***$p < .001$.

To further test for age group effects and interaction effects of age group and initial recall, and to comprehensively look at all three recall conditions, several factorial 2 × 3 between-subjects ANOVAs were conducted with initial recall condition in the first session (SAI, wFR, no) and age group as fixed factors and each of the nine total variables from the second session free recall as dependant variable, respectively.

In exploring the assumptions of an ANOVA, the Kolmogorov-Smirnov test and Levene's test were again conducted. The Kolmogorov-Smirnov test was significant for most variables, indicating they were not normally distributed. Levene's test was significant for some variables, indicating that variances were significantly different in these cases. However, as was previously discussed, the F-statistic is said to be fairly robust to violations of normality as well as homogeneity of variances and can still perform accurately when group sizes are equal (Field, 2009) and thus, analyses used the original, non-transformed data. The results are shown in table 3.6.

Table 3.6. Factorial ANOVAs at T2 in the interview free recall

	Recall condition (R)			Age group (A)			Interaction R x A		
Variable	F^a		Part. η^2	F^b		Part. η^2	F^a		Part. η^2
Accuracy rate	6.54	**	.09	28.76	***	.17	2.72		.04
Correct details	8.11	***	.11	9.06	**	.06	2.44		.03
Incorrect details	0.13		.00	42.83	***	.24	4.41	*	.06
Confab. details	0.83		.01	0.70		.01	1.52		.02
Person details	6.88	***	.09	7.33	**	.05	2.63		.04
Action details	4.24	*	.06	12.34	***	.08	4.58	*	.06
Object details	0.22		.00	7.20	**	.05	1.46		.02
Setting details	5.13	**	.07	18.91	***	.12	1.74		.03

[a]$df = 2, 138$. [b]$df = 1, 138$.
*$p < .05$. **$p < .01$. ***$p < .001$.

Main effects for recall condition were found on the variables *accuracy rate*, $F\,(2, 138) = 6.54$, $p < .01$, *correct details*, $F\,(2, 138) = 8.11$, $p < .001$, as well as *person*, $F\,(2, 138) = 6.88$, $p < .001$, *action*, $F\,(2, 138) = 4.24$, $p < .05$, and *setting details*, $F\,(2, 138) = 5.13$, $p < .01$. Effect sizes for significant recall condition effects are indicated by partial eta squared and are between $\eta^2_{partial} = .06$ and $\eta^2_{partial} = .11$, which represent medium-sized effects (Cohen, 1988). The Tukey HSD *post hoc* test revealed that participants who had filled in the SAI performed significantly better in

the recall part of the witness interview after one week compared to those who did not have an initial recall option ($p < .001$ for correct and person details; $p < .01$ for accuracy rate and setting details; and $p < .05$ for action details). Participants in the written free recall condition neither performed significantly worse than those in the SAI condition, nor significantly better than those without an initial recall. Results are displayed in figure 3.4. for correct details, and in figure 3.5. for person, action and setting details.

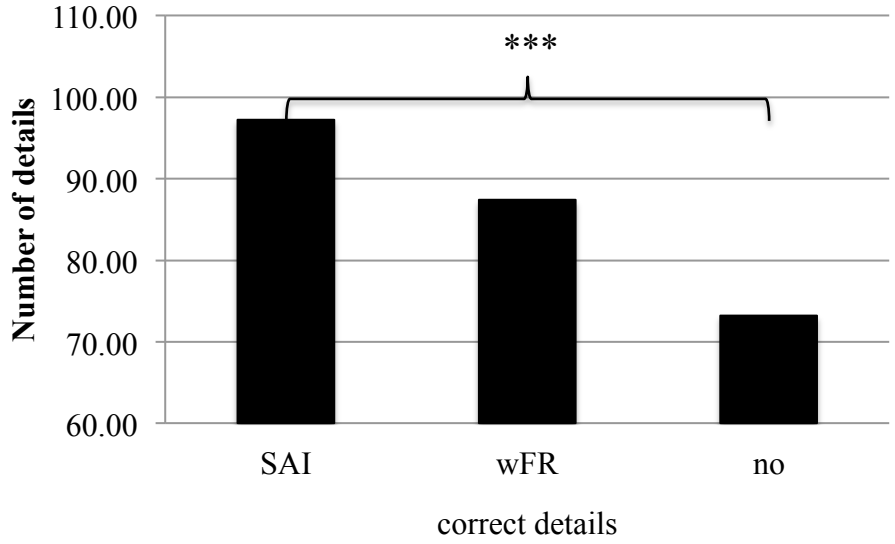


Figure 3.4. Average number of correct details as a function of recall condition at T2 in the interview free recall.

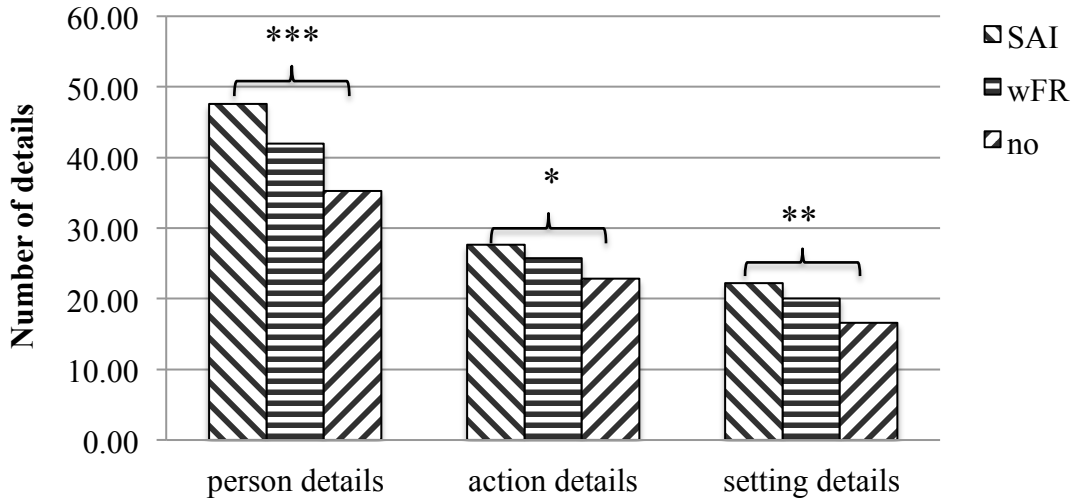


Figure 3.5. Average number of person, action and setting details as a function of recall condition at T2 in the interview free recall.

Main effects for age group were observed for all total variables except total confabulated details, i.e. the *accuracy rate*, F (1, 138) = 28.76, $p < .001$, *correct details*, F (1, 138) = 9.06, $p < .01$, *incorrect details*, F (1, 138) = 42.83, $p < .001$, *and person*, F (1, 138) = 7.33, $p < .01$, *action*, F (1, 138) = 12.34, $p < .001$, *object*, F (1, 138) = 7.20, $p < .01$, and *setting details*, F (1, 138) = 18.91, $p < .001$. Contrary to the prediction, older adults remembered *more* correct details (M = 93.28, SD = 29.80), person (M = 45.31, SD = 16.53), action (M = 27.78, SD = 8.16), object (M = 8.10, SD = 3.32) and setting details (M = 22.75, SD = 9.37) than young adults (M = 78.60, SD = 32.12; M = 37.94, SD = 17.75; M = 23.00, SD = 8.91; M = 6.44, SD = 4.02; M = 16.50, SD = 8.44, respectively). In accordance with the prediction, younger adults had an overall higher accuracy rate (M = 0.94, SD = 0.04) and produced less incorrect details (M = 4.60, SD = 3.81) than older adults (M = 0.89, SD = 0.06 and M = 9.78, SD = 5.68). Effect sizes for significant age group effects range between $\eta^2_{partial} = .05$ and $\eta^2_{partial} = .24$, which represent small to large effects.

There were interaction effects between the recall condition and age of the participants on the variables *incorrect details*, F (2, 138) = 4.41, $p < .05$, and *action details*, F (2, 138) = 4.58, $p < .05$. This indicates that older and young adults were affected differently by the recall condition. Specifically, for the number of incorrect details, it can be seen that older adults (M = 11.00, SD = 7.41) produced more incorrect details than young adults (M = 3.88, SD = 2.74) in the no condition as well as in the wFR condition (M = 10.46, SD = 5.12 and M = 3.92, SD = 3.15, respectively); however, the number of incorrect details was similar for older (M = 7.88, SD = 3.54) and young adults (M = 6.00, SD = 4.93) in the SAI condition. So whereas older adults produced similarly high levels of incorrect details in the no and wFR condition, in the SAI condition they improved and produced fewer incorrect details. This was the opposite for young adults, who produced similarly low levels of incorrect details in the no and wFR condition, but slightly worsened in the SAI condition where they produced more incorrect details. This finding represents a medium-sized effect of $\eta^2_{partial} = .06$ and is displayed in figure 3.6.

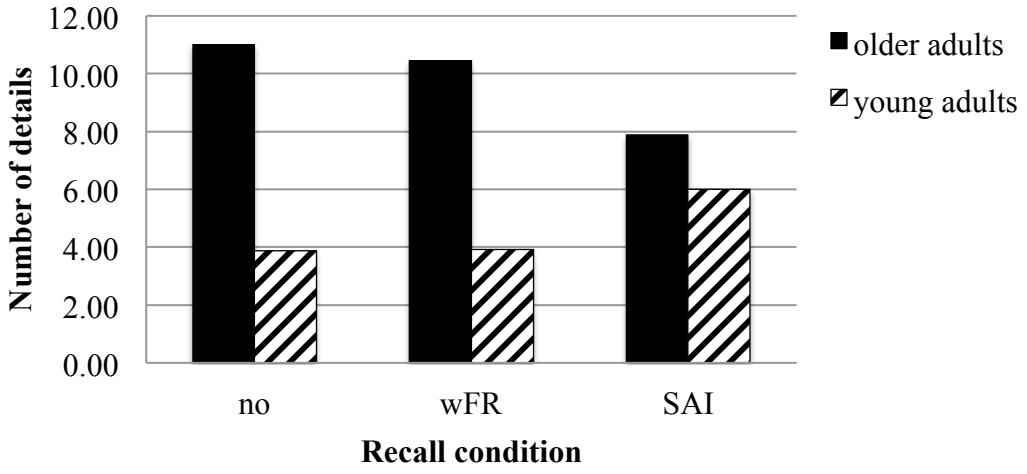


Figure 3.6. Number of incorrect details as a function of recall condition and age group at T2 in the interview free recall.

For action details, performance was higher in older adults (M = 27.58, SD = 10.54) than young adults (M = 65.33, SD = 21.18) in the no condition; performance was also higher for older adults (M = 28.38, SD = 6.37) than young adults (M = 23.08 SD = 9.16) in the wFR condition; however, performance was similar for older (M = 27.38, SD = 7.32) and young adults (M = 27.88, SD = 8.39) in the SAI condition. This means the performance increased across conditions for the young adults, whereas older adults performed on a similar level regardless of whether they did not have an initial recall, gave a written free recall or filled in the SAI. This finding represents a medium-sized effect of $\eta^2_{partial}$ = .06 and is displayed in figure 3.7.

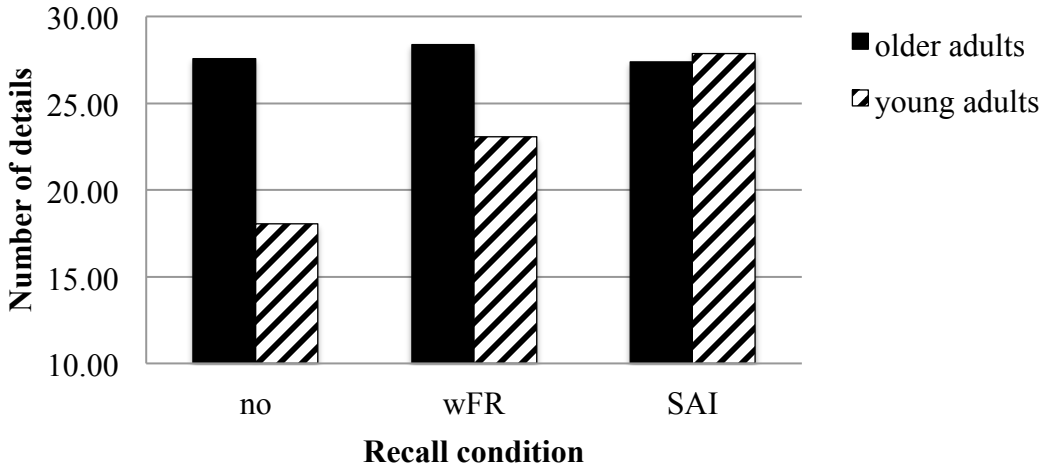


Figure 3.7. Number of action details as a function of recall condition and age group at T2 in the interview free recall.

Thus, analyses of the witness interview recall part fully supported the *interview transfer hypothesis*. Participants who had filled in the SAI a week earlier produced a more detailed and more accurate account compared to those who did not have an initial recall opportunity. Participants in the written free recall group neither performed significantly better than the no group, nor significantly worse than the SAI group. The *age group hypothesis* was only partly supported. Contrary to the prediction, older adults gave a more detailed account with also more correct details in the free recall compared to young adults. In accordance with the prediction, older adults also produced more incorrect details and had an overall lower accuracy rate than young adults. The *interaction hypothesis* was not supported. Contrary to the prediction, the benefit of the SAI was higher for young adults than for older adults.

3.2.2.3. Interview Question Part

In this section, analyses for the question part of the second session are presented, this means a comparison of details produced in answers to questions of the witness interview after one week, depending on recall condition in the first session. A complete table of means and standard deviations across all conditions and variables for the question part can be found in appendix O. Lastly, analyses for the specific questions (duration, yes/no and suggestive questions) will be presented. The hypotheses that are relevant to this part of the analyses are:

Interview transfer hypothesis

Participants will give more detailed and accurate accounts in an eyewitness interview at time 2 when having first completed the Self-Administered Interview at time 1 compared to a written free recall or no initial recall at time 1.

Age group hypothesis

Older adults give less detailed and less accurate accounts in the eyewitness interview at time 2 compared to young adults.

Interaction hypothesis

The benefit of the SAI for older adults exceeds that of young adults.

From a practitioner's point of view it is again interesting to see a comparison of participants who had completed the SAI ("SAI" condition) and participants who did not have an initial recall ("no" condition, which is the current standard in police investigation). Thus, to start off with, t-tests were conducted for the details produced in the specific interview questions for the SAI vs. the no condition. Table 3.7. comprises means, standard deviations, t-tests and corresponding effect sizes for the total details. Results are shown for the total sample as well as for the older and young adults separately.

Table 3.7. Group differences at T2 in the interview question part by recall condition (SAI vs. no), for total sample and subsamples

Group	SAI		no				
Variable	M	SD	M	SD	t		d
Total sample (df = 94)							
Accuracy rate	0.88	0.03	0.85	0.07	3.09	**	0.56
Correct details	94.56	23.21	82.69	23.38	2.50	*	0.51
Incorrect details	9.94	4.46	11.31	6.31	-1.23		0.25
Confabulated details	2.75	1.91	3.19	3.36	-0.78		0.16
Person details	55.08	16.39	47.71	13.68	2.39	*	0.49
Action details	23.06	5.41	21.63	7.20	1.11		0.22
Object details	10.46	3.10	10.52	4.54	-0.79		0.02
Setting details	18.65	4.37	17.33	4.97	1.37		0.28
Older adults (df = 46)							
Accuracy rate	0.87	0.03	0.84	0.08	1.72		0.50
Correct details	87.42	17.76	89.04	25.10	-0.26		0.07
Incorrect details	10.88	4.49	13.79	7.40	-1.65		0.48
Confabulated details	2.67	1.47	3.00	4.36	-0.36		0.10
Person details	51.33	13.07	52.46	14.49	-0.28		0.08
Action details	21.21	4.15	24.33	7.31	-1.82		0.52
Object details	9.83	3.09	10.96	5.26	-0.90		0.26
Setting details	18.58	4.87	18.08	4.79	0.36		0.10
Young adults (df = 46)							
Accuracy rate	0.90	0.04	0.86	0.05	2.90	**	0.88
Correct details	101.71	26.05	76.33	20.06	3.78	***	1.09
Incorrect details	9.00	4.31	8.83	3.69	0.14		0.04
Confabulated details	2.83	2.30	3.38	2.00	-0.87		0.26
Person details	58.83	18.67	42.96	11.18	3.57	***	1.03
Action details	24.92	5.95	18.92	6.11	3.45	***	0.99
Object details	11.08	3.05	10.08	3.75	1.01		0.29
Setting details	18.71	3.92	16.58	5.14	1.61		0.47

*$p < .05$, **$p < .01$, ***$p < .001$.

Starting with the total sample, participants who had completed the SAI a week earlier produced significantly more *correct details* ($M = 94.56$, $SD = 23.21$), and *person details* ($M = 55.08$, $SD = 16.39$) in the interview question part after a one week delay compared to those who did not have an initial recall ($M = 82.69$, $SD = 23.38$; $M = 47.71$, $SD = 13.68$), $t = 2.50$, $p < .05$; $t = 2.39$, $p < .05$, respectively. The SAI group also had a significantly higher *accuracy rate* ($M = 0.88$, $SD = 0.03$) compared to the no group ($M = 0.85$, $SD = 0.07$), $t = 3.09$, $p < .01$. Effect sizes indicated by *Cohens' d* range between $d = 0.49$ and $d = 0.56$ and thus represent medium-sized

effects throughout these significant differences in the total sample. It may again be noted, that participants overall produced significantly more correct details without increasing the number of incorrect or confabulated details in the SAI versus the no condition.

When separating the sample by age-group, none of the variables differed significantly between older adults who had completed the SAI a week earlier and those who had not had any initial recall. In the young adults subsample however, participants who had completed the SAI a week earlier produced significantly more *correct* (M = 101.71, SD = 26.05), *person* (M = 58.83, SD = 18.67) and *action details* (M = 24.92, SD = 5.97) in the interview question part compared to those who had not had an initial recall (M = 76.33, SD = 20.06; M = 42.96, SD = 11.18; *and* M = 18.92, SD = 6.11), t = 3.78, $p < .001$; t = 3.57, $p < .001$; and t = 3.45, $p < .001$, respectively. The young adults in the SAI group also had a significantly higher *accuracy rate* (M = 0.90, SD = 0.04) compared to those in the no group (M = 0.86, SD = 0.05), t = 2.90, $p < .01$. Effect sizes indicated by *Cohens' d* range between d = 0.88 and d = 1.09 and thus represent large effects throughout these significant differences in the young adults subsample.

To further test for age group effects and interaction effects of age group and initial recall, and to comprehensively look at all three recall conditions, several factorial 2 × 3 between-subjects ANOVAs were conducted with initial recall condition in the first session (SAI, wFR, no) and age group as fixed factors and each of the nine total variables from the second session question part as dependant variable, respectively.

In exploring the assumptions of an ANOVA, the Kolmogorov-Smirnov test and Levene's test were again conducted. The Kolmogorov-Smirnov test was significant for most variables, indicating they were not normally distributed. Levene's test was significant for some variables, indicating that variances were significantly different in these cases. However, as was previously discussed, the F-statistic is said to be fairly robust to violations of normality as well as homogeneity of variances and can still perform accurately when group sizes are equal (Field, 2009) and thus, analyses used the original, non-transformed data. The results are shown in table 3.8.

Table 3.8. Factorial ANOVAs at T2 in the interview question part for total details

	Recall condition (R)			Age group (A)			Interaction R x A		
Variable	F^a		Part. η^2	F^b		Part. η^2	F^a		Part. η^2
Accuracy rate	5.59	**	.08	8.51	**	.06	0.12		.00
Correct details	4.04	*	.06	0.09		.00	4.91	**	.07
Incorrect details	1.13		.02	13.25	***	.09	1.55		.02
Confab. details	0.93		.01	0.08		.00	0.09		.00
Person details	3.73	*	.05	0.03		.00	4.89	**	.07
Action details	0.74		.01	0.34		.00	7.41	***	.10
Object details	0.02		.00	0.00		.00	0.98		.01
Setting details	1.28		.02	4.21	*	.03	1.74		.03

[a]*df* = 2, 138. [b]*df* = 1, 138.
*p < .05. **p < .01. ***p < .001.

Main effects for recall condition were found on the variables *accuracy rate*, F (2, 138) = 5.59, p < .01, *correct details*, F (2, 138) = 4.04, p < .05, and *person details*, F (2, 138) = 3.73, p < .05. Effect sizes for significant recall condition effects are indicated by partial eta squared and are between $\eta^2_{partial}$ = .05 and $\eta^2_{partial}$ = .08, which represent small to medium-sized effects (Cohen, 1988). The Tukey HSD *post hoc* test revealed that participants who had filled in the SAI performed significantly better in the question part of the witness interview after one week compared to those who did not have an initial recall option (p < .01 for accuracy rate; and p < .05 for correct and person details). Participants in the written free recall condition neither performed significantly worse than those in the SAI condition, nor significantly better than those without an initial recall. Results are displayed in figure 3.8. for correct and person details.

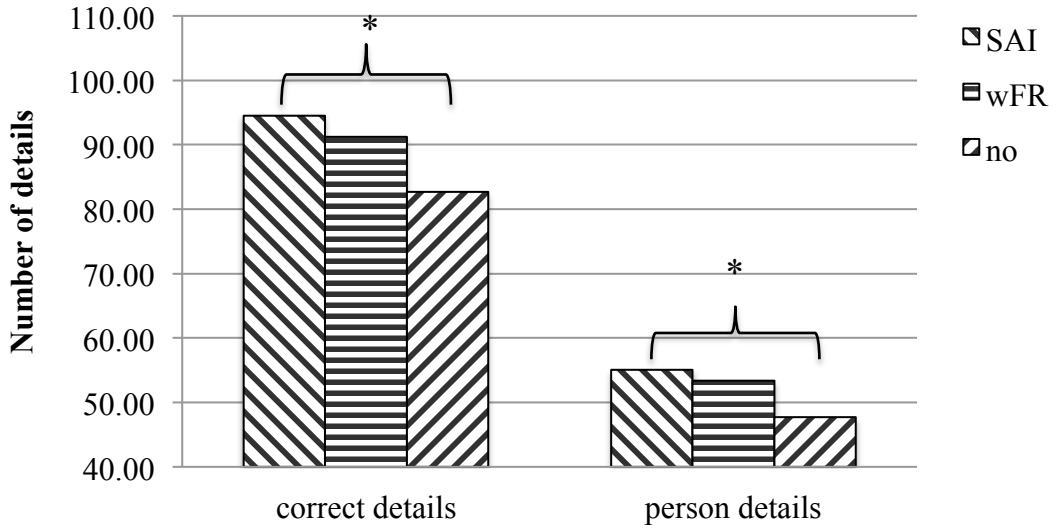


Figure 3.8. Average number of correct details and person details as a function of recall condition at T2 in the interview question part.

Main effects for age group were observed for *accuracy rate*, F (1, 138) = 8.51, $p < .01$, *incorrect details,* F (1, 138) = 13.25, $p < .001$, and *setting details*, F (1, 138) = 4.21, $p < .05$. In accordance with the prediction, younger adults produced a higher accuracy rate (M = 0.88, SD = 0.05) and remembered *less* incorrect details, (M = 8.99, SD = 4.31), than older adults (M = 0.85, SD = 0.05; M = 11.93, SD = 5.39, respectively). However, contrary to the prediction, older adults produced *more* setting details (M = 19.03, SD = 5.17), than young adults (M = 17.42, SD = 4.28). Effect sizes for significant age group effects range between $\eta^2_{partial}$ = .03 and $\eta^2_{partial}$ = .09, which represent small to medium-sized effects.

There were interaction effects between the recall condition and age of the participants on the variables *correct details*, F (2, 138) = 4.91, $p < .01$, *person details*, F (2, 138) = 4.89, $p < .01$, and *action details*, F (2, 138) = 7.41, $p < .001$. This indicates that older and young adults were affected differently by the recall condition. Specifically, for the number of correct details performance was higher in older adults (M = 89.04, SD = 25.10) than young adults (M = 76.33, SD = 20.06) in the no condition; performance was similar for older adults (M = 90.46, SD = 16.79) and young adults (M = 92.00, SD = 19.12) in the wFR condition; and then lower for older (M = 87.42, SD = 17.76) than young adults (M = 101.71, SD = 26.05) in the SAI condition. So whereas performance increased across conditions for the young adults, older adults performed on a similar level regardless of whether they did not have an

initial recall, gave a written free recall or filled in the SAI. This finding represents a medium-sized effect of $\eta^2_{partial} = .07$ and is displayed in figure 3.9.

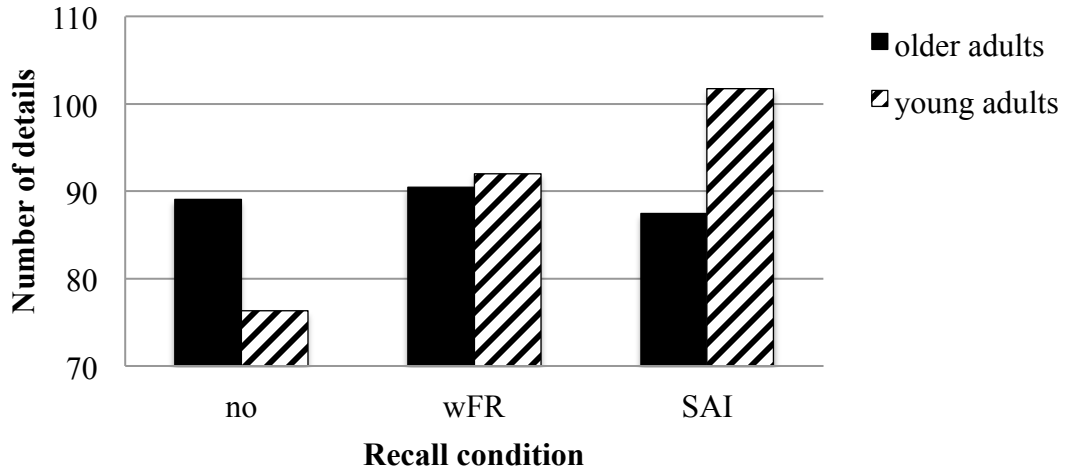


Figure 3.9. Number of correct details as a function of recall condition and age group at T2 in the interview question part.

For the number of person details, performance was higher in older adults ($M = 52.46$, $SD = 14.50$) than young adults ($M = 42.96$, $SD = 11.18$) in the no condition; however, performance was lower for older adults ($M = 51.79$, $SD = 9.86$) than young adults ($M = 54.92$, $SD = 14.01$) in the wFR condition; and also lower for older ($M = 51.33$, $SD = 13.07$) than young adults ($M = 58.83$, $SD = 18.67$) in the SAI condition. Again, whereas performance increased across conditions for the young adults, older adults performed on a similar level regardless of whether they did not have an initial recall, gave a written free recall or filled in the SAI. This finding also represents a medium-sized effect of $\eta^2_{partial} = .07$ and is displayed in figure 3.10.

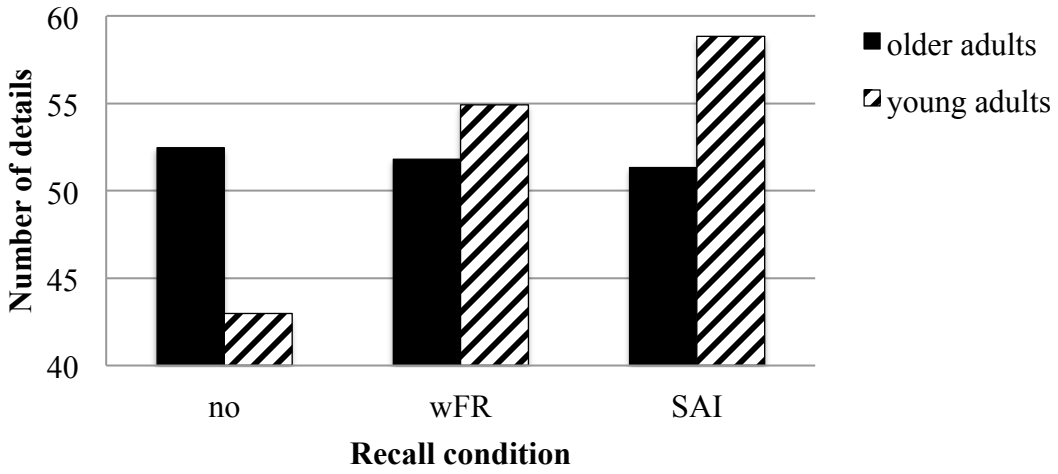


Figure 3.10. Number of person details as a function of recall condition and age group at T2 in the interview question part.

And lastly, for the number of action details, performance was higher in older adults (M = 24.33, SD = 7.31) than young adults (M = 18.92, SD = 6.11) in the no condition; performance was almost identical for older adults (M = 22.50, SD = 5.22) and young adults (M = 22.50, SD = 5.83) in the wFR condition; and then lower for older (M = 21.21, SD = 4.15) than young adults (M = 24.92, SD = 5.95) in the SAI condition. For the young adults, performance increased across conditions, whereas older adults' performance decreased across conditions. This finding represents a medium-sized effect of $\eta^2_{partial}$ = .10 and is displayed in figure 3.11.

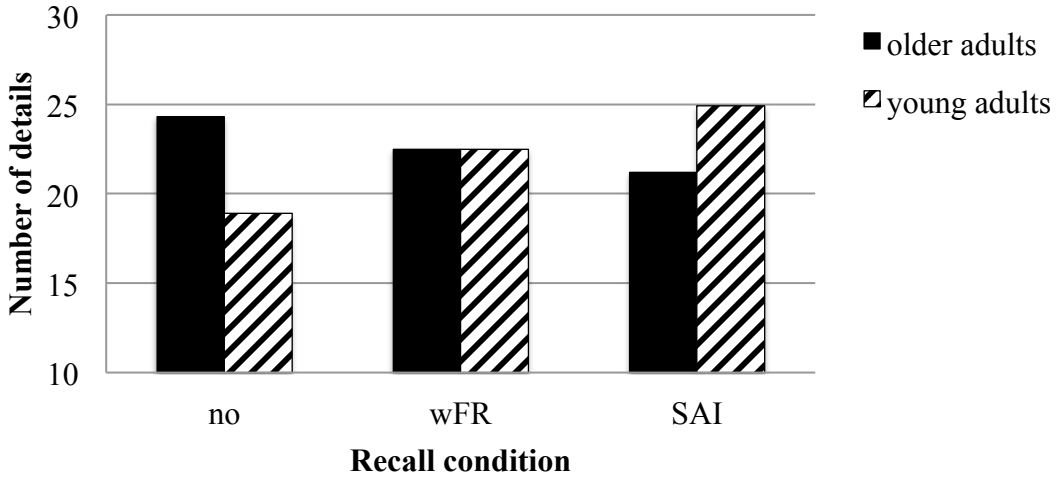


Figure 3.11. Number of action details as a function of recall condition and age group at T2 in the interview question part.

Thus, analyses of the witness interview question part fully supported the *interview transfer hypothesis*. Participants who had filled in the SAI a week earlier produced a more detailed and more accurate account compared to those who did not have an initial recall opportunity. Participants in the written free recall group neither performed significantly better than the no group, nor significantly worse than the SAI group. The *age group hypothesis* was mainly supported. In accordance with the prediction, older adults also produced more incorrect details and had an overall lower accuracy rate than young adults. The only findings contradicting the prediction were that older adults gave a more correct setting details compared to young adults. The *interaction hypothesis* was not supported: Contrary to the prediction, the benefit of the SAI was higher for young adults than for older adults.

3.2.2.4. Examination of Possible Alternative Hypotheses

In order to control the data for potential moderators, multivariate analyses of covariance (MANCOVA) on the total details variables of the initial recall and of the witness interview were conducted. The variables *ethnicity, university degree* and *alertness rating* differed significantly in the demographic analyses by age-group and were thus considered as covariates[16]. As they did not differ significantly in the demographic analyses by recall condition, they showed independence from the experimental effect and thus agreed with the theoretical assumption of MANCOVA (Field, 2009). Therefore, three sets of MANCOVAs were conducted with recall condition and age group as independent variables, the nine total variables for the initial recall at T1 and for the witness interview free recall part and question part at T2 as dependent variables, and *ethnicity*[17]*, university degree*[18] and *alertness rating*[19] as covariates, respectively. Results can be found in tables 3.9. – 3.11.

Starting with *ethnicity* as covariate (table 3.9.), there was a significant effect of the covariate on four variables in the initial recall of the first session, *total details*, $F(1, 91) = 4.37$, $p < .05$, *total correct,* $F(1, 91) = 5.04$, $p < .05$, *total person*, $F(1, 91) = 4.00$, $p < .05$, and *total action* details, $F(1, 91) = 4.23$, $p < .05$; as well as on one variable each in the free recall and question part of the second session, *total action* details, $F(1, 137) = 5.02$, $p < .05$, and *total correct* details, $F(1,137) = 4.15$, $p < .05$, respectively, with those participants who indicated to be of white ethnicity performing better than those who indicated to be of another ethnic background. This may largely be due to the fact that all non-white participants ($n = 6$) were students who were non-native speakers, and thus may have experienced more difficulties

[16] The variables MEQ, CESD-R and SUDOKU also differed significantly between age groups, but were not considered as covariates in this analysis as these measurements merely functioned as filler tasks rather than demographic indicators.

[17] Although there is no prior research suggesting that there would be differences by ethnicity, this analysis was included for completeness' sake.

[18] University degree was included in the analysis as a proxy measure for cognitive ability. For example, it has been shown that memory decline is faster in less educated people (Schmand et al., 1997).

[19] Alertness was included in the analysis as people who feel more awake may perform better on a memory task than those who feel sleepy (e.g. Buela Casal et al., 1990).

describing the event in comprehensive detail.[20] Effect sizes are indicated by partial eta squared and range between $\eta^2_{partial} = .03$ and $\eta^2_{partial} = .05$, which represent small effects for all significant effects of the covariate *ethnicity*. After controlling for ethnicity, there were still significant main effects for both recall condition, $F(1, 91) = 11.40$, $p < .01$, and age group, $F(1, 91) = 6.63$, $p < .05$, as well as a significant interaction effect, $F(1, 91) = 5.78$, $p < .05$, for the variable *total details* of the first session, with medium-sized effects ranging between $\eta^2_{partial} = .06$ and $\eta^2_{partial} = .11$. The same was true for the variables *total correct* details, $F(1, 91) = 11.40$, $p < .01$ (main effect recall condition), $F(1, 91) = 9.60$, $p < .01$ (main effect age group), $F(1, 91) = 5.74$, $p < .05$ (interaction effect) and *total person* details of the first session, $F(1, 91) = 39.36$, $p < .001$ (main effect recall condition), $F(1, 91) = 11.10$, $p < .01$ (main effect age group), $F(1, 91) = 5.59$, $p < .05$ (interaction effect), with medium-sized to large effects ranging between $\eta^2_{partial} = .06$ and $\eta^2_{partial} = .30$. For the variable *total action* details of the first session there was still a significant main effect for age group, $F(1, 91) = 5.93$, $p < .05$, and a significant interaction effect, $F(1, 91) = 6.18$, $p < .05$, with medium-sized effects of $\eta^2_{partial} = .06$ for each. And similarly for the second session, after controlling for ethnicity, there were still significant main effects for both recall condition, $F(2, 137) = 4.25$, $p < .05$, and age group, $F(1, 137) = 7.11$, $p < .01$, as well as a significant interaction effect, $F(2, 137) = 4.45$, $p < .05$, for the variable *total action* details of the free recall, with medium-sized effects of $\eta^2_{partial} = .06$ for each; as well as a significant main effect for recall condition, $F(2, 137) = 4.11$, $p < .05$, and a significant interaction effect, $F(2, 137) = 4.86$, $p < .01$, for the variable *total correct* details of the question part, with medium-sized effects of $\eta^2_{partial} = .06$ and $\eta^2_{partial} = .07$, respectively. Thus, the covariate *ethnicity* had an impact on testimony performance, especially in the first session initial recall. However, on all these variables, there were still significant experimental effects, both main and interaction effects, that were also larger in comparison.

When looking at *university degree* as covariate (table 3.10.), there was only one significant effect of the covariate, which was on the variable *total setting* details of the second session question part, $F(1, 137) = 4.84$, $p < .05$, with those participants

[20] It may be noted, that some participants who indicated to be of white ethnicity were also non-native speakers, especially among the young adults group.

who attended university performing better than those who did not attend a university. The effect size of $\eta^2_{partial}$ = .03 represents a small effect. After controlling for university degree, there was still a significant main effect for age group on this variable, F (1, 137) = 7.85, p < .01, with a small effect of $\eta^2_{partial}$ = .05.

And finally, when looking at *alertness* as covariate (table 3.11.), there was a significant effect of the covariate on one variable in each account, *total action* details in the initial recall of the first session, F (1, 91) = 4.06, p < .05, *total action* details in the free recall of the second session, F (1, 137) = 4.10, p < .05, and *total object* details in the question part of the second session, F (1, 137) = 4.87, p < .05, with those participants who indicated to feel fairly or completely awake performing better than those who indicated to feel a little or quite sleepy. Effect sizes range between $\eta^2_{partial}$ = .03 and $\eta^2_{partial}$ = .04, which represent small effects for all significant effects of the covariate *alertness.* After controlling for alertness, there was still a significant interaction effect, F (1, 91) = 6.36, p < .05, for the variable *total action* details of the first session, with a medium-sized effect of $\eta^2_{partial}$ = .07. There were also still significant main effects for both recall condition, F (2, 137) = 4.57, p < .05, and age group, F (1, 137) = 16.25, p < .001, as well as a significant interaction effect, F (2, 137) = 4.52, p < .05, for the variable *total action* details of the second session free recall, with medium-sized effects ranging between $\eta^2_{partial}$ = .06 and $\eta^2_{partial}$ = .11. For the variable *total object* details of the second session question part, there were no significant main or interaction effects, as in the previous ANOVA. Thus, the covariate *alertness* had an impact on testimony performance, albeit small and only sporadically.

Table 3.9. Factorial MANCOVAs at T1 and T2 for total details with ethnicity as covariate.

	Ethnicity			Recall condition (R)			Age group (A)			Interaction R x A		
Variable	df	F	Part. η^2	df	F	Part. η^2	df	F	Part. η^2	df	F	Part. η^2
T1 initial recall	1, 91			1, 91			1, 91			1, 91		
Total details		4.37 *	.05		11.40 **	.11		6.63 *	.07		5.78 *	.06
Accuracy rate		1.66	.02		0.62	.01		27.53 ***	.23		0.99	.01
Total correct		5.04 *	.05		11.40 **	.11		9.60 **	.10		5.74 *	.06
Total incorrect		0.38	.00		7.30 **	.07		0.60	.01		2.62	.03
Total confab.		0.00	.00		0.02	.00		6.83 *	.07		1.33	.01
Total person		4.00 *	.04		39.36 ***	.30		11.10 **	.11		5.59 *	.06
Total action		4.23 *	.04		3.25	.04		5.93 *	.06		6.18 *	.06
Total object		2.47	.03		0.28	.00		4.27 *	.05		2.25	.02
Total setting		2.54	.03		0.15	.00		0.17	.00		2.98	.03
T2 free recall	1, 137			2, 137			1, 137			2, 137		
Total details		3.64	.03		6.63 **	.09		9.45 **	.07		3.06	.04
Accuracy rate		0.14	.00		6.51 **	.09		26.92 ***	.16		2.76	.04
Total correct		3.86	.03		8.12 ***	.11		5.10 *	.04		2.26	.03
Total incorrect		0.15	.00		0.14	.00		36.76 ***	.21		4.28 *	.06
Total confab.		0.08	.00		0.81	.01		0.49	.00		1.45	.02
Total person		3.25	.02		6.82 **	.09		4.06 *	.03		2.45	.04
Total action		5.02 *	.04		4.25 *	.06		7.11 **	.06		4.45 *	.06
Total object		3.01	.02		0.26	.00		4.05 *	.03		1.44	.02
Total setting		0.42	.00		5.07 **	.07		15.28 ***	.10		1.61	.02
T2 question part	1, 137			2, 137			1, 137			2, 137		
Total details		3.65	.03		2.61	.04		0.03	.00		5.13 **	.07
Accuracy rate		1.27	.01		5.55 **	.08		9.78 **	.07		0.10	.00
Total correct		4.15 *	.03		4.11 *	.06		0.87	.01		4.86 **	.07
Total incorrect		0.69	.01		1.13	.02		10.13 **	.07		1.57	.02
Total confab.		1.35	.01		0.78	.01		0.01	.00		0.13	.00
Total person		3.85	.03		3.81 *	.05		0.61	.00		4.97 **	.07
Total action		2.45	.02		0.72	.01		0.00	.00		7.37 **	.10
Total object		1.30	.01		0.01	.00		0.18	.00		0.90	.01
Total setting		0.36	.00		1.30	.02		3.06	.02		1.61	.02

*p < .05. **p < .01. ***p < .001.

Table 3.10. Factorial MANCOVAs at T1 and T2 for total details with university degree as covariate.

	University degree			Recall condition (R)			Age group (A)			Interaction R x A		
Variable	df	F	Part. η^2	df	F	Part. η^2	df	F	Part. η^2	df	F	Part. η^2
T1 initial recall	1, 91			1, 91			1, 91			1, 91		
Total details		0.31	.00		12.97 **	.13		2.61	.03		6.06 **	.06
Accuracy rate		0.31	.00		0.28	.00		20.28 ***	.18		0.87	.01
Total correct		0.37	.00		13.09 ***	.13		4.18 **	.04		6.02 **	.06
Total incorrect		0.06	.00		7.75 **	.08		1.09	.01		2.64	.03
Total confab.		0.24	.00		0.05	.00		5.65 **	.06		1.51	.02
Total person		0.07	.00		40.35 ***	.31		6.16 **	.06		6.09 **	.06
Total action		0.31	.00		4.31 **	.05		2.23	.02		6.45 **	.07
Total object		1.14	.01		0.78	.01		1.26	.01		2.13	.02
Total setting		0.36	.00		0.01	.00		0.11	.00		3.10	.03
T2 free recall	1, 137			2, 137			1, 137			2, 137		
Total details		1.08	.01		6.49 **	.09		15.46 ***	.10		3.22 *	.05
Accuracy rate		0.29	.00		6.29 **	.08		21.63 ***	.14		2.79	.04
Total correct		1.22	.01		7.93 **	.10		10.24 **	.07		2.41	.03
Total incorrect		0.00	.00		0.13	.00		35.13 ***	.20		4.37 *	.06
Total confab.		0.05	.00		0.84	.01		0.72	.01		1.42	.02
Total person		2.20	.02		6.77 **	.09		9.54 **	.07		2.61	.04
Total action		0.23	.00		4.14 *	.06		11.48 **	.08		4.57 *	.06
Total object		0.68	.01		0.17	.00		7.74 **	.05		1.53	.02
Total setting		0.03	.00		5.06 **	.07		16.12 ***	.11		1.67	.02
T2 question part	1, 137			2, 137			1, 137			2, 137		
Total details		1.30	.01		2.38	.03		0.80	.01		5.34 **	.07
Accuracy rate		0.00	.00		5.52 **	.08		6.95 **	.05		0.12	.00
Total correct		0.98	.01		3.85 *	.05		0.02	.00		5.02 **	.07
Total incorrect		0.57	.00		1.23	.02		13.12 ***	.09		1.68	.02
Total confab.		0.23	.00		0.82	.01		0.00	.00		0.06	.00
Total person		1.05	.01		3.52 *	.05		0.08	.00		5.11 **	.07
Total action		0.04	.00		0.72	.01		0.37	.00		7.38 **	.10
Total object		0.02	.00		0.02	.00		0.00	.00		0.97	.01
Total setting		4.84 *	.03		0.98	.01		7.85 **	.05		1.25	.02

*p < .05. **p < .01. ***p < .001.

Table 3.11. Factorial MANCOVAs at T1 and T2 for total details with alertness as covariate.

	Alertness			Recall condition (R)			Age group (A)			Interaction R x A		
Variable	df	F	Part. η^2	df	F	Part. η^2	df	F	Part. η^2	df	F	Part. η^2
T1 initial recall	1, 91			1, 91			1, 91			1, 91		
Total details		1.32	.01		11.63 **	.11		2.20	.02		6.20 *	.06
Accuracy rate		0.24	.00		0.48	.01		21.35 ***	.19		0.79	.01
Total correct		1.87	.02		11.55 **	.11		3.51	.04		6.14 *	.06
Total incorrect		0.30	.00		8.04 **	.08		0.62	.01		3.04	.03
Total confab.		0.01	.00		0.02	.00		6.56 *	.07		1.36	.02
Total person		0.03	.00		40.16 ***	.31		6.60 *	.07		6.38 *	.07
Total action		4.06 *	.04		3.18	.03		1.21	.01		6.36 *	.07
Total object		2.07	.02		0.28	.00		1.16	.01		2.35	.03
Total setting		2.02	.02		0.14	.00		0.34	.00		3.12	.03
T2 free recall	1, 137			2, 137			1, 137			2, 137		
Total details		0.77	.01		6.74 **	.09		15.30 ***	.10		3.17 *	.04
Accuracy rate		1.24	.01		6.80 **	.09		21.73 ***	.14		2.88	.04
Total correct		1.20	.01		8.27 ***	.11		10.27 **	.07		2.34	.03
Total incorrect		0.77	.01		0.15	.00		34.13 ***	.20		4.52 *	.06
Total confab.		0.07	.00		0.84	.01		0.48	.00		1.54	.02
Total person		0.12	.00		6.87 **	.09		7.05 **	.05		2.58	.04
Total action		4.10 *	.03		4.57 *	.06		16.25 ***	.11		4.52 *	.06
Total object		2.10	.02		0.32	.01		9.17 **	.06		1.46	.02
Total setting		0.00	.00		5.09 **	.07		16.61 ***	.11		1.72	.02
T2 question part	1, 137			2, 137			1, 137			2, 137		
Total details		1.02	.01		2.69	.04		0.61	.00		5.13 **	.07
Accuracy rate		0.62	.01		5.45 **	.07		9.04 **	.06		0.14	.00
Total correct		0.54	.00		4.13 *	.06		0.00	.00		4.86 **	.07
Total incorrect		1.33	.01		1.02	.02		14.58 ***	.10		1.60	.02
Total confab.		0.43	.00		1.00	.01		0.00	.00		0.07	.00
Total person		0.72	.01		3.85 *	.05		0.02	.00		4.89 **	.07
Total action		0.12	.00		0.75	.01		0.44	.00		7.35 **	.10
Total object		4.87 *	.03		0.01	.00		0.48	.00		0.90	.01
Total setting		0.04	.00		1.29	.02		3.96 *	.03		1.69	.02

*p < .05. **p < .01. ***p < .001.

Another question that was addressed was, whether an increase in the amount of total details was at the same time accompanied with an increase in incorrect and confabulated details. Several correlations were run with the variables *total details* and *total incorrect* and *total confabulated details* for the initial recall at T1 and for the witness interview free recall part and question part at T2, both for the total sample and the young and older adults, respectively. Results can be found in table 3.12.

Table 3.12. Pearson correlation coefficients for the total sample and older and young adults separately at T1 and T2.

Total details	Total incorrect	Total confabulated
Total sample		
T1 initial recall (n = 96)	.74 ***	.18
T2 free recall (N = 144)	.54 ***	.02
T2 question part (N = 144)	.37 **	.13
Older adults		
T1 initial recall (n = 48)	.51 ***	.24
T2 free recall (n = 72)	.38 **	-.00
T2 question part (n = 72)	.38 **	.12
Young adults		
T1 initial recall (n = 48)	.85 ***	.30 *
T2 free recall (n = 72)	.64 ***	-.01
T2 question part (n = 72)	.38 **	.15

*p < .05. **p < .01. ***p < .001.

For the total sample, the correlations show that the amount of *total details* and the amount of *total incorrect details* were significantly positively related in all three account-giving opportunities (r = .74, p < .001 for the initial recall at T1, r = .54, p < .001 for the free recall at T2 and r = .37, p < .01 for the question part at T2). However, the amount of *total details* and the amount of *total confabulated details* were not significantly related in any of the three accounts. The same was true when looking at the older adults only: Again, the amount of *total details* and the amount of *total incorrect details* were significantly positively related in all three accounts (r = .51, p < .001 for the initial recall at T1, r = .38, p < .01 for the free recall at T2 and r = .38, p < .01 for the question part at T2), whereas the amount of *total details* and *total confabulated details* were not significantly related at all. For the young adults, again the amount of *total details* and the amount of *total incorrect details* were significantly positively related in all three accounts (r = .85, p < .001 for the initial

recall at T1, $r = .64$, $p < .001$ for the free recall at T2 and $r = .38$, $p < .01$ for the question part at T2). Also, the amount of *total details* and *total confabulated details* was significantly positively related in the initial recall at T1, $r = .30$, $p < .05$.

To get further insight into this question, more analyses were conducted for the three recall conditions separately. Several correlations were run with the variables *total details* and *total incorrect* and *total confabulated details* for the initial recall at T1 and for the witness interview free recall part and question part at T2, both for the total sample and the young and older adults, and for the SAI, wFR and no recall group, respectively. Results can be found in table 3.13.

Table 3.13. Pearson correlation coefficients for total sample and older and young adults at T1 and T2, separated by recall condition.

	SAI		wFR		no	
Total details	Total incorr.	Total confab.	Total incorr.	Total confab.	Total incorr.	Total confab.
Total sample ($n = 48$)						
T1 initial recall	.79 **	.21	.34 *	.22	n.a.	n.a.
T2 free recall	.65 **	.10	.59 **	.21	.53 **	-.10
T2 question part	.45 **	.22	.45 **	.22	.34 *	.05
Older adults ($n = 24$)						
T1 initial recall	.58 **	.32	.31	.20	n.a.	n.a.
T2 free recall	.45 *	.36	.40	.20	.42 *	-.33
T2 question part	.71 **	.04	.42 *	.19	.22	.10
Young adults ($n = 24$)						
T1 initial recall	.88 **	.31	.36	.21	n.a.	n.a.
T2 free recall	.78 **	-.16	.62 **	-.12	.11	.34
T2 question part	.41 *	.30	.49 *	.24	.30	.01

*$p < .05$. **$p < .01$.

For the Self-Administered Interview, correlations show that the amount of *total details* and the amount of *total incorrect details* was significantly positively related across all accounts in the total sample ($r = .79$, $p < .01$ for the initial recall at T1, $r = .65$, $p < .01$ for the free recall at T2 and $r = .45$, $p < .01$ for the question part at T2) as well as young and older adults separately ($r = .58$, $p < .01$ at T1, $r = .45$, $p < .05$ for the free recall at T2 and $r = .71$, $p < .01$ for the question part at T2 for older adults; and $r = .88$, $p < .01$ at T1, $r = .78$, $p < .01$ for the free recall at T2 and $r = .41$, $p < .05$ for the question part at T2 for young adults, respectively).

For the written free recall, correlations show that the amount of *total details* and the amount of *total incorrect details* was again significantly positively related

across all accounts in the total sample (r = .34, p < .05 for the initial recall at T1, r = .59, p < .01 for the free recall at T2 and r = .45, p < .01 for the question part at T2). However, for older adults they were only significantly correlated in the question part at T2, r = .42, p < .05; and for young adults also only at T2 (r = .62, p < .01 for the free recall and r = .49, p < .05 for the question part).

For the no recall condition, correlations show that the amount of *total details* and the amount of *total incorrect details* was again significantly positively related across all possible accounts in the total sample[21] (r = .53, p < .01 for the free recall at T2 and r = .34, p < .05 for the question part at T2). For older adults they were only significantly correlated in the free recall at T2, r = .62, p < .01, and for young adults they were not significantly correlated at all.

Across all three recall conditions, there was no significant correlation between the amount of *total details* and the amount of *confabulated details,* neither in the total sample, nor in the subsamples by age group.

To sum up, according to the overall analyses, the greater the amount of total details was that a witness has given, the more incorrect information was given at the same time, regardless of the age group. When split by recall condition, positive correlations were found for the SAI across all account-giving opportunities in the total sample and in both age groups. For the wFR and the no recall groups, positive correlations were still found across all account-giving opportunities in the total sample, but only for some accounts in the older and young adult subsamples. However, an increase in the total amount of details was not accompanied by an increase in confabulated details in any of the recall conditions.

These results may seem contradictory to the previous t-test and ANOVA results in which mostly no significant effects for an increase in incorrect details were found. This can be explained by the different focus of these analyses. Whereas t-tests and ANOVAs test for differences in group averages, correlations make suggestions about the connection between two variables at the level of the individual. The t-tests and ANOVAs compared group differences across the three recall conditions. Here there was no increase for incorrect details in the SAI as a group compared to the wFR and no recall conditions. However, correlations looked *within each recall condition* at

[21] Since participants did not provide any T1 recall in the no-recall condition, correlations could only be calculated for the interview at T2.

whether more details were related to more incorrect details. Thus these findings can be reconciled in the following way: While within each recall group those participants who reported more details also reported more incorrect details, when comparing the SAI group against the wFR and no recall group there was no significant increase in the number of incorrect details.

Finally, a closer look at the older adult subgroup was taken. In the second session after a one-week delay, older adults in this sample performed surprisingly well in that they produced more overall and more correct details than younger adults (albeit also producing more incorrect details). This raised the question whether the cut-off point for 'older' adults was maybe chosen too low resulting in a large variance within this age group. To address the question whether performance does decline with older age and whether the age of the older adults can contribute significantly to the model, several multiple regression analyses were conducted with recall condition and age as predictors using the enter method and the total variables from the second session recall and question part as the outcome variables. The assumptions of regression include that predictors need to be continuous or categorical with only two categories. As the predictor *recall condition* has three categories for the second session (no, wFR, SAI), a reduction to only two of the categories was considered. Given that the contrast between no initial recall and the SAI is the most interesting one from a practitioner's point of view, only participants who fitted those two groups were included in the analyses (n = 48). The mean age for the remaining older adults was M = 68.44, SD = 5.39 with a range from 60 to 83 years. In total 18 multiple regression analyses were run, 9 each for the recall and question part *total details, accuracy rate, total correct/ incorrect/ confabulated details and total person/ action/ object/ setting details*. Results can be found in table 3.14.

When looking at recall condition, analyses show that whether older adults filled in the SAI or did not have an initial recall only significantly predicted one outcome variable, *accuracy rate* in the free recall, after a one-week delay, β = .48, $p < .01$. Recall condition (no vs. SAI) did not impact any other of the total variables in the witness interview after one week within the older adults subgroup.

When looking at age, analyses show that the age of older participants significantly predicted the *total details*, β = -.35, $p < .05$, and the number of *total correct*, β = -.35, $p < .05$, *total person*, β = -.33, $p < .05$, and *total action*

details, $\beta = -.35$, $p < .05$, in the free recall after a one-week delay, as well as the number of *total person details*, $\beta = -.31$, $p < .05$, in the interview question part after a one-week delay. The negative β-values indicate that with increasing age, the performance levels declined and participants produced significantly fewer total, correct, person and action details.

Table 3.14. Multiple regression analyses summary for recall condition and age predicting total variables at T2.

	Recall condition				Age				
Variable	B	SE B	β		B	SE B	β		R^2
T2 free recall									
Total details	14.23	9.84	.21		-2.23	0.92	-.35	*	.13
Accuracy rate	0.06	0.02	.48	**	-0.00	0.00	-.17		.22
Total correct	17.43	9.29	.27		-2.10	0.87	-.35	*	.14
Total incorrect	-2.67	1.75	-.23		-0.16	0.16	-.14		.09
Total confab.	-0.53	0.49	-.17		0.03	0.05	.08		.03
Total person	8.28	5.25	.23		-1.11	0.49	-.33	*	.12
Total action	1.54	2.59	.09		-0.59	0.24	-.35	*	.12
Total object	-0.43	0.98	-.07		-0.14	0.09	-.22		.06
Total setting	4.84	2.88	.25		-0.40	0.27	-.22		.08
T2 question part									
Total details	-1.34	6.84	-.03		-1.20	0.64	-.28		.08
Accuracy rate	0.03	0.02	.26		-0.00	0.00	-.05		.06
Total correct	1.53	6.38	.04		-1.07	0.60	-.27		.07
Total incorrect	-2.71	1.86	-.22		-0.07	0.17	-.06		.06
Total confab.	-0.16	0.98	-.03		-0.06	0.09	-.10		.01
Total person	1.17	4.01	.04		-0.78	0.38	-.31	*	.09
Total action	-2.34	1.75	-.19		-0.27	0.17	-.24		.12
Total object	-1.24	1.31	-.15		0.04	0.12	.05		.02
Total setting	1.06	1.44	.11		-0.19	0.14	-.21		.05

Note. This analysis contains the older adult subgroup only, $n = 72$.
*p < .05. **p < .01.

3.2.2.5. Interview Question Part: Suggestibility

In addition to the analyses of details in the interview question part, the answers to the specific questions (duration in seconds, yes/no and suggestive questions) were also analysed, the results of which will be presented in the following. The hypotheses that are relevant to this part of the analyses are:

Suggestibility hypothesis

Participants are less susceptible to suggestions at time 2 when first being presented the Self-Administered Interview at time 1 than when being presented with a written free recall or no initial recall at time 1.

Age group hypothesis

Older adults are more susceptible to suggestions in the eyewitness interview compared to young adults.

Interaction hypothesis

The benefit of the SAI for older adults exceeds that of young adults.

From a practitioner's point of view it is again interesting to see a comparison of participants who had completed the SAI ("SAI" condition) and participants who did not have an initial recall ("no" condition, which is the current standard in police investigation). Thus, to start off with, t-tests were conducted for the scores produced in the specific questions for the SAI vs. the no condition. Table 3.15. comprises means, standard deviations, t-tests and corresponding effect sizes for these scores. Results are shown for the total sample as well as for the older and young adults separately.

Starting with the total sample, participants who had completed the SAI a week earlier produced a significantly lower *suggestibility* score ($M = 7.08$, $SD = 3.13$) in the interview question part after a one week delay compared to those who did not have an initial recall ($M = 10.19$, $SD = 4.87$), $t = -3.71$, $p < .001$. The effect size indicated by *Cohens' d* is $d = 0.76$ and thus represents a medium-sized effect. When separating the sample by age-group, none of the variables differed significantly between older adults who had completed the SAI a week earlier and those who had not had any initial recall. In the young adults subsample however, participants who had completed the SAI a week earlier produced a significantly lower *suggestibility* score ($M = 6.42$,

SD = 3.19) in the interview question part after a one week delay compared to those who did not have an initial recall (*M* = 10.54, *SD* = 4.85), *t* = -3.48, *p* < .001. The effect size indicated by *Cohens' d* is *d* = 1.00 and thus represents a large effect.

Table 3.15. Group differences at T2 in the interview question part for specific questions by recall condition (SAI vs. no), for total sample and subsamples.

Group	SAI		no				
Variable	M	SD	M	SD	t		d
Total sample (df = 94)							
Duration	316.25	156.68	296.88	152.22	0.62		0.13
Yes/No questions	5.54	1.60	5.25	1.52	0.92		0.19
Suggestibility	7.08	3.13	10.19	4.87	-3.71	***	0.76
Older adults (df = 46)							
Duration	337.50	139.39	293.75	138.46	1.09		0.31
Yes/No questions	5.29	1.27	4.75	1.48	1.36		0.39
Suggestibility	7.75	3.00	9.83	4.97	-1.76		0.51
Young adults (df = 46)							
Duration	295.00	172.60	300.00	167.79	-0.10		0.03
Yes/No questions	5.79	1.87	5.75	1.42	0.09		0.02
Suggestibility	6.42	3.19	10.54	4.85	-3.48	***	1.00

***$p < .001$.

To further test for age group effects and interaction effects of age group and initial recall, and to comprehensively look at all three recall conditions, several factorial 2 × 3 between-subjects ANOVAs were conducted with initial recall condition in the first session (SAI, wFR, no) and age group as fixed factors and the three specific questions variables from the second session question part as dependant variable, respectively.

In exploring the assumptions of an ANOVA, the Kolmogorov-Smirnov test and Levene's test were again conducted. The Kolmogorov-Smirnov test was significant for these three variables, indicating they were not normally distributed. Levene's test was *non*-significant for the three variables, indicating that variances were equal in these cases. As was previously discussed, the F-statistic is said to be fairly robust to violations of normality and can still perform accurately when group sizes are equal (Field, 2009) and thus, analyses used the original, non-transformed data. The results are shown in table 3.16.

Table 3.16. Factorial ANOVAs at T2 in the interview question part for specific questions.

	Recall condition (R)			Age group (A)		Interaction R x A	
Variable	F^a		Part. η^2	F^b	Part. η^2	F^a	Part. η^2
Duration	0.23		.00	0.00	.00	0.59	.01
Yes/No questions	0.78		.01	0.78	.01	1.18	.02
Suggestive questions	8.33	***	.11	1.80	.01	1.51	.02

[a]*df* = 2, 138. [b]*df* = 1, 138.
****p* < .001.

There was one main effect for recall condition on the variable *suggestive questions*, F (2, 138) = 8.33, $p < .001$. The effect sizes indicated by partial eta squared is $\eta^2_{partial} = .11$, which represents a medium-sized effect (Cohen, 1988). The Tukey HSD *post hoc* test revealed that participants who had filled in the SAI were significantly less suggestible in the question part of the witness interview after one week compared to those who did not have an initial recall option ($p < .001$). Participants who had filled in the written free recall were also significantly less suggestible compared to those who did not have an initial recall option ($p < .01$). Participants' performance in the SAI group and in the written free recall group did not differ significantly from each other.

There were no main effects for age group and no interaction effects between recall condition and age group on any of the three variables.

With regards to suggestibility, analyses of the witness interview question part mainly supported the *suggestibility hypothesis*. Participants who had filled in the SAI a week earlier were indeed less susceptible to suggestions in the eyewitness interview compared to those who did not have an initial recall opportunity. However, participants in the written free recall group were also less susceptible to suggestions than the no group, but did not differ from the SAI group. The *age group hypothesis* and *interaction hypothesis* were not supported. Older adults were not found to be more susceptible to suggestions than young adults. They also did not benefit from the SAI more than young adults.

3.2.2.6. Summary

In this testimony results section, first, analyses of accounts given in the first experimental session were presented, i.e. a comparison of details produced in the Self-Administered Interview and in the written free recall. Results mainly supported the *initial recall hypothesis*, in that participants produced a more detailed and accurate account in the Self-Administered Interview compared to the simple written free recall. However, they also gave more incorrect details. Results fully supported the *age group hypothesis*, as older adults gave less detailed and less accurate accounts in the SAI and the wFR compared to young adults. The *interaction hypothesis* was not supported as results did not show a greater benefit of the SAI for older adults than for young adults. Young and older adults performed similarly in the wFR, but young adults outperformed older adults significantly in the SAI.

Then the witness accounts given in the second experimental session after a one week delay were analyzed, separated by free recall, question part and answers to the specific questions. Results of the free recall fully supported the *interview transfer hypothesis*. Participants who had filled in the SAI a week earlier produced a more detailed and more accurate account compared to those who did not have an initial recall opportunity. Participants in the written free recall group neither performed significantly better than the no group, nor significantly worse than the SAI group. The *age group hypothesis* was only partly supported. Contrary to the prediction, older adults gave a more detailed account with also more correct details in the free recall compared to young adults. In accordance with the prediction, older adults also produced more incorrect details and had an overall lower accuracy rate than young adults. The *interaction hypothesis* was not supported. Contrary to the prediction, the benefit of the SAI was higher for young adults than for older adults.

When analyzing the witness' answers to interview questions, it was found that results fully supported the *interview transfer hypothesis*. Participants who had filled in the SAI a week earlier produced a more detailed and more accurate account compared to those who did not have an initial recall opportunity. Participants in the written free recall group neither performed significantly better than the no group, nor significantly worse than the SAI group. The *age group hypothesis* was mainly supported. In accordance with the prediction, older adults also produced more incorrect details and had an overall lower accuracy rate than young adults. The only findings contradicting

the prediction were that older adults gave a more correct setting details compared to young adults. The *interaction hypothesis* was not supported. Contrary to the prediction, the benefit of the SAI was higher for young adults than for older adults.

And finally, some further considerations and analyses were presented. It was shown that the chosen covariates *ethnicity, university degree* and *alertness* did have an impact on testimony performance. However, this impact was small and only sporadically relevant compared to the persistent experimental effects of recall condition and age group. It was furthermore shown that an increase in the amount of details a witness presents is unfortunately coupled with an increase in incorrect details as well. And finally, a closer look at the older adult subgroup was taken and established that with increasing age, performance levels declined and participants produced significantly fewer total, correct, person and action details.

In the end, when looking at the specific questions, it was found that results of the question part also mainly supported the *suggestibility hypothesis*. Participants who had filled in the SAI a week earlier were indeed less susceptible to suggestions in the eyewitness interview compared to those who did not have an initial recall opportunity. However, participants in the written free recall group were also less susceptible to suggestions than the no group, but did not differ from the SAI group. With regard to suggestibility, the *age group hypothesis* and *interaction hypothesis* were not supported. Older adults were not found to be more susceptible to suggestions than young adults. They also did not benefit from the SAI more than young adults.

3.2.3. Person Identification

In this section, first the results of the lineup performance are presented. This is followed by analyses of the confidence-accuracy relationship as well as the impact of the SAI on the confidence-accuracy relationship. In the end, some further analyses will be illustrated and a brief summary again completes this section.

3.2.3.1. Lineup Performance

The hypotheses that are relevant to the lineup performance part of the results chapter are:

Identification hypothesis

Participants perform better in the person identification task at time 2 when first being presented the Self-Administered Interview at time 1 compared to a written free recall or no initial recall at time 1.

Age-group hypothesis

Older adults perform worse in the person identification task compared to young adults.

Interaction hypothesis

The benefit of the SAI for older adults exceeds that of young adults.

Frequencies and percentages of the lineup performance for the total sample broken down by recall condition and target presence are shown in table 3.17. Target present (TP) lineups are differentiated by correct identification, false identification and false rejection, whereas target absent (TA) lineups are differentiated by correct rejection and false identification. Two things may be worth noting: In target present lineups, 25.0% of participants who had filled in the SAI a week earlier correctly identified the perpetrator, whereas only 10.4% in the wFR and in the no group, respectively, correctly identified him. This difference is significant, χ^2 (1) = 4.46, $p < .05$, for both comparisons SAI vs. wFR and SAI vs. no. And secondly, when the target was not in the lineup, participants in all three recall conditions performed on a similar level (20.8%, 22.9% and 20.8% correct rejection rates for SAI, wFR and no group, respectively), χ^2 (1) = 0.09, p = .50 (SAI vs. wFR) and χ^2 (1) = 0.00, p = .62 (SAI vs. no).

Table 3.17. Lineup performance by recall condition and target presence for total sample.

	SAI		wFR		no		total	
Target present								
Correct identification	12	(25.0%)	5	(10.4%)	5	(10.4%)	22	(15.3%)
False identification	5	(10.4%)	12	(25.0%)	4	(8.3%)	21	(14.6%)
False rejection	7	(14.6%)	7	(14.6%)	15	(31.3%)	29	(20.1%)
Target absent								
Correct rejection	10	(20.8%)	11	(22.9%)	10	(20.8%)	31	(21.5%)
False identification	14	(29.2%)	13	(27.1%)	14	(29.2%)	41	(28.5%)
Total								
Correct	22	(45.8%)	16	(33.3%)	15	(31.3%)	53	(36.8%)
Incorrect	26	(54.2%)	32	(66.7%)	33	(68.8%)	91	(63.2%)

Note. Data is shown in frequencies, with percentages in brackets.

Frequencies and percentages of older and young adults' lineup performance separately, broken down by recall condition and target presence are shown in table 3.24. Again, TP lineups are differentiated by correct identification, false identification and false rejection, whereas TA lineups are differentiated by correct rejection and false identification. Two things may be worth noting. First, young adults significantly outperformed the older adults in that their overall rate of correct choices across recall conditions was 45.8%, whereas older adults had a total correct of only 27.8%, χ^2 (1) = 5.05, p < .05. Secondly, both age group subsamples performed on a similar level across the three recall conditions in target absent lineups, i.e. *older adults'* correct rejection rates were 16.7%, 12.5% and 20.8% for the SAI, wFR and no recall group, respectively, χ^2 (1) = 0.20, p = .50 (SAI vs. wFR) and χ^2 (1) = 0.18, p = .50 (SAI vs. no). Similarly, *young adults*' correct rejection rates were 25.0%, 33.3% and 20.8% for the SAI, wFR and no recall group, respectively, χ^2 (1) = 0.69, p = .34 (SAI vs. wFR) and χ^2 (1) = 0.17, p = .50 (SAI vs. no). Figure 3.12. shows the percentage of correct choice by age group and recall condition to further illustrate the findings.

Table 3.18. Lineup performance by recall condition and target presence for older and young adults separately.

	SAI		wFR		no		total	
Older adults								
Target present								
Correct identification	5	(20.8%)	2	(8.3%)	1	(4.2%)	8	(11.1%)
False identification	4	(16.7%)	7	(29.2%)	4	(16.7%)	15	(20.8%)
False rejection	3	(12.5%)	3	(12.5%)	7	(29.2%)	13	(18.1%)
Target absent								
Correct rejection	4	(16.7%)	3	(12.5%)	5	(20.8%)	12	(16.7%)
False identification	8	(33.3%)	9	(37.5%)	7	(29.2%)	24	(33.3%)
Total								
Correct	9	(37.5%)	5	(20.8%)	6	(25.0%)	20	(27.8%)
Incorrect	15	(62.5%)	19	(79.2%)	18	(75.0%)	52	(72.2%)
Young adults								
Target present								
Correct identification	7	(29.2%)	3	(12.5%)	4	(16.7%)	14	(19.4%)
False identification	1	(4.2%)	5	(20.8%)	0	(0.0%)	6	(8.3%)
False rejection	4	(16.7%)	4	(16.7%)	8	(33.3%)	16	(22.2%)
Target absent								
Correct rejection	6	(25.0%)	8	(33.3%)	5	(20.8%)	19	(26.4%)
False identification	6	(25.0%)	4	(16.7%)	7	(29.2%)	17	(23.6%)
Total								
Correct	13	(54.2%)	11	(45.8%)	9	(37.5%)	33	(45.8%)
Incorrect	11	(45.8%)	13	(54.2%)	15	(62.5%)	39	(54.2%)

Note. Data is shown in frequencies, with percentages in brackets.

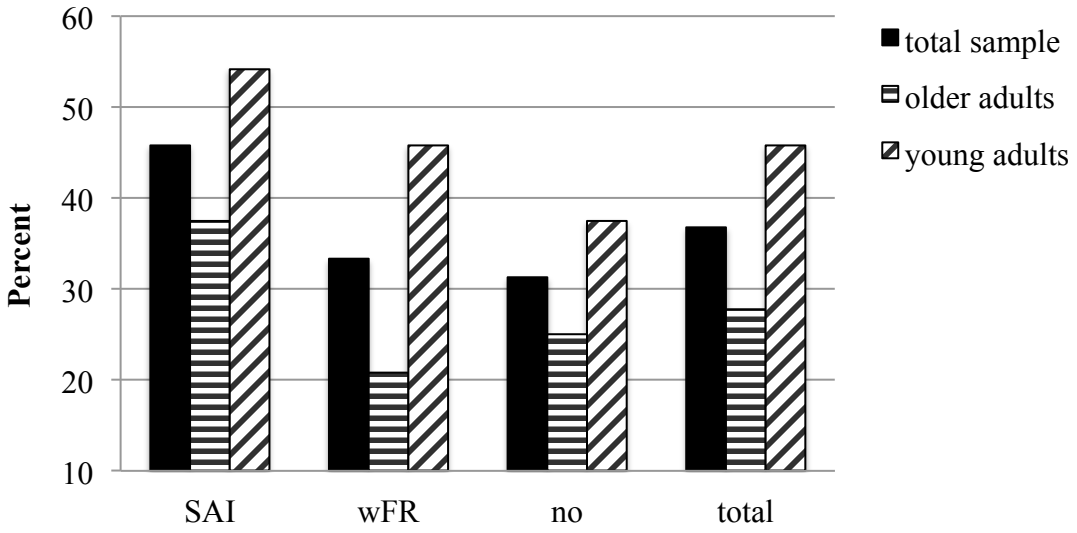


Figure 3.12. Percentage of correct choice by age group and recall condition.

To examine the effect of recall condition, age group and target presence on lineup performance a hierarchical log-linear analysis (HILOG, saturated hierarchical backward elimination method) was conducted. This is the standard statistical procedure to deal with three or more categorical variables (Field, 2009) as it combines features of standard chi-square tests (i.e. determining the fit between observed and expected cell counts) with features of ANOVA (i.e. simultaneous testing of main effects and interactions in multi-factorial designs). It is therefore frequently used in eyewitness identification research (see e.g. Memon & Gabbert, 2003; Wilcock & Bull, 2010). The model in this study included recall condition (SAI/wFR/no), age group (older/young adults), target presence (target present/absent) and lineup performance (correct/incorrect). Correct lineup performance meant here that it included correct identifications in a TP lineup and correct rejections in a TA lineup. Incorrect performance included false rejections and false identifications in a TP lineup and false identifications in a TA lineup. The statistical assumptions for loglinear analysis were met, which means that all cells of the contingency table were independent (one person only contributed to one cell), and the expected frequencies in every cell were greater than 1 with less then 20% being smaller than 5 (16.67%; Field, 2009).

The four-way loglinear analysis produced a final model with a likelihood ratio χ^2 (18) = 9.61, p = .94. This non-significance indicated a good fit between the expected frequencies generated by the model and the observed frequencies, meaning they were not significantly different and that the model is a good fit of the data. The final model revealed that the main effects of recall condition, age group, target presence and lineup performance were approaching significance, χ^2 (5) = 10.15, p = .07. No interaction effects were observed. Although these effects were not significant per se, it is still important to follow-up HILOG with chi-square analyses (Field, 2009). Table 3.19. thus shows the percentage of correct choice by recall condition, and the corresponding χ^2 and odds ratios for the total sample, young and older adults, target present and absent lineups. Furthermore, the table comprises data for the subsample of "choosers" and "non-choosers". This illustrates a different approach in analyzing lineup data as proposed by Sporer et al. (1995) and differentiates between participants who made a choice, i.e. in TP lineups a correct identification or a false identification and in TA lineups a false identification, and

participants who did not make a choice, i.e. in TP lineups falsely rejected the lineup and in TA lineups correctly rejected the lineup. Odds ratios are given for the comparison of the SAI versus the no initial recall group. As previously presented in the testimony results section, this is the most interesting comparison from a practitioner's point of view – the "no" condition being the current standard in police investigation and the SAI being the proposed new tool.

Table 3.19. Percentage of correct choice by recall condition for the total sample and several subsamples.

Group	SAI	wFR	no	χ^2 (2)		SAI vs. no OR (95% CI)	
Total (N = 144)	45.8%	33.3%	31.3%	2.57		1.86	(0.81-4.28)
Older adults (n = 72)	37.5%	20.8%	25.0%	1.80		1.80	(0.52-6.22)
Young adults (n = 72)	54.2%	45.8%	37.5%	1.34		1.97	(0.62-6.23)
Target present (n = 72)	50.0%	20.8%	20.8%	6.42	*	3.80	(1.07-13.52)
Target absent (n = 72)	41.7%	45.8%	41.7%	0.11		1.00	(0.32-3.15)
Choosers (n = 88)	39.4%	19.4%	20.8%	3.93		2.27	(0.67-7.75)
Non-choosers (n = 56)	60.0%	58.8%	41.7%	1.73		2.14	(0.61-7.51)

*$p < .05$.

Generally, the observed lineup performance reflects the hypothesized tendency that participants perform better when having been presented the SAI compared to having been presented with a written free recall or no initial recall. This tendency is true for the total sample, the older and young adults subsamples, the target present subsample and both the choosers and non-choosers subsample; the only exception being the target absent subsample, in which participants in all three recall conditions performed on a similar level. A significant effect was found for target present lineups, χ^2 (2) = 6.42, $p < .05$. Here, the overall odds of correctly identifying the perpetrator were 3.8 times higher if participants had been given the SAI (50.0%) compared to a written free recall (20.8%) or no initial recall (20.8%; see figure 3.13.).

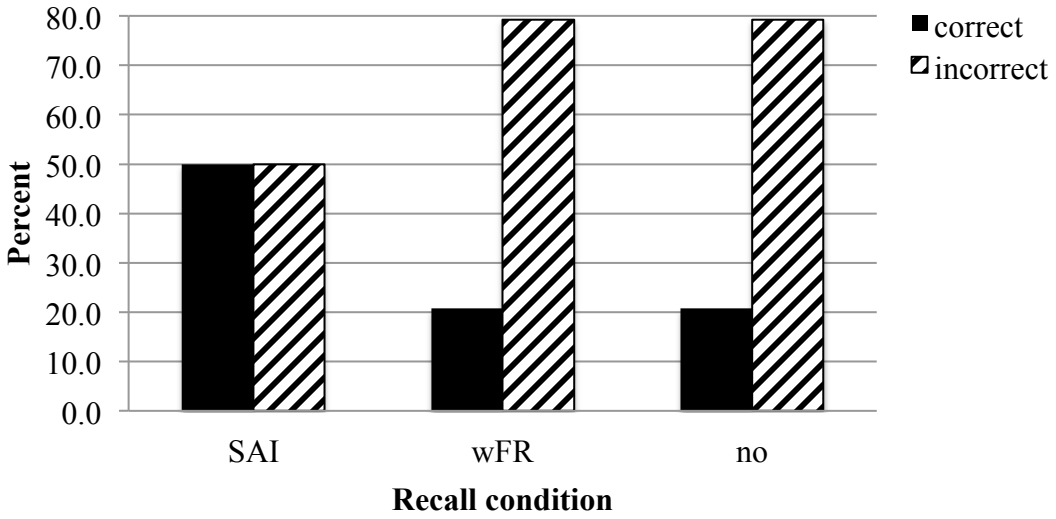


Figure 3.13. Percentage of correct and incorrect choices in target present lineups across the total sample by recall condition.

For further examination of the lineup performance, data from both control groups (wFR and no recall) were collapsed and compared with the SAI group. This gives a better understanding of the potential of the SAI over any control condition. Table 3.20. shows the percentage of correct choice for the SAI vs. any control condition, and the corresponding χ^2 and odds ratios for the total sample, young and older adults, target present and absent lineups, choosers and non-choosers. In addition to mirroring the findings above for the three recall groups separately, collapsing the two control conditions resulted in a significant effect for the chooser subsample, $\chi^2(1) = 3.91$, $p < .05$. This means that if a witness has chosen someone from the lineup, the overall odds of having correctly identified the perpetrator were 2.6 times higher if participants had been given the SAI (39.4%) compared to any of the control groups (20%).

Table 3.20. Percentage of correct choice by recall condition (SAI vs. any control) for the total sample and several subsamples.

Group	SAI	Control	χ^2 (1)		OR (95% CI)
Total (N = 144)	45.8%	32.3%	2.52		1.77 (0.87-3.61)
Older adults (n = 72)	37.5%	22.9%	1.70		2.02 (0.70-5.86)
Young adults (n = 72)	54.2%	41.7%	1.01		1.65 (0.62-4.44)
Target present (n = 72)	50.0%	20.8%	6.42	*	3.80 (1.32-10.98)
Target absent (n = 72)	41.7%	43.8%	0.03		0.92 (0.34-2.48)
Choosers (n = 88)	39.4%	20.0%	3.91	*	2.60 (0.99-6.80)
Non-choosers (n = 56)	60.0%	48.8%	0.55		1.58 (0.47-5.23)

Note. 'Control' means data is collapsed over wFR and no recall conditions.
*$p < .05$.

Thus, analyses of the person identification part of this study mainly supported the *identification hypothesis*. Participants who had filled in the SAI a week earlier performed better in a lineup compared to those who did not have an initial recall opportunity or who had filled in a written free recall. In target present lineups, the advantage of the SAI over no initial recall made a significant difference in correctly choosing the perpetrator. However, there was no performance difference between recall conditions in target absent lineups. The *age group hypothesis* was also supported in that older adults performed worse in the lineup task, both in target present and in target absent conditions, compared to young adults. The *interaction hypothesis* was tentatively supported. Although there was no significant evidence of a stronger benefit of the SAI for older adults compared to young adults, the data showed a tendency in this direction.

3.2.3.2. Confidence-Accuracy Relationship

Another focus of this study was to examine the confidence-accuracy relationship of both older and young eyewitnesses. The relevant hypothesis here is the *confidence-accuracy hypothesis,* stating that *p*ost-identification confidence is not related to accuracy of identification.

Table 3.21. comprises the results of several group comparisons for participants' post-identification confidence ratings by lineup performance (correct/incorrect) for the total sample and the subsamples older and young adults, target present, absent, choosers and non- choosers.

Table 3.21. Group differences for lineup confidence by lineup performance for the total sample and several subsamples.

	Correct		Incorrect			
Group	M	SD	M	SD	df	t
Total (N = 144)	4.00	1.35	3.89	1.49	142	0.44
Older adults (n = 72)	3.60	1.19	3.85	1.58	70	-0.63
Young adults (n = 72)	4.24	1.39	3.95	1.40	70	0.89
Target present (n = 72)	4.05	1.33	4.04	1.55	70	0.01
Target absent (n = 72)	3.97	1.38	3.71	1.42	70	0.78
Choosers (n = 88)	4.04	1.27	3.91	1.46	86	0.40
Non-choosers (n = 56)	3.97	1.43	3.85	1.61	54	0.28

Note. All t-values are non-significant.

There was no significant difference in participants' post-identification confidence between those who performed correctly and those who did not. Neither across the total sample, nor in any of the subsamples, were people who correctly identified the perpetrator or correctly rejected the lineup more confident than those who falsely identified a filler or falsely rejected the lineup. In fact, among those participants who rated themselves to be "*confident*" about their choice (i.e. they selected 5, 6 or 7 on a Likert scale from 1 to 7, $n = 48$), twice as many had made an incorrect choice ($n = 32$, $M = 5.56$, $SD = 0.76$) compared to those who had made the correct choice ($n = 16$, $M = 5.63$, $SD = 0.72$). Even more interestingly, when only looking at older adults who rated themselves to be confident about their choice ($n = 20$), only 2 ($M = 6.00$, $SD = 1.41$) actually made the correct choice, whereas 18 ($M = 5.61$, $SD = 0.78$) made an incorrect choice from the lineup. Within the confident

young adults subsample (n = 28) on the other hand, the distribution is equal with 14 confident young adults having made the correct (M = 5.57, SD = 0.65) and 14 having made an incorrect choice (M = 5.50, SD = 0.76). Figure 3.14. shows the frequencies of lineup performance among confident participants for the total sample, and older and young adults subsamples.

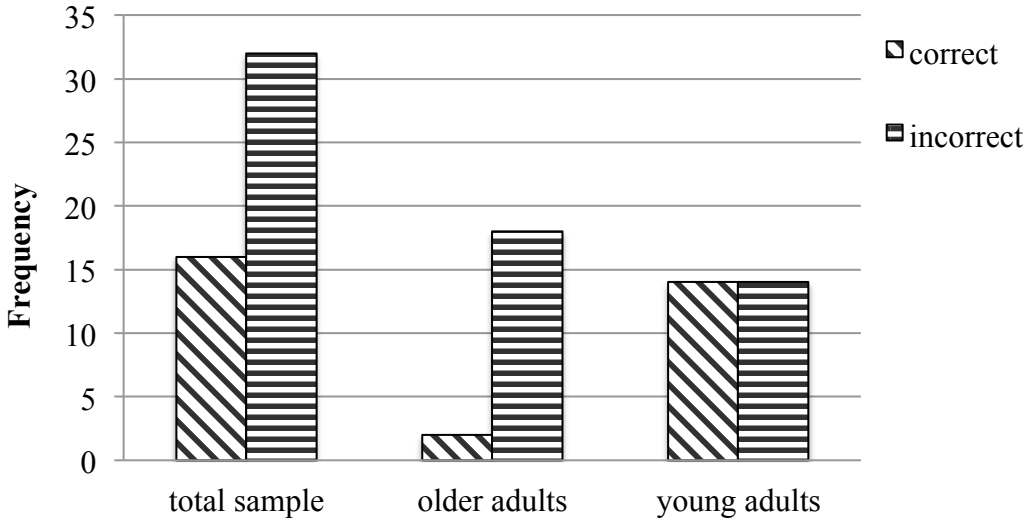


Figure 3.14. Frequencies of lineup performance among confident participants for the total sample, and older and young adults subsamples.

When separating the confident participants into those who viewed a target present (n = 29) and those who viewed a target absent lineup (n = 19), the rate is again 2:1. In target present lineups, more than twice as many participants (n = 20) falsely identified a filler from the lineup or falsely rejected the lineup while being confident they had actually chosen the perpetrator (M = 5.60, SD = 0.75); compared to those who correctly identified the perpetrator while being confident that this choice was correct (M = 5.33, SD = 0.50, n = 9). In target absent lineups, n = 12 participants falsely identified a filler from the lineup while being confident they had actually chosen the perpetrator (M = 5.50, SD = 0.80); whereas only n = 7 participants correctly rejected the lineup while being confident that this choice was correct (M = 6.00, SD = 0.82). When looking at participants who identified someone from the lineup with confidence (choosers, n = 32), only 9 (M = 5.33, SD = 0.50) actually correctly identified the perpetrator, whereas 23 (M = 5.52, SD = 0.73) falsely identified a filler (from either a TP or a TA lineup). Among confident non-choosers

(n = 16), the distribution is more equal. Here, 7 confident participants correctly rejected the lineup (M = 6.00, SD = 0.82) and 9 confident participants incorrectly rejected the lineup (M = 5.67, SD = 0.87). Figure 3.15. shows the frequencies of lineup performance among confident participants for target present, target absent, choosers and non-choosers subsamples.

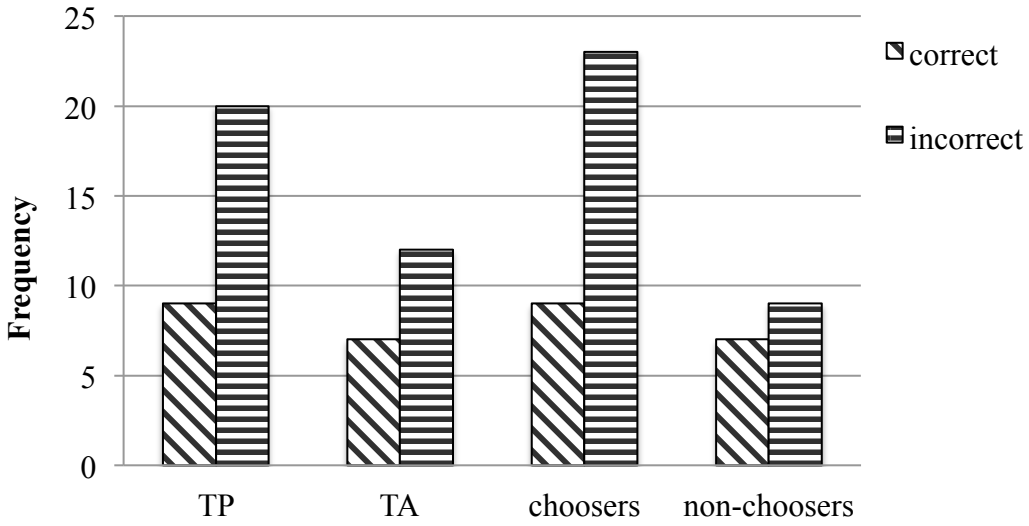


Figure 3.15. Frequencies of lineup performance among confident participants for target present, target absent, choosers and non-choosers subsamples.

Thus, analyses regarding the post-identification confidence and accuracy of identification fully supported the *confidence-accuracy hypothesis*. There was no significant difference in participants' post-identification confidence between those who performed correctly and those who did not.

3.2.3.3. The SAI and the Confidence-Accuracy Relationship

Finally, to test whether the SAI had any impact on the confidence-accuracy relationship, several factorial 2 × 3 between-subjects ANOVAs were run with lineup performance (correct/incorrect) and recall condition (SAI, wFR, no) as fixed factors and lineup confidence as dependent variable. In total seven ANOVAs were conducted, one for the total sample, older adults, young adults, target present, target absent, choosers and non-choosers subsample each. The relevant hypothesis here is the *SAI confidence-accuracy hypothesis,* stating that the Self-Administered Interview does not impact the confidence-accuracy relationship.

In exploring the assumptions of each ANOVA, the Kolmogorov-Smirnov test and Levene's test were conducted. The Kolmogorov-Smirnov test was significant for the dependent variable in all ANOVAs ($p < .001$), indicating that lineup confidence was not normally distributed in the total sample or any of the subsamples. Levene's test of equality of variances however was non- significant ($p > .05$) in all ANOVAs, indicating roughly equal variances. With the F-statistic being fairly robust to violations of normality (Field, 2009), analyses used the original, non-transformed data. Results of the factorial ANOVAs are presented in table 3.22.

No main effects were found for recall condition or lineup performance, as well as no interaction effects between the two factors.

Thus, analyses of the impact of the SAI on the confidence-accuracy relationship fully support the *SAI confidence-accuracy hypothesis*. Participants who had filled in the SAI a week earlier and who subsequently performed correctly on the lineup were *not* more confident than those who had filled in a written free recall or did not have any initial recall and still performed correctly.

Table 3.22. Means, standard deviations and ANOVA results for lineup confidence as dependent variable and recall condition and lineup performance as fixed factors for the total sample and several subsamples.

Group	Correct		Incorrect		ANOVA, F-value		
Condition	M	SD	M	SD	RC[a]	LP[b]	RC x LP
Total					0.02	0.20	0.57
SAI	3.86	1.46	4.12	1.31			
wFR	4.13	1.36	3.75	1.59			
no	4.07	1.22	3.85	1.56			
Older adults					0.86	0.40	1.27
SAI	3.33	1.12	4.27	1.28			
wFR	3.20	0.84	3.58	1.81			
no	4.33	1.37	3.78	1.56			
Young adults					0.38	0.65	0.25
SAI	4.23	1.59	3.91	1.38			
wFR	4.55	1.37	4.00	1.23			
no	3.89	1.17	3.93	1.62			
Target present					0.08	0.05	0.75
SAI	3.83	1.47	4.17	1.53			
wFR	4.60	1.14	3.79	1.62			
no	4.00	1.23	4.21	1.55			
Target absent					0.20	0.60	0.61
SAI	3.90	1.52	4.07	1.14			
wFR	3.91	1.45	3.69	1.60			
no	4.10	1.29	3.36	1.50			
Chooser					0.23	0.33	0.82
SAI	3.85	1.41	4.20	1.20			
wFR	4.50	1.05	3.84	1.63			
no	4.00	1.23	3.68	1.49			
Non-chooser					0.36	0.18	0.11
SAI	3.89	1.62	3.83	1.72			
wFR	3.90	1.52	3.43	1.51			
no	4.10	1.29	4.07	1.69			

Note. All F-values are non-significant.
[a] = Recall condition, [b] = Lineup performance.

3.2.3.4. Examination of Possible Alternative Hypotheses

In order to examine data for possible alternative hypotheses, the impact of several demographic and background variables on identification performance was tested. Group differences are shown in table 3.23. and were investigated using the appropriate statistical significance testing (independent t-test for metric data, Mann-Whitney Test for ordinal data and Chi-Square Test for nominal data, see Field, 2009). None of the variables tested revealed a significant difference between participants who performed correctly and those who performed incorrectly and showed again that randomization appears to have been successful in establishing equal groups.

Table 3.23. Demographic variables and background characteristics by lineup performance.

Variable	Lineup performance		Significance test
	Correct (n = 53)	Incorrect (n = 91)	
Age	M = 42.04 (SD = 22.57)	M = 49.31 (SD = 22.76)	t = -1.85
N male/female[b]	19/34	33/58	χ^2 = 0.00
Ethnicity (% white)[b]	91	91	χ^2 = 0.02
N home visit yes/no	3/50	3/88	χ^2 = 0.47
N university yes/no[b]	46/7	78/13	χ^2 = 0.03
Health rating	Mdn = 3.00	Mdn = 4.00	U = 2094.00
N sports yes/no[b]	49/4	84/7	χ^2 = 0.00
Sports frequency	Mdn = 6.00	Mdn = 6.00	U = 2142.50
Alertness rating	Mdn = 4.00	Mdn = 5.00	U = 2232.00
MMSE[c,d]	M = 15.45 (SD = 0.89)	M = 15.46 (SD = 0.61)	t = -0.06
MEQ[a]	M = 51.91 (SD = 10.71)	M = 55.21 (SD = 10.11)	t = -1.85
CESD-R[a]	M = 9.43 (SD = 8.12)	M = 7.20 (SD = 5.37)	t = 1.79
SUDOKU	Mdn = 6.00	Mdn = 5.00	U = 2396.50

Note. All significance test values are non-significant.
[d]older adults only.
[a]df = 142. [b]df = 1. [c]df = 70.

3.2.3.5. Summary

In this person identification results section, first, analyses of the lineup performance were presented. Results mainly supported the *identification hypothesis*, in that participants who had filled in the Self-Administered Interview performed better in a lineup after a one-week delay compared to those who did not have an initial recall opportunity or who had filled in a written free recall. The *age group hypothesis* was also supported in that older adults performed worse in the lineup task, both in target present and in target absent conditions, compared to young adults. However, there was no significant evidence of a stronger benefit of the SAI for older adults compared to young adults, despite the data showing a tendency in this direction and thus only tentatively supporting the *interaction hypothesis*.

This was followed by analyses of the confidence-accuracy relationship as well as the impact of the SAI on the confidence-accuracy relationship. Results fully supported the *confidence-accuracy hypothesis* in that there was no significant difference in participants' post-identification confidence between those who performed correctly and those who did not. Results also fully supported the *SAI confidence-accuracy hypothesis*. Participants who had filled in the SAI a week earlier and who subsequently performed correctly on the lineup were *not* more confident than those who had filled in a written free recall or did not have any initial recall and still performed correctly.

In the end, further analyses revealed no significant difference on demographic variables between participants who performed correctly and those who performed incorrectly and showed again that randomization appears to have been successful in establishing equal groups.

3.3. Discussion

In order to examine the potential benefits of the Self-Administered Interview for older adult witnesses and test its impact on both witness testimony and person identification performance, a between-subjects laboratory experiment with 144 participants was conducted. Half of the participants were 60 years and over and half were between 18 and 30 years old. They were all tested individually in two sessions. In the first session, they watched a film with a staged crime and filled in some questionnaires as filler tasks. After a 15minutes interval, depending on which study condition the participant was randomly assigned to, they could either go home, or were either asked to fill in a written free recall form or the SAI and then could go home. In the second session after one week, all participants were interviewed individually. They were first asked to give a free recall of what they remember and then also asked a series of specific questions including suggestive questions. They then filled in some questionnaires as filler tasks, and after a 15minutes interval, half of them was presented with either a target present or a target absent simultaneous photo lineup, respectively.

For the *testimony* part of this study, the above-mentioned design resulted in a 3 (recall condition) x 2 (age group) factorial design with 24 participants in each group. For the *person identification* part of this study, this theoretically resulted in a 3 (recall condition) x 2 (age group) x 2 (target presence) factorial design with 12 participants in each group. However, analyses were conducted in a 3 (recall condition) x 2 (*either* age group *or* target presence) factorial design in order to retain a high statistical power of 0.97.

3.3.1. The Impact of the SAI on Testimony

In the following, a number of questions reflecting the testimony results will be discussed, starting with whether the Self-Administered Interview may be beneficial compared to not having any comprehensive initial recall, which represents the current police practice. The second question focuses on whether the SAI may be superior to a simple written free recall instruction. The third question looks at the performance of older adults compared to young adult witnesses, independent of recall condition. The fourth question discusses the impact the SAI may have on older adults, and the last

question focuses on whether the SAI can reduce susceptibility to suggestions in witness interviews.

3.3.1.1. Is there a Benefit of the SAI over Current Police Practice?

The current practice in British police forces is to give witnesses a brief initial interview after a crime and then call them into the station later for the investigative interview. Research has shown that this initial questioning can have a detrimental effect on testimony elicited in a subsequent interview (Marsh et al., 2005; Tversky & Marsh, 2000). Furthermore, surveys with police officers have revealed that there is likely to be a delay before a full interview can be administered (e.g. Kebbell et al., 1999). Depending on the length of the delay this leaves a possibly large window for memory to fade and post-event information to disturb the information that was originally encoded. The SAI was set out to gather a comprehensive initial statement and, more importantly, to help protect and strengthen a witnesses' memory for a later recall. Thus, one aim of this study was to examine whether the SAI can improve testimony in a delayed interview and thus replicate previous findings.

In a direct comparison between those participants who had filled in the SAI and those who did not have an initial recall opportunity, the SAI group showed a clear advantage in the investigative interview. In the *free recall*, SAI participants remembered significantly more total details, especially more correct, person, action and setting details than no-recall participants. Importantly, filling in the SAI did not increase the amount of incorrect or confabulated details in the later interview and thus, the SAI group had an overall higher accuracy rate than the no-recall group.

In the subsequent *question part* of the investigative interview, again the SAI group remembered significantly more total details, especially more correct and person details compared to the no-recall group, without increased rates of incorrect details, which resulted in a higher accuracy rate overall.

This finding is consistent with prior studies on the Self-Administered Interview that also found lasting effects after a delay. Gabbert et al. (2009) showed in their first study on the SAI that it significantly increased correct details, but not incorrect details after one week and resulted in a higher accuracy rate. Paterson et al. (2015) also found an increase in correct details after two weeks, but no increase in incorrect ones with the SAI compared to no initial recall.

It is well established, that retrieval of details from episodic memory increases the likelihood of recalling these details in subsequent attempts. It seems however important that the first retrieval that is generated is as detailed as possible and of good quality, as this facilitates later recall (Marsh et al., 2005; Pansky & Nemets, 2012; Shaw et al., 1995). The so-called associative networks of memory (J. R. Anderson, 1983) provide a theoretical basis for these findings. According to this theory, memory is seen as a network consisting of nodes that represent concepts and share associative links. Importantly, the quality of the initial coding determines the strength of the associative link, and subsequent retrieval then further strengthens these links across episodic memory. When the strength falls below a certain threshold, the information contained in the respective node can no longer be accessed. It follows, that a more extensive retrieval attempt, as it is achieved with the Self-Administered Interview, leads to increased activation levels of the encoded details and also the associations between details, and thus supports subsequent retrieval of details.

This is a highly encouraging finding for police and practitioners, as it clearly shows that in comparison to the current practice, filling in the Self-Administered Interview is beneficial for memory performance and police investigations. Moreover, the content of the SAI can be an additional source of information for the investigative process and help determine who may be a key witness or remembers critical information. This may however also be achieved with a simpler written free recall and leads to the question whether the SAI is superior to that.

3.3.1.2. Is the SAI Superior to a Written Free Recall?

Another aim of this study was to test whether the Self-Administered Interview holds any benefit over another initial recall tool, namely a written free recall, both in an initial comparison of the details elicited in both tools, and in a comparison of the testimony gained in a subsequent interview. From a practitioners' point of view, even easier than handing out the SAI booklet to witnesses, would be to just give them a piece of paper and instruct them to write down everything they can remember. There would be no need to carry (a potentially large number of) SAI booklets with them. Moreover, filling in a wFR would arguably be quicker than a SAI. Would such a written free recall be sufficient to enhance memory performance?

When comparing the testimony produced in the Self-Administered Interview with that produced in a written free recall, i.e. the content of both tools, results showed that participants remembered more details in total, and specifically more person and action details. This means that they described what the perpetrator, victims and other witnesses looked like and what they were doing in more depth in the SAI than in a wFR. In doing that, they produced more correct details, but also more incorrect details. The increase in correct details was however larger than the increase in incorrect details and, importantly, the overall accuracy rate was unaffected by recall condition.

Some previous studies showed similar patterns of a simultaneous increase in correct and incorrect details (e.g. Hope et al., 2014), whereas others only found an increase in correct details (Gabbert et al., 2009; Krix et al., 2016). The accuracy rate seems however consistently unaffected by whether the participants were given a SAI or a wFR (Gabbert et al., 2009; Hope et al., 2014; Krix et al., 2016). The increase in correct details suggests that the Self-Administered Interview and its specific instructions and prompts facilitate recall immediately after the event. Again, this may be due to an increase in activation levels of the encoded details and their associations within the associative memory network (J. R. Anderson, 1983). The results further suggest that a simple written free recall task does not lead to increased activation levels, at least not to the same extent as the SAI.

More interesting still is of course the transfer to the second session and the effect of the SAI on a delayed investigative interview in comparison to the effect of a simple written free recall instruction. In the *interview free recall*, wFR-participants did not perform significantly worse than SAI-participants. However, they did not perform significantly better than the no-recall participants either, whereas SAI-participants did. In other words, participants who had filled in the SAI remembered more total details, correct details and person details after one week than participants who did not fill in any initial recall. Participants who had filled in the wFR failed to remember more details than the no-recall ones. The same pattern was found for the details given in the *interview question part*. It may also be noted that in contrast to the initial comparison after the event, in the delayed investigative interview, there was no increase in incorrect details for the SAI participants.

It can thus be concluded that filling in the SAI immediately after a crime leads to more correctly remembered details than giving a simple written free recall, both in

the content of the SAI and in the testimony given after one week. An increase in incorrectly remembered details was also found in an immediate comparison; however, this effect did not transfer to the investigative interview after one week. This finding supports the note that only a high quality and comprehensive initial recall facilitates and benefits subsequent recall attempts, such as in an investigative interview. This is also a promising result in light of other interviewing tools that are frequently found to increase the amount of incorrect details, such as the Cognitive Interview (see Memon et al., 2010 for a meta-analysis). Whereas the CI has been found to increase errors in remembering details of a crime during the investigative interview (in which the CI is applied), the SAI seems to only increase correctly remembered details. Completing the SAI thus supports subsequent retrieval of correct details that might otherwise have become inaccessible and forgotten.

3.3.1.3. Old versus Young – Some Surprises

Thus far the discussion focused on results referring to both young and older adults combined. Typically, older adults perform worse in memory tasks than young adults. Thus, one aim of this study was to examine to which extent there would be a difference between older adult witnesses and young adult witnesses in remembering the details of a staged crime when recalling the event initially and after one week.

Immediately after watching the staged crime film, older adults remembered significantly fewer total details, person details and correct details than young adults, confabulated more details and had an overall lower accuracy rate. This finding is consistent with the broader literature on memory decline in older adults (Balota et al., 2000) and with findings specifically on eyewitness accounts (Coxon & Valentine, 1997; Yarmey, 2001).

In the investigative interview after one week the results were more mixed. In the free recall older adults surprisingly remembered *more* total details, total correct details and total person, action, object and setting details than young adults. They had however an overall lower accuracy rate. In the question part, again older adults had a lower accuracy rate and remembered more total incorrect details, but they also remembered more setting details than young adults.

The focus on setting details may be explained by the inhibition theory, according to which older adults may be more distracted by irrelevant information (e.g.

environmental details) and thus have more difficulties in memory tasks (e.g. Hartman & Hasher, 1991; Hasher & Zacks, 1988). Older adults thus may have had difficulties suppressing this less relevant information and instead remembered more setting details compared to young adults. Overall however, it seems that older adults performed better than would be expected, which may be due to two factors: The older adults in this study were quite active (almost 90% stated to engage in sports or related activities at least once per week) and well educated (70% stated to hold at least a Bachelor's degree or equivalent). This could indicate a cognitively high-functioning sample. This theory is also supported by results from further analyses that revealed an impact of *university degree* on the amount of setting details that was reported. It may also be that older adults were more motivated to perform well in a memory task compared to their student counterparts. Lastly, older adults may have been too young still with a mean age of 69 years. Further analyses of the older adult group established that with increasing age, performance levels declined and participants produced significantly fewer total, correct, person and action details. In future studies it may therefore be worthwhile to set a higher age cut-off, starting at 65 or 70 years.

3.3.1.4. Is the SAI Helpful for Older Adult Witnesses?

The question that was initially raised with this experiment is whether the SAI is an effective tool to specifically aid older witnesses. When comparing the testimony older adults produced in the SAI with that they produced in a wFR, i.e. the *content of both tools*, results showed that for most variables, there was no significant difference between the two recall tools. Older adults did however remember significantly more person details in the SAI, which represented a large effect. This was reflected in an increase in both correct and incorrect person details. Interestingly, older and young adults were affected differently by recall condition, i.e. by whether they have been given the SAI or the wFR. Whereas young adults showed a distinct increase in the total amount of details, as well as total person, action and total correct details, for older adults this increase was only small. In other words, older and young adults performed similarly well in the wFR, but young adults significantly outperformed older adults in the SAI. It therefore seems that the positive effect of the Self-Administered Interview in a content comparison derived from young adults and that older adults did no benefit as much.

When looking at the *transfer to the second session* and comparing the effect of the SAI on a delayed investigative interview to a wFR or no initial recall, older adults did benefit from using the SAI: They achieved a higher accuracy rate in the free recall and produced fewer incorrect action details in both the free recall and the question part compared to the no-recall older adults. In the free recall of the interview, again, young and older adults were affected differently by recall condition, i.e. by whether they have been given the SAI, the wFR or no recall. Interestingly, older adults performed on a similarly *high* level throughout and were relatively unaffected by recall condition (e.g. for the total details, total action and correct action details), whereas young adults' performance increased across conditions to match the older ones in the SAI condition. Thus, the SAI seemed to help raise young adults' performance to a level at which the older adults' performed to begin with. There was a similar pattern for the interview question part, in that older adults performed on a similar level throughout, relatively unaffected by recall condition, whereas young adults' performance increased across conditions (e.g. for total correct and total person details). This time however, older and young adults performed similar in the written free recall condition and the SAI seemed to help raise young adults' performance above the older adults' level.

From a theoretical perspective, Craik (1986) suggested that older adults would benefit from more environmental support in a memory task compared to young adults. For instance, Craik and McDowd (1987) showed that older adults performed much better on a recognition task than on a recall task and that recall had a more detrimental effect on older adults performance compared to young adults. Thus, whereas young adults seem to be relatively successful at initiating memory retrieval strategies themselves, older adults seem to be more dependent on external support and therefore benefit more from it. Following from that, the SAI with its additional prompts and cues in comparison to a free recall would provide the necessary environmental support for older adults and aid them in their recall of events. However, a greater impact of the SAI on older adults as opposed to young adults was not observed. Instead, young adults benefitted from having environmental support in the form of a comprehensive early recall opportunity more so than older ones.

This should however not derogate from the positive effects that were found for older adults. In accordance with previous literature on the Self-Administered Interview, results show that it works very well for young adults. Furthermore, the

present study was able to show that there is evidence for a benefit for older adults, too. Older adults did show the same trend as young adults in that they produced slightly more details, and significantly more person details. Notably, the accuracy rate remained stable. This is consistent with one prior study on the SAI for older adults: Gawrylowicz, Memon, Scoboria, et al. (2014) also found an increase in correct and incorrect details, while maintaining a similar accuracy rate compared to a written free recall. In contrast to Gawrylowicz's study (2014b), the present study provides an additional comparison to a young adult sample and was able to show that, while the SAI did aid older adults' witness accounts to an extent, the SAI effect was even larger for young adults. It was also able to provide more evidence that the SAI has a positive effect on older adults' testimony, both in an immediate recall and in a delayed interview. Given that older adults are particularly vulnerable to offences that involve a high degree of face-to-face contact between the victim and the offender (McMahon, 2000), a significant increase in person description details with the SAI shows its great potential for police investigations.

Another interesting aspect is the fact that in this study, older adults performed very well throughout, i.e. whereas the SAI managed to elevate young adults' performances above the wFR or no recall groups, the older adults performed on that level to begin with, irrespective of the recall group. Again, this could be due to a motivational aspect, that the student sample was lacking, or to the fact that the older adult sample was quite active and probably cognitively high-functioning. It may also be noted that the lack of improvement across conditions for the older adults could not be attributed to difficulties in applying the Self-Administered Interview. When debriefing the older adults after the second study session, informal feedback was gathered regarding comprehension and usability of the SAI. None of that feedback pointed to any problems in using the SAI correctly. This is also consistent with results described by Gawrylowicz, Memon, Scoboria, et al. (2014), who found the SAI to be clear, and helpful in aiding concentration and in facilitating a complete account for older adults.

3.3.1.5. Can the SAI Reduce Suggestibility?

Another aim of this study was to examine whether the Self-Administered Interview would be helpful in reducing susceptibility to suggestions made in an investigative interview. Leading or misleading questions, often asked unintentionally

during a police interview, are a very common way of contaminating and altering a witness' memory (Loftus, 2005). This can have a disastrous effect on the memory itself, the direction of the police investigation and also on the outcome of the trial.

Results showed that participants who had filled in the SAI were indeed less susceptible to suggestions after one week compared to those who did not have an initial recall opportunity. This is consistent with other studies on the SAI: Gabbert et al. (2012) also found SAI participants to be less suggestible to those who did not have an initial recall opportunity. Moreover, in studies by Gittins et al. (2015) and McPhee et al. (2014) SAI participants showed greater rejection of misinformation compared to no-recall participants. Between participants who filled in the SAI and participants who completed a written free recall, evidence regarding suggestibility is less clear. In this study, participants in the wFR group were also less susceptible to suggestions than the no recall group, but did not differ from the SAI group. Other studies did not find a difference between SAI and wFR groups, either (Mauer et al., 2013; Stephan et al., 2013), whereas Schoof et al. (2014) found some support for a greater reduction in suggestibility with the SAI. According to Loftus (2005), having a strong original memory helps witnesses to detect and ultimately reject discrepant or contradictory information. The SAI seems successful in strengthening the original memory trace and thus helpful in inoculating against suggestions.

With regards to older adult witnesses, results showed no difference in the susceptibility to suggestions between older and young adults. This means that older adults were not found to be more suggestible than young adults, which is consistent with other studies on suggestibility and age group effects (e.g. Gabbert et al., 2003). However, there are also some studies supporting the notion that older adults may be more susceptible to suggestions (e.g. Mitchell et al., 2003). Again, one possible explanation for the good performance of older adults in this study could be the relative mental and physical fitness of this sample. Results further revealed that there was no interaction between age group and recall condition, meaning that older adults were not affected differently by recall conditions. The Self-Administered Interview therefore proved to be helpful in reducing suggestibility in older adults as well as young adults.

3.3.2. The Impact of the SAI on Person Identification

The second major aim of this study was to examine any potential impact of the Self-Administered Interview on person identification performance – both in young and in older adults. Thus, a number of questions reflecting the lineup results will be discussed in the following, starting with whether the SAI can improve witnesses' lineup performance. The second question focuses on the performance of older adults compared to young adult witnesses, independent of recall condition. The third question discusses the impact the SAI may have on older adults, and the last question focuses on the confidence-accuracy relationship in older adults.

3.3.2.1. Can the SAI Improve Witnesses' Lineup Performance?

All studies on the Self-Administered Interview to date have focused on its potential impact on witness testimony. There are however reasons to assume that it could also benefit witnesses in their person identification decisions. Through strengthening the memory for the event and specifically for characteristics of the perpetrator, witnesses may find it easier to identify the correct person or correctly say that the perpetrator is not in the lineup. Moreover, given that other interviewing tools, especially elaborate ones such as the Cognitive Interview, have been found to impact negatively on a subsequent identification (see Meissner & Brigham, 2001a, for a meta-analysis), even a neutral effect for the SAI would be of an advantage in comparison.

Results showed that overall, participants who had filled in the Self-Administered Interview performed better in a lineup after a one-week delay compared to those who did not have an initial recall opportunity *and* to those who had filled in a written free recall. There was a general *tendency* throughout that participants perform better when having been presented the SAI compared to having been presented with a written free recall or no initial recall. Whereas almost half (45.8%) of the SAI participants made a correct choice overall (i.e. they correctly identified the perpetrator in target present or correctly rejected the lineup in target absent lineups), only a third of the wFR and no-recall participants (33.3% and 31.3%, respectively) chose correctly. Most notably, the effect was significant for target present lineups, showing that the overall odds of correctly identifying the perpetrator were 3.8 times higher if

participants had been given the SAI compared to a written free recall or no initial recall. Performance in target absent lineups remained stable across recall conditions, meaning that participants in the SAI group correctly rejected the lineup as often as those in the wFR or the no-recall group.

This represents a novel and highly encouraging finding for police and practitioners. It means that the SAI not only does not hinder a subsequent identification (even if the identification follows a rather elaborate witness interview as in this study), but it even improves identification performance, especially when the perpetrator is among the people in the lineup. Moreover, when the actual perpetrator was not in the lineup, the SAI did not 'encourage' witnesses to choose a person anyhow. This is, again, a noteworthy outcome as it shows that the SAI does not increase false identification rates. And finally, the SAI was superior to a more simple initial recall tool in that participants who had been given the SAI outperformed those who had been given a written free recall form. Thus it can be concluded that an early and comprehensive recall, as it is facilitated in the Self-Administered Interview, is beneficial for person identification after one week.

3.3.2.2. Old versus Young – No Surprises?

Another aim of this study was to examine any difference in the person identification performance between older and young adults. Whereas almost half of the young adults (45.8%) made a correct choice overall (i.e. they correctly identified the perpetrator in target present or correctly rejected the lineup in target absent lineups), only one quarter of older adults (27.8%) chose correctly. For target absent lineups only, results represent an average drop in correct rejections of .20, namely from .53 for young adults to .33 for older adults. This means that only one third of older adults correctly indicated that the perpetrator was not in the lineup. For target present lineups, there was an average drop in correct identifications of .17 (from .39 for young adults to .22 for older adults), which means that only 1 in 5 older adults correctly identified the perpetrator from the lineup.

This is consistent with literature on older adult witnesses, showing that they generally perform more poorly compared to young adults (Havard & Memon, 2009; Searcy et al., 1999; Wilcock et al., 2007). Furthermore, results from this study are almost identical to those found in Bartlett & Memon (2007), in which they averaged data over 10 studies comparing young and older adult witnesses. They reported an

average drop of the proportion of correct rejections in TA lineups of .22, namely from .53 for young adults to .31 for older ones.

To account for these age-related differences especially in false identifications, Searcy et al. (1999) put forward the context recollection theory, which is based on the work of Mandler (1980). It accounts for the problem of recognizing that a face is familiar without being able to identify it. This is due to familiarity being based solely on the characteristics of the face, but not the context in which it was encountered. As such, filler faces in lineups are based on perceived familiarity with the perpetrator mugshot. Identification on the other hand relies on recollection of context, which requires detailed information on the face and the relationship between face and context. According to the context recollection theory, older adults have greater problems with recollecting contextual information required for identification, but not with the perceptual processes required for familiarity. As a result, they rely on familiarity in lineup procedures, i.e. a face standing out as being familiar to them, more so than young adults, and therefore make more mistakes in choosing a filler face from a lineup. Memon et al. (2002) were able to support this theory in their study.

The context recollection theory also offers options for interventions that improve context recall and consequently might help to reduce false identifications in older adult witnesses. One such intervention could be the Self-Administered Interview, as it strengthens the original memory and thus may help put the face of the perpetrator in the right context for later retrieval in a lineup task.

3.3.2.3. Can the SAI Aid Older Adults' Person Identification?

The question that was initially raised with this experiment is whether the SAI is an effective tool to specifically aid older witnesses – not only in improving their witness testimony but also their person identification performance. If an early and high-quality recall as it is facilitated in the SAI can strengthen the original memory and thus provide a stronger context for the source of where a face was encoded, it should decrease false identifications for older adults.

However, there was no significant evidence of a stronger benefit of the SAI for older adults compared to young adults. Especially in the target absent condition, older adults as well as young adults performed on a similar level throughout, independent of whether they had been given the SAI, a wFR or no recall a week earlier. In the target present condition, there was no significant benefit of the SAI for

older adults either. It should be noted that overall correct identification rates were quite low (only 8 older adults correctly identified the perpetrator in target present lineups, of which 5 were in the SAI group, 2 in the wFR group and 1 in the no recall group), which is detrimental to finding significant differences in performance.

Although results do not support the context recollection theory, this should not be viewed as disconfirming the model. It may be that the SAI does not support context recollection enough to be able to overcome false choosing errors after a delay of one week. More research is needed to address this issue and to specifically focus on older adult witnesses. Literature suggests that it seems generally difficult to improve person identification in older witnesses. For example, Rose et al. (2005) found that older adults were less able to remember the instructions given to them prior to a lineup compared to young adults and, more importantly, participants who failed to remember the information made also more false identifications from the lineup. Furthermore, enhanced lineup instructions especially for older adults found no effect on lineup performance either (Wilcock et al., 2005).

This is the first study on the impact of the Self-Administered Interview on person identification performance in general, and in specific with an older adult group. Despite the fact that it did not significantly reduce false identifications for older adults, it neither increased them. At the very least it can thus be concluded that applying the SAI to older adults witnesses is of no disadvantage for their lineup performance and can be applied without the risk of a negative effect.

3.3.2.4. The Confidence-Accuracy Relationship in Older Adults

Finally, another aim of this study was to increase the evidence-base regarding the relationship between witnesses' post-identification confidence and their accuracy of identification. Analyses revealed that there was no significant difference in participants' post-identification confidence between those who performed correctly and those who did not. Neither across the total sample, nor in any of the subsamples (young or older adults, target present or target absent), were people who correctly identified the perpetrator or correctly rejected the lineup more confident than those who falsely identified a filler or falsely rejected the lineup. In fact, among those participants who rated themselves to be "*confident*" about their choice (n = 48), twice as many had made an incorrect choice (n = 32) compared to those who had made the correct choice (n =16). Even more interestingly, when only looking at older adults

who rated themselves to be confident about their choice ($n = 20$), only 2 actually made the correct choice, whereas 18 chose incorrectly from the lineup.

This is consistent with previous research on the relationship in general eyewitness samples (see Sporer et al., 1995, for a meta-analysis). Specifically regarding older adults, Scogin, Calhoon & D'Errico (1994) found no significant correlation between lineup accuracy and the confidence of participants either. Even if an older witness was very confident to have chosen the correct person this did not mean they actually were correct (Wilcock et al., 2008). This finding was replicated in several other studies (Memon et al., 2003; Memon et al., 2002; Wilcock et al., 2007).

In previous research it was further suggested that the confidence-accuracy relationship may be moderated by whether or not the witness has chosen someone from the lineup. In a meta-analysis, Sporer et al. (1995) found a higher confidence-accuracy correlation for choosers than non-choosers, suggesting that once witnesses have identified someone from a lineup, their confidence may be a stronger predictor of their accuracy. However, in this study there was no difference for choosers or non-choosers regarding their confidence-accuracy relationship. Among those participants who identified someone from the lineup with confidence (choosers, $n = 32$), only 9 actually correctly identified the perpetrator, whereas 23 falsely identified a filler (from either a TP or a TA lineup). This shows that again, confidence is no reliable indicator for a witnesses' accuracy in a lineup task.

The probably most important finding however is, that the Self-Administered Interview did not impact the confidence-accuracy relationship. Participants who had filled in the SAI a week earlier and who subsequently performed correctly on the lineup were *not* more confident than those who had filled in a written free recall or did not have any initial recall and still performed correctly. This means that the SAI did not 'artificially' enhance a witness' confidence. Given that confidence is no reliable indicator for whether or not a witness has chosen the correct person from a lineup, it is important not to enhance their perceived confidence.

3.3.3. Limitations of this Study

Naturally, this study is not without limitations, regarding the characteristics of this specific experiment, the wider characteristics of laboratory experiments and also concerning the Self-Administered Interview in general. Starting with the latter, the SAI booklet format may be limiting to several groups of witnesses, such as those with language problems or literacy difficulties. This may be especially problematic in regions with a high percentage of non-native speakers that are also high in crime. Here, it may be desirable to have translations of the SAI in the most common languages spoken in that particular region. Another alternative to applying a SAI and still facilitating a recall as soon as possible may be to let witnesses narrate or type their testimony. Indeed, McPhee et al. (2014) provide first support for an equally effective spoken recall (in response to the same prompts as in the SAI) compared to a written SAI. A second and potentially large group of witnesses that may find the SAI format difficult are vulnerable witnesses. Victims of sexual offences, human trafficking, child victims or other traumatized victims for instance are often very dependent on social support during interviewing, that the SAI cannot provide. This limitation may however not be of practical importance, as the SAI was designed particularly for situations where there is a large number of witnesses present, rather than for isolated, traumatised victims. On the other hand one could argue that it is these cases in particular that could benefit from a high quality, comprehensive first witness account. Future studies and field trials should therefore focus more on victim witnesses and examine ways in which the SAI could be successfully adapted.

A limitation more specific to this study could be the selection of the older adult sample. As was mentioned before, older adults reported to be quite active still with 90% doing sports at least once a week, and 70% holding at least one university degree. This could suggest an above-average functioning sample and explain why older adults performed so well in the testimony part compared to young adults. Nevertheless, recall condition did impact on this sample's cognitive performance, showing that a) even presumably above-average functioning older adults make mistakes in remembering details of a crime and in subsequent person identification, and b) that they still benefit from filling in the SAI as an initial recall. For example,

older adults who had completed the SAI were shown to have a higher accuracy rate in a recall after one week.

A further limiting factor could be the chosen length of intervals. A relatively short initial interval of approximately 15 minutes was chosen between witnessing the crime and administering the SAI. The second interval between filling in the SAI and recalling the event in the witness interview was one week. Although these intervals are in accordance with prior research on the SAI (e.g. Gabbert et al., 2009; Hope et al., 2014; Krix, Sauerland, Lorei, et al., 2015), they may not represent the most ecologically valid study design. In real life, especially with large-scale incidents, it may be several hours until the SAI is distributed to witnesses. Two studies have started addressing this and delayed administration of the SAI for 24 hours (Mackay & Paterson, 2015; Paterson et al., 2015) and even one week (Paterson et al., 2015). They found that delayed administration impacted negatively on recall accuracy and that the efficacy of the SAI was dependent on administering it within 24 hours. However, there is no research as of yet that examined intervals between 15 minutes and 24 hours. As for the second interval, it may either be only a couple of days, or much longer than one week until a witness is called in for the police interview. Only three studies have used intervals that were longer (and none that were shorter): af Hjelmsäter et al. (2012) and Paterson et al. (2015) administered the SAI after two weeks, and Gabbert et al. (2012) used a 3-weeks delay. More research is needed in ascertaining typical intervals between crime, initial police contact, and statement taken, to then adjust study intervals accordingly.

An important issue that inevitably is raised concerning eyewitness research in general is that of external validity of laboratory and even staged real-life studies. Due to ethical, legal and other constraints, the amount of violence that is displayed, the nature of the events that can be staged, or the persons affected by the event differ crucially from real-life situations (Kapardis, 2010). The findings from artificial studies are accordingly questionable with regards to generalizability to the real world. Having artificial stimuli may either lead to the participants paying less attention, being less motivated and less stressed due to the minor personal importance compared to a real-life situation, or it may even lead to participants being extra focused and motivated because they want to show how well older adults can perform. However, compared to laboratory research, archival and single case studies pose constraints such as the lack of systematically altered variables and lack of knowledge about the

actual event (Kapardis, 2010). But put one way or another, generally performance most certainly differs between laboratory and real life. And plausible as these criticisms may seem, they represent the common gap between experimental research and forensic practice, or methodologically speaking, the trade-off between internal and external validity. However, limited as laboratory eyewitness research may be, it has provided evidence base and informed policy, and it is indispensable to draw knowledge from both experimental and real-world studies to gain the best understanding of witness performance.

Due to ethical concerns, previous research mostly used non-violent crimes, as in this study, to address eyewitness investigations. This is of course an understandable limitation to safeguard participants, although one could raise the question whether there are any differences in the impact of the SAI and the applicability to different types of crimes. Furthermore, participants in this study knew they were part of an eyewitness study and would have to remember details about the event and also the perpetrator's face. However, they were blind to the hypothesis tested, i.e. they did not know that an early recall should help them remember better. In fact, they did not know that there were different recall groups at all. Still, another approach could be to design an unexpected eyewitness interview with a film. The film with the staged crime could be used as a filler task while the participants believe they are actually tested on something different. This approach certainly includes the deception of the participants, and therefore should be thoroughly considered. However, using a real-life event with an element of deception has been done for other eyewitness studies (e.g. Mueller-Johnson & Ceci, 2004).

From a few studies that have used the SAI, there is some support that the recall tool works for real-life and more violent crimes as well. For example, Krix et al. (2016) used a non-violent, but staged real-life event (a theft of a cellphone) and found a very large effect (d = 1.62 for correct details) immediately after the event. Colomb and Gabbert (2013) used a staged film depicting some violence (a robbery with punches and threatening with a gun) and found a very large effect (d = 1.46 for correct details) after one week. Furthermore, Gawrylowicz, Memon, and Scoboria (2014) and Gawrylowicz, Memon, Scoboria, et al. (2014) used a staged video of a date-rape scene for half of their participants and found a medium-sized (d = 0.72) and a large (d = 1.21) effect for correct details, respectively, immediately after the event. The most extreme stimulus material so far was used by Gittins et al. (2015) and

Mackay and Paterson (2015), who used a trauma-analogous film depicting the aftermath of a real car accident, that was shown to increase PTSD and anxiety symptoms. They found medium-sized effects (both $d = 0.50$) for an increase in correct details after one week. These examples are encouraging and future studies should extend the focus on real-life events as well as more violent scenes (within ethical boundaries) to widen the evidence-base around the Self-Administered Interview.

Another possible concern is a potential bias of the photographic stimuli used for the lineup in this study. Biases in the lineup always pose a problem, in research as well as in real-life. An impressive example was described by Ellison & Buckhout (1981): in a case in which the suspect was described as a black man U.S. police used a six-person lineup with one black suspect and five white persons as fillers. Although unfairness in this case was very obvious, for many lineups it is less clear. Therefore in the present study measurements were conducted to make sure the lineups were not biased, neither towards the perpetrator, nor towards the target replacement (proportions technique), and that there were no effects of the order of photo placement (Latin square design arrangements). Thus although bias may often be a problem in lineups, it was not the case in this study.

3.3.4. Implications

The present study replicated the SAI effect found in prior studies, that is the increase of correctly recalled details immediately after an event and after one week. It was also able to show a benefit of the Self-Administered Interview for older adult witnesses, albeit smaller than that found for young adults. Further, this is the first study to examine and show a benefit of the SAI not only for testimony, but also for person identification. This finding has implications for police investigations and thus criminal justice. It simplifies police application of the SAI as it transfers the effect to a wider witness group. All adult witnesses should be given the SAI in order to conserve their witness memory, regardless of what (adult) age group they belong to *and* regardless of which witness task they will be facing later on in the investigational process. Further, this study was able to provide additional support for the SAI being able to inoculate against misinformation. It thus reduces the risk of inappropriate interview strategies such as leading questions, not just at the initial questioning after the event, but as shown in this study, also in an interview after one week.

More research however is needed to further address its use for lineup procedures. When looking at the current police practice in the UK, more and more sequential video lineups are conducted instead of a) photo lineups and b) simultaneous lineups. In the present study, a simultaneous photo lineup was used and thus it would be important to examine the effect of the SAI on sequential lineups, and with regards to the current police practice, on video parades. Furthermore it would of course be useful to see if the SAI is beneficial for children's lineup performance. Taking not only young and older adults, but also children into account would moreover increase the sampling frame for a potential field experiment. If a beneficial effect of the Self-Administered Interview could be shown in all these scenarios, this would represent substantial evidence for a way to improve accuracy of eyewitness identification decisions.

In sum, this study provides further evidence-base for the benefit of adopting the Self-Administered Interview in everyday investigative interviewing. It can obtain high-quality evidence without increasing the workload for police officers. Especially when a large number of witnesses is present, it can even reduce resource problems by the police. Police forces not just throughout the UK, but internationally, should

strongly consider including the SAI in their witness treatment protocols. The Self-Administered Interview has the potential to significantly impact police investigations and trial outcomes, and thus provide justice for a large number of affected individuals.

4. CONCLUDING REMARKS

This thesis examined the effectiveness of the Self-Administered Interview through a meta-analytic review and comprehensive empirical study with the aim to extend prior research on the SAI and on age-related differences in eyewitness performance.

Meta-analyses for several outcome measures covered 38 experimental comparisons from 22 empirical studies representing 1712 interviewees. Results indicated a strong benefit of the SAI both immediately after the witnessed crime (d = 1.20 for the increase in correct details compared to other forms of initial recall) and in a delayed recall (d = 0.92 for correct details compared to no initial recall) after one to three weeks. Further, an experiment investigated the effectiveness of the SAI for older witnesses' testimony, suggestibility and lineup performance. 144 participants, half of which were 60 years or older and half aged 18-30 years, took part in two sessions. In the first session, they were shown a film of a staged crime and either filled in the SAI, gave a written free recall or no initial recall. In the second session after one week they were then asked to give a free recall of what they remembered, answer questions including suggestive questions, and also to identify the perpetrator from the film from a 6-person simultaneous photo lineup. Results confirmed the standard SAI effect for young adults, i.e. an increase in correct and incorrect details immediately after the event with an unaffected accuracy rate, and an increase in correct, but not incorrect details in a later interview, with a higher accuracy rate compared to no SAI. Results also showed a small beneficial effect for older adults and further indicated that witnesses who were given the SAI were less suggestible than other witnesses. Most excitingly, results showed a beneficial effect of the SAI on lineup performance (OR = 3.8 for correct identifications from target present lineups).

To the author's knowledge this is the first study examining the effects of the Self-Administered Interview on witnesses' lineup performance and only the second one on the effects on older adult witnesses. It is further the first meta-analytic review conducted on this investigative interviewing tool. Overall, results indicate a strong benefit of applying the SAI as opposed to no initial recall and even other initial recall

tools, at least in laboratory settings. The evidence gathered in this thesis can be summed up as follows:

1. The SAI elicits a comprehensive and accurate initial recall.
2. The SAI preserves memory for a later recall after 1 to 3 weeks.
3. The SAI inoculates against suggestibility.
4. The SAI improves correct identifications from lineups.
5. The SAI benefit is strong for young adults.
6. The SAI is of no disadvantage for older adults.

Following these findings, it would be negligent not to use the Self-Administered Interview in everyday investigative police work, and especially in cases with multiple witnesses and major catastrophic events. It is an easily implemented tool for improving criminal justice for young and older adults and can further increase accuracy and credibility of witness testimony. More real-world applications and field trial evaluations are of course needed to strengthen the evidence-base, but there should be no delay in integrating the SAI as an additional tool for police forces throughout the UK.

5. REFERENCES

Af Hjelmsäter, E. R., Strömwall, L. A., & Granhag, P. A. (2012). The Self-Administered Interview: a means of improving children's eyewitness performance? *Psychology, Crime & Law, 18*(10), 897-911. doi:10.1080/1068316X.2011.582844

Alogna, V. K., Attaya, M. K., Aucoin, P., Bahník, Š., Birch, S., Birt, A. R., . . . Zwaan, R. A. (2014). Registered Replication Report:Schooler and Engstler-Schooler (1990). *Perspectives on Psychological Science, 9*(5), 556-578. doi:10.1177/1745691614545653

Anastasi, J., & Rhodes, M. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review, 12*(6), 1043-1047. doi:10.3758/bf03206441

Anderson, J. R. (1983). A spreading activation theory of memory. *Journal of verbal learning and verbal behavior, 22*(3), 261-295. doi:10.1016/S0022-5371(83)90201-3

Anderson, N. D., & Craik, F. I. M. (2000). Memory in the aging brain. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 411-425). Oxford, England: Oxford University Press.

Armstrong, R. A. (2014). When to use the Bonferroni correction. *Ophthalmic and Physiological Optics, 34*(5), 502-508. doi:10.1111/opo.12131

Aschermann, E., Mantwill, M., & Köhnken, G. (1991). An independent replication of the effectiveness of the cognitive interview. *Applied Cognitive Psychology, 5*, 489-495. doi:10.1002/acp.2350050604

Bachman, R., & Meloy, M. L. (2008). The epidemiology of violence against the elderly. *Journal of Contemporary Criminal Justice, 24*(2), 186-197. doi:10.1177/1043986208315478

Baddeley, A. D. (1995). *Working memory*. Oxford, England: Clarendon Press.

Balota, D. A., Dolan, P. O., & Duchek, J. M. (2000). Memory changes in healthy older adults. In E. Tulving & F. I. M. Craik (Eds.), *The Oxford Handbook of Memory* (pp. 395-409). Oxford, England: Oxford University Press.

Bartlett, J. C., & Memon, A. (2007). Eyewitness memory in young and older adults. In D. S. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *The Handbook of Eyewitness Psychology. Volume II: Memory for People* (pp. 309-338). Mahwah, NJ: Lawrence Erlbaum Associates.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the royal statistical society. Series B (Methodological)*, 289-300.

Boon, R. (2012). *Enhancing witness statements. Examining the effects of reading a Self-Administered Interview on interviewer behaviour and reliability of witness statements.* (M.Sc.), Politie Kennemerland, NL.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). Comprehensive Meta-analysis Version 2. Biostat. Englewood, NJ.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009) *Introduction to Meta-Analysis*. Chichester, United Kingdom: John Wiley & Sons, Ltd.

Bornstein, H. B. (1995). Memory processes in elderly eyewitnesses: What we know and what we don't know. *Behavioral Sciences and the Law, 13*(3), 337-348. doi:10.1002/bsl.2370130303

Bössenrodt, E., Hewig, J., Kraus, U., & Paelecke, M. (2011). *Von der Praxis zur Theorie und wieder zurück: Das Self-Administered Interview in einer praktischen Anwendung an einer deutschsprachigen Stichprobe.* (M.Sc. Thesis), University of Würzburg.

Boyce, M., Beaudry, J. L., & Lindsay, R. C. L. (2007). Belief of eyewitness identification evidence. In R. C. L. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of Eyewitness Psychology. Volume II: Memory for People*. Mahwah, NJ: Lawrence Erlbaum Associates.

Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior, 26*(3), 353-364. doi:10.1023/a:1015380522722

Brigham, J. C., & Pfeiffer, J. E. (1994). Evaluating the fairness of lineups. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult Eyewitness Testimony: Current Trends and Developments* (pp. 201-222). Cambridge, England: Cambridge University Press.

Bruce, V. (1982). Changing faces: Visual and non-visual coding processes in face recognition. *British Journal of Psychology, 73*(1), 105-116. doi:10.1111/j.2044-8295.1982.tb01795.x

Bruce, V., Burton, M., & Hancock, P. (2007). Remembering faces. In D. S. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of Eyewitness Psychology. Volume II: Memory for People* (pp. 87-100). Mahwah, NJ: Lawrence Erlbaum Associates.

Bruce, V., & Young, A. (1986). Understanding face recognition. *British Journal of Psychology, 77*(3), 305-327. doi:10.1111/j.2044-8295.1986.tb02199.x

Buela Casal, G., Caballo, V. E., & García Cueto, E. (1990). Differences between morning and evening types in performance. *Personality and Individual Differences, 11*(5), 447-450. doi:10.1016/0191-8869(90)90056-W

Bull, R., & Corran, E. (2002). Interviewing child witnesses: Past and future. *International Journal of Police Science & Management, 4*(4), 315-322. doi:10.1350/ijps.4.4.315.10872

Burt, D. B., Zembar, M. J., & Niederehe, G. (1995). Depression and memory impairment: a meta-analysis of the association, its pattern, and specificity. *Psychological Bulletin, 117*(2), 285-305.

Cabeza, R., Anderson, N. D., Locantore, J. K., & McIntosh, A. R. (2002). Aging Gracefully: Compensatory Brain Activity in High-Performing Older Adults. *NeuroImage, 17*(3), 1394-1402. doi:https://doi.org/10.1006/nimg.2002.1280

Ceci, S. J., Toglia, M. P., & Ross, D. F. (1988). On remembering more or less: A trace strength interpretation of developmental differences in suggestibility. *Journal of Experimental Psychology: General, 117*(2), 201-203. doi:10.1037/0096-3445.117.2.201

Cerella, J. (1985). Information processing rates in the elderly. *Psychological Bulletin, 98*, 67-83. doi:10.1037/0033-2909.98.1.67

Chiroro, P., & Valentine, T. (1995). An Investigation of the Contact Hypothesis of the Own-race Bias in Face Recognition. *The Quarterly Journal of Experimental Psychology Section A, 48*(4), 879-894. doi:10.1080/14640749508401421

Clarke, C., & Milne, R. (2001). *A national evaluation of the PEACE Investigative Interviewing Course*: Home office London.

Clifford, B. R., & Gwyer, P. (1999). The effects of the cognitive interview and other methods of context reinstatement on identification. *Psychology, Crime & Law, 5*(1-2), 61-80. doi:10.1080/10683169908414994

Clifford, B. R., & Richards, V. J. (1977). Comparison of recall by policemen and civilians under conditions of long and short durations of exposure. *Perceptual and Motor Skills, 45*(2), 503-512. doi:10.2466/pms.1977.45.2.503

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2. ed.). Hillsdale, NJ: Erlbaum.

Colcombe, S., & Kramer, A. F. (2003). Fitness effects on the cognitive function of older adults: A meta-analytic study. *Psychological Science, 14*(2), 125-130. doi:10.1111/1467-9280.t01-1-01430

Colomb, C., & Gabbert, F. (2013). *The Impact of the Self-Administered Interview on Eyewitness Cooperation (working title)*. Unpublished manuscript.

Congdon, N., O'Colmain, B., Klaver, C. C. W., Klein, R., Muñoz, B., Friedman, D. S., . . . Eye Diseases Prevalence Research, G. (2004). Causes and prevalence of visual impairment among adults in the United States. *Archives of ophthalmology (Chicago, Ill. : 1960), 122*(4), 477-485. doi:10.1001/archopht.122.4.477

Coxon, P., & Valentine, T. (1997). The effects of the age of eyewitnesses on the accuracy and suggestibility of their testimony. *Applied Cognitive Psychology, 11*, 415-430. doi:10.1002/(SICI)1099-0720(199710)11:5<415::AID-ACP469>3.0.CO;2-A

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendorf (Eds.), *Human Memory and Cognitive Capabilities, Mechanisms and Performances* (pp. 409-422). North Holland: Elsevier.

Craik, F. I. M., & McDowd, J. M. (1987). Age differences in recall and recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13*, 474-479. doi:10.1037/0278-7393.13.3.474

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 43*, 245-265. doi:10.1348/0144665031752934

Curtis, E. J. (2013). *The Self Administered Interview: The role of co-witnesses in minimising the misinformation effect in adolescent eyewitnesses.* (M.Sc.), University of Portsmouth.

Cutler, B. L., Berman, G. L., Penrod, S., & Fisher, R. P. (1994). Conceptual, practical, and empirical issues associated with eyewitness identification test media. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult Eyewitness Testimony: Current Trends and Developments* (pp. 163-181). Cambridge, England: Cambridge University Press.

Dando, C. J., Wilcock, R., & Milne, R. (2009). Novice police officers application of the Cognitive Interview procedure. *Psychology, Crime, & Law, 15*(8), 679-696.

Davis, S. L., & Bottoms, B. L. (2002). Effects of social support on children's eyewitness reports: a test of the underlying mechanism. *Law and Human Behavior, 26*(2), 185. doi:10.1023/A:1014692009941

Deffenbacher, K. (1980). Eyewitness accuracy and confidence: Can we infer anything about their relationship? *Law and Human Behavior, 4*(4), 243-260. doi:10.1007/bf01040617

Dickersin, K., Chan, S., Chalmersx, T., Sacks, H., & Smith, H. (1987). Publication bias and clinical trials. *Controlled clinical trials, 8*(4), 343-353. doi:10.1016/0197-2456(87)90155-3

Division, U. N. P. (2017). World Population Prospects. The 2017 Revision, Key Findings and Advance Tables. *ESA/P/WP/248*.

Doob, A. N., & Kirshenbaum, H. M. (1973). Bias in police lineups - partial remembering. *Journal of Police Science and Administration, 1*, 287-293.

Dunning, D., & Perretta, S. (2002). Automaticity and eyewitness accuracy: A 10- to 12-second rule for distinguishing accurate from inaccurate positive identifications. *Journal of Applied Psychology, 87*(5), 951-962. doi:10.1037/0021-9010.87.5.951

Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology, 67*(5), 818-835. doi:10.1037/0022-3514.67.5.818

Duval, S., & Tweedie, R. (2000a). A nonparametric "Trim and Fill" method of accounting for publication bias in meta-analysis. *Journal of the American Statistical Association, 95*(449), 89-98. doi:10.1080/01621459.2000.10473905

Duval, S., & Tweedie, R. (2000b). Trim and Fill: A simple funnel-plot–based method of testing and adjusting for publication bias in meta-analysis. *Biometrics, 56*(2), 455-463. doi:10.1111/j.0006-341X.2000.00455.x

Eaton, W. W., Muntaner, C., Smith, C., Tien, A., & Ybarra, M. (2004). Center for Epidemiologic Studies Depression Scale: Review and revision (CESD and CESD-R). In M. E. Maruish (Ed.), *The Use of Psychological Testing for Treatment Planning and Outcomes Assessment* (3rd ed., pp. 363-377). Mahwah, NJ: Lawrence Erlbaum.

Erickson, W. B., Lampinen, J. M., & Moore, K. N. (2016). Eyewitness Identifications by Older and Younger Adults: A Meta-Analysis and Discussion. *Journal of Police and Criminal Psychology, 31*(2), 108-121. doi:10.1007/s11896-015-9176-3

Farrington, D. P., & Welsh, B. C. (2005). Randomized experiments in criminology: What have we learned in the last two decades? *Journal of Experimental Criminology, 1*(1), 9-38. doi:10.1007/s11292-004-6460-0

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*(2), 175-191. doi:10.3758/BF03193146

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). London: Sage.

Field, A., & Hole, G. (2008). *How to Design and Report Experiments*. London: Sage.

Finger, K., & Pezdek, K. (1999). The effect of cognitive interview on face identification accuracy: Release from verbal overshadowing. *Journal of Applied Psychology, 84*(3), 340-348. doi:10.1037/0021-9010.84.3.340

Fisher, R. P., & Geiselman, R. E. (1992). *Memory-enhancing Techniques for Investigative Interviewing*. Springfield, IL: Charles C. Thomas.

Fitzgerald, R. J., & Price, H. L. (2015). Eyewitness identification across the life span: A meta-analysis of age differences. *Psychological Bulletin, 141*(6), 1228-1265. doi:10.1037/bul0000013

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). "Mini-mental state". A practical method for grading the cognitive state of patients for the clinican. *Journal of Psychiatric Research, 12*(3), 189-198.

Gabbert, F., Hope, L., & Fisher, R. P. (2009). Protecting eyewitness vidence: examining the efficacy of a self-administered interview tool. *Law and Human Behavior, 33*(4), 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*(4), 568-575. doi:10.1002/acp.2828

Gabbert, F., Memon, A., & Allan, K. (2003). Memory conformity: Can eyewitnesses influence each other's memories for an event? *Applied Cognitive Psychology, 17*(6), 703-714. doi:10.1002/acp.885

Gawrylowicz, J., Memon, A., & Scoboria, A. (2014). Equipping witnesses with transferable skills: the Self-Administered Interview©. *Psychology, Crime & Law, 20*(4), 315-325. doi:10.1080/1068316X.2013.777961

Gawrylowicz, J., Memon, A., Scoboria, A., Hope, L., & Gabbert, F. (2014). Enhancing older adults' eyewitness memory for present and future events with the Self-Administered Interview. *Psychology and Aging, 29*(4), 885-890. doi:10.1037/a0038048

Geiselman, R. E., Fisher, R. P., MacKinnon, D. P., & Holland, H. L. (1986). Enhancement of eyewitness memory with the Cognitive Interview. *American Journal of Psychology, 99*, 385-401. doi:10.2307/1422492

Gittins, C. B., Paterson, H. M., & Sharpe, L. (2015). How does immediate recall of a stressful event affect psychological response to it? *Journal of Behavior Therapy and Experimental Psychiatry, 46*, 19-26. doi:10.1016/j.jbtep.2014.07.006

Glisky, E. L., Rubin, S. R., & Davidson, P. S. R. (2001). Source memory in older adults: An encoding or retrieval problem? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 27*(5), 1131-1146. doi:10.1037/0278-7393.27.5.1131

Gould, J. B., & Leo, R. A. (2010). One hundred years later: Wrongful convictions after a century of research. *The Journal of Criminal Law & Criminology, 100*(3), 825-868.

Greathouse, S. M., & Kovera, M. B. (2009). Instruction bias and lineup presentation moderate the effects of administrator knowledge on eyewitness identification. *Law and Human Behavior, 33*(1), 70-82. doi:10.1007/s10979-008-9136-x

Greuel, L. (2008). Zeugenvernehmung. In R. Volbert & M. Steller (Eds.), *Handbuch der Rechtspsychologie* (pp. 221-231). Göttingen: Hogrefe.

Gwyer, P., & Clifford, B. R. (1997). The effects of the Cognitive Interview on recall, identification, confidence and the confidence/accuracy relationship. *Applied Cognitive Psychology, 11*(2), 121-145. doi:10.1002/(sici)1099-0720(199704)11:2<121::aid-acp443>3.0.co;2-l

Hancock, P., Bruce, V., & Burton, A. (2000). Recognition of unfamiliar faces. *Trends in Cognitive Sciences, 4*(9), 330-337. doi:10.1016/s1364-6613(00)01519-9

Hartman, M., & Hasher, L. (1991). Aging and suppression: Memory for previously irrelevant information. *Psychology and Aging, 6*, 587-594. doi:10.1037/0882-7974.6.4.587

Hasher, L., & Zacks, R. T. (1988). Working memory, comprehension, and aging: A review and a new view. In G. H. Bower (Ed.), *The Psychology of Learning and Motivation* (Vol. 22, pp. 193-225). Orlando, FL: Academic Press.

Havard, C., & Memon, A. (2009). The influence of face age on identification from a video line-up: A comparison between older and younger adults. *Memory, 17*(8), 847-859. doi:10.1080/09658210903277318

Hedges, L., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis*. Orlando, US: Academic Press.

Heubrock, D., & Palkies, P. (2008). The Rapport Basics and application of tactic communication methods in accused and witness interrogation. *KRIMINALISTIK, 62*(11), 602-608.

Higgins, J. P. T., & Green, S. (2011). Cochrane Handbook for Systematic Reviews of Interventions. 5.1.0. Retrieved from http://handbook.cochrane.org

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ : British Medical Journal, 327*(7414), 557-560. doi:10.1136/bmj.327.7414.557

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*(2), 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*(3), 304-313. doi:10.1002/acp.2984

Horne, J. A., & Östberg, O. (1976). A self-assessment questionnaire to determine morningness-eveningness in human circadian rhythms. *International Journal of Chronobiology, 4*, 97-110.

Horne, J. A., & Östberg, O. (1977). Individual differences in human circadian rhythms. *Biological Psychology, 5*(3), 179-190. doi:10.1016/0301-0511(77)90001-1

Huff, M. J., & Umanath, S. (2017). Evaluating Suggestibility to Additive and Contradictory Misinformation Following Explicit Error Detection in Younger and Older Adults. *Journal of Experimental Psychology: Applied*, No Pagination Specified-No Pagination Specified. doi:10.1037/xap0000138

Huff, M. J., Weinsheimer, C. C., & Bodner, G. E. (2016). Reducing the Misinformation Effect Through Initial Testing: Take Two Tests and Recall Me in the Morning? *Applied Cognitive Psychology, 30*(1), 61-69. doi:10.1002/acp.3167

Hutton, G., Johnston, D., & Sampson, F. (2005). *Blackstone's Police Investigator's Manual*. Oxford, England: Oxford University Press.

Innocence Project. (2013). DNA Exonerations Worldwide Factsheet. Retrieved from http://www.innocenceproject.org/Content/DNA_Exonerations_Nationwide.php

Innocence Project. (2016). Celebrating 25 Years of Freedom and Justice. The Innocence Project Annual Report 2016. Retrieved from https://www.innocenceproject.org/wp-content/uploads/2018/01/IP-AR2016_16sm.pdf

Karpel, M. E., Hoyer, W. J., & Toglia, M. P. (2001). Accuracy and qualities of real and suggested memories: Nonspecific age differences. *Journal of Gerontology, 56B*(2), 103-110. doi:10.1093/geronb/56.2.P103

Kassin, S. M., & Neumann, K. (1997). On the power of confession evidence: An experimental test of the fundamental difference hypothesis. *Law and Human Behavior, 21*(5), 469-484. doi:10.1023/a:1024871622490

Kebbell, M. R., Milne, R., & Wagstaff, G. F. (1999). The cognitive interview: A survey of its forensic effectiveness. *Psychology, Crime and Law, 5*(1-2), 101-115. doi:10.1080/10683169908414996

Kemp, R. I., Paterson, H. M., & Yu, A. (2016). *Comparing the Enhanced Cognitive Interview and the Self-Administered Interview (working title)*. Manuscript in preparation.

Köhnken, G., Milne, R., Memon, A., & Bull, R. (1999). The Cognitive Interview: A meta-analysis. *Psychology, Crime & Law, 5*(1-2), 3-27. doi:10.1080/10683169908414991

Köhnken, G., Schimossek, E., Aschermann, E., & Höfer, E. (1995). The Cognitive Interview and the assessment of the credibility of adults' statements. *Journal of Applied Psychology, 80*(6), 671-684. doi:10.1037/0021-9010.80.6.671

Krix, A. C., Sauerland, M., Lorei, C., & Rispens, I. (2015). Consistency across repeated eyewitness interviews: Contrasting police detectives' beliefs with actual eyewitness performance. *PLoS ONE, 10*(2), e0118641. doi:10.1371/journal.pone.0118641

Krix, A. C., Sauerland, M., Merckelbach, H., Gabbert, F., & Hope, L. (2015). How effective is retrieval support for witnesses with different levels of working and source memory? *Journal of Cognitive Psychology, 27*(3), 335-348. doi:10.1080/20445911.2014.1003219

Krix, A. C., Sauerland, M., Merckelbach, H. L. G. J., Gabbert, F., & Hope, L. (September, 2011). *Wie Zeugen sich selbst vernehmen können: Erste Ergebniss zur Deutschen und Niederländischen Form des Self-Administered Interview*. Paper presented at the 14. Fachgruppentagung Rechtspsychologie der Deutschen Gesellschaft für Psychologie (DGPs) - Prävention und Intervention, Münster, Germany.

Krix, A. C., Sauerland, M., Raymaekers, L. H. C., Memon, A., Quaedflieg, C. W. E. M., & Smeets, T. (2016). Eyewitness evidence obtained with the Self-Administered Interview© is unaffected by stress. *Applied Cognitive Psychology, 30*(1), 103-112. doi:10.1002/acp.3173

Lanier, D. N., & Dietz, T. L. (2009). Elder criminal victimization: Its relative rate compared to non-elders. *The Social Science Journal, 46*(3), 442-458. doi:10.1016/j.soscij.2009.03.003

Leippe, M. R., & Eisenstadt, D. (2007). Eyewitness confidence and the confidence-accuracy relationship in memory for people. In D. S. Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Handbook of Eyewitness Psychology. Volume II: Memory for People* (pp. 377-426). Mahwah, NJ: Lawrence Erlbaum Associates.

Lewis, M. B. (2006). Eye-witnesses should not do cryptic crosswords prior to identity parades. *Perception, 35*, 1433-1436. doi:10.1068/p5666

Lin, F. R., Niparko, J. K., & Ferrucci, L. (2011). Hearing loss prevalence in the united states. *Archives of Internal Medicine, 171*(20), 1851-1853. doi:10.1001/archinternmed.2011.506

Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology, 70*(3), 556-564. doi:10.1037/0021-9010.70.3.556

Lipsey, M. W., & Wilson, D. B. (2001). Applied social research methods series; Vol. 49. Practical meta-analysis: Thousand Oaks, CA: Sage Publications.

List, J. A. (1986). Age and schematic differences in the reliability of eyewitness testimony. *Developmental Psychology, 22*(1), 50-57. doi:10.1037/0012-1649.22.1.50

Loftus, E. F. (2005). Planting misinformation in the human mind: A 30-year investigation of the malleability of memory. *Learning & Memory, 12*(4), 361-366. doi:10.1101/lm.94705

Loftus, E. F., & Greenspan, R. L. (2017). If I'm Certain, Is It True? Accuracy and Confidence in Eyewitness Memory. *Psychological Science in the Public Interest, 18*(1), 1-2. doi:10.1177/1529100617699241

Mackay, T. L., & Paterson, H. M. (2015). How does timing of recall affect eyewitness memory and psychological distress? *Journal of Police and Criminal Psychology, 30*(4), 242-253. doi:10.1007/s11896-014-9156-z

Maddox, L., Lee, D., & Barker, C. (2011). Police empathy and victim PTSD as potential factors in rape case attrition. *Journal of Police and Criminal Psychology, 26*(2), 112-117. doi:10.1007/s11896-010-9075-6

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66*(4), 482-489. doi:10.1037/0021-9010.66.4.482

Malpass, R. S., Tredoux, C. G., Schreiber Compo, N., McQuiston-Surrett, D., Maclin, O. H., Zimmerman, L. A., & Topp, L. D. (2008). Study space analysis for policy development. *Applied Cognitive Psychology, 22*(6), 789-801. doi:10.1002/acp.1483

Mandler, G. (1980). Recognizing: The judgment of previous occurrence. *Psychological Review, 87*(3), 252. doi:10.1037/0033-295X.87.3.252

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*(5), 693-701. doi:10.1002/acp.3055

Marche, T., Jordan, J. J., & Owre, K. P. (2002). Younger adults can be more suggestible than older adults: The influence of learning differences on misinformation reporting. *Canadian Journal on Aging, 21*(1), 85-93. doi:10.1017/S0714980800000660

Marsh, E. J., Tversky, B., & Hutson, M. (2005). How eyewitnesses talk about events: Implications for memory. *Applied Cognitive Psychology, 19*(5), 531-544. doi:10.1002/acp.1095

Mauer, J., Hewig, J., & Kraus, U. (2013). *Das Selbstadministrierte Interview und der Einfluss von Suggestion und positiven Teststrategien auf die Zeugenaussage.* (M.Sc. Thesis), University of Würzburg.

McCabe, K. A., & Gregory, S. S. (1998). Elderly victimization. *Research on Aging, 20*(3), 363-372. doi:10.1177/0164027598203005

McCloskey, M., & Zaragoza, M. (1985). Misleading postevent information and memory for events: Arguments and evidence against memory impairment hypotheses. *Journal of Experimental Psychology, 114*(1), 1-16. doi:10.1037/0096-3445.114.1.1

McDonald, J. H. (2009). *Handbook of biological statistics* (Vol. 2): Sparky House Publishing Baltimore, MD.

McMahon, M. (2000). The effect of the enhanced cognitive interview on recall and confidence in elderly adults. *Psychiatry, Psychology and Law, 7*(1), 9-32. doi:10.1080/13218710009524968

McPhee, I., Paterson, H. M., & Kemp, R. I. (2014). The power of the spoken word: Can spoken-recall enhance eyewitness evidence? *Psychiatry, Psychology and Law, 21*(4), 551-566. doi:10.1080/13218719.2013.848001

Mecacci, L., & Zani, A. (1983). Morningness-eveningness preferences and sleep-waking diary data of morning and evening types in student and worker samples. *Ergonomics, 26*(12), 1147 - 1153.

Meissner, C. A., & Brigham, J. C. (2001a). A meta-analysis of the verbal overshadowing effect in face identification. *Applied Cognitive Psychology, 15*(6), 603-616. doi:10.1002/acp.728

Meissner, C. A., & Brigham, J. C. (2001b). Thirty years of investigating the own-race bias in memory for faces: A meta-analytic review. *Psychology, Public Policy, and Law, 7*(1), 3-35. doi:10.1037/1076-8971.7.1.3

Memon, A. (2008). Eyewitness Research: Theory and Practice. In D. Canter & R. Zukauskiene (Eds.), *Psychology and Law: Bridging the Gap*. Aldershot, UK: Ashgate Publishing Ltd.

Memon, A., Bartlett, J. C., Rose, R., & Gray, C. (2003). The aging eyewitness: Effects of age on face, delay, and source-memory ability. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 58*(6), P338-P345. doi:10.1093/geronb/58.6.P338

Memon, A., Gabbert, F., & Hope, L. (2004). The ageing eyewitness. In J. R. Adler (Ed.), *Forensic Psychology. Concepts, Debates and Practice* (pp. 96-112). Cullompton, England: Willan.

Memon, A., Hope, L., Bartlett, J. C., & Bull, R. (2002). Eyewitness recognition errors: The effects of mugshot viewing and choosing in young and old adults. *Memory & Cognition, 30*(8), 1219-1227. doi:10.3758/BF03213404

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*(4), 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

Menec, V. H. (2003). The relation between everyday activities and successful aging: A 6-year longitudinal study. *J Gerontol B Psychol Sci Soc Sci, 58*(2), S74-82.

Mitchell, K. J., Johnson, M. K., & Mather, M. (2003). Source monitoring and suggestibility to misinformation: adult age-related differences. *Applied Cognitive Psychology, 17*(1), 107-119. doi:10.1002/acp.857

Moulin, C. J. A., Thompson, R. G., Wright, D. B., & Conway, M. A. (2007). Eyewitness memory in older adults. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *Handbook of Eyewitness Psychology. Volume I: Memory for Events* (pp. 627-646). Mahwah, NJ: Lawrence Erlbaum Associates.

Mueller-Johnson, K., & Ceci, S. J. (2004). Memory and suggestibility in older adults: Live event participation and repeated interview. *Applied Cognitive Psychology, 18*, 1109-1127.

Mueller-Johnson, K., & Ceci, S. J. (2007). The elderly eyewitness: A review and prospectus. In M. P. Toglia, J. D. Read, D. F. Ross, & R. C. L. Lindsay (Eds.), *Handbook of Eyewitness Psychology. Volume I: Memory for Events* (pp. 577-603). Mahwah, NJ: Lawrence Erlbaum Associates.

Multhaup, K. S., de Leonardis, D. M., & Johnson, M. K. (1999). Source memory and eyewitness suggestibility in older adults. *Journal of General Psychology, 126*(1), 74-84. doi:10.1080/00221309909595352

Neuschatz, J. S., Preston, E. L., Burkett, A. D., Toglia, M. P., Lampinen, J. M., Fairless, A. H., . . . Goodsell, C. A. (2005). The effects of post -identification feedback and age on retrospective eyewitness memory. *Applied Cognitive Psychology, 19*(4), 435-453. doi:10.1002/acp.1084

Nilsson, L.-G. (2003). Memory function in normal aging. *Acta Neurologica Scandinavica, 107*, 7-13. doi:10.1034/j.1600-0404.107.s179.5.x

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical assessment, research & evaluation, 9*(6), 1-12.

Östberg, O. (1973). Interindividual differences in circadian fatigue patterns of shift workers. *British Journal of Industrial Medicine, 30*, 341-351. doi:10.1136/oem.30.4.341

Pansky, A., & Nemets, E. (2012). Enhancing the quantity and accuracy of eyewitness memory via initial memory testing. *Journal of Applied Research in Memory and Cognition, 1*(1), 2-10. doi:10.1016/j.jarmac.2011.06.001

Paterson, H. M., Eijkemans, H., & Kemp, R. I. (2015). Investigating the impact of delayed administration on the efficacy of the Self-Administered Interview. *Psychiatry, Psychology and Law, 22*(2), 307-317. doi:10.1080/13218719.2014.947670

Peirce, J. W. (2007). PsychoPy - Psychophysics Software in Python. *Journal of Neuroscience Methods, 162*(1-2), 8-13. doi:10.1016/j.jneumeth.2006.11.017

Pezdek, K., & Roe, C. (1995). The effect of memory trace strength on suggestibility. *Journal of Experimental Child Psychology, 60*(1), 116-128. doi:10.1006/jecp.1995.1034

Pfeil, K., & Mueller-Johnson, K. (2016). *The effectiveness of the Self-Administered Interview©. A meta-analytic review and empirical study with older adult witnesses.* (Ph.D. Dissertation), University of Cambridge.

Phillips, M. R., McAuliff, B. D., Kovera, M. B., & Cutler, B. L. (1999). Double-blind photoarray administration as a safeguard against investigator bias. *Journal of Applied Psychology, 84*(6), 940-951. doi:10.1037/0021-9010.84.6.940

Powell, M. B., Fisher, R. P., Wright, R., Brewer, N., & Williams, K. (2005). Investigative interviewing. *Psychology and law: An empirical perspective*, 11-42.

Rabinowitz, J. C., Craik, F. I. M., & Ackerman, B. P. (1982). A processing resource account of age differences in recall. *Canadian Journal of Psychology, 36*, 325-344. doi:10.1037/h0080643

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement, 1*, 385-401. doi:10.1177/014662167700100306

Rhodes, M. G., & Anastasi, J. S. (2012). The own-age bias in face recognition: A meta-analytic and theoretical review. *Psychological Bulletin, 138*(1), 146-174. doi:10.1037/a0025750

Rose, R., Bull, R., & Vrij, A. (2005). Non-biased lineup instructions do matter - a problem for older witnesses. *Psychology, Crime and Law, 11*(2), 147-159. doi:10.1080/10683160512331316307

Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin, 86*(3), 638-641. doi:10.1037/0033-2909.86.3.638

Ryan, L., Hatfield, C., & Hofstetter, M. (2002). Caffeine reduces time-of-day effects on memory performance in older adults. *Psychological Science, 13*(1), 68-71. doi:10.1111/1467-9280.00412

Sala-Llonch, R., Junqué, C., Arenaza-Urquijo, E. M., Vidal-Piñeiro, D., Valls-Pedret, C., Palacios, E. M., . . . Bartrés-Faz, D. (2014). Changes in whole-brain functional networks and memory performance in aging. *Neurobiology of*

Aging, 35(10), 2193-2202. doi: https://doi.org/10.1016/j.neurobiolaging.2014.04.007

Salthouse, T. A. (1996). The processing-speed theory of adult age differences in cognition. *Psychological Review, 103*, 403-428. doi:10.1037/0033-295X.103.3.403

Salthouse, T. A., Mitchell, D. R. D., Skovronek, E., & Babcock, R. L. (1989). Effects of adult age and working memory on reasoning and spatial abilities. *Journal of Experimental Psychology: Learning, Memory and Cognition, 15*, 505-516. doi:10.1037/0278-7393.15.3.507

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods, 8*(2), 597 – 599.

Saywitz, K. J., Wells, C. R., Larson, R. P., & Hobbs, S. D. (2016). Effects of Interviewer Support on Children's Memory and Suggestibility:Systematic Review and Meta-Analyses of Experimental Research. *Trauma, Violence, & Abuse, 0*(0), 1524838016683457. doi:10.1177/1524838016683457

Schacter, D. L., Koutstaal, W., & Norman, K. A. (1997). False memories and aging. *Trends in Cognitive Sciences, 1*(6), 229-236. doi: https://doi.org/10.1016/S1364-6613(97)01068-1

Schmand, B., Smit, J., Lindeboom, J., Smits, C., Hooijer, C., Jonker, C., & Deelman, B. (1997). Low education is a genuine risk factor for accelerated memory decline and dementia. *Journal of Clinical Epidemiology, 50*(9), 1025-1033. doi:10.1016/S0895-4356(97)00121-2

Schneider, B. A., & Pichora-Fuller, M. K. (2000). Implications of perceptual deterioration for cognitive aging research. In F. I. M. Craik & T. A. Salthouse (Eds.), *The Handbook of Aging and Cognition* (2nd ed., pp. 155-219). Mahwah, NJ: Lawrence Erlbaum Associates.

Schoof, M. C., Hewig, J., & Kraus, U. (2014). *Das Selbstadministrierte Interview bei hoher und niedriger Suggestibilität.* (M.Sc. Thesis), University of Würzburg.

Schultz-Larsen, K., Lomholt, R. K., & Kreiner, S. (2007). Mini-Mental Status Examination: A short form of MMSE was as accurate as the original MMSE in predicting dementia. *Journal of Clinical Epidemiology, 60*(3), 260-267. doi:10.1016/j.jclinepi.2006.06.008

Scogin, F., Calhoon, S. K., & D'Errico, M. (1994). Eyewitness confidence and accuracy among three age cohorts. *Journal of Applied Gerontology, 13*(2), 172-184. doi:10.1177/073346489401300205

Searcy, J., Bartlett, J. C., & Memon, A. (1999). Age differences in accuracy and choosing in eyewitness identification and face recognition. *Memory & Cognition, 27*(3), 538-552. doi:10.3758/BF03211547

Searcy, J., Bartlett, J. C., & Memon, A. (2000). Influence of post-event narratives, line-up conditions and individual differences on false identification by young and older eyewitnesses. *Legal and Criminological Psychology, 5*(2), 219-235. doi:10.1348/135532500168100

Shaw, J. S., Bjork, R. A., & Handal, A. (1995). Retrieval-induced forgetting in an eyewitness-memory paradigm. *Psychonomic Bulletin & Review, 2*(2), 249-253. doi:10.3758/BF03210965

Shepherd, E., & Griffiths, A. (2013). *Investigative Interviewing: The Conversation Management Approach*: OUP Oxford.

Smith, C. S., Reilly, C., & Midkiff, K. (1989). Evaluation of three circadian rhythm questionnaires with suggestions for an improved measure of morningness. *Journal of Applied Psychology, 74*, 728-738. doi:10.1037/0021-9010.74.5.728

Souchay, C., Isingrini, M., & Espagnet, L. (2000). Aging, episodic memory feeling-of-knowing, and frontal functioning. *Neuropsychology, 14*(2), 299-309. doi:10.1037/0894-4105.14.2.299

Sporer, S. L., Penrod, S., Read, J. D., & Cutler, B. L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118*(3), 315-327. doi:10.1037/0033-2909.118.3.315

St-Yves, M. (2013). The psychology of rapport: Five basic rules. In T. Williamson (Ed.), *Investigative interviewing. Rights, research, regulation.* (pp. 87-106). London: Routledge.

Steblay, N., Dysart, J., Solomon, F., & Lindsay, R. C. L. (2001). Eyewitness accuracy rates in sequential and simultaneous lineup presentations: A meta-analytic comparison. *Law and Human Behavior, 25*(5), 459-473. doi:10.1023/A:1012888715007

Stephan, A., Hewig, J., & Kraus, U. (2013). *Suggestion und das Selbstadministrierte Interview.* (M.Sc. Thesis), University of Würzburg.

Sternberg, K. J., Lamb, M. E., Orbach, Y., Esplin, P. W., & Mitchell, S. (2001). Use of a structured investigative protocol enhances young children's responses to free-recall prompts in the course of forensic interviews. *Journal of Applied Psychology, 86*(5), 997-1005. doi:10.1037/0021-9010.86.5.997

Szpitalak, M., & Polczyk, R. (2016). Reinforced Self-affirmation and Interrogative Suggestibility. *Psychiatry, Psychology and Law, 23*(4), 512-520. doi:10.1080/13218719.2015.1081088

Tombaugh, T. N., McDowell, I., Kristjansson, B., & Hubley, A. M. (1996). Mini-Mental State Examination (MMSE) and the Modified MMSE (3MS): A psychometric comparison and normative data. *Psychological Assessment, 8*(1), 48-59. doi:10.1037/1040-3590.8.1.48

Troje, N. F., & Kersten, D. (1999). Viewpoint-dependent recognition of familiar faces. *Perception, 28*(4), 483-487. doi:10.1068/p2901

Tversky, B., & Marsh, E. J. (2000). Biased retellings of events yield biased memories. *Cognitive Psychology, 40*(1), 1-38. doi:10.1006/cogp.1999.0720

United Nations Population Division. (2008). World population prospects: The 2008 revision. Retrieved from http://www.un.org/esa/population/unpop.htm

Valentine, T. (1991). A unified account of the effects of distinctiveness, inversion, and race in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 43*(2), 161-204. doi:10.1080/14640749108400966

Valentine, T., Lewis, M. B., & Hills, P. J. (2016). Face-space: A unifying concept in face recognition research. *The Quarterly Journal of Experimental Psychology, 69*(10), 1996-2019. doi:10.1080/17470218.2014.990392

Van Gerven, P. W. M., & Guerreiro, M. J. S. (2016). Selective Attention and Sensory Modality in Aging: Curses and Blessings. *Frontiers in Human Neuroscience, 10*, 147. doi:10.3389/fnhum.2016.00147

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, 47*, 1063–1070. doi:10.1037/0022-3514.54.6.1063

Wells, G. L. (1978). Applied eyewitness-testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 36*(12), 1546-1557. doi:10.1037/0022-3514.36.12.1546

Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist, 48*(5), 553-571. doi:10.1037/0003-066x.48.5.553

Wells, G. L., & Bradfield, A. L. (1998). Good, you identified the suspect : Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83*(3), 360-376. doi:10.1037/0021-9010.83.3.360

Wells, G. L., Lindsay, R. C., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology, 64*(4), 440-448. doi:10.1037/0021-9010.64.4.440

Wells, G. L., Malpass, R. S., Lindsay, R. C. L., Fisher, R. P., Turtle, J. W., & Fulero, S. M. (2000). From the lab to the police station: A successful application of eyewitness research. *American Psychologist, 55*(6), 581-598. doi:10.1037/0003-066x.55.6.581

Welsh, B. C., & Farrington, D. P. (2007). Evidence-Based Crime Prevention. In B. C. Welsh & D. P. Farrington (Eds.), *Preventing Crime: What Works for Children, Offenders, Victims and Places* (pp. 1-17). New York, NY: Springer New York.

Weuve, J., Kang, J. H., Manson, J. E., Breteler, M. M. B., Ware, J. H., & Grodstein, F. (2004). Physical activity, including walking, and cognitive function in older women. *JAMA, 292*(12), 1454-1461. doi:10.1001/jama.292.12.1454

Wilcock, R. A., & Bull, R. (2010). Novel lineup methods for improving the performance of older eyewitnesses. *Applied Cognitive Psychology, 24*(5), 718-736. doi:10.1002/acp.1582

Wilcock, R. A., Bull, R., & Milne, R. (2008). *Witness Identification in Criminal Cases: Psychology and Practice*. Oxford, England: Oxford University Press.

Wilcock, R. A., Bull, R., & Vrij, A. (2005). Aiding the performance of older eyewitnesses: Enhanced non-biased line-up instructions and line-up presentation. *Psychiatry, Psychology and Law, 12*(1), 129-140. doi:10.1375/pplt.2005.12.1.129

Wilcock, R. A., Bull, R., & Vrij, A. (2007). Are old witnesses always poorer witnesses? Identification accuracy, context reinstatement, own-age bias. *Psychology, Crime & Law, 13*(3), 305-316. doi:10.1080/10683160600822212

Wilson, D. B. (2001). Practical Meta-Analysis Effect Size Calculator. Retrieved from http://www.campbellcollaboration.org/resources/effect_size_input.php

Wilson, D. B. (2010). Meta-analysis macros for SAS, SPSS, and Stata. Retrieved from http://mason.gmu.edu/~dwilsonb/ma.html

Wixted, J. T., & Ebbesen, E. B. (1991). On the form of forgetting. *Psychological Science, 2*(6), 409-415. doi:10.1111/j.1467-9280.1991.tb00175.x

Wixted, J. T., & Wells, G. L. (2017). The Relationship Between Eyewitness Confidence and Identification Accuracy: A New Synthesis. *Psychological Science in the Public Interest, 18*(1), 10-65. doi:10.1177/1529100616686966

Wright, A. M., & Holliday, R. E. (2007a). Enhancing the recall of young, young-old and old-old adults with cognitive interviews. *Applied Cognitive Psychology, 21*(1), 19-43.

Wright, A. M., & Holliday, R. E. (2007b). Interviewing cognitively impaired older adults: How useful is a Cognitive Interview? *Memory, 15*(1), 17-33.

Wright, D. B., & Stroud, J. N. (2002). Age differences in lineup identification accuracy: People are better with their own age. *Law and Human Behavior, 26*(6), 641-654. doi:10.1023/A:1020981501383

Yarmey, A. D. (2001). The older eyewitness. In M. B. Rothman, B. D. Dunlop, & P. Entzel (Eds.), *Elders, Crime and the Criminal Justice System* (pp. 127-148). New York: Springer.

Yarmey, A. D., Jones, H. T., & Rashid, S. (1984). Eyewitness memory of elderly and young adults. In D. J. Muller, D. E. Blackman, & A. J. Chapman (Eds.), *Psychology and Law* (pp. 215-228). Chichester, England: Wiley.

Yarmey, A. D., & Kent, J. (1980). Eyewitness identification by elderly and young adults. *Law and Human Behavior, 4*(4), 359-371. doi:10.1007/BF01040627

Zander, M. (1990). *The Police and Criminal Evidence Act*. London: Sweet and Maxwell.

Zaragoza, M. S., & Lane, S. M. (1994). Source misattributions and the suggestibility of eyewitness memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(4), 934. doi:10.1037/0278-7393.20.4.934

Zeier, F., Hewig, J., Kraus, U., & Wagner, W. (2016). *Does the SAI trump an initial police officer's questioning?* Manuscript under review.

6. APPENDIX

Appendix A. Overview of moderator coding for all studies and subgroups in the meta-analyses.

Table A1. Moderator coding for all studies and subgroups included in the meta-analyses.

Study code	Subgroup	Laboratory	Published	Age group	Delay	Control	T2 recall type	T2 recall modality
Af Hjelmsaeter 2012 (1)	T2 FR, SAI vs. no	0	1	0	2	0	0	1
Af Hjelmsaeter 2012 (2)	T2 FR, SAI vs. FR	0	1	0	2	1	0	1
Boessenrodt 2011 (1)	T2 FR, SAI vs. FR	0	0	1	1	1	0	1
Colomb 2013	T2 FR, SAI vs. no	0	0	1	1	0	0	1
Gabbert 2009 (1)	T1, SAI vs. FR	1	1	1	0	1	-	-
Gabbert 2009 (2)	T2 FR, SAI vs. no	1	1	1	1	0	0	0
Gabbert 2012 (1)	T2 FR, SAI vs. no	1	1	1	1	0	0	0
Gabbert 2012 (2)	T2 CR, SAI vs. no	1	1	1	3	0	1	0
Gawrylowicz 2014a	T1, SAI vs. FR	0	1	1	0	1	-	-
Gawrylowicz 2014b	T1, SAI vs. FR	0	1	0	0	1	-	-
Gittins 2015	T2 CR, SAI vs. no	0	1	1	1	0	1	0
Hope 2014 (1)	T1, SAI vs. FR	1	1	1	0	1	-	-
Hope 2014 (2)	T2 FR, SAI vs. no	1	1	1	1	0	0	1
Hope 2014 (3)	T2 FR, SAI vs. FR	1	1	1	1	1	0	1
Kemp 2016	T1, SAI vs. FR	0	0	1	0	1	-	-
Krix 2015a (1)	T1, SAI vs. FR	0	1	1	0	1	-	-
Krix 2015a (2)	T2 CR, SAI vs. FR	0	1	1	1	1	1	0
Krix 2015b	T1, SAI vs. FR	1	1	1	0	1	-	-
Krix 2016	T1, SAI vs. FR	0	1	1	0	1	-	-
Mackay 2015	T2 CR, SAI vs. no	0	1	1	1	0	1	0
Maras 2014 (1)	T1, SAI vs. SR	0	1	1	0	1	-	-
Maras 2014 (2)	T2 CR, SAI vs. SR	0	1	1	1	1	1	0

Mauer 2013 (1)	T1, SAI vs. FR	0	0	1	0	1	-	-
Mauer 2013 (2)	T2 FR, SAI vs. FR	0	0	1	1	1	0	1
McPhee 2014	T2 CR, SAI vs. no	0	1	1	1	0	1	1
Paterson 2015	T2 FR, SAI vs. no	0	1	1	2	0	0	0
Pfeil 2016 (1)	T1, SAI vs. FR	0	0	1	0	1	-	-
Pfeil 2016 (2)	T1, SAI vs. FR	0	0	0	0	1	-	-
Pfeil 2016 (3)	T2 FR, SAI vs. no	0	0	1	1	0	0	1
Pfeil 2016 (4)	T2 FR, SAI vs. no	0	0	0	1	0	0	1
Pfeil 2016 (5)	T2 FR, SAI vs. FR	0	0	1	1	1	0	1
Pfeil 2016 (6)	T2 FR, SAI vs. FR	0	0	0	1	1	0	1
Schoof 2014 (1)	T1, SAI vs. FR	0	0	1	0	1	-	-
Schoof 2014 (2)	T2 FR, SAI vs. FR	0	0	1	1	1	0	1
Stephan 2013 (1)	T1, SAI vs. FR	0	0	1	0	1	-	-
Stephan 2013 (2)	T2 FR, SAI vs. FR	0	0	1	1	1	0	1
Zeier 2016 (1)	T1, SAI vs. SR	0	0	1	0	1	-	-
Zeier 2016 (2)	T2 CR, SAI vs. SR	0	0	1	1	1	1	0

Note. Laboratory: 1 = Gabbert/Hope, 0 = other. Published: 1 = yes, 0 = no. Age group: 1 = adults, 0 = children and elderly. Delay: in weeks. Control: 0 = no recall, 1 = other recall. T2 recall type: 0 = free recall, 1 = cued recall. T2 recall modality: 1 = oral, 0 = written.

SAI Self Administered Interview

SECTION A: What happened?

Please read through the information below and follow the instructions

Before recording any of your memories of the event in this booklet, take a few moments to picture in your mind where you were, what you saw, what you were thinking and how you were feeling at the time

Why should you do this?

Thinking about the event before writing down your memories will help you remember more details.

What should you do now?

Give yourself plenty of time to concentrate, and visualise what happened in your mind.

Think about the following things:

- Where you were
- What you were doing
- Who you were with
- How you were feeling
- What was happening
- Who was involved
- What you could see
- What you could hear

It may help to shut your eyes while you remember the event.

Once you are satisfied that you have managed to fully remember the event, please turn the page to continue.

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.

SAI Self Administered Interview

Please read through the information below and follow the instructions

In the space provided, write down everything that you can remember about the event and the people who were involved.

Write things down as you remember them. It doesn't matter if you remember things out of the order in which they happened.

Do not leave out any details, but **do not guess** about details that you cannot remember.

Feel free to use full sentences or bullet points – but please make sure your report is as **complete** and **accurate** as possible.

Remember: Please complete this booklet alone without seeking the assistance of others. We are interested only in your own memories of the event.

Start here:

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.

SAI Self Administered Interview

Continue Here:

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.

SAI Self Administered Interview

Continue Here:

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.

Continue Here:

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.

SAI Self Administered Interview

SECTION B: Who committed the crime?

Please read through the information below and follow the instructions

Please provide as much detail as you can about the offender (the person who committed the crime or who was involved in the incident) on the page opposite. If there was more than one offender, begin with the person you had the best view of or remember most clearly.

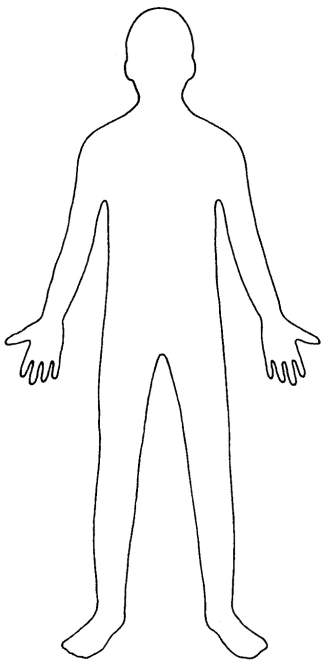
If you have already provided a description elsewhere in this booklet, take a few moments to consider whether there are any further details you can recall now.

If possible provide the following information about each person. **Do not guess** about anything you are unsure of.

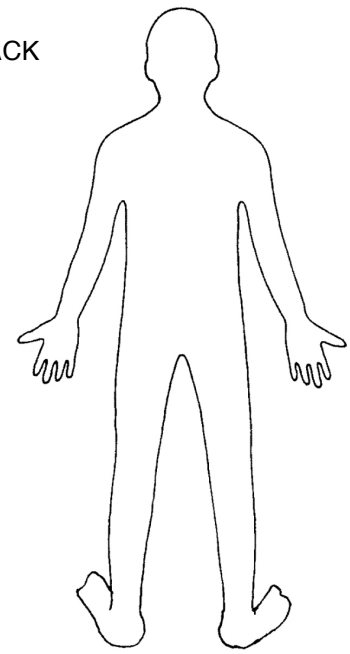
- Gender
- Apparent Age
- Height
- Ethnic origin
- Weight / Build
- Eyes / Ears / Mouth / Nose / etc.
- Hair Colour
- Facial Hair
- Complexion
- Clothing / Shoes
- Accent
- Glasses
- Jewellery
- Accessories
- Scars / Marks / Tattoos
- Any other details about the offender(s) that we have not asked about?

If you like you can use the diagram figures below to add further information. You can write and/or draw on the figures to show additional details or information (e.g. position of any accessories, etc.)

FRONT



BACK



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.

Provide description(s) here

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.

SAI Self Administered Interview

SECTION C: The scene

Please read through the information below and follow the instructions

Making a sketch of the scene may help you remember more details – and provide further information about the event.

Please use the space below, and/or on the next page, to sketch the scene as you remember it.

You should include details of where you were, and where other people were. You can use arrows to indicate the movement of yourself and other people you saw.

You can use labels and notes within your sketch to indicate features of the scene, or to indicate if you are not certain of something.

This is not a test of your drawing ability – we are only interested in the layout of the scene, e.g., what you saw, and where.

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.

Sketch here:

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.

SAI Self Administered Interview

SECTION D: Were other people present who saw what happened?

YES (complete this section) **NO** (go to Section E)

Please provide a description of anyone else who was present, and who might have also seen what happened, but who was not involved in the incident (i.e., any other potential witnesses)

Start Here:

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.

Continue Here:

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.

SAI Self Administered Interview

SECTION E: Were there any vehicles involved?

YES (complete this section) **NO** (go to Section F)

Please read through the information below and follow the instructions

Please provide as much detail as you can about the vehicle(s) involved in the incident.

If you have already provided a description elsewhere in this booklet, take a few moments to consider whether there are any further details you can recall now.

If possible provide the following information about each vehicle. **Do not guess** about anything you are unsure of.

- Size
- Shape
- Colour
- Make / Model
- Number of Doors
- Registration Number
- Driving style
- Speed
- Anything else?

Start Here:

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.

SAI Self Administered Interview

SECTION F: How well did you see the incident?

Please answer the following questions

- How long did you witness the event for?

- What were the weather conditions like at the time?

- What time of day did the event occur?

- Were there any obstructions to your view?

- Are there any particular reasons for remembering the event, or the offender(s)?

- Was anyone involved that you know, or who you have seen before? (If so, where and when?)

- How much time has past between witnessing the event and completing this booklet?

- It is often helpful to have a description of the witness and their clothing to assist with viewing CCTV etc. If you think that is relevant in these circumstances could you provide a description of yourself and what you were wearing when you witnessed this incident.

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.

SAI Self Administered Interview

SECTION G: Is there any other information you would like to tell us about the event that we have not asked you about? If so, write it in the space below.

Start Here:

PLEASE RETURN TO:

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.

Appendix C. Written free recall form.

Free Recall

Now please think back to the film you watched earlier.
Please write down everything you can remember about the film in as much detail as possible. You can provide information about the persons shown in the film and their actions. You can also include descriptions of the surrounding. Please report everything you can remember even if it seems unimportant to you. However, do not guess about details that you cannot remember.

Feel free to use full sentences or bullet points.

Appendix D. Eyewitness Interview Form.

1. FREE RECALL PART

Now please think back to the film you watched a week ago.

Please tell me everything you can remember about the film in as much detail as possible.

- Let participant talk until they don't remember anything else.
- Then clarify anything that wasn't clear to you/ or was ambiguous in their report.
- Is there anything else you can remember or would you like to add any more detail?

2. QUESTION PART

I am now going to ask you some more questions. It is possible that you have covered some of this already, but if so, please go through it again. I am asking all participants the same questions so that I can better compare people's answers.

1.S_P. Please describe the surroundings in as much detail as possible.

2. S_P. What time of day did the event occur?

_____

3. S_P. What were the weather conditions like at the time?

_____

4. P_C. How many people were directly involved in the crime?

_____

5. P_C. Were they men or women?

_____

6. Can you now please describe these 3 *(tell number)* people in detail, one by one? I would like to know what they were wearing, what their appearances were and what they were doing. For the first one you mentioned, please describe his/her clothing.

Perpetrator (Sex):	**Victim 1 (Sex):**	**Victim 2 (Sex):**
Clothing: - Legs - Upper body - Shoes - Accessories	*Clothing:* - Legs - Upper body - Shoes - Accessories	*Clothing:* - Legs - Upper body - Shoes - Accessories
Appearance: - Age - Ethnic origin - Height - Build - Hair length - Hair colour - Distinguishing features	*Appearance:* - Age - Ethnic origin - Height - Build - Hair length - Hair colour - Distinguishing features	*Appearance:* - Age - Ethnic origin - Height - Build - Hair length - Hair colour - Distinguishing features
Action:	*Action:*	*Action:*

Any other details about the person we have not asked about?

I am now going to ask some more specific questions about these people.

7. P_C_SU. What did the perpetrator's tattoo look like?

_____

8. A_C. What exactly did the perpetrator do?

_____

9. P_C_SU_FC. Did the perpetrator wear a blue or a yellow sweater?

_____

10. A_C. What happened to the first victim?

_____

11. A_C_SU. How did the first victim react after noticing the perpetrator?

_____

12. O_C_SU. What did the perpetrator steal from the first victim?

_____

13. A_C_SU_YN. Did the first victim scream? Yes/No

14. A_C. What happened to the second victim?

_____

15. S_C. From which direction did the perpetrator come for the second time?

_____

16. A_C. What was the second victim doing before the perpetrator stole her bag?

_____

17. O_C. What did the stolen bag look like (colour/ type/ size)

_____

18. A_C_SU. What did the perpetrator say to the second victim?

_____

19. O_C_SU. What kind of weapon did the perpetrator use to threaten the second victim?

_____

20. A_C_SU. How did the second victim react after having fallen down to the ground?

_____

21. A_C. What did the second victim do after she was robbed?

_____

22. A_C. How did the perpetrator escape?

_____

23. S_C. In which direction did he escape?

_____

24. A_C_SU. What was the first thing the policeman did when he arrived?

_____

25. A_C_YN. Did the perpetrator push the second victim? Yes/No

26. A_C_YN. Did the second victim scream? Yes/No

27. O_C_SU_YN. Was any of the victims' handbags red? Yes/No

Now we have covered the people directly involved in the crime. Next I would like you to focus on potential witnesses.

28. P_P. Was there anyone who could have witnessed the crime and/or the perpetrator?

_____

29. Can you now please describe anyone who could have witnessed the event or seen the perpetrator in detail, one by one? I would like to know what they were wearing, what their appearances were and what they were doing. For the first one you mentioned, please describe his/her clothing.

Person 1 (Sex):	**Person 2 (Sex):**	**Person 3 (Sex):**
Clothing: - Legs - Upper body - Shoes - Accessories	*Clothing:* - Legs - Upper body - Shoes - Accessories	*Clothing:* - Legs - Upper body - Shoes - Accessories
Appearance: - Age - Ethnic origin - Height - Build - Hair length - Hair colour - Distinguishing features	*Appearance:* - Age - Ethnic origin - Height - Build - Hair length - Hair colour - Distinguishing features	*Appearance:* - Age - Ethnic origin - Height - Build - Hair length - Hair colour - Distinguishing features

Action:	*Action:*	*Action:*

Again, I am now going to ask you some more specific questions about these people.

30. P_C_SU. Who chased the perpetrator?

_____

31. P_P. What pattern was on the shirt of the man helping the second victim?

_____

32. P_P_SU. What colour was the hat of the man helping the second victim?

_____

33. A_P_YN. Did the man trying to help the second victim call someone? Yes/No

34. P_P_SU_YN Did anyone challenge or say anything to the perpetrator? Yes/No

So far you have described the perpetrator and the victims and you have described the people who could have witnessed the crime. In the last section I would now like you to focus on any other people in the scene.

35. Can you now please describe anyone else you can remember, one by one? I would like to know what they were wearing, what their appearances were and what they were doing. For the first one you mentioned, please describe his/her clothing.

Person 1 (Sex):	**Person 2 (Sex):**	**Person 3 (Sex):**
Clothing: - Legs - Upper body - Shoes - Accessories	*Clothing:* - Legs - Upper body - Shoes - Accessories	*Clothing:* - Legs - Upper body - Shoes - Accessories
Appearance: - Age	*Appearance:* - Age	*Appearance:* - Age

- Ethnic origin - Height - Build - Hair length - Hair colour - Distinguishing features	- Ethnic origin - Height - Build - Hair length - Hair colour - Distinguishing features	- Ethnic origin - Height - Build - Hair length - Hair colour - Distinguishing features
Action:	*Action:*	*Action:*

Again, I am now going to ask you some more specific questions about these people.

36. S_P_SU. From which direction did the man in the wheelchair come?

_____

37. O_P. Which colour had the man's umbrella?

_____

38. S_P. What was exhibited in the shop window?

_____

39. O_P_SU. What did the man with the umbrella buy in the shop?

_____

40. S_P_SU. Where did the person walking the dog go to?

_____

41. P_P_SU_FC. Did the running lady wear black or blue jeans?

_____

42. O_P_SU. What colour was the child's bike?

_____

43. P_P_YN. Was there a man with a newspaper? Yes/No

44. P_P_YN. Was the girl with the red rucksack the first person to walk by? Yes/No

45. S_P_YN. Was the supermarket called Savers? Yes/No

46. S_P. How long did the whole film take?

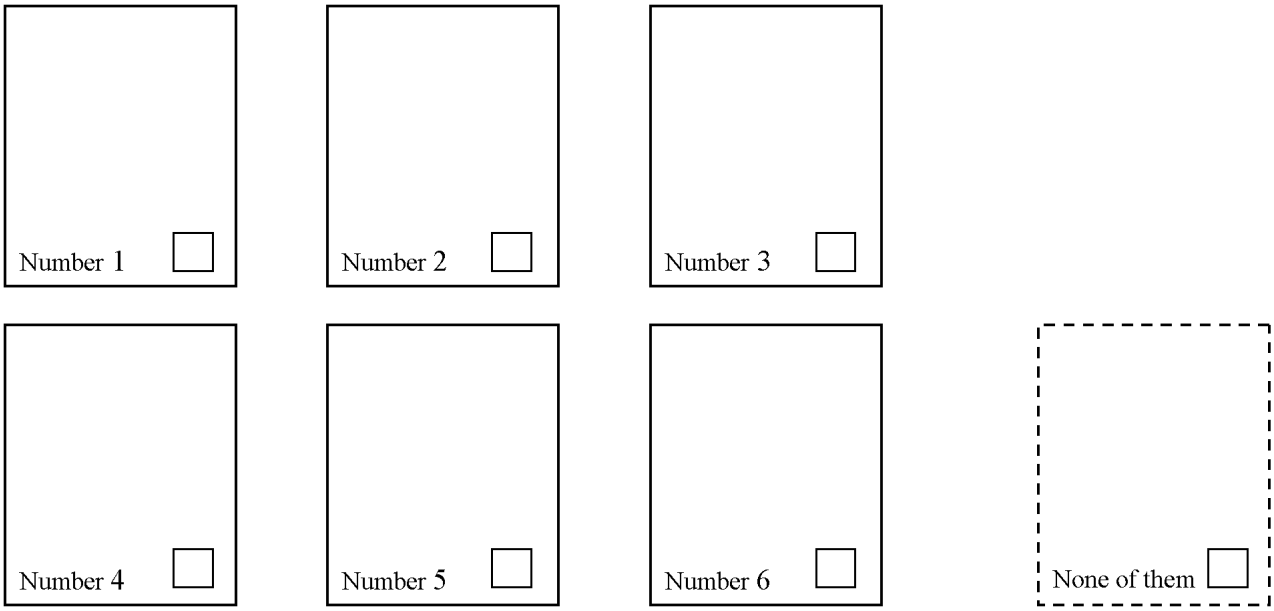
_____

Appendix E. Lineup answer sheet.

You will be presented with 6 photographs on the screen in front of you. Below you find boxes representing each photo. Tick the box on this form that corresponds either to the photo showing the perpetrator from the film or "none of them" if the perpetrator is not present. As with a real police line-up, the perpetrator may or may not be present.

1. Please look carefully at each photograph and take your time.
2. As soon as you have made a decision, **please press the spacebar**.
3. Then tick the corresponding box below.

If you have any questions, please do not hesitate to ask the experimenter now. To start, please press the spacebar.



4. Please now indicate on the scale below how confident you are about the choice you have made.

1	2	3	4	5	6	7

Not at all reasonably highly confident

Appendix F. Minimental State Exam (modified short version).

MMSE	SCORE
1. What is the year/month/day of the week?	1/1/1
2. What city/building are we in?	1/1
3. I'm going to say three words. You say them back when I've finished.	
APPLE COIN CHAIR	1/1/1
Keep those words in mind. I'm going to ask you to say them again in a few minutes.	
4. Now I'd like you to subtract 7 from 100. Keep subtracting 7 from each answer until I tell you to stop. What is 100 take away 7? (alternative: WORLD backwards)	
Record responses: ___ ___ ___ ___ ___	1/1/1/1/1
5. What were those three words I asked you to remember?	1/1/1

Total Score: ____ / 16

Appendix G. Demographic questionnaire.

1.Your sex: ☐ male ☐ female

2.Your age: _____ years

3.What is your subject/occupation or was your profession before retirement? _____

4.What is your highest educational degree?

☐PhD ☐MA ☐BA ☐A-Level ☐O-Level ☐other

5.How would you best describe your ethnic group?

☐ White ☐ Black ☐ Asian ☐ Other, please specify: _____

6.How would you rate your general health?

☐ Excellent ☐ Good ☐ Fair ☐ Poor

7.Do you do any sport or activities to keep fit? (walking, fitness, swimming etc)

☐ Yes ☐ No

8.**If yes**, how frequently do you do sport or activities to stay fit?

☐ Every day
☐ 3 times or more per week
☐ Twice a week
☐ Once a week
☐ Every two weeks
☐ Once a month
☐ On and off but not regularly

9. How awake do you feel now?

☐ Completely awake
☐ Fairly awake
☐ A little tired
☐ Quite tired
☐ Very tired

10.Have you taken anything in the last couple of hours that helps you to feel awake (e.g. coffee, tea, other caffeinated drinks, medication)?

☐ Yes ☐ No

11.Have you taken anything today that makes you feel sleepy? (e.g. medication)

☐ Yes ☐ No

Appendix H. Positive and Negative Effect Schedule.

PANAS Questionnaire

This scale consists of a number of words that describe different feelings and emotions. Read each item and then list the number from the scale below next to each word. Indicate to what extent you feel this way right now, that is, at the present moment.Use the following scale to record your answers.

1	2	3	4	5
very slightly or not at all	A little	Moderately	Quite a bit	Extremely

_____ 1. Interested
_____ 2. Distressed
_____ 3. Excited
_____ 4. Upset
_____ 5. Strong
_____ 6. Guilty
_____ 7. Scared
_____ 8. Hostile
_____ 9. Enthusiastic
_____ 10. Proud
_____ 11. Irritable
_____ 12. Alert
_____ 13. Ashamed
_____ 14. Inspired
_____ 15. Nervous
_____ 16. Determined
_____ 17. Attentive
_____ 18. Jittery
_____ 19. Active
_____ 20. Afraid

Appendix I. Center for Epidemiologic Studies Depression Scale Revised.

Center for Epidemiologic Studies Depression Scale – Revised (CESD-R)

Below is a list of the ways you might have felt or behaved. Please check the boxes to tell me how often you have felt this way in the past week or so.	Last Week				Nearly every day for 2 weeks
	Not at all *or* Less than 1 day	1 - 2 days	3 - 4 days	5 - 7 days	
My appetite was poor.	0	1	2	3	4
I could not shake off the blues.	0	1	2	3	4
I had trouble keeping my mind on what I was doing.	0	1	2	3	4
I felt depressed.	0	1	2	3	4
My sleep was restless.	0	1	2	3	4
I felt sad.	0	1	2	3	4
I could not get going.	0	1	2	3	4
Nothing made me happy.	0	1	2	3	4
I felt like a bad person.	0	1	2	3	4
I lost interest in my usual activities.	0	1	2	3	4
I slept much more than usual.	0	1	2	3	4
I felt like I was moving too slowly.	0	1	2	3	4
I felt fidgety.	0	1	2	3	4
I wished I were dead.	0	1	2	3	4
I wanted to hurt myself.	0	1	2	3	4
I was tired all the time.	0	1	2	3	4
I did not like myself.	0	1	2	3	4
I lost a lot of weight without trying to.	0	1	2	3	4
I had a lot of trouble getting to sleep.	0	1	2	3	4
I could not focus on the important things.	0	1	2	3	4

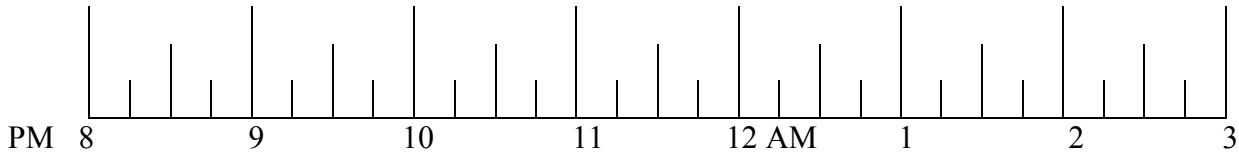
Appendix J. Morningness-Eveningness Questionnaire.

1. Please read each question very carefully before answering.
2. Answer ALL questions.
3. Answer questions in numerical order.
4. Each question should be answered independently of others. Do NOT go back and check your answers.
5. All questions have a selection of answers. For each question place a cross alongside ONE answer only. Some questions have a scale instead of a selection of answers. Place a cross at the appropriate point along the scale.
6. Please answer each question as honestly as possible. Both your answers and the results will be kept in strict confidence.
7. Please feel free to make any comments in the sections provided below each question.

---

1. Considering only your own "feeling best" rhythm, at what time would you get up if you were entirely free to plan your day?

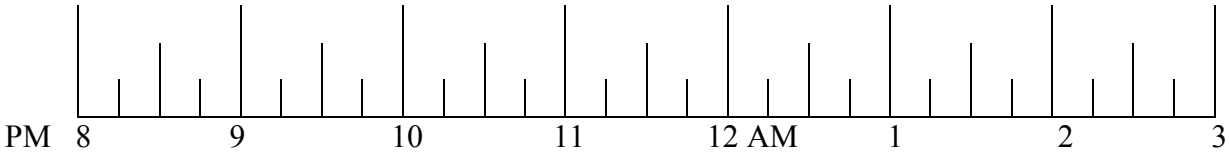
2. Considering only your own "feeling best" rhythm, at what time would you go to bed if you were entirely free to plan your evening?



3. If there is a specific time at which you have to get up in the morning, to what extent are you dependent on being woken up by an alarm clock?	☐ Not at all dependent ☐ Slightly dependent ☐ Fairly dependent ☐ Very dependent
4. Assuming adequate environmental conditions, how easy do you find getting up in the mornings?	☐ Not at all easy ☐ Not very easy ☐ Fairly easy ☐ Very easy
5. How alert do you feel during the first half hour after having woken in the morning?	☐ Not at all alert ☐ Slightly alert ☐ Fairly alert ☐ Very alert

6. How is your appetite during the first half-hour after having woken in the mornings?	☐ Very poor ☐ Fairly poor ☐ Fairly good ☐ Very good
7. During the first half-hour after having woken in the morning, how tired do you feel?	☐ Very tired ☐ Fairly tired ☐ Fairly refreshed ☐ Very refreshed
8. When you have no commitments the next day, at what time do you go to bed compared to your usual bedtime?	☐ Seldom or never later ☐ Less than one hour later ☐ 1-2 hours later ☐ More than two hours later
9. You have decided to engage in some physical exercise. A friend suggests that you do this one hour twice a week and the best time for him is between 7.00 and 8.00 AM. Bearing in mind nothing else but your own "feeling best" rhythm how do you think you would perform?	☐ Would be on good form ☐ Would be on reasonable form ☐ Would find it difficult ☐ Would find it very difficult

10. At what time in the evening do you feel tired and as a result in need of sleep?



11. You wish to be at your peak performance for a test which you know is going to be mentally exhausting and lasting for two hours. You are entirely free to plan your day and considering only your own "feeling best" rhythm which ONE of the four testing times would you choose?	☐ 8.00 - 10.00 AM ☐ 11.00 AM – 1.00 PM ☐ 3.00 PM – 5.00 PM ☐ 7.00 – 9.00 PM
12. If you went to bed 11.00 PM at what level of tiredness would you be?	☐ Not at all tired ☐ A little tired ☐ Fairly tired ☐ Very tired
13. For some reason you have gone to bed several hours later than usual, but there is no need to get up at any particular time the next morning. Which ONE of the following events are you most likely to experience?	☐ Will wake up at usual time and will NOT fall asleep again ☐ Will wake up at usual time and will doze thereafter ☐ Will wake up at usual time but will fall asleep again ☐ Will not wake up until later than usual

14. One night you have to remain awake between 4:00 AM and 6:00 AM in order to carry out a night watch. You have no commitments the next day. Which ONE of the following alternatives will suit you best?	☐ Would NOT go to bed until watch was over ☐ Would take a nap before and sleep after ☐ Would take a sleep before and nap after ☐ Would take all sleep before watch
15. You have to do two hours of hard physical work. You are entirely free to plan your day and considering only your own "feeling best" rhythm which ONE of the following times would you choose?	☐ 8.00 - 10.00 AM ☐ 11.00 AM – 1.00 PM ☐ 3.00 PM – 5.00 PM ☐ 7.00 – 9.00 PM
16. You have decided to engage in hard physical exercise. A friend suggests that you do this for one hour twice a week and the best time for him is between 10.00-11.00 **PM**. Bearing in mind nothing else but your own "feeling best" rhythm how well do you think you would perform?	☐ Would be on good form ☐ Would be on reasonable form ☐ Would find it difficult ☐ Would find it very difficult

17. Suppose that you can choose your own work hours. Assume that you worked a FIVE hour day (including breaks) and that your job was interesting and paid by results. Which FIVE consecutive hours would you select?

12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12

Midnight Noon Midnight

18. At what time of the day do you think that you reach your "feeling best" peak?

12	1	2	3	4	5	6	7	8	9	10	11	12	1	2	3	4	5	6	7	8	9	10	11	12

Midnight Noon Midnight

19. One hears about "morning" and "evening" types of people. Which ONE of these types do you consider yourself to be?	☐ Definitely a "morning" type ☐ Rather more a "morning" type than an "evening" type ☐ Rather more an "evening" than a "morning" type ☐ Definitely an "evening" type

Appendix K. SUDOKU.

Sudoku is a popular logic-based number placement puzzle. The 9×9 board has 9 rows, 9 columns and 9 sections of 3×3 cells. The objective is to fill the board so that each row, each column and each section contains the digits from 1 to 9.

Easy

			4	6	5		8	
		9					1	5
	5		9	1				3
3	8			7	4		5	
9	1		6		8		3	
5		6	2					
						8		
7				8	9		2	6
		3						4

Easy

			4	3	1		5	
		8					6	1
	1		8	6				9
9	5			2	4		1	
8	6		3		5		9	
1		3	7					
						5		
2				5	8		7	3
		9						4

Medium

			2	4				
6				3	9		8	2
			6			7		
8	7					5		6
		1		2		9		
3		2					4	1
		5			2			
2	8		5	9				3
				6	3			

Medium

					6			
	2	8	9			6		7
6	3	9	5	7	8			
5			7	6		1		
7			1	5	4		6	
2		1	8		3	4		
		7	6				3	
		2			9			
				8				

1. Have you ever completed a Sudoku before? Yes / No
2. Please indicate on the scale below how difficult it was to solve the Sudoku:

1	2	3	4	5	6	7

Not at all reasonably highly difficult

Appendix L. Testimony coding instructions.

1) Person details

- Girl with backpack
  - Black/dark (1-P) jacket (1-P)
  - Blue/green (1-P) jeans (1-P)
  - Black/dark (1-P) trainers (1-P)
  - Red (1-P) backpack (1-P)
  - Orange (1-O) shopping-bag (Sainsbury's) (1-O) in right (1-P) hand (1-P)
  - Medium long, (1-P) dark (brunette to black) (1-P), wavy/curly hair (1-P), pinned up hair (1-P)
  - Face not visible
  - Glasses (1-P)
  - Caucasian (1-P)
  - 20-30 years (1-P)
  - 5'5"-5'7" = 1.65-1.70m (1-P)
  - Slim/medium build (1-P)

- Man with both hands in his pockets
  - White/grey (1-P) t-shirt (1-P) with dark stripes (1-P), dark neckline and cuffs (1-P)
  - Dark (1-P) trousers (1-P)
  - Brown (1-P) sandals (1-P) with white (1-P) socks (1-P) OR brown (1-P) and white (1-P) trainers (1-P)
  - Medium long (1-P) , dark blonde/brunette (1-P), wavy/curly hair (1-P), hair is pulled back (1-P)
  - Unshaved (1-P)
  - Caucasian (1-P)
  - 25-35 years (1-P)
  - 5'11"-6'1" = 1.80-1.85m (1-P)
  - Medium build (1-P)
  - Has both hands in his pockets (1-P)

- Man with newspaper
  - Black/dark (1-P) t-shirt (1-P)
  - Blue (1-P) jeans (1-P) , colour slightly faded out (1-P)
  - Brown/grey/beige (1-P) shoes (1-P)
  - Short (1-P) Blonde to dark blonde (1-P), wavy, tousled hair (1-P), receding hairline (1-P)
  - Sideburns, full beard (1-P)
  - Caucasian (1-P)
  - 45-55 years (1-P)
  - 5'11"-6'1" = 1.80-1.85m (1-P)
  - Medium build (1-P)
  - Holds newspaper (1-O) in both hands (1-P)

- Man eating
    - Red (1-P) t-shirt (1-P) with a small white (1-P) logo (1-P) on the right (1-P)
    - Black/dark (1-P) trousers (1-P)
    - Black/dark (1-P) shoes (1-P)
    - Short (1-P) blonde (1-P) hair (1-P)
    - Wears black (1-O) bag (1-O) with a grey (1-O) flap (1-O) on his left (1-P) shoulder (1-P), diagonal across body (1-P)
    - Caucasian (1-P)
    - 20-25 years old (1-P)
    - 6'1"-6'3" = 1.85-1.90m (1-P)
    - Slim to medium build (1-P)
    - Sandwich (1-O) in both hands (1-P)

- Victim 1
    - Dark blue (1-P) jeans (1-P)
    - Black/dark (1-P) jumper/jacket (1-P) , dark (1-P) t-shirt (1-P) with a grey/silver (1-P) logo (1-P) worn underneath (1-P)
    - Dark (1-P) shoes (1-P)
    - Wears black (1-O) bag (1-O) on her left (1-P) shoulder (1-P) and holds the shoulder strap (1-O) (1-P) with her left (1-P) hand (1-P)
    - Bag is open (1-O)
    - Dark (1-P) sunglasses (1-P) worn in hair (1-P)
    - Shoulder long (1-P), blonde (1-P), wavy hair (1-P) , worn loose/not pulled back (1-P)
    - Caucasian (1-P)
    - 5'5"-5'7" = 1.65m-1.70m (1-P)
    - Medium build (1-P)
    - 45-50 years (1-P)
    - Newspaper (1-O) and bottle (1-O) in bag

- Perpetrator
    - Red(1-P) hooded (1-P) sweater (1-P) with buttons (1-P) in the front (1-P), a (grey inside) (1-P) and white details (neckline, ribbons, logo) (1-P)
    - Dark green (1-P) t-shirt (1-P) with a yellow (1-P) lettering/print(1-P) seen underneath (1-P)
    - Dark blue (1-P) jeans (1-P)
    - Black (1-P) and white (1-P) trainers (like Converse Chucks) (1-P)
    - Short (1-P) dark blonde (1-P) hair (1-P)
    - Unshaved (1-P)
    - narrow nose (1-P) and lips (1-P)
    - sharp facial features (1-P)
    - Caucasian (1-P)
    - 25-30 years (1-P)
    - Medium build (1-P)
    - 5'11"-6'1" = 1.80-1.85m (1-P)

- Girl running
    - Black (1-P) strapless (1-P) , knee-length(1-P) tube dress (1-P)
    - Black (1-P) , flat (1-P) sandals (1-P)
    - Long (1-P) dyed (1-P) dark red (1-P) hair(1-P), hair worn loose/not pulled back (1-P)
    - Caucasian (1-P)
    - 5'3"-5'5" = 1.60-1.65 (1-P)
    - 20-25 years (1-P)
    - Slim build (1-P)
    - Piece of paper (1-O) in left (1-P) hand (1-P)

- Girl leaving the shop
    - White (1-P) tank-top (1-P)
    - Golden, bronze (1-P) shiny (1-P) leggings (1-P)
    - Light coloured (1-P) flat (1-P) sandals (1-P)
    - Brown (1-O) handbag (1-O) worn on right (1-P) shoulder (1-P)
    - Black (1-P) hair (1-P), pulled up (1-P)
    - Darker skin (1-P)
    - 5'3"-5'5" = 1.60-1.65m (1-P)
    - 20-25 years (1-P)
    - Slim build (1-P)
    - Holding wallet/purse (1-O) in hands (1-P)

- Man with umbrella
    - Black/dark (1-P) coat (1-P)
    - Black/dark/blue (1-P) jeans (1-P)
    - Back (1-P) trainers (1-P) with white sole (1-P)
    - Short (1-P) dark (1-P) hair (1-P), greying (1-P) and bolding (1-P)
    - Carries Black (1-O) umbrella (1-O) in right (1-P) hand (1-P)
    - Caucasian (1-P)
    - 60-65 years (1-P)
    - Medium build (1-P)
    - 5'9"-5'11" = 1.75-1.80m (1-P)

- Victim 2
    - Grey (1-P) tank top (1-P)
    - Black (1-P) , knee-length (1-P) skirt (1-P)
    - Small (1-O) black (1-O) handbag (1-O) on left(1-P) shoulder (1-P)
    - Long (1-P) dark (1-P) hair (1-P), pulled back (1-P)
    - Caucasian (1-P)
    - Slim to medium build (1-P)
    - Pale coloured (1-P) shoes (1-P) with wedge heels (1-P)
    - 5'3"-5'5" = 1.60-1.65m (1-P)
    - 25-30 years (1-P)

- Man helping Victim 2
    - White (1-P) shirt (1-P) with dark (1-P) stripes (1-P), worn under a dark (1-P) leather (1-P) jacket (1-P)
    - Brown (1-P) trousers (1-P)
    - Short (1-P) dark (1-P) blonde (1-P) hair (1-P)
    - Brown (1-P) shoes (1-P) with red (1-P) shoelace (1-P) and pale coloured soles (1-P)
    - Corpulent (1-P)
    - Caucasian (1-P)
    - 35-40 years (1-P)
    - 5'9"-5'11" = 1.75-1.80m (1-P)
    - Mobile phone (1-O)

- Family
    - Mother
        - Black (1-P) t-shirt (1-P) with a white (1-P) heart-shaped (1-P) print (1-P)
        - Black (1-P) trousers (1-P)
        - Black (1-P) sandals (1-P)
        - Shoulder long (1-P) dark (1-P) hair (1-P)
        - Silver (1-P) watch (1-P) on left (1-P) wrist (1-P)
        - Silver (1-P) bangle (1-P) on right (1-P) wrist (1-P)
        - Blue (1-O) plastic (1-O) bag (1-O) with a red (1-O) logo (1-O) on left (1-P) hand (1-P)
        - 30-35 years (1-P)
        - 5'3"-5'5" = 1.60-1.65m (1-P)
        - Big build (1-P)

    - Father
        - White/grey (1-P) t-shirt (1-P) with a dark grey (1-P) print (1-P)
        - Grey (1-P) shorts (1-P) with camouflage pattern (1-P)
        - Blue (1-O) plastic (1-O) bag (1-O) with a red (1-O) logo (1-O) in each hand (1-P)
        - Very short hair or bold (1-P)
        - Black (1-P) and white (1-P) trainers (1-P)
        - 5'5"-5'7" = 1.65-1.70m (1-P)

    - Daughter
        - Jeans (1-P) jacket (1-P)
        - Pink/ (1-P) white (1-P) t-shirt (1-P) underneath
        - Pink (1-P) capri (1-P) pants (1-P)
        - 5 years (1-P)
        - 3'7" = 1.10m (1-P)
        - Long (1-P) dark (1-P) hair (1-P)

- Woman with white shopping bag
    - (Light) blue (1-P) t-shirt (1-P)
    - Black (1-P) sandals (1-P) with a white (1-P) detail (1-P)
    - Black/dark (1-P) skirt (1-P)
    - Bag (1-O) on left (1-P) shoulder (1-P)
    - White (1-O) plastic (1-O) shopping bag (1-O) in left (1-P) hand (1-P)
    - Silver (1-P) watch (1-P) on left (1-P) wrist (1-P)
    - 5'3"-5'5" = 1.60-1.65m (1-P)
    - 40-45 years (1-P)
    - Tanned (1-P)
    - Caucasian (1-P)
    - Medium build (1-P)
    - Black (1-P), curly (1-P) shoulder long (1-P) hair (1-P)

- Girl with mother
    - Black (1-P) tank-top (1-P) with a pink (1-P) and with flower prints (1-P)
    - Black (1-P) leggings (1-P)
    - Black (1-P) flats (1-P)
    - Long (1-P), dark (1-P), curly (1-P) hair (1-P) with blonde (1-P) hair-ends (1-P)
    - Caucasian (1-P)
    - 15-20 years (1-P)
    - Medium to corpulent build (1-P)
    - 5'5"-5'7" = 1.65-1.70m (1-P)
    - Phone (1-O) in right (1-P) hand (1-P)
    - Pale coloured (1-O) bag (1-O) on left (1-P) shoulder (1-P)
    - Plaster (1-P) on right (1-P) hand (1-P)

2) Action details

- Girl with Backpack
    - Comes (1-A) on screen from the right corner (1-S)
    - walks across (1-A) the market square (1-S)
    - passes (1-A) Savers (1-S) slowly and looks briefly (1-A) to the shop window (1-S)
    - carries (1-A) a shopping bag (1-O)
    - leaves (1-A) screen on the left (1-S)

- Man with both hands in his pockets
    - Comes (1-A) on screen from the right (1-S)
    - Heads (1-A) to Savers (1-S)
    - Stops (1-A) in front of the right shop window (1-S)
    - Looks (1-A) at the goods displayed (1-O)in the shop window (1-S)
    - Turns around (1-A) to the shop's front door (1-S)
    - Enters the shop (1-S) while grabbing (1-A) something (1-O) from the bargain bin (1-S) ((?) in the entrance (1-S)

- Man with newspaper
  - Leaves (1-A) the shop (1-S)
  - Stops (1-A) at the bargain bin (?)(1-S)
  - Takes out (1-A) a newspaper/advertising supplement (1-O)
  - Walks across (1-A) market square (1-S) while unfolding (1-A) the newspaper (1-O), looking(1-A) at front page and folding it (1-A)
  - Leaves (1-A) screen on the right (1-S)

- Man eating
  - Only seen in the background (1-S)
  - Comes (1-A) on screen from the left (1-S)
  - Walks across (1-A) the market square (1-S) while eating (1-A)
  - Turns head (1-A) sideways (1-S) and back to Savers (1-S) several times
  - Leaves (1-A) screen on the right (1-S)

- Victim 1
  - Comes (1-A) on screen from the right (1-S)
  - Walks straight (1-A) to Savers' right shop window (1-S) and stops (1-A) in front of it (1-S)
  - Bends forward (1-A) to have a closer look (1-A) at the goods displayed (1-O) in the shop window (1-S)
  - Thrusts (1-A) her handbag aside (1-O)
  - Unbends (1-A)
  - Does not realize she is the victim of an attempted theft (1-A)
  - Turns away (1-A) from the shop window (1-S)
  - Leaves (1-A) screen on the left (1-S)

- Perpetrator
  - Comes (1-A) on screen from the left/around the corner (1-S)
  - Looks around (1-A)
  - Walks towards (1-A) Victim 1 (1-S)
  - Stops (1-A) behind the Victim (1-S)
  - Recognizes that her handbag is open (1-A)
  - Looks inside the Victim's handbag (1-A)
  - Lifts (1-A) an object (1-O) in order to "examine" the content of her handbag (1-A)
  - Puts (1-A) the object (1-O) back in place when Victim 1 unbends (1-A) and leaves her (1-A)
  - Walks straight towards (1-A) the camera/the viewer (1-S) and passes (1-A) it on its right (1-S)

- Girl running
  - Leaves (1-A) the shop opposite to the camera (1-S)
  - First walks (1-A) then runs (1-A) towards Savers (1-S) (diagonal across the market square)
  - Enters (1-A) the shop (1-S)

- Girl leaving the shop
    - Leaves (1-A) Savers (1-S)
    - Walks across (1-A) the market square (1-S)
    - Leaves (1-A) screen on the right (1-S)

- Man with umbrella
    - Comes (1-A) on screen from behind the camera (1-S)
    - Walks past (1-A) Savers (1-S)
    - Takes a few looks (1-A) at the shop windows (1-S)
    - Then turns (1-A) to the right and walks across (1-A) the market square (1-S)
    - Leaves (1-A) screen on the right (1-S)

- Victim 2
    - Comes (1-A) on screen from the right (1-S)
    - Walks straight to (1-A) Savers' right shop window (1-S) and stops (1-A) in front of it (1-S)
    - Has a closer look (1-A) at the goods displayed (1-O) in the shop window (1-S)

- Perpetrator
    - Re-enters (1-A) the screen (comes on screen from behind the camera) (1-S)
    - Walks (1-A) towards Victim 2 (1-S)
    - Grabs/steals (1-A) her handbag (1-O)
    - Victim 2 realizes what is going on (1-A), turns around (1-A) and tries to keep it (1-A)
    - Perpetrator and Victim 2 jostle (1-A) for the handbag (1-O)
    - Perpetrator pushes (1-A) Victim2 away (1-S)
    - Victim 2 screams (1-A)
    - Perpetrator runs away (1-A) with handbag (1-O)
    - Victim 2 seems to be shocked according to her facial expression and gestures (hands to mouth) (1-A)
    - Perpetrator leaves screen (1-A) on the left (1-S)

- Man helping Victim 2
    - Witnesses the theft (1-A)
    - Comes (1-A) on screen from the right (1-S)
    - Rushes (1-A) to Victim 2 (1-S)
    - Looks after (1-A) the Perpetrator (1-S)
    - Stops (1-A) at the corner (1-S)
    - Turns(1-A) to Victim 2 (1-S)
    - Puts (1-A) out his phone (1-O)
    - Talks to Victim 2 (1-A)
    - Makes a phone call (1-A)

- Family
    - Family (1-P)
    - 1-P per mentioned family member (1-P)
        - *family with mom, dad and a little child → 3-P*
        - *a family (1-P) (…). The mother (1-P)*
    - Come (1-A) on screen from the left (background) (1-S)
    - Look (1-A) at Victim 2 (1-S) and the man trying to help her (1-A)
    - Woman stops (1-A) and turns (1-A) to the left (1-S)
        - Screaming (1-A) something ("they're filming") (1-A)
    - They have not witnessed the theft (1-A) but they see that something is going on (1-A)
    - They do not offer their help (1-A)
    - All leave (1-A) the screen on the right (1-S)

- Mother and daughter
    - Leave (1-A) the shop (1-S)
    - Look briefly (1-A) to Victim 2 (1-S) and the man trying to help her (1-S)
    - Leave (1-A) screen passing (1-A) the camera (on the left) (1-S)

3) Object details

- Different Shopping bags C
    - White (1-O) plastic (1-O)
    - Blue (1-O) plastic (1-O) with red (1-O) logo (1-O)
    - Orange (1-O) plastic (1-O)
- Goods displayed (1-O)
- Newspaper/ advertising supplement (1-O)
- Mobile phone (1-O)
- Sandwich (1-O)
- Bottles (1-O)
- Piece of paper (1-O)
- Wallet (1-O)
- Black (1-O) umbrella (1-O)

4) Setting details

- Market Square (1-S) in a rather small village (1-S)
    - Paved with red bricks (1-S)
    - Surrounded by four buildings(1-S)
    - Two streets leading to the square can be seen (1-S)
        - Junction (1-S)
        - One street runs to the left behind Savers (1-S)
    - Monument on the left side of the square (1-S)

- Lanterns (1-S) and flowers (1-S)
- A few/some people (1-P) strolling up and down/walking around (1-A)
- A few/some pigeons (1-S) walking around (1-A)
- Other buildings opposite to the camera (1-S)
    - Red brick house (1-S)
    - Sandstone building (1-S)
    - Each other building is mentioned (1-S)

- Building left to the market square
    - Savers (1-S)
        - Supermarket/general store/newsagent (1-S) located on ground floor (1-S)
        - Entrance is located in the middle between (1-S) two shop windows (1-S)
        - Several goods (1-O) are displayed in the shop windows (1-S)
            - Bottles (1-O) displayed in the left shop window (1-S)
        - Notice/advertisement (1-S)displayed in the shop windows: (1-S) “Low Prices” (1-S), written on yellow paper (1-S)
        - Entrance door is kept open (1-S) by a white bargain bin (?)(1-S)
        - Savers’ logo (1-S): blue “Savers” lettering, except for the letter “v” which is pink; white background (1-S)
    - “Abbeygate” (1-S) sign (1-S) above Savers (1-S)
        - White background (1-S)
        - Blue flower print above the lettering (1-S)

- Camera angle
    - Camera angle constant (1-S)
    - Zoom mode (in) when Victim 1 shows up (1-S)
    - Slow motion mode (1-S) and zoom mode (in) (1-S) when perpetrator is walking towards the camera
    - Zoom mode (out) (1-S) when perpetrator left the screen for the first time and slow motion mode is finished (1-S)
        - Camera concentrated on perpetrator (1-S)

- Weather
    - Blue sky (1-S), cloudy/overcast (1-S), no rain (1-S)
    - Bright day (1-S)

- Other details
    - No conversations (1-S)
    - Not much sound except Victim 2 screams (1-S)

Appendix M. Informed Consent Form.

UNIVERSITY OF CAMBRIDGE
INSTITUTE OF CRIMINOLOGY
Sidgwick Avenue, Cambridge, CB3 9DA

You are invited to participate in a study on eyewitness identification performance, conducted by Katrin Pfeil at the Institute of Criminology, University of Cambridge.

This study consists of two sessions, which take place one week apart from each other. In the first session you will watch a short film about a staged crime and be asked to fill in several questionnaires, e.g. on demographic information. In the second session you will then be asked to give a free recall, answer questions about the film and to identify the perpetrator you saw in the film from a photo line-up. You will also be asked to fill in questionnaires, e.g. about your preferred time of day.

Any findings from this study will be reported in group form only so that it will not be possible to trace anybody's individual answers. Findings from this study will be used for the student's thesis and for academic publication.

Participation in this study should take about 60 minutes for each of the two sessions. **Your participation in this study is completely voluntary**. You are free to withdraw from the session at any time without having to give a reason.

If you have any questions, feel free to ask them now. If you have any questions after you have left today's session, please feel free to contact **Katrin Pfeil,** at **07552 064264**, the address on the letterhead, or using email: **witnessresearch@gmail.com**. The study has been approved by the Institute of Criminology's ethic committee, and it is conducted under the supervision of Dr. Katrin Mueller-Johnson. If you have any concerns about this study you can contact her by email: kum20@cam.ac.uk or by phone: 01223 767 184. You will be given a copy of this consent form for your records.

I acknowledge that I have been informed of, and understand, the nature of the study and I freely consent to participate.

Name (print): Date:

Signature:

Appendix N. Means and standard deviations across all conditions and variables for the interview free recall part.

Table N1. Means and standard deviations at T2 in the interview free recall analysed by recall condition and age-group.

	SAI[a]		wFR[a]		no[a]		Older adults[b]		Young adults[b]	
Variable	M	SD	M	SD	M	SD	M	SD	M	SD
Totals										
Total details	104.83	34.23	95.23	30.83	81.67	34.54	103.93	31.48	83.89	34.40
Accuracy rate	0.93	0.04	0.92	0.05	0.90	0.06	0.89	0.06	0.94	0.04
Total person details	47.60	19.01	42.00	14.75	35.27	16.54	45.31	16.53	37.94	17.75
Total action details	27.63	7.79	25.73	8.25	22.81	9.86	27.78	8.16	23.00	8.91
Total object details	7.42	3.43	7.42	4.57	6.98	3.21	8.10	3.32	6.44	4.02
Total setting details	22.19	9.13	20.08	8.80	16.60	9.65	22.75	9.37	16.50	8.44
Total correct	97.19	31.45	87.38	27.82	73.25	31.72	93.28	29.80	78.60	32.12
Total incorrect	6.94	4.35	7.19	5.35	7.44	6.59	9.78	5.68	4.60	3.81
Total confabulated	0.71	1.30	0.67	1.04	0.98	1.51	0.88	1.50	0.69	1.06
Subtotals										
Person correct	44.42	17.23	39.04	13.47	32.44	15.14	41.64	15.22	35.63	16.34
Person incorrect	2.85	2.58	2.60	2.50	2.40	2.88	3.29	2.92	1.94	2.16
Person confabulated	0.33	0.81	0.35	0.56	0.44	0.80	0.38	0.74	0.38	0.72
Action correct	25.98	7.53	24.06	7.86	20.46	9.12	25.13	7.96	21.88	8.69
Action incorrect	1.52	1.27	1.54	1.57	2.21	2.54	2.51	2.21	1.00	1.07
Action confabulated	0.13	0.39	0.13	0.49	0.15	0.41	0.14	0.45	0.13	0.41
Object correct	5.73	3.17	5.63	4.33	5.04	2.67	5.49	2.86	5.44	3.97
Object incorrect	1.46	1.24	1.63	1.39	1.58	1.70	2.28	1.51	0.83	0.93
Object confabulated	0.23	0.63	0.17	0.43	0.35	0.67	0.33	0.67	0.17	0.47
Setting correct	21.06	8.49	18.65	8.40	15.31	9.02	21.03	8.80	15.65	8.22
Setting incorrect	1.10	1.36	1.42	1.90	1.25	1.82	1.69	1.98	0.82	1.24
Setting confabulated	0.02	0.14	0.02	0.14	0.04	0.20	0.03	0.17	0.03	0.17

[a]n = 48. [b]n = 72.

Table N2. Means and standard deviations at T2 in the interview free recall analysed by recall condition × age-group (continued).

	SAI OA		SAI YA		wFR OA		wFR YA		no OA		no YA	
Variable	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Totals												
Total details	105.63	30.59	104.04	38.17	108.17	25.14	82.29	31.01	98.00	37.85	65.33	21.18
Accuracy rate	0.92	0.03	0.94	0.03	0.89	0.05	0.95	0.03	0.87	0.07	0.92	0.05
Total person details	47.00	16.33	48.21	21.70	46.92	12.77	37.08	15.21	42.00	19.92	28.54	8.24
Total action details	27.38	7.32	27.88	8.39	28.38	6.37	23.08	9.16	27.58	10.54	18.04	6.33
Total object details	7.67	3.36	7.17	3.56	8.13	3.43	6.71	5.47	8.50	3.27	5.46	2.36
Total setting details	23.58	10.11	20.79	8.00	24.75	8.29	15.42	6.65	19.92	9.32	13.29	8.97
Total correct	97.17	28.66	97.21	34.64	96.71	23.32	78.04	29.26	85.96	35.94	60.54	20.66
Total incorrect	7.88	3.54	6.00	4.93	10.46	5.12	3.92	3.15	11.00	7.41	3.88	2.74
Total confabulated	0.58	1.44	0.83	1.17	1.00	1.29	0.33	0.56	1.04	1.76	0.92	1.25
Subtotals												
Person correct	43.71	15.03	45.13	19.49	42.96	11.67	35.13	14.23	38.25	18.31	26.63	7.93
Person incorrect	3.08	2.48	2.63	2.70	3.50	2.55	1.71	2.14	3.29	3.68	1.50	1.32
Person confabulated	0.21	0.66	0.46	0.93	0.46	0.66	0.25	0.44	0.46	0.88	0.42	0.72
Action correct	25.46	7.24	26.50	7.93	25.83	6.23	22.29	9.00	24.08	10.12	16.83	6.34
Action incorrect	1.83	1.27	1.21	1.22	2.33	1.71	0.75	0.90	3.38	3.03	1.04	1.08
Action confabulated	0.08	0.28	0.17	0.48	0.21	0.66	0.04	0.20	0.13	0.34	0.17	0.48
Object correct	5.50	2.86	5.96	3.51	5.42	2.92	5.83	5.45	5.54	2.93	4.54	2.34
Object incorrect	1.88	1.26	1.04	1.08	2.38	1.38	0.88	0.95	2.58	1.82	0.58	0.72
Object confabulated	0.29	0.75	0.17	0.48	0.33	0.56	0.00	0.00	0.38	0.71	0.33	0.64
Setting correct	22.50	9.47	19.63	7.29	22.50	8.28	14.79	6.69	18.08	8.18	12.54	9.12
Setting incorrect	1.08	1.25	1.13	1.48	2.25	2.31	0.58	0.78	1.75	2.11	0.75	1.33
Setting confabulated	0.00	0.00	0.04	0.20	0.00	0.00	0.04	0.20	0.08	0.28	0.00	0.00

Note. All groups n = 24.

Appendix O. Means and standard deviations across all conditions and variables for the interview question part.

Table O1. Means and standard deviations at T2 in the interview question part analysed by recall condition and age-group.

	SAI[a]		wFR[a]		no[a]		Older adults[b]		Young adults[b]	
Variable	M	SD	M	SD	M	SD	M	SD	M	SD
Totals										
Total details	107.25	25.30	104.92	19.83	97.19	24.70	104.00	21.55	102.24	25.68
Accuracy rate	0.88	0.03	0.87	0.05	0.85	0.06	0.85	0.05	0.88	0.05
Total person details	55.08	16.38	53.35	12.09	47.71	13.68	51.86	12.45	52.24	16.22
Total action details	23.06	5.41	22.50	5.47	21.63	7.20	22.68	5.77	22.11	6.38
Total object details	10.46	3.10	10.38	3.64	10.52	4.54	10.43	4.02	10.47	3.55
Total setting details	18.65	4.37	18.69	4.99	17.33	4.97	19.03	5.17	17.42	4.28
Total correct	94.56	23.21	91.23	17.81	82.69	23.37	88.97	19.98	90.01	24.07
Total incorrect	9.94	4.46	10.13	4.22	11.31	6.30	11.93	5.39	8.99	4.31
Total confabulated	2.75	1.91	3.56	3.21	3.19	3.36	3.10	3.12	3.24	2.68
Subtotals										
Person correct	46.73	14.58	44.60	10.34	39.40	11.76	42.46	10.50	44.69	14.47
Person incorrect	7.79	3.63	7.71	3.84	7.42	4.29	8.56	3.90	6.72	3.71
Person confabulated	0.56	1.37	1.04	2.05	0.90	2.28	0.85	1.99	0.82	1.89
Action correct	21.25	5.27	20.71	5.40	19.04	7.42	20.28	5.92	20.39	6.39
Action incorrect	0.44	0.74	0.38	0.73	1.38	1.65	1.03	1.47	0.43	0.77
Action confabulated	1.38	1.00	1.42	1.47	1.21	1.49	1.38	1.43	1.29	1.24
Object correct	8.75	2.94	8.42	3.60	8.04	3.99	8.26	3.78	8.54	3.27
Object incorrect	1.19	1.12	1.46	1.22	1.81	1.58	1.71	1.33	1.26	1.32
Object confabulated	0.52	0.74	0.50	0.83	0.67	0.86	0.46	0.73	0.67	0.87
Setting correct	17.83	4.37	17.50	4.69	16.21	4.92	17.97	4.98	16.39	4.26
Setting incorrect	0.52	0.80	0.58	0.79	0.71	1.07	0.64	0.97	0.57	0.82
Setting confabulated	0.29	0.68	0.60	1.38	0.42	0.92	0.42	1.28	0.46	0.73

[a]n = 48. [b]n = 72.

Table O2. Means and standard deviations at T2 in the interview question part analysed by recall condition × age-group (continued).

	SAI OA		SAI YA		wFR OA		wFR YA		no OA		no YA	
Variable	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
Totals												
Total details	100.96	20.82	113.54	28.14	105.21	18.11	104.63	21.81	104.26	25.02	88.54	20.80
Accuracy rate	0.87	0.03	0.90	0.03	0.86	0.04	0.88	0.05	0.84	0.08	0.86	0.05
Total person details	51.33	13.07	58.83	18.67	51.79	9.86	54.92	14.01	51.57	14.13	42.96	11.18
Total action details	21.21	4.15	24.92	5.95	22.50	5.22	22.50	5.83	23.83	7.02	18.92	6.11
Total object details	9.83	3.09	11.08	3.05	10.50	3.46	10.25	3.87	10.91	5.38	10.08	3.75
Total setting details	18.58	4.87	18.71	3.92	20.42	5.72	16.96	3.47	17.96	4.86	16.58	5.14
Total correct	87.42	17.76	101.71	26.05	90.46	16.79	92.00	19.12	87.48	24.44	76.33	20.06
Total incorrect	10.88	4.48	9.00	4.31	11.13	3.04	9.13	5.00	13.65	7.54	8.83	3.69
Total confabulated	2.67	1.46	2.83	2.30	3.63	2.90	3.50	3.55	3.13	4.41	3.38	2.00
Subtotals												
Person correct	42.42	10.73	51.04	16.75	42.42	8.70	46.79	11.52	41.83	11.98	36.25	10.60
Person incorrect	8.29	3.62	7.29	3.65	8.71	3.14	6.71	4.26	8.43	4.86	6.17	3.23
Person confabulated	0.63	1.06	0.50	1.64	0.67	0.96	1.42	2.72	1.30	3.21	0.54	0.66
Action correct	19.21	3.88	23.29	5.74	20.38	4.59	21.04	6.19	20.65	8.00	16.83	5.70
Action incorrect	0.63	0.71	0.25	0.74	0.21	0.51	0.54	0.88	2.30	1.89	0.50	0.66
Action confabulated	1.38	1.01	1.38	1.01	1.92	1.69	0.92	1.02	0.87	1.36	1.58	1.56
Object correct	7.96	2.69	9.54	3.01	8.42	3.83	8.42	3.44	8.30	4.76	7.67	3.21
Object incorrect	1.33	1.17	1.04	1.08	1.71	1.23	1.21	1.18	2.13	1.52	1.54	1.64
Object confabulated	0.54	0.78	0.50	0.72	0.38	0.71	0.63	0.92	0.48	0.73	0.88	0.95
Setting correct	17.83	4.78	17.83	4.03	19.25	5.22	15.75	3.35	16.70	4.88	15.58	5.04
Setting incorrect	0.63	0.92	0.42	0.65	0.50	0.72	0.67	0.87	0.78	1.24	0.63	0.92
Setting confabulated	0.13	0.34	0.46	0.88	0.67	1.83	0.54	0.72	0.48	1.20	0.38	0.58

Note. All groups n = 24.