Situational and personal determinants influencing eyewitness evidence.

Meta-analytical syntheses

Inaugural-Dissertation
zur Erlangung des Doktorgrades
der Philosophie des Fachbereichs 06 Psychologie
der Justus-Liebig-Universität Gießen

Vorgelegt von
Kerstin Kocab (geb. Wilhelm)

aus Marburg

2013

Dekan: Prof. Dr. Marco Ennemoser

1.Berichterstatter: Prof. Siegfried L. Sporer, Ph.D.

2.Berichterstatter: Prof. Dr. Joachim Stiensmeier-Pelster

Tag der Disputation: 18.12.2013

Dank

An dieser Stelle möchte ich mich bei einer ganzen Reihe von Personen bedanken, die mich bei der Erstellung dieser Arbeit auf ganz unterschiedliche Weise unterstützt haben.

Bedanken möchte ich mich bei meinem Betreuer Herrn Prof. Siegfried L. Sporer für seine hilfreichen fachlichen Rückmeldungen. Auch meinem Zweitgutachter Herrn Prof. Dr. Joachim Stiensmeier-Pelster gilt mein besonderer Dank.

Herrn Prof. Dr. Clemens Lorei danke ich für den gewinnbringenden fachlichen Austausch und interessante Anregungen hinsichtlich der vorliegenden Arbeit.

Bedanken möchte ich mich auch bei meinen (ehemaligen) Kolleginnen Maike Breuer, Perdita Müller, Kristina Kaminski und Valerie Hauch für die gute Zusammenarbeit und ihr allzeit offenes Ohr. Valerie Hauch danke ich außerdem für ihre hilfreichen Rückmeldungen und Anregungen im Hinblick auf diese Arbeit. Auch Natalie Martschuk möchte ich danken, für die gemeinsame Arbeit und Unterstützung bei der Kodierung einer der vorliegenden Metaanalysen.

Ein besonderer Dank gilt meinen Eltern Edith und Hans-Jürgen Wilhelm, welche mir erst das Studium der Psychologie ermöglicht und mich auch im Rahmen meiner Promotion in jeder nur erdenklichen Weise unterstützt haben.

Meinem Ehemann Jens Kocab möchte ich danken für die vielen unterstützenden und motivierenden Worte, sowie die interessanten Anregungen, die sicherlich zu einem besseren Gelingen der vorliegenden Arbeit beigetragen haben.

Auch für seine unermessliche Geduld im Hinblick auf schier unlösbare EDV-Probleme möchte ich ihm herzlich danken.

Vor allem für die emotionale Unterstützung möchte ich mich auch bei den folgenden Personen herzlich bedanken: Nina Wiche, Kiki Finger, Eva Lüllmann, Katharina Gaudlitz, Peter Heibach und meiner Schwester Yvonne Casella.

Für die hingebungsvolle und herzliche Betreuung unserer Tochter Leni während der arbeitsreichen Zeit möchte ich mich vor allem bei meinen Eltern und meinen Schwiegereltern Ute und Dieter Kocab bedanken.

Mein besonderer Dank gilt schließlich unserer Tochter Leni, die mir durch ihre kindliche Unbefangenheit immer wieder Optimismus und Fröhlichkeit vermittelt hat.

Gießen, im Mai 2013 *Kerstin Kocab*

Abstract

This dissertation describes two meta-analyses that are concerned with factors affecting eyewitness testimony. Meta-analysis 1 investigated the so-called *weapon focus effect* (WFE). This describes the phenomenon that a person witnessing a criminal event where the perpetrator carried a weapon is later worse able to describe or identify the target person compared to eyewitnesses not being confronted with a weapon. A total of 23 research articles met the inclusion criteria. A significant WFE could be calculated for description accuracy of the target person, g_u = 0.568, 95% CI [0.490, 0.647], k = 29, while results for identification performance failed to reach significance. A parallel analysis was conducted with studies that tested the notion that the unusualness of an object is responsible for this effect. Unusualness also affected person descriptions but not correct identifications. The effect on false identifications could not be investigated due to insufficient data within the studies included. With regard to extensive consequences an identification of an innocent person has, future research should necessarily focus on this.

In meta-analysis 2, the effect of *eyewitness' age* was examined, comparing elderly with young witnesses. Overall, 22 studies tested the hypothesis that older eyewitnesses are worse in identifying a target after observing a crime compared to younger adults. Significant effects in favor for the younger age groups were found for all dependent measures under investigation. Largest effects were found for foil identifications for both target-present (*OR* = 2.453 [1.794, 3.356]) and target-absent lineups (*OR* = 3.074 [2.310, 4.091]), indicating that older eyewitnesses were 2.5 to 3 times more likely to choose a wrong person from a lineup. Similar results were found for old-age faces. Cognitive impairments as well as a more liberal choosing behavior of older people are discussed when interpreting the results.

To conclude, the present meta-analyses demonstrated that both the presence of a weapon as well as the age of someone observing a crime affect eyewitness performance. However, it should be noted that experiments are far from real cases as they do not reflect the stress levels real crime victims (or bystanders) are likely to experience. Hence, even larger effects as they were found in the present syntheses are expected in real cases.

In both cases, factors under investigation were estimator variables and are therefore not modifiable by the police within relevant investigations. Nevertheless, policy makers and decision makers could be informed to arrive at better evidence-based decisions. Practical implications are also discussed.

In the case of WFE, a lineup consisting of different weapons could be helpful as analyses demonstrated eyewitnesses to focus on central details like the weapon carried by the perpetrator. Information about the weapon could be additional evidence in police investigations. One possibility to reduce or even prevent a WFE could be to train people (e.g., employees of a bank or gas station) not to be affected as much by the phenomenon. Some studies demonstrated a higher description accuracy when witnesses were warned beforehand (e.g., Pickel, French, & Betts, 2003).

For the age of witnesses, interventions should be investigated to make the choosing behavior of older persons more cautious as they are 2.3 times more likely to choose someone from a lineup in comparison to younger adults. Although, age differences in correct identifications are smaller than in foil identifications higher choosing rates increase both. Appropriate instructions as well as pre-identification procedures could therefore be helpful for the elderly to adopt a stricter decision criterion (Dysart & Lindsay, 2001; Wilcock & Bull, 2010).

Finally, methodological aspects of meta-analyses, which became very popular in the last decades, are critically discussed.

Table of Contents

Introduction

> "How do eleven years pass when you are locked up for a crime you didn't commit? I couldn't begin to imagine. For me, there were eleven years measured in birthdays, first days of school, Christmas mornings. Ronald Cotton and I were exactly the same age, and he had none of those things because I picked him" (Thompson-Cannino, Cotton & Torneo, 2009, p.237).

The case Ronald Cotton is surely one of the most famous cases of wrongful convictions in the US albeit it is one out of hundreds (www.innocenceproject.org). In November 1987, the innocent suspect Ronald Cotton was convicted for two rapes and two counts of burglary. Cotton was arrested for eleven years until DNA evidence proved his innocence. Improper forensic science (a flashlight that resembled that one used by the assailant and rubber from Cotton's tennis shoe that was consistent with rubber found at one of the crime scenes) and the identification by the victim Jennifer Thompson were the main causes for his arrest. The North Carolina Superior Court sentenced him to life plus fifty-four years. This happened although another victim did not choose Cotton from the lineup but a foil, and although a man in prison, Bobby Poole, told another inmate he had committed the crimes Cotton got arrested for.

By now it is widely known that eyewitness evidence is often erroneous. In fact, eyewitness misidentification is one of the greatest causes of wrongful convictions in the USA, playing a role in more than 75% of convictions overturned through DNA testing (see www.innocenceproject.org). Nevertheless, witness evidence is still crucial in many criminal investigations as in most cases DNA samples are not available. Eyewitness evidence varies from person description to construction of a composite or an identification of the suspect or his voice (e.g., Kempen & Tredoux, 2012; Sporer & Martschuk, in press; Wilcock, Bull, & Milne, 2008). All of these tasks deliver specific information, require different cognitive efforts and are dependent from each other in a specific way. In most crimes the victim or a witness is first asked by the police to give a description about what happened as well as the perpetrator. This is what Jennifer Thompson did just a few hours after the assault. On the basis of this description the police tries to locate the perpetrator frequently with the help of composites. When investigations led the police to a certain suspect, the police is conducting a lineup, again using the description as they select filler persons for the parade. The witness is than confronted with the lineup and has to decide whether or not the perpetrator is part of the lineup.

Consequently, person descriptions are usually the basis of police investigations although descriptions are often nondistinct and can frequently apply to many people (Meissner, Sporer, & Schooler, 2007). Usually, witnesses give few characteristics (M = 9.81; Sporer, 1992), which refer to more general information (height, age or ethnicity), clothing (which is not permanent) or characteristics of the face (hair or eye color which can also be changed with the help of contact lenses for example). More detailed information about inner characteristics of the face (e.g., chin, cheek, eye shape) are rather rare (Sporer, 1992). Consequently, on the basis of this superficial description the police might locate an innocent person. Even though, Jennifer Thompson tried to intensively study the face of her perpetrator as she for example allured him to illuminated parts of her apartment, her description given to the police might have applied to many black young men in the US. Why is it so difficult to describe a person seen within a crime? Besides factors during the crime impairing the memory for the perpetrator and event, the task itself is not so easy as often meant. When a person is confronted with an unknown face, he/she will encode facial information holistic rather than feature-based. Faces are not processed as sets of separate features but as interactive systems of features including interfeatural properties like distances between features, relative sizes, and other topographical information (Wells, 1993; Wells & Hryciw, 1984). In an identification task the witness is usually confronted with a lineup of six persons and the witness has to compare each person of the lineup with the picture in his/her head to decide whether one person fits the picture or not. Both phases, encoding and recognition of the face, are therefore characterized through holistic-based processes. In contrast, when witnesses are asked to describe the target person their task is to leach single characteristics from the person, which were originally encoded holistically (Meissner, Sporer, & Susa, 2008). Consequently, it is more difficult to describe different parts of the target's face than to recognize him/her due to different underlying processes. This is primarily critical because the description usually provides the basis for further police investigations. Consequently, there is a small but significant correlation between description and identification accuracy (r = 0.14, Meissner, Sporer, & Susa, 2008).

Nevertheless, it does not imply, that recognizing a target person within a lineup is an easy job (see above, 75% of wrong identifications within exonerated cases). There are multiple possibilities of making mistakes with more or less dramatic consequences (see Appendix A). When the lineup contains the perpetrator research calls it a target-present (TP) lineup. A correct identification therefor is the witness' choice of the suspect. If he/she chooses another person from the lineup it is called a foil identification. This choice is less dramatic as the police frequently select police officers as filler persons. No choice presents a false rejection. On the other hand, when the suspect located by the police is innocent research calls it a target-absent (TA) lineup. When the witness does not choose anybody he/she correctly rejects the lineup. A choice would therefore be a mistake in any case. But to pick a filler person would again be less dramatic than to choose the innocent suspect. Unfortunately, research often does not separate between the identification of an innocent suspect or a filler person in TA lineups subsuming it to false identifications. Due to the consequences the different choices have, those outcomes should be investigated separately in future research.

Nevertheless, in real life the police do not know if the suspect is guilty or not. But when an eyewitness points to the suspect and says confidently, “That’s him, I am sure”, there are hardly reasons not to believe the person who witnessed the crime. Jennifer Thompson was confronted with a photo lineup and later with a live lineup. In both cases she confidently identified Ronald Cotton. Even though, when Bobby Poole was momentary suspect of committing the crimes, Jennifer rejected a lineup where Poole was part of it. In the event, Jennifer made a false identification in the TA lineup and a false rejection within the TP lineup. Because the choice of an innocent person has dramatic consequences there are numerous methods investigated apparently reducing the rate of false identifications (e.g., sequential lineup testing, context reinstatement methods). On the other side there is considerable doubt if this happens to the expense of less correct identifications as witnesses use stricter decision criteria. Strategies should therefore focus on the *no-cost view*, delivering lower false identification rates with little or no reduction in hit rates (Clark, 2012). To investigate methods improving eyewitness evidence psychological research has intensively investigated variables affecting eyewitness evidence. Wells (1978) divided these factors into *estimator* and *system variables*.

Estimator variable research investigates variables that affect eyewitness accuracy but are not under control of the criminal justice system. Although these variables can be manipulated under experimental conditions, they cannot be controlled for in actual criminal cases (Wells, 1978). For this reason their impact can only be estimated. *Estimator variables* can be arranged into four wide categories: characteristics of the witness, characteristics of the event, characteristics of the testimony, and abilities of the testimony evaluators to distinguish between accurate and inaccurate witness testimony (Wells & Olson, 2003). One example for an estimator variable a lot attention was paid to is the ethnicity of the perpetrator relative to the ethnicity of the witness, known as the cross-race effect (Meissner & Brigham, 2001; Sporer, 2001a, 2001b) where an eyewitness is better in identifying suspects from the same ethnicity compared to other ethnic group members. As in the case of Ronald Cotton diverse ethnicities of perpetrator and victim might be one reason forwarding the misidentification. Further witness variables receiving considerable attention but ecologically difficult to assess, are stress and arousal experienced while watching or being involved in a crime (e.g., Deffenbacher, Bornstein, Penrod, & McGorty, 2004; Price, Lee & Read, 2009; Valentine & Mesout, 2009). Variables of the event are lightening conditions, exposure time of the perpetrator and distance to the perpetrator, for example.

System variables on the other hand are those that also influence eyewitness accuracy but over which the criminal system has (or can have) control. In eyewitness identification research this contains all variables being linked with the lineup task. These factors are also sorted into four categories: instructions, content, presentation method, and behavioral influence (Wells & Olson, 2003).

The distinction between those two types of variables is important because the first group should help the criminal system to increase the probability to discriminate between accurate and inaccurate witnesses whereas the second one helps to prevent inaccurate identifications from occurring in the first place (Wells & Olson, 2003).

This dissertation investigates two estimator variables. The first meta-analysis concerns the so-called WFE, which describes the phenomenon that a witness observing a crime where a weapon was involved is later less able to identify the target in an identification task. In 1992, Steblay published the first meta-analysis demonstrating a small effect (h = .13) for lineup identification accuracy and a moderate effect (h = .55) for feature accuracy in person description tasks. Steblay included 19 independent hypothesis tests from 1976 to 1991. Attentional factors as well as physical arousal were discussed as reasons for the effect. Around the turn of the millennium a new approach trying to explain the effect found its way in the literature, namely the influence of the unusualness of a weapon in a certain scenario. Pickel (1998, 1999) postulated that any object that stands out of a given scenario in some manner could produce a similar effect.

Eventually, it was one aim of the current meta-analysis to extend and update the 1992 published meta-analysis as there were new studies to include in the synthesis. Overall, it was expected to confirm the findings of Steblay's meta-analysis and to demonstrate the significant role of attention, stress and unusualness when explaining the existence of the WFE.

In the case of Ronald Cotton, it remains unclear how Jennifer was affected by methods of interrogation or lineup instruction but she was confronted with a weapon as Poole threatened her with a knife on her throat throughout the rape. Even though, she intensively tried to memorize the perpetrators face, as she for example allured him to illuminated parts of the apartment her encoding might be disturbed by the presence of the weapon.

Meta-analysis 2 investigated the *age of an* eyewitness. As to our knowledge, there is no meta-analysis investigating the identification performance of older eyewitnesses although one examining children as witnesses exists (Pozzulo & Lindsay, 1998) and one investigating face recognition across the lifespan (Rhodes & Anastasi, 2012). The influence of the *witness' age* was predominantly investigated within two different research paradigms, namely facial recognition and eyewitness paradigms. The difference between these two approaches is how recognition is tested, delivering two different types of effect measures. Participants in facial recognition studies have to encode a certain amount of faces, 12 for example, and are later asked to recognize these 12 faces out of a larger amount of faces, for example 24 faces, indicating at each face if it was 'old' or 'new'. On the contrary, in eyewitness studies participants normally are exposed with a staged crime scenario (via slides, video or live) and afterwards they are asked to identify the perpetrator in a lineup parade. Because the following work especially aimed to give practical implications on how to deal with factors influencing eyewitness evidence, only those eyewitness studies were included in the meta-analysis due to their higher ecological validity. It was expected that older eyewitnesses would be worse witnesses compared to their younger counterparts. As it is often observed that elders make more false identifications in both target present (TP) and target absent (TA) lineups (e.g., Memon & Bartlett, 2002; Wilcock, Bull, & Vrij, 2005, 2007) it was also important to investigate a potential *response bias*. Possibly, older eyewitnesses choose more often someone from a lineup either out of pressure to help (even though the pressure in laboratory studies should be smaller than in real life) or because they might not be aware of the option that the perpetrator could also *not* be in the lineup. Furthermore, the possibility of an *own-age bias* where witnesses are better with persons of their own age will be investigated.

Jennifer Thompson and Bobby Poole were almost the same age but they were of diverse ethnicities. Similar to the *own-age bias* the *own-race bias* describes the phenomenon that witnesses demonstrate better eyewitness performance when the perpetrator was of the same ethnicity. One reason seem to be that persons have more contact to people of their own ethnicity and are therefore experts as they are more familiar with specific characteristics (Meissner & Brigham, 2001; Sporer, 2001a, 2001b). Same processes might be responsible for a possible *own-age bias*.

Why meta-analyses?

As demonstrated above, there are numerous studies investigating the effect of different variables on eyewitness abilities. But practical conclusions often remain unclear as some studies did find an effect of a certain phenomenon whereas others did not. Why do researchers generally investigate the same phenomenon leading to a large amount of study findings? Reasons could be that researchers are skeptical of what others are doing, they are unaware of others' work, or want to replicate their findings. At some point it is important to combine all results to draw conclusions.

But in fact, no study resembles another in all details. According to this, techniques summarizing those findings without neglecting differences influencing the results are needed. Meta-analysis is a technique to summarize individual study findings in a systematical manner.

Strengths of meta-analyses are as follows (see Lipsey & Wilson, 2001). First, if done thoroughly, a good meta-analysis is conducted as a structured research technique requiring documentations of each step and therefore delivering an explicit and systematic process and conclusions. Second, meta-analysis represents key study findings more differentiated and sophisticated than conventional review techniques that rely on qualitative summaries. Third, meta-analysis is capable of analyzing the relationships between study characteristics and study findings.

Furthermore, meta-analysis produces synthesized effect estimates with considerably more statistical power than individual studies (Cohn & Becker, 2003; Hunter & Schmidt, 1990; Matt & Cook, 2009). Fifth, meta-analysis can handle information from a large number of studies with almost unlimited capabilities of coding and storing.

The purpose of the following meta-analyses was to summarize the findings of WFE and old age effects and to provide estimates of effect sizes. Besides theoretical explanations practical advice for the justice system can be derived on how to deal with eyewitnesses being affected by the one or the other variable under investigation. At last, the current work demonstrates the structured and sophisticated process of accomplishing meta-analyses.

References

Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods, 8*, 243-253. doi: 10.1037/1082-989X.8.3.243

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science, 7*, 238-259. doi: 10.1177/1745691612439584Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28*, 687-706. doi: 10.1007/s10979-004-0565-x

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

Kempen, K., & Tredoux, C.G. (2012). ‘Seeing is believing’: The effect of viewing and constructing a composite on identification performance. *South African Journal of Psychology, 42*, 434-444.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-Analysis*. Thousand Oaks, CA: SAGE Publications.

Matt, G. E., & Cook, T. D. (2009). Threats to the validity of research synthesis. In H. Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 503-520). New York: Russell Sage Foundation.

Meissner, C. A., & Brigham, J. C. (2001). Thirty years of investigating the own-race bias in memory for faces: A meta-analytical review. *Psychology, Public Policy, and Law, 7*, 3-35. doi: 10.1002/acp.728

Meissner, C.A., Sporer, S.L., & Schooler, J.W. (2007). Person descriptions as eyewitness evidence. In R. Lindsay, D. Ross, J. Read & M.Toglia (Eds.). Handbook of Eyewitness Psychology: Memory for People (pp.3-34). Mahwah, NJ: Lawrence Erlbaum & Associates.

Meissner, C.A., Sporer, S.L., & Susa, R.S. (2008). A theoretical review and meta-analysis of the description-identification relationship in memory for faces. *European Journal of Cognitive Psychology, 20*, 414-455.

Memon, A., & Bartlett, J. (2002). The effects of verbalization on face recognition in young and older adults. *Applied Cognitive Psychology, 16*, 635-650. doi: 10.1002/acp.820

Pickel, K. L. (1998). Unusualness and threat as possible causes of "Weapon Focus". *Memory, 6*, 277-295. doi: 10.1080/741942361

Pickel, K. L. (1999). The influence of context on the *Weapon Focus* effect. *Law and Human Behavior, 23*, 299-311. doi: 10.1023/A:1022356431375

Pozzulo, J. D., & Lindsay, R. C. L. (1998). Identification accuracy of children versus adults: A meta-analysis. *Law and Human Behavior, 22*, 549-570. doi: 10.1023/A:1025739514042

Price, H. L., Lee, Z., & Read, J. D. (2009). Memory for committing a crime: Effects of arousal, proximity, and gender. *American Journal of Psychology, 122*, 75-88.

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

Sporer, S.L. (1992). *Das Wiedererkennen von Gesichtern* [Recognizing faces]. Weinheim, Germany: Beltz/ Psychologie Verlags Union.

Sporer, S. L. (2001a). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law, 7*, 36-97. doi: 10.1037/1076-8971.7.1.36

Sporer, S. L. (2001b). The cross-race bias: Beyond recognition of faces in the laboratory. *Psychology, Public Policy, and Law, 7*, 170-200. doi: 10.1037/1076-8971.7.1.170

Sporer, S.L. & Martschuk, N. (in press). The reliability of eyewitness identification by the elderly: An evidence-based review. To appear in: M. Toglia, D. Ross, J. Pozzulo, & E. Pica (Eds.). The elderly witness in court. New York: Psychology Press.

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16*, 413-424. Retrieved from http://www.jstor.org/stable/1394272

Thompson-Cannino, J., Cotton, R., & Tonero, E. (2009). *Picking Cotton: Our memory of injustice and redemption*. New York: St. Martin's Press.

Valentine, T., & Mesout, J. (2009). Eyewitness identification under stress in the London Dungeon. *Applied Cognitive Psychology, 23*, 151-161. doi: 10.1002/acp.1463

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

Wells, G. L., & Olsen, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54*, 277-295. doi: 10.1146/annurev.psych.54.101601.145028

Wilcock, R.A., Bull, R., & Milne, R. (2008). Witness Identification in Criminal Cases: Psychology and Practice. New York: 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*, 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*, 305-316. doi: 10.1080/10683160600822212

Meta-analysis 1

The Weapon Focus Effect: A Meta-analysis

In most crimes DNA-rich biological traces are not available as evidence. Hence, eyewitness is often still crucial for investigating and prosecuting crimes (Wells & Olson, 2003). Especially in cases of murders, drive-by shootings, convenience store robberies, muggings, and other common crimes perpetrators almost never leave DNA trace evidence (Wells et al., 2000). Despite this important role, some legal scholars have argued that eyewitness misidentification is the most important cause of wrongful convictions in the US, playing a role in more than 75% of convictions overturned through DNA testing (Scheck, Neufeld, & Dwyer, 2001). Although the justice system relies heavily on eyewitness memories, many factors were investigated within the last decades demonstrating the error-proneness of person identifications. One of these factors presumably having a great impact on how people remember witnessed crimes is the presence of a weapon.

The so-called weapon focus effect (WFE) describes the phenomenon that eyewitnesses observing a crime where a perpetrator carries a weapon are less accurate in describing or identifying the suspect in a lineup compared to crimes with no weapons involved. Since the late 70s, numerous studies were conducted to investigate the phenomenon (e.g., Johnson & Scott, 1976; Cutler, Penrod, & Martens, 1987a, 1987b; Shaw & Skolnick, 1999; Hope & Wright, 2007)--repeatedly but not consistently--demonstrating the existence of the WFE. Hence, most eyewitness experts are convinced that the effect is a reliable phenomenon: 42 of 63 experts surveyed by Kassin, Tubb, Hosch, and Memon (2001) thought that this phenomenon is "generally reliable" or "very reliable", and 87% thought it to be reliable enough to testify in court (97% on a research basis). The present meta-analysis can shed light on whether this is justified. Note, however, that the willingness to testify about a topic should not be confused with the judgment that a phenomenon is generally true (Read & Desmarais, 2009), which can only be answered by the available evidence.

We synthesize the evidence that the presence of a weapon leads to a deterioration of eyewitness performance regarding both the identification as well as the description of a target. Furthermore, we test under which conditions the weapon focus effect is more or less likely to occur.

Theoretical Explanations

Three main explanatory approaches are discussed in the literature to explain the WFE. First, the stress and arousal approach assumes that the perceived threat induced by a weapon produces a general reduction in eyewitness performance (Deffenbacher, 1983; Deffenbacher, Bornstein, Penrod, & McGorty, 2004; Easterbrook, 1959). Second, and closely related to the first approach, the attentional focus approach more specifically addresses the presence of a weapon by distinguishing between central (the weapon) and peripheral details of an emotional event (Christianson & E. F. Loftus, 1991). Third, more recently researchers following the unusualness approach have argued that arousal is not the single cause, but that a weapon is unexpected in certain situations (Pickel, 1998, 2009). Even if we know that bank robberies might happen from time to time, few of us think about this possibility when entering a bank and hence do not expect the presence of a weapon. Hence, attention is focused on any unusual object present at the expense of processing other information.

Emotional Stress and Arousal

Emotional stress is a general explanation for the WFE, drawing on more than 100 years of research on the effects of stress and arousal on (memory) performance (for historical treatments, see Deffenbacher, 1983, 1991; Sporer, 1982, 2008). Whereas a negative emotional event is defined as something new, unexpected and potentially threatening, emotional stress describes a psychological experience with concurrent autonomic-hormonal changes as a consequence to a negative emotional event (Christianson, 1992). This can range from moderate to excessive levels of stress. In this context, emotional stress and arousal are often treated as the same, although the latter construct has a more physiological connotation (e.g., Neiss, 1988). Here, stress and arousal will be used synonymously following the majority of researchers on this topic.

Until the 1990s there was considerable controversy concerning whether increases in arousal facilitate or inhibit memory performance. According to Easterbrook's hypothesis (1959), there is a progressive restriction of the range of cues used or attended to as a function of an increase in emotional arousal, and consequently memory performance declines. The so-called range of cue utilization is defined as the "total number of environmental cues in any situation that an organism observes, maintains an orientation towards, responds to, or associates with a response" (Easterbrook, 1959, p. 183). This is different from the eye fixation approach discussed below in so far as Easterbrook assumed a general reduction of information processed in an emotionally arousing situation and did not distinguish central and peripheral information as in the attentional approach. Findings regarding eyewitness performance under stress were mixed, some showing better performance under emotional arousal, others demonstrating worse memories.

In a systematic review of the literature, Deffenbacher (1983) hypothesized that variations in stressor intensity affect performance level following an inverted-U function according to the Yerkes-Dodson law (Yerkes & Dodson, 1908). Supposedly, in studies showing a facilitation effect on eyewitness memory with increased arousal participants were operating on the ascending portion of the inverted-U curve, whereas in studies showing a reduction in memory with increased arousal participants were operating on the descending portion. A problem with this approach is that it is difficult to estimate ex post facto where a witness was located on this curve at the time of the crime (see also Deffenbacher, 1991).

Christianson (1992) criticized this simple unimodal explanation, arguing that high states of emotional arousal are always accompanied by a decrease in available processing capacity and less efficient information processing. Contrary to

Deffenbacher, Christianson noted that high emotions seem to be associated with relatively accurate memory for central details but relatively inaccurate memory for peripheral details. However, one of the problems with this explanation is that it is difficult to assign in advance, or in retrospect, which details are central and which are not, especially in field studies (Christianson, 1992; Heuer & Reisberg, 1990; Ibabe & Sporer, 2004). Studies distinguishing between central and peripheral details of information found that central aspects of an event are relatively well retained in memory, while memory is impaired for many other specific details, especially the peripheral ones (Christianson & Loftus, 1991; Loftus & Burns, 1982). These results are consistent with the *von Restorff effect* (1933; see below), due to a fixation on central details.

In summary, the arousal approach predicts that higher levels of arousal will produce poorer eyewitness memories compared to lower levels of arousal. This effect should be observable across different measures of eyewitness memory, that is, both person identifications and person descriptions. Unfortunately, few studies included in the current meta-analysis directly evaluated arousal levels, and most of them gathered self-reports after watching a video (Mitchell, Livorsky, & Mather, 1998; Kramer, Buckhout, & Eugenio, 1990; Pickel, Ross, & Truelove, 2006). Also, participants watching a video will rarely experience the (high) levels of arousal characteristic of victims or bystanders of actual crimes. Additionally, self-reported arousal levels are not objective, compared to physiological measures like heart rate or Galvanic skin response. Overall, reported levels of arousal were moderate rather than high. Also, some participants in the weapon condition of a staged event simulation stated feeling more agitated or angry (Maass & Köhnken, 1989) rather than feeling threatened or frightened.

Due to the lack of information in most studies regarding participants' experienced stress levels, we used variables likely to be correlated with arousal as moderator variables in our analyses. For instance, the most important moderator presumably being associated with different levels of stress could be the type of weapon used in a study. We expected that guns produce higher levels of arousal and consequently lower performance regarding identification and description of the target compared to other weapons like knives or meat cleavers. Higher risks of injury could be responsible for the effect.

Attentional Focus while Watching a Crime

One aspect common to both explanatory approaches is the fixation of the object, be it threatening or unusual. The *Von Restorff effect*, also called the *isolation effect*, predicts that an item that "stands out" is more likely to be remembered than other items (e.g., G. R. Loftus & Mackworth, 1978; von Restorff, 1933; Wallace, 1965). Performance enhancement of central information and impairment of peripheral information can occur whenever a target event or target object is *distinctive*. Early studies on memory for central and peripheral information examined this hypothesis by investigating eye movements and fixation times (E. F. Loftus, G. R. Loftus, & Messo, 1987). As participants focused longer and more often on a weapon compared to a neutral object, performance regarding the perpetrator was worse as participants were less able to describe (and identify [in Exp. 2]) the target person compared to participants in the control group.

As the duration of fixation on the critical item could be related to memory accuracy (G. R. Loftus, 1972), Christianson, E. F. Loftus, Hoffman, and G. R. Loftus (1991) held the fixation constant by allowing only a single eye fixation on the critical slide (by presenting it for 180 ms [Experiment 1] or 150 ms [Experiment 2]). Nevertheless, memory for central details was still better in the emotional compared to the neutral condition. The authors argued that enhanced memory for central details of arousing events does not occur solely because more attention is devoted to the emotional information.

Christianson and E. F. Loftus (1991) pointed out that arousing events as well as distinctive or unusual events often produce *von Restorff*-like data (e.g., Detterman, 1975; Tulving, 1969). But when comparing an emotionally arousing condition (a woman having an accident with a bicycle) with an unusual condition (the woman carrying a bicycle on her shoulder) both events produced basically the same poor performance with respect to memory for peripheral details (such as memory for an orange car in the background). On the other hand, the emotionally aroused group showed significantly better memory concerning central details (such as the blue coat of the woman) compared to the neutral group. These differences were not found in the unusual group (Christianson & E. F. Loftus, 1991).

Friedman (1979) argued that "since automatized recognition procedures should require few or no resources, while interactive procedures are resource expensive, it is likely that when interactive procedures are necessary for object identification, the result is a relatively more elaborate representation with respect to descriptive information" (p. 326). Hence, objects either threatening or unexpected will lead to interactive procedures and consequently to better memories regarding these objects. As visual information processing is presumed to take place during fixation (Latour, 1962; G. R. Loftus, 1972), and in line with the *von Restorff effect*, we expected participants in the current analysis to be more accurate in describing the weapon compared to neutral objects held by the target. These findings could help to understand attentional processes while observing a crime.

In general, we expect a deterioration in performance as a function of weapon presence for both person identification and description results although researchers have not addressed possible differences in outcomes as a function of dependent measures used in their theoretical arguments or predictions.

Unusualness of an Object or Weapon

A more recent explanatory approach assumes the unusualness of a weapon as the mechanism responsible for the phenomenon (Pickel, 1998, 1999, 2009). Kramer et al. (1990) explained unusualness with feelings of surprise, like not expecting a robbery in a store even though one knows that it might happen from time to time. Pickel (1998) postulated that "weapon" focus could be caused by any object other than a weapon, as long as it is unusual within a given context. She defined unusualness in terms of something being "unexpected or out of place given the business establishment setting" (Pickel, 1999, p. 301). The effects of contextual violations and semantic inconsistencies on object identification has been demonstrated in numerous studies (e.g., De Graef, Christiaens, & d'Ydewalle, 1990; Henderson, Week, & Hollingworth, 1999). In this regard, Friedman (1979) explained that objects which are *nonobligatory* to a frame are likely to be forgotten in terms of both descriptive as well as episodic information, while visual details of *unexpected* objects should be forgotten less frequently because they are "stuck on" the frame. They are what makes a situation "interesting" or different from previous experiences. In the case of a weapon it is unexpected in so far as it is not part of daily interactions in a bank or on the street.

Regarding operationalizations, some authors used unusual stimulus materials like a raw chicken (Pickel, 1998) or a plastic flamingo (Mansour, Lindsay, & Munhall, 2008), while others still showed weapons but in completely unusual situations, like a weapon carried by a priest compared to one carried by a police officer (Pickel, 1999). Most studies also used traditional weapon conditions in addition to these unusual objects. Findings were mixed as some authors found a larger effect for weapon conditions (Hope & Wright, 2007; Mitchell, Livorsky, & Mather, 1998, Exp. 1), whereas others found a smaller effect compared to an unusual only condition (Mitchell et al., 1998, Exp. 2; Pickel, 1998, 1999, 2009). Nonetheless, in contrast to the neutral control group, an unusualness effect has been demonstrated several times (e.g., Pickel, 1998; Hope & Wright, 2007; Pickel et al., 2006). Some studies also used (unusual) weapons in the unusualness condition, thus testing the effects of unusualness and weapon focus as additive (or interactive) components.

In summary, we expected unusualness to have similar effects on eyewitness memory with regard to both target descriptions and identifications. Because these studies used both weapon and unusualness conditions but only one control group, effect size estimates for weapon presence and unusualness are not independent from each other (cf. Gleser & Olkin, 2009). Consequently, the impact of unusualness was calculated in a separate meta-analysis.

Hypotheses

The purpose of the present meta-analysis was to statistically synthesize past studies on the WFE and to test rival theoretical approaches to account for the source of deterioration in memory performance when a witness is confronted with a weapon. Although the hypotheses are stated in a way that the WFE is supposed to indicate a deterioration in witness performance, thus implying a negative direction of effect, we coded the effects obtained with an inverse sign, so that *positive* values (and odds ratios > 1) denote *better* performance in the *control group.*

Mean weighted effect sizes using both fixed- and random-effects models were calculated for all dependent variables concerning identification as well as description accuracy. The following hypotheses were tested:

Hypotheses regarding person identifications:

(1) A target holding a weapon is identified correctly less frequently in a target-present (TP) lineup in comparison to a target carrying either nothing or a neutral object.

(2) In target-absent (TA) lineups, witnesses who observe a perpetrator with a weapon make more false identifications compared to those who observe a target without a weapon.

(3) Taken together, persons make fewer correct decisions across TP and TA lineups when the perpetrator carries a weapon than when he/she does not.

Predictions regarding recall:

(4) The presence of a weapon leads to less accurate descriptions of the perpetrator compared to conditions where no weapon is present.

(5) In contrast, we expect better recall of the weapon (central information) compared to other objects carried by the target in the control condition.

Hypotheses regarding unusualness:

We expected the factor of unusualness to have a similar impact on memory performance to that of the WFE. Thus, when a target carries an unusual object, or a weapon in an unusual context, the target should be identified less often or described less accurately compared to an empty-handed target or a target with a neutral object.

Moderator variables

Presumably, the most important moderator variable with respect to theoretical explanations and practical applications is the arousal a person experiences while witnessing a crime. Hence, we assume perceived threat to moderate the WFE. Because most studies did not report arousal levels of participants, we used the type of weapon as a proxy to presumed levels of arousal. Hence, we expected a larger weapon focus effect when a gun was used compared to any other weapon, like a knife or a screwdriver. We further investigated whether the WFE was associated with mode of presentation (live, video, or slides) and type of crime (no crime vs. crime).

The effects of unusualness may be stronger or moderated when a weapon is used in an unusual context compared to an unusual object held by the target.

Method

Search Strategies Used

Several strategies recommended in the literature (Lipsey & Wilson, 2001; White, 2009; Sporer & Cohn, 2011) were pursued to identify potential studies, including: (a) an electronic literature search in the databases Web of Science (including articles listed in the Social Sciences Citation Index (SSCI) post 1956, the Science Citation Index post 1945, and the Arts and Humanities Index post 1975), PsycINFO (post 1967), Dissertation Abstracts (post 1980) from the first available year to August 2009. The keywords identified via relevant studies and therefore used in the electronic database searches were: "weapon focus (effect)", "armed robber*", "reliability AND eyewitness identification", "attention AND eyewitness identification", "weapon* visibility" in title and topic; (b) a citation search in the SSCI for the first meta-analysis on this topic (Steblay, 1992), (c) a manual search of conference programs (American Psychology-Law Society, European Association of Psychology and Law, Conference of the German Psychological Society, International Congress of Psychology, Psychology and Law, International Conference); (d) a manual search in topically relevant journals (*Law and Human Behavior*, *Journal of Applied Social Psychology, Psychology, Crime, and Law*, *Journal of Applied Psychology*, *Applied Cognitive Psychology*), (e) a complete search of relevant references cited in any of the previously located articles, reviews or meta-analyses on the WFE; and (f) contacting first authors of included studies by e-mail to send any relevant work, published or unpublished. To reduce the effect of *publication bias*, that is, the tendency for studies with non-significant findings to be less likely to be submitted, or if submitted, to be published (Lipsey & Wilson, 1993; Sporer & Goodman-Delahunty, 2011; Sutton, 2009), we also sought to include unpublished data as well.

Studies were first screened by their titles, then by abstracts, and, if necessary, a full text reading to decide about final inclusion.

Inclusion Criteria

For inclusion in the current analyses, studies had to meet the following criteria: (a) the study design included at least one condition with the presence of a weapon, where a weapon was clearly visible or used, and one condition, where no weapon was present, visible or used; and (b) either identification accuracy or recall of physical features concerning the target was reported as dependent variable(s); (c) sufficient statistical information was reported to calculate effect sizes (see below); (d) study information was available either in English, Spanish, or German.

One study was excluded that used a face recognition paradigm (Tooley, Brigham, Maass, & Bothwell, 1987) although it had been included in Steblay's (1992) meta-analysis, because proportions from continuous distributions cannot be considered equivalent to identification data with dichotomous outcomes. Whereas in lineup studies participants normally have to remember one perpetrator, which they later have to identify in a lineup, in face recognition studies participants are shown a series of targets, which they later have to recognize in a larger set of faces. Furthermore, because the focus of all but one study was on adults' eyewitness performance, one study was excluded which examined only children (Davies, Smith, & Blincoe, 2008). Two further studies had to be excluded because the authors were not prepared to send the information necessary for inclusion (Lorenz & Yan, 2008), or could not be contacted (Kenrick & Mallard, 2006).

A total of 23 research articles met the inclusion criteria (vs. 11 in Steblay's, 1992, meta-analysis). All studies included and the respective dependent variables extracted from them are marked with an asterisk in the reference list.

Coding of the Studies

Two coders completed all coding independently by means of a coding protocol, following Wilson's (2009) recommendations that information about study characteristics, as well as information about dependent variables, should be double coded. The main moderator variables coded are listed in Table 1. These moderator variables had originally been coded into more fine-grained categories, which were later collapsed into broader categories to avoid too small cell sizes.

Inter-coder reliabilities were estimated using Cohen's *kappa* for categorical variables, as it controls for chance agreement, and the intraclass correlation coefficient (r_I) for continuous variables which also takes systematic differences between coders into account (Orwin & Vevea, 2009; Sporer & Cohn, 2011). Overall, inter-coder agreement was highly satisfactory, with all coefficients indicating either perfect agreement, or $r_I \geq .74$ (see Table 2). The few remaining disagreements were resolved among the authors by discussion.

Effect sizes

For the current meta-analysis, weighted mean effect sizes and their inverse variance weights were calculated, following the recommendations in the respective chapters in the *Handbook of Research Synthesis* (Borenstein, 2009; Fleiss & Berlin, 2009; Shadish & Haddock, 2009; Konstantopoulos & Hedges, 2009; Raudenbusch, 2009).

Table 1

Major Study Characteristics of all Weapon Focus Studies Included

Study	Date	Exp.	Weapon	Total *N*	Arousal	Mode	Interval	Target Exposure in s	Weapon Exposure in s	Crime/ Event
Bothwell	1991	2	Gun	137	--	Video	Imm.	20	20	No
		3	Gun	56	--	Video	Imm.	20	20	No
		3	Knife	58	--	Video	Imm.	20	20	No
Cutler & Penrod	1988		Gun	175	--	Video	Imm.	75	--	Robbery
Cutler et al.	1986	1	Gun	320	--	Video	Imm.	--	--	Varied
		2	Gun	287	--	Video	Varied	--	--	Robbery
Cutler et al.	1987a		Gun	165	Varied	Video	Varied	75	--	Robbery
Cutler et al.	1987b		Gun	290	--	Video	Varied	--	--	Robbery
Hope & Wright	2007		Gun	30	--	Slides	< 1 hour	--	4	No
Hulse & Memon	2006		Gun	70	Varied	Live	< 1 hour	30	5	Domestic dispute
Johnson & Scott	1976		Opener	48	Varied	Live	Varied	4	4	Exchange overheard
Kramer	1990		Cleaver	62	--	Slides	Imm.	18	3	No
Kramer et al.	1990	1	Bottle	64	Varied	Slides	Imm.	17	12	Assault
		2	Cleaver	64	--	Slides	Imm.	18	18	No
		3	Cleaver	32	Low	Slides	Imm.	12	18	No
		4	Cleaver	48	Low	Slides	Imm.	18	12	No
		5	Cleaver	42	Low	Slides	Imm.	18	3	No
Loftus et al.	1987	1	Gun	36	--	Slides	< 1 hour	--	6	Robbery
		2	Gun	80	--	Slides	< 1 hour	--	6	Robbery
Maass & Köhnken	1989		Syringe	86	Varied	Live	< 1 hour	20	20	Threat
Mitchell & Livosky	1998	1	Gun	41	Low	Video	Imm.	7	5	Robbery
		2	Gun	120	Low	Video	Imm.	7	5	Robbery
O'Rourke et al.	1989		Gun	132	--	Video	1 week	75	--	Robbery

Table 1 (continued)

Authors	Date	Exp.	Weapon	Total *N*	Arousal	Mode	Interval	Target Exposure in s	Weapon Exposure in s	Crime/ Event
Pickel	1998	1	Gun	91	--	Video	< 1 hour	--	--	No
		2	Knife	105	--	Video	Imm.	--	--	No
Pickel	1999	1	Gun	129	--	Video	< 1 hour	20	--	Varied
Pickel	2009	1	Gun	127	Mod.	Video	Imm.	30	--	Robbery
		2	Knife	181	Mod.	Video	Imm.	30	--	Robbery
		3	Knife	255	--	Video	Imm.	--	--	Robbery
Pickel et al.	2003	1	Gun	217	--	Video	< 1 hour	--	--	Intruder/ Stalker
		2	Knife	140	Mod.	Video	< 1 hour	--	--	Intruder/ Stalker
Pickel et. al	2006	1	Gun	61	Mod.	Live	Imm.	30	30	Classroom intruder
		2	Gun	56	Mod.	Live	Imm.	15	15	Classroom intruder
Pickel et al.	2008	1	Gun	261	--	Video	Imm.	--	--	Robbery
		2	Gun	113	--	Video	Imm.	--	--	Robbery
Shaw & Skolnick	1999		Gun	120	--	Video	< 1 hour	30	30	Classroom intruder
Stanny & Johnson	2000	1	Gun	40	--	Video	Imm.	16	2	Domestic dispute
		2	Gun	29	Varied	Video	Imm.	--	2	Domestic dispute

Note. Exp. = Number of experiment; Mod. = Moderate; Imm. = Immediately; Varied = Manipulated across conditions; -- = not reported.

Table 2

Study Descriptors for Weapon Focus Effect and Unusualness Effect Studies

	Weapon focus studies						**Unusualness studies**					
Variable	*k*	%	*M*	*SD*	*kappa*	r_I	*k*	%	*M*	*SD*	*kappa*	r_I
PARTICIPANT CHARACTERISTICS												
Sample					1.000						1.000	
Students	38	90.5					17	100				
Community residents	1	2.4										
Police officers	3	7.1										
Sample sizes	42		111.45	78.48		.979	17		65.88	23.84		1.000
METHOD CHARACTERISTICS												
Duration of event/film (s)	37		76.92	55.92		.739	16		313.59	581.06		.799
Target visible (s)	27		25.92	19.20		.898	12		66.67	88.02		.957
Weapon/object visible (s)	24		12.73	9.35		.994	3		193.33	150.11		na
Mode of presentation					1.000						1.000	
Video	27	64.3					12	70.6				
Slides	9	21.4					1	5.9				
Live	6	14.3					4	23.4				

Table 2 (continued)

	Weapon focus studies						Unusualness studies					
Variable	*k*	%	*M*	*SD*	*kappa*	r_I	*k*	%	*M*	*SD*	*kappa*	r_I
Staged crime					1.000						1.000	
Staged crime	28	66.6					8	47.1				
No staged crime	14	33.3					9	52.9				
Retention interval					.954						1.000	
Immediate	25	59.5					17	100				
< 1 hour	12	28.6										
< 1 week	1	2.4										
Mixed[a]	4	9.5										
Arousal					.979						.941	
Not reported	23	57.2					6	35.3				
Low	5	11.9					4	23.5				
Medium	7	16.7					4	23.5				
High												
Varied[a]	6	14.3					3	17.6				

Note: k = number of hypothesis tests where data were available; *kappa* = Cohen's *kappa*; r_I = intraclass correlation; na = there were not enough cases to calculate a reliability coefficient; duration of event/film = full length of event/film presented to participants; target visible = time in s the target was visible throughout the event/film presented to participants; weapon/object visible = time in s the weapon/unusual object was visible throughout the event/film presented to participants; mode of presentation = form in which the stimulus material was presented; staged crime = event/film contained a staged crime to participants; retention interval = time between exposure of stimulus material and identification/description task; arousal = arousal experienced by the participants; varied = manipulated across conditions.

Analyses were carried out separately for person identification and person description data.[1]

---

[1] Unfortunately, we were not aware that another research group was also conducting a meta-analysis on the WFE until all our analyses were completed (Fawcett, Russell, Peace, & Christie, 2011). The major difference between their meta-analysis and ours is that they extracted a single effect size per study (or averaged effect sizes within studies to arrive at a single effect size per study) calling it an overall effect on memory performance. Although we do not intend to criticize this procedure, we justify our separate effect size calculations as follows: (1) It has been argued for a long time that different memory mechanisms underlie recall and recognition. Some authors have even argued that there may be two distinct underlying memory systems (e.g., Tulving, 1985). (2) Recall and recognition are differentially affected by some independent variables, for example the steeper drop in performance in recall compared to recognition as a function of retention interval. (3) There appear to be only a low (if any) correlations between identification accuracy and various measures of person descriptions (see the meta-analysis by Meissner, Sporer, & Susa, 2008). (4) In the identification literature, it is customary to separately report analyses of TP and TA outcomes, presumably because the decision processes differ. (5) Although different authors have proposed different formulae to convert effect sizes for standardized mean differences and odds-ratios and vice versa, these conversions are not without problems, depending on a variety of factors (Lipsey & Wilson, 2001; Sanchez-Meca, Chacón-Moscoso, & Marín-Martínez, 2003). (6) Finally, we believe that the problem to establish "ground truth" in archival studies does not allow us to calculate effect sizes for identification or description outcomes which can be integrated with effect sizes from experimental (simulation) studies. Nonetheless, we will incorporate comparisons with outcomes from archival analyses in our final discussion.

The effect size of choice regarding *identification accuracy* (a binary dependent variable), when testing differences between proportions, is the odds ratio (Fleiss & Berlin, 2009; Lipsey & Wilson, 2001; Sporer & Cohn, 2011). The odds ratio (*OR*) is an effect size statistic that compares two groups in terms of relative odds of a status or event, not to be mistaken with the ratio of two probabilities (rate ratio). For example, when the target is correctly identified by 75% of the participants (.75), the odds of a successful outcome, here the correct identification of the perpetrator, are 3 to 1 (three successes to one failure), whereas the probability of a successful outcome is 3 in 4 (three successes in four cases). Accordingly, the odds ratio is the ratio of two odds,

(1) $$OR = \frac{p_{EG}\left(1 - p_{CG}\right)}{p_{CG}\left(1 - p_{EG}\right)}$$

(Lipsey & Wilson, 2001, p. 53, Equation 3.28), where p_{EG} indicates the probability of the occurrence of an event (e.g., a hit) in the experimental (weapon) group, and p_{CG} the probability of this event in the control group. The odds ratio is applicable to research findings that use dichotomous variables, which are presented in the form of relative frequencies and proportions. It is customary to perform all analyses on the natural log of the odds ratio (*LOR*) which is centered around 0. When interpreting results, *LOR*s are backtransformed (Fleiss & Berlin, 2009; Lipsey & Wilson, 2001). As already indicated, we calculated the effect sizes in such a way that *OR*s > 1 (and LORs > 0) would indicate better performance in the control group.

According to the continuous nature of recall measures like *description accuracy* of target or object characteristics, where correct or false descriptions are usually expressed in terms of means and standard deviations, we used the standardized mean difference, Cohen's *d* (Cohen, 1988; Hedges & Olkin, 1985),

(2) $$d = \frac{M_{EG} - M_{CG}}{s_p}$$

(Lipsey & Wilson, 2001, p. 48, Equation 3.21), where M_{EG} is the mean of the experimental group, M_{CG} indicates the mean of the control group and s_p is the standard deviation pooled across both groups, defined as

(3) $$s_p = \sqrt{\frac{(n_{EG}-1)s_{EG}^2 + (n_{CG}-1)s_{CG}^2}{(n_{EG}-1)+(n_{CG}-1)}}$$

(Lipsey & Wilson, 2001, p. 47, Equation 3.20), where s_{EG} is the standard deviation for the experimental group, s_{CG} is the standard deviation for the control group and n_{EG} and n_{CG} the number of participants for each group.

The effect size *d* was transformed to obtain an unbiased estimator Hedges g_u, using the following formula:

(4) $$g_u = Jd\ ,\ J \approx 1 - \frac{3}{4df-1}\ ,\ df = n_{EG} + n_{CG} - 2$$

(Borenstein, 2009, p. 226, Equation 12.15, 12.16), where *J* is the correction factor and *d* is Cohen's *d* as defined above. If studies did not directly report means and standard deviations, *t*- or *F*-values were used to calculate Hedges g_u, taking unequal sample sizes into account whenever possible (Borenstein, 2009). Table 3 demonstrates approximate transformations between the used effect sizes as well as Point-biseral *r*.

Outlier analyses

Before weighting the single effects within fixed and random effect models, we conducted outlier analyses to detect extreme cases following the procedures recommended by Hedges and Olkin (1985). Extreme effect sizes might be unrepresentative of the results and have disproportionate influence on means, variances, and other statistics used in meta-analysis, resulting in misleading conclusions (Hedges & Olkin, 1985).

Table 3

Approximate Transformations between Odds Ratios (OR, LOR) , Cohen's d (ESd) and Point-biserial r (ESr)

OR	LOR	ESd	ESr
1.00	0.00	0.00	0.00
1.50	**0.41**	**0.22**	**0.11**
2.00	0.69	0.38	0.19
2.50	**0.92**	**0.51**	**0.24**
3.00	1.10	0.61	0.29
3.50	1.25	0.69	0.33
4.00	1.39	0.76	0.36
4.50	**1.50**	**0.83**	**0.38**
5.00	1.61	0.89	0.41
5.50	1.70	0.94	0.43
6.00	1.79	0.99	0.44

Note. "Small", "medium", and "large" effect sizes according to Cohen's (1988) recommendations are marked in bold face.

In a first step, graphical techniques were used for displaying effect size estimates, like the distribution of single effect sizes and their confidence intervals (see Sporer & Cohn, 2011), as well as box plots demonstrating the median, upper and lower quartiles and possible deviant cases (see Meissner, Sporer, & Susa, 2008).

In a further step, we calculated the standardized residuals *e* (difference to the weighted average effect size) and the homogeneity test statistic *Q*, for all *k*-1 effect sizes, after removing one study at a time (Hedges & Olkin, 1985), which is sometimes referred to as the leave-one-out method. Due to the assumed standard normal distribution of the residuals, "residuals larger than about 2.00 in absolute magnitude occur only about 5 percent of the time when the effect sizes are homogeneous" (Hedges & Olkin, 1985, p. 255). Usually, the homogeneity index *Q* shows a dramatic change when an outlier is removed.

If such a noticeable difference was found and no methodological peculiarity of the study could be identified that would explain the outlier status, the effect size for that study was approximated using a procedure called *Winsorizing* which is also recommended for meta-analyses by Lipsey and Wilson (2001). As a safety check, we calculated models with and without the outliers in order to eliminate any possibility of affecting substantive conclusions.

Fixed and random effects models

We used fixed and random effects models (Konstantopoulos & Hedges, 2009; Lipsey & Wilson, 2001; Raudenbush, 2009), where studies are weighted by the inverse of their variance (plus the random effects variance component in the random effects model). The test statistic *Q* was used as a test of homogeneity (Shadish & Haddock, 2009). Because the homogeneity test has little statistical power (Harwell, 1997; Hedges & Pigott, 2001; Jackson, 2006), as a further indicator of heterogeneity the descriptive statistic I^2 was used to quantify the proportion of total variation between effect sizes, where 25% indicates small variation, 50% medium and 75% large heterogeneity. Compared to *Q*, the index I^2 has the advantage of being independent of sample size and the effect size metric calculated and is therefore recommended to be used as a supplement to *Q* to test for heterogeneity (Higgins & Thompson, 2002; Sutton & Higgins, 2008).

Moderator analyses

One approach is to block the studies into two or more subgroups. Problems occur where grouping variables themselves are confounded with each other (Lipsey, 2003). As is evident by inspection of Table 1, this was frequently the case. To detect such dependencies, cross tabulation analyses as well as inter-correlations between predictor variables were calculated to avoid small (or even empty) cells in blocking. In addition, meta-regression analyses were conducted as weighted hierarchical/sequential procedures (Lipsey & Wilson, 2001; Pigott, 2012; Sporer & Cohn, 2011).

Results

All effect sizes coded are summarized in Table 4. We primarily describe the results of the fixed-effects model (FEM), only noting when there are large discrepancies between FEM and random-effects model (REM). Table 5 contains both FEM and REM results.

Lineup Identifications

Correct identification in target-present lineups. Outlier analyses revealed two studies as problematic (see Figure 1). An unpublished study by Bothwell (1991) used two different weapon conditions (handgun or knife) vs. a no weapon control condition. The handgun resulted in an extreme logarithmized odds ratio (*LOR*) of -1.706 (*OR* = 0.182), a clear outlier, whereas the comparison knife vs. empty-handed, *LOR* = -0.402 (*OR* = 0.669), was a less extreme outlier. Both effect sizes point to an improvement in identification when a weapon was present. As the effect sizes of these two conditions are statistically dependent due to the common control group (cf. Gleser & Olkin, 2009), they cannot both be used. Hence, we only used the effect size for the more typical gun condition.

Table 4
Effect Sizes for Weapon Focus Effect Studies

Study	Date	Exp.	Weapon	Total *N*	*OR* Hits	*OR* Overall Correct Decisions	*OR* False Alarm (TA)	Hedges g_u Target Description Accuracy	*OR* Correct Object Naming	Hedges g_u Object Description Accuracy
Bothwell	1991	2[a]	Gun	137				0.000		
		3	Gun	56	0.178*	0.267*	-1.145			
		3[a]	Knife	58	0.669		-0.134			
Cutler & Penrod	1988		Gun	175		0.713				
Cutler et al.	1986	1	Gun	320		1.000				
		2	Gun	287		1.429				
Cutler et al.	1987a		Gun	165		1.829				
Cutler et al.	1987b		Gun	290		0.882				
Hope & Wright	2007		Gun	30				1.100	9.337	1.073
Hulse & Memon	2006		Gun	70			-.655	-.280		
Johnson & Scott	1976[a]		Opener	48	2.000					
Kramer	1990[a]		Cleaver	62	1.143					
Kramer et al.	1990	1	Bottle	64	7.714			.801		
		2	Cleaver	64	1.137			1.235		
		3	Cleaver	32	1.000			1.495*		
		4	Cleaver	48	2.363			.915		
		5	Cleaver	42	0.825			.307		
Loftus et al.	1987	1	Gun	36	5.089			.240	1.000	
		2	Gun	80	3.053			.451		
Maass & Köhnken	1989		Syringe	86			.850			.495[b]
Mitchell & Livosky	1998	1	Gun	41				.487		
		2	Gun	120				.548		
O'Rourke et al.	1989		Gun	132		1.598				

Table 4 (continued)

Study	Date	Exp.	Weapon	Total N	*OR* Hits	*OR* Overall Correct Decisions	*OR* False Alarm (TA)	Hedges g_u Target Description Accuracy	*OR* Correct Object Naming	Hedges g_u Object Description Accuracy
Pickel	1998	1	Gun	91	1.777			.347	115.700	
			Scissors					.436		
		2	Knife	105	2.259			.457	87.182	
			Screwdriver					.561		
Pickel	1999	1	Gun	129	0.544			-.089		
		2	Gun	122	1.039			-.372		
Pickel	2009	1	Gun	127				1.165	1.000	
		2	Knife	181				.827	1.000	
		3	Knife	255				1.403	0.738	
Pickel et al.	2003	1	Gun	217		1.398		.363	19.240	
		2	Knife	140		0.824		.505	6.699	
Pickel et al.	2006	1	Gun	61				1.035		(1.472[c])
		2	Gun	56/53				1.175		1.564
Pickel et al.	2008	1	Gun	261				2.317*		
		2	Gun	113				.484		
Shaw & Skolnick	1999		Gun	120/132[d]				.143		-.249
Stanny & Johnson	2000	1	Gun	40				1.410		
		2	Gun	29						

Note. Exp. = Number of experiment; *LOR* = logged odds ratio; * = outlier. [a] = Unpublished. [b]Description of hand area.

[c]This value was estimated after correcting the *SD* reported. [d]Estimated *N* = 120 or 132.

Table 5

Mean Weighted Effect Sizes for the Weapon Focus Effect for Correct Identifications in Target-Present Lineups, Overall Correct Identification Decisions and Accuracy of Target Descriptions

DV	Model	*k*	*N*	*LOR*	*OR*	95% CI *LL*	95% CI *UL*	g_u	*Z*	p_Z	*VC*	*Q*	p_Q	I^2
Correct Identifications in TP Lineups	FEM	14	952	0.252	1.287	0.971	1.706		1.753	.080		21.46	.064	.394
	REM	14	952	0.297	1.346	0.921	1.968		1.535	.125	.193			
Overall Correct Identification Decisions	FEM	9	676	0.097	1.101	0.795	1.525		0.581	.561		9.10	.334	.121
	REM	9	676	0.081	1.084	0.764	1.539		0.453	.650	.035			
Target Description Accuracy	FEM	29	2649			0.490	0.647	0.568	14.189	<.001		155.70	<.001	.820
	REM	29	2649			0.427	0.809	0.618	6.350	<.001	.215			

Note. *LOR* = logged odds ratio; *OR* = odds ratio, g_u = Hedges g_u; *CI* = confidence interval; *LL* = lower limit, *UL* = upper limit; *Z* = significance test; p_Z = significance level for Z-test; *VC* = variance component in REM; *Q* = heterogeneity test statistic; p_Q = significance level for *Q*-test; I^2 = indicator of heterogeneity; Target Description Accuracy included number of correct details, scores calculated by subtracting incorrect details from correct ones, or proportion accuracy (correct details/[correct + incorrect details]).

The study by Pickel (2009, Exp. 1) was also an outlier, *LOR* = -0.608, *OR* = 0.544. No remarkable feature in this study could be detected.

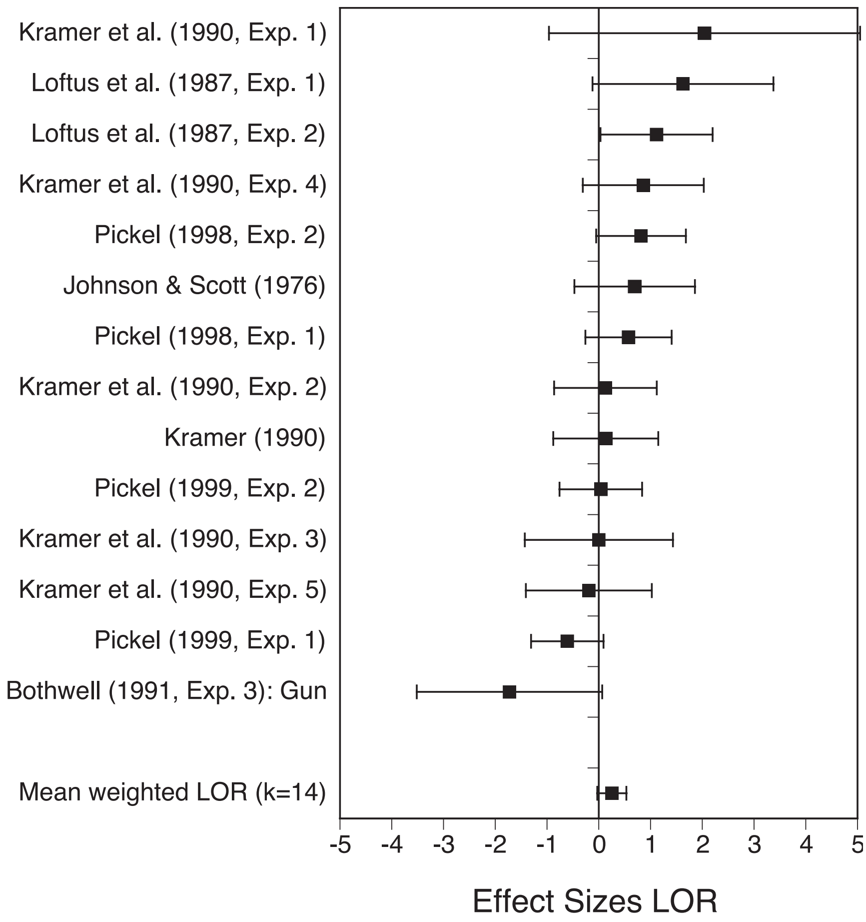

Figure 1. Effect sizes (*LOR*) and 95% confidence intervals fort he WFE on correct identification decision in perpetrator-present lineups.

When both published and unpublished studies were included (*k* = 14, *N* = 952), there was no significant WFE effect, *OR* = 1.287 [0.971, 1.706] for the FEM or for the REM (see Table 4). In other words, participant-witnesses were only 1.29 more likely to identify the perpetrator correctly in TP lineups when no weapon was present compared to the weapon condition but this odds ratio did not differ from chance. After exclusion of the unpublished studies by Bothwell (1991, Exp. 3), which involved showups, not lineups, and by Johnson and Scott (1976) and Kramer (1990; as cited in Steblay, 1992) the effect remained nonsignificant both in the FEM, *OR* = 1.337 [0.983, 1.819], and the REM, *OR* = 1.441 [0.956, 2.172].

Correct decisions. Some authors did not report separate outcomes but only correct decision rates across both TP and TA lineups, yielding k = 9 independent hypothesis tests (see Table 4). No outliers were detected and results can be considered homogeneous, $Q(8) = 9.10$, $p = .334$, $I^2 = 0$. The WFE showed no significant effect, neither for the FEM, *OR* = 1.101 [0.795, 1.525], nor for the REM (see Table 5). Participants were only 1.10 more likely to make a correct lineup decision when no weapon was present compared to the weapon condition. Even when the study by Bothwell (1991, Exp. 3) was excluded, the effect was far from being significant, *OR* = 1.231 [0.878, 1.726], $Z = 1.21$, $p = .228$.

False identifications in target-absent lineups. Only three studies reported results for target-absent lineups (Bothwell, 1991, Exp. 3; Maass & Köhnken, 1989; Hulse & Memon, 2006), yielding effect sizes of *OR* = 1.129, 95% CI [.848, 2.294] for the handgun condition, and *OR* = .396, 95% CI [.182, .861] for the knife condition in the study by Bothwell (1991). For the latter two studies, the effect sizes were *OR* = 2.373, 95% CI [1.524, 3.690] and *OR* = .550, 95% CI [.340, .887], respectively. For this small, heterogeneous set of studies, no meta-analytic synthesis is meaningful.

Eyewitness Recall

Correct naming of weapon or objects. A total of k = 15 (N = 1494) studies reported whether the weapon/object was correctly recalled. In the weapon condition, on average 93.8% of the participants correctly recalled (named) the weapon (a clear ceiling effect), compared to 72.0% recall of the control object. In fact, in 9 of the 15 studies *all* participants correctly named the weapon. We consider this dependent variable simply as a manipulation check that the weapon was actually noticed by participants watching the film (or staged event).

Description of the perpetrator. In terms of the ability to later describe the perpetrator, different authors used different measures of description accuracy. While some authors used number of correct details, others used different types of accuracy scores. Despite these differences, we combined the respective effect sizes in one model for our main analyses because they all report some measure reflecting the ability to describe the target.

Preliminary analyses regarding different types of operationalizations of description accuracy measures support our decision to pool results. Two moderator analyses were conducted, one regarding the type of operationalizations used, and a second one regarding the definition of accuracy. The majority of studies (k = 16) reported "feature accuracy" as a combined measure of free and cued recall as well as multiple choice questions, which resulted in a weighted mean effect size of g_u = 0.586 [0.634, 0.680], Z = 12.15, p < .001. Another k = 8 studies used either free or cued recall measures, with g_u = 0.533 [0.360, 0.707], Z = 6.01, p < .001. Finally, a small group of studies (k = 5) used multiple choice questions only, with g_u = 0.523 [0.282, 0.764], Z = 4.25, p < .001. A blocked analysis revealed that the three groups of studies did not differ in outcome, $Q(2)$ = .419, p = 811.

However, there were significant differences between groups of studies, depending on whether or not incorrect details were considered in the definitions of description accuracy, $Q(2)$ = 33.298, p < .001. In the few studies (k = 4) that reported results on correct details only, the WFE was strongest, g_u = 0.965 [0.804, 1.126], Z = 11.73, p < .001, followed by studies, in which incorrect details were simply subtracted from correct details, g_u = 0.380 [0.262, 0.498], Z = 6.32, p < .001, k = 11. Only medium to small effects were reported in the majority of studies, where accuracy was defined as a relative proportion measure, that is, "accuracy = (correct details/(correct+incorrect details)", g_u = 0.535 [0.396, 0.674], Z = 7.55, p < .001, k = 14. Because there were only four studies from which we used correct details only as a measure (Pickel, 2009, Exp. 1, 2, and 3; Shaw & Skolnick, 1999), we felt justified to ignore these differences. The larger effect sizes in the Pickel (2009) experiments are likely a function of the manipulation of gender as a target, which resulted in higher effect sizes for female (presumably due to expectancy violation) compared to male perpetrators.

Overall, 29 independent hypothesis tests with a total of *N* = 2649 participants were located examining descriptions of the target (28 published, 1 unpublished). Outlier analyses revealed a highly heterogeneous set of effect sizes, $Q(28) = 175.71$, $p < .001$, $I^2 = .841$, ranging from $g_u = -0.372$ to 2.317 (see Figure 2). Even after winsorizing one extreme value from the study by Pickel (2009, Exp. 3, mail carrier), heterogeneity was still very large, $Q(28) = 155.70$, $p < .001$, $I^2 = 0.820$, suggesting that moderator analyses should be used to account for this heterogeneity (see Table 5).

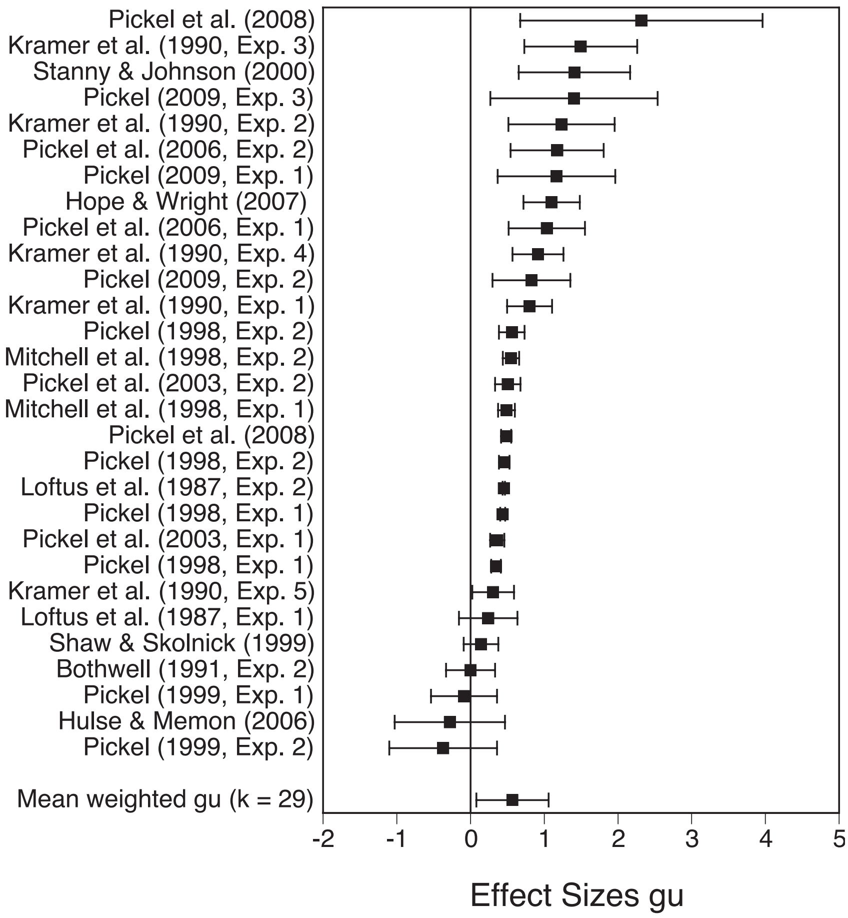

Figure 2. Effect sizes (g_u) and 95% confidence intervals for the WFE on the accuracy of perpetrator descriptions in WFE studies.

The weighted mean effect size was g_u = 0.568, 95% CI [0.490, 0.647], Z = 14.19. As expected, confidence intervals were larger in the REM (see Table 5). As predicted, descriptions of the target without a weapon were more accurate than with a weapon.[2]

Consistent with expectations, there was evidence in four out of five studies that participants in the weapon group were more accurate in describing the weapon (or the hand area) than participants in the no weapon control group (see Figure 3). As these five studies were highly heterogeneous, $Q(4) = 43.95$, $p < .001$, $I^2 = .909$, no mean effect size was calculated.

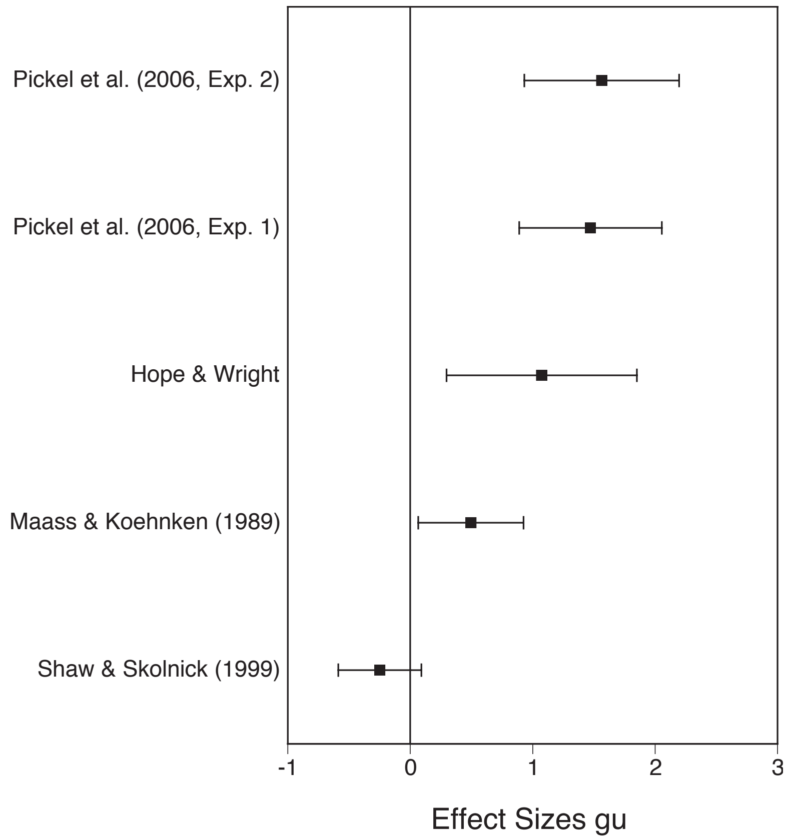

Figure 3. Effect sizes (g_u) and 95% confidence intervals for the WFE on the accuracy of object descriptions in WFE studies.

---

[2] Excluding the unpublished studies by Bothwell (1991, Exp. 2) resulted in a weighted mean effect size of 0.602 [0.521, 0.682] which is also highly significant.

Moderator Analyses. Based on the theoretical approaches described, we included three methodological variables presumably associated with emotional arousal: mode of presentation while encoding (slides vs. video/live), the presence of a staged crime (present vs. absent) and the type of weapon (gun vs. others). Predictors were largely uncorrelated (type of weapon and presence of a staged crime $\varphi(27) = -.30$, $p = .127$; type of weapon and mode of presentation $\varphi(27) = -.30$, $p = .127$), whereas mode of presentation was significantly associated with presence of a staged crime ($\varphi(27) = .49$, $p < .05$), but not too large to be concerned about multi-collinearity. As blocking analyses would be inconclusive due to these partial confounds, meta-regression was conducted with these three predictors which explained 24% of the variance (see Table 6).

While mode of presentation did not affect the results, both the presence of a crime and weapon dangerousness were significantly associated with the accuracy of person descriptions. While the WFE was larger when a crime occurred than without a crime, the dangerousness of the weapon showed an effect opposite to what we had expected. When the target had a knife (or a meat cleaver), the WFE was larger, $g_u = 0.773$, 95% CI [0.656, 0.890], $Z = 12.96$, $p < .001$, $k = 14$, than when he had a gun, $g_u = 0.400$, 95% CI [0.294, 0.506], although the latter effect was also significant, $Z = 7.41$, $p < .001$, $k = 15$.

In the following, we describe the method and results of unusualness analogously to the traditional WFE studies.

Table 6

Hierarchical Weighted Fixed-Effect Meta-regression for Accuracy of Target Descriptions in Weapon focus Effect Studies (k = 29)

Study characteristics	*B*	*beta*	*p*	R^2
Step 1:				
Model $Q(1) = 4.08$, $p = .043$				.026
Residual $Q(27) = 151.62$, $p < .001$				
Mode of presentation				
(slides vs. video/live)	-.227	-.162	.043	
Step 2:				
Model $Q(3) = 52.47$, $p < .001$				.337
Residual $Q(25) = 103.24$, $p < .001$				
Mode of presentation				
(slides vs. video/live)	-.202	-.144	.078	
Crime category				
(no crime vs. crime)	.343	.328	<.001	
Type of weapon				
(gun vs. others)	.492	.490	<.001	
Regression constant	.006			

Note. *Q* statistics shown are from fixed-effects meta-regression models and test whether (1) the methodological variable mode of presentation, and (2) crime category and type of weapon used account for an additional portion of the variance in effect sizes.

Unusualness Effect

Method

In- and exclusion criteria as well as analyses were the same as for the traditional WFE. Unusualness had been manipulated either as object type, which was unusual within a given context, or according to stereotype-dependent expectations ("expectancy violations").

Study characteristics. Overall, 16 independent hypothesis tests could be identified examining the factor of unusualness, 6 for identification decisions, and 16 for target description accuracy. Due to the small set of hypothesis tests for other dependent variables only the results of these two dependent variables, viz. correct identifications in target-present lineups and description accuracy, were analyzed.

Studies were conducted between 1998 and 2009, including *N* = 1102 students as participants. All studies used between-participants designs, either comparing unusual objects with neutral ones, varying the target carrying a weapon (e.g., a priest vs. a police officer), presenting unusual weapon types (e.g., a 19th-century weapon), or presenting the weapon in an unusual context compared to an usual one (e.g., shooting range vs. baseball field; see Table 7 for further details and effect sizes).

Results

Table 8 gives an overview of the meta-analytic syntheses in the FEM and REM.

Correct identification of the target in target-present lineups. Six independent effect sizes could be identified comparing an unusual group with a control condition. Outlier analyses revealed no extreme cases and the homogeneity test did not reach significance, $Q(5) = 4.36$, $p = .499$, $I^2 = 0$.

Table 7

Study Characteristics and Effect Sizes for Hits in Target-Present Lineups and for Accuracy of Target Descriptions for Unusualness Studies that did or did not also Contain a Weapon

Study	Date	Exp	Object	Total N	*OR* Hits	Weapon Present (0/1)	Hedges g_u Target Description Accuracy
Hope & Wright	2007		Feather duster	30	--	0	0.452
Mansour et al.	2008	1	Plastic flamingo	40	1.416	0	--
		2	Plastic flamingo	40	1.000	0	--
Mitchell et al.	1998	1	Celery	42	--	0	0.588
		2	Celery	80	--	0	0.392
Pickel	1998	1	Raw, whole chicken	92	1.415	0	0.897
		2	Doughboy	104	2.829	0	0.746
Pickel	1999	1	Gun at shooting range/baseball field	129	1.035	1	0.594
		2	Priest (vs. police officer) with gun	122	0.710	1	0.687
Pickel	2009	1a	Man with gun	63	--	1	0.841
		1b	Woman with gun	64	--	1	1.894
		2a	Man with knitting needles	60	--	0	1.089
		2b	Woman with knitting needles	60	--	0	0.651
		3a	CG: Man with gun	65	--	1	0.880
		3b	EG: Woman with gun	64	--	1	3.400
		3c	CG: Man with gun	64	--	1	0.048
		3d	EG: Woman with gun	64	--	1	0.128
Pickel et al.	2006	1	19th-century pistol	61	--	1	1.129

Table 8

Mean Weighted Effect Sizes for the Unusualness Effect for Correct Identifications in TP Lineups and Accuracy of Target Descriptions

DV	Model	*k*	*N*	*LOR*	*OR*	95% CI		g_u	*Z*	p_Z	*VC*	*Q*	p_Q	I^2
						LL	*UL*							
Correct Identifications in TP Lineups	FEM	6	465	0.284	1.329	0.884	1.997		1.369	.171		4.36	.499	.147
	REM	6	465	0.277	1.319	0.918	1.894		1.498	.134	.037			
Target Description Accuracy	FEM	16	1102			0.639	0.884	0.761	12.184	<.001		50.77	<.001	0.705
	REM	16	1102			0.569	1.026	0.798	6.837	<.001	0.150			

Note. LOR = logged odds ratio; *OR* = odds ratio, g_u = Hedges g_u; *CI* = confidence interval; *LL* = lower limit, *UL* = upper limit; *Z* = significance test; p_Z = significance level for *Z*-test; *VC* = variance component in REM; *Q* = heterogeneity test statistic; p_Q = significance level for *Q*-test; I^2 = indicator of heterogeneity.

A FEM analysis resulted in a nonsignificant small effect, *OR* = 1.329, 95% CI [0.884, 1.997] (see Table 7), demonstrating that participants in the control group had only a 1.33 times better chance to identify the target in a later lineup compared to participants in the unusual condition. Results for the REM were similar.[3]

Description accuracy of the target. For the 16 comparisons examining accuracy of target-descriptions, outlier analyses revealed one extreme effect size (Pickel, 2009, Exp. 3, No Stereotype condition; cf. Figure 4). Even after this value (as well as one value at the other end of the distribution) were winsorized to a more moderate level, homogeneity tests still indicated large variability, $Q(15) = 50.78$, $p < .001$, $I^2 = 0.705$.

A FEM analysis showed a significant effect for unusualness on description accuracy, $g_u = 0.761$, 95% CI [0.639, 0.884], $Z = 12.18$; $p < .001$, $k = 16$ (see Table 7). As predicted, participants in the control condition were better able to describe the target compared to participants watching a target with an unusual object.

Moderator Analyses. Because some studies used weapons presented in an unusual way we calculated a blocked moderator analysis with weapon presence as a classification variable for description accuracy. There was no significant effect of weapon presence on target description accuracy, $Q_{between}(1) = 0.63$, $p = .423$. Studies using a weapon (e.g., a gun wrapped with a fluorescent yellow band), produced an effect size, $g_u = 0.805$, 95% CI [0.642, 0.968], $Z = 9.67$; $p < .001$, $k = 9$, to those studies varying unusualness by using unusual objects (e.g., a plastic flamingo), $g_u = 0.705$, 95% CI [0.520, 0.890], $Z = 7.46$; $p < .001$, $k = 7$; $Q_{total}(15) = 50.78$, $p < .001$; $Q_{within}(14) = 50.15$, $p < .001$.

---

[3] When the two unpublished hypothesis tests by Mansour et al. (2008, Exp.1 and 2) were excluded, results remained virtually identical (*OR* = 1.33, 95% CI [0.871, 2.038].

Studies involving an expectancy violation (e.g., a priest vs. police officer with a gun) resulted in a higher mean effect size, g_u = 1.050, 95% CI [0.833, 1.267], Z = 9.48; $p < .001$, k = 5, than other studies, g_u = 0.627, 95% CI [0.478, 0.775], Z = 8.28; $p < .001$, k = 11; $Q_{between}(1) = 9.96$, p = .002; $Q_{total}(15) = 50.78$, $p < .001$; $Q_{within}(14)$ = 40.82, $p < .001$.

A meta-regression analysis with both variables entered simultaneously confirmed these findings, demonstrating a significant effect for expectancy violation, *beta* = .467, p = .002, but not for weapon presence, *beta* = -.064, p = .670.

Publication Bias

Last but not least, we briefly address the problem of publication bias. Despite our appreciation of Rosenthal's (1979) pioneering work to point out the file-drawer problem, we refrain from calculating and reporting a fail-safe number due to the recent criticisms of this index (Becker, 2005; Sutton, 2009). We have made an extensive effort to locate unpublished studies (usually conference presentations) and have carefully scrutinized them for potential methodological flaws. Furthermore, we have conducted sensitivity analyses to check whether or not our conclusions would change with or without unpublished studies included. As indicated above, the few unpublished studies included did not substantially change any of the conclusions drawn. As an additional check, we plotted the effect sizes g_u against the standard error of g_u in Figure 5. Rather than using traditional funnel plots with effect sizes plotted against sample size, we used the standard error on the Y-axis as recommended in the recent literature (see Sterne & Egger, 2001; Sterne, Becker, & Egger, 2005; Sutton, 2009).

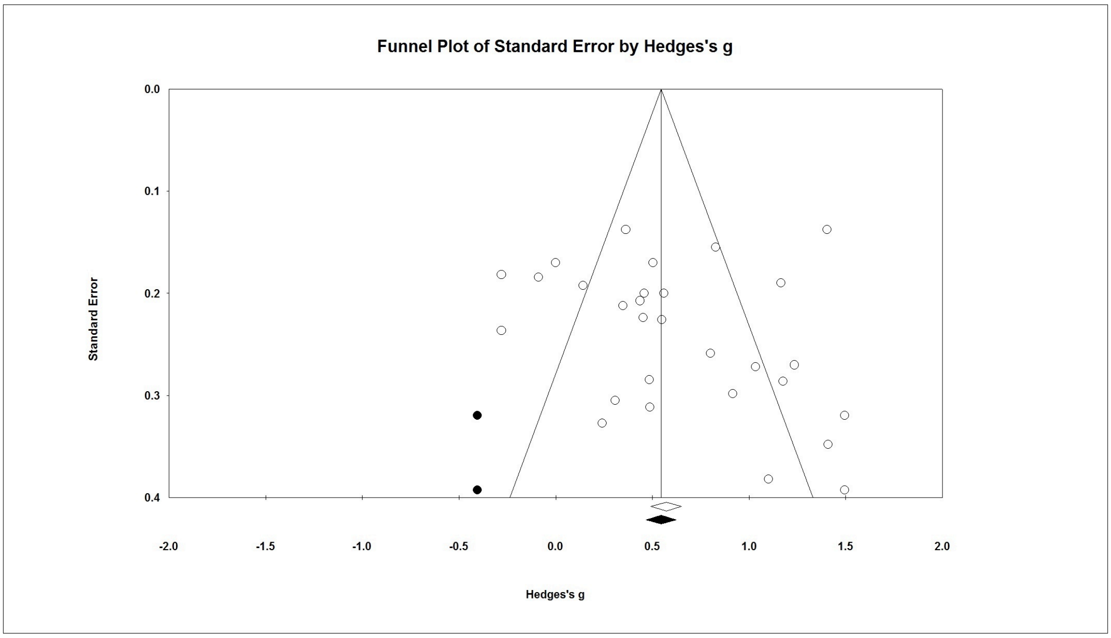

Figure 5. Funnel plot of standardized mean difference (g_u) of accuracy of perpetrator descriptions and the standard error.

For the accuracy of perpetrator description studies (*k* = 29), the (slight) asymmetry of the funnel plot in Figure 5 gives the impression of a negative relationship between standard error and effect size. In other words, only when the standard error was rather small (due to larger sample sizes),were small, or even negative, effects more likely to have been reported in the literature and could effect sizes be calculated. Conversely, smaller studies with larger standard errors showed larger WFEs (at the bottom right of the graph). While it should be clear that perception and interpretation of funnel asymmetry is highly subjective and not conclusive evidence for the presence of publication bias (Sterne et al., 2005; Sutton, 2009), publication bias ought to be considered as a possible alternative explanation here. Hence, we employed more formal statistical tests, using the software *Comprehensive Meta-Analysis 2* (see Borenstein, 2005).

Begg and Mazumdar (1994; see also Begg, 1994; Sterne et al., 2005; Sutton, 2009) have proposed a rank correlation method to assess a potential publication bias. For the accuracy of perpetrator descriptions, Kendall's *tau* (without continuity correction) was .249, Z = 1.89, p = .058 (two-tailed). As this test has rather low power, Begg (1994) recommends using it with a "very liberal significance level" (p. 403).

We also employed the Egger, Smith, Schneider, and Minder (1997) regression method which is considered to be somewhat more likely to detect "small study effects" (for a critical discussion, see Sterne & Egger, 2005; Sutton, 2009). Surprisingly, Egger's linear regression method did not yield a significant effect, intercept = 1.538 [95% CI: -1.694, 4.769], SE = 1.575, $t(27)$ = 0.98, p = .337, although this test is considered more powerful.

We also applied the trim and fill method (Duval & Tweedie, 2000a, 2000b; see Duval, 2005) to the perpetrator description data, which resulted in only two studies that would have to be added to the left side of the distribution. The trim and fill method adds effect sizes to asymmetrical funnel plots and, by an iterative procedure, reestimates the overall effect size from the resulting more symmetrical distribution. This new estimate resulted in an overall effect size of g_u = 0.546, 95% CI [0.468, 0.623], which is only slightly lower than the original observed effect size, g_u = 0.568, 95% CI [0.490, 0.647].

Thus, there is some concern that the data may have been affected by a publication bias. However, these more formal tests do not necessarily indicate that publication bias may have been a serious problem in this meta-analytic literature. In fact, quite a few studies were published despite their results being nonsignificant, particularly regarding identifications (see Figure 1). Similarly, the studies reviewed included several nonsignificant findings regarding the WFE effect on perpetrator descriptions (see Figure 2), so we can be reasonably confident that these studies represent reliable effects.

A funnel plot for the unusualness studies yielded a more symmetrical distribution. Parallel analyses with the unusualness studies (k = 16) resulted in a Kendall's *tau* = .271, Z = 1.44, p = .150, for the Begg-Mazumdar test, and also a nonsignificant effect for Egger's linear regression method, intercept = 3.565 [95% CI:

-2.139, 9.269], *SE* = 2.656, 95% CI: -2139, *t*(14) = 1.34, *p* = .201 (two-tailed). Duval and Tweedie's trim and fill method indicated no values to be trimmed. Although the number of available studies may be considered too small for these formal tests to be conclusive, the unusualness studies seem even less likely to have been affected by publication bias than the traditional WFE hypothesis tests.

Space limitations prevent us from modeling the potential for publication bias by other recent modeling techniques (Rothstein et al., 2005; Sutton, 2009), which might be more sensitive and yield more conclusive evidence. However, these techniques are still evolving and should be consulted in future meta-analyses, in particular with larger data sets.

Discussion

The goal of the present meta-analysis was not only to bring an earlier meta-analysis by Steblay (1992) up to date by including newer studies conducted in the last 20 years but also to take advantage of contemporary meta-analytic methods (see Borenstein, Hedges, Higgins, & Rothstein, 2009; Lipsey & Wilson, 2001; Sporer & Cohn, 2011; Valentine et al., 2009). In particular, we choose the odds-ratio (rather than Cohen's *h*) for dichotomous dependent variables which is widely endorsed in the recent meta-analytic literature as the effect size of choice (Borenstein et al., 2009; Fleiss & Berlin, 2009; Fleiss, Levin, & Paik, 2003,). For continuous dependent variables, we used Cohen's *d* (adjusted for small sample bias; for a similar strategy in analyzing dichotomous and continuous variables, see Loesel & Schmucker, 2005; Pirelli, Gottdiener, & Zapf, 2011). Secondly, we meticulously checked for outliers via graphical and via statistical methods (adapting the methods advocated by Hedges & Olkin, 1985) and tested for homogeneity before interpreting any mean effect sizes (which would be meaningless if results are clearly heterogeneous). To further guard against premature conclusions we calculated both fixed-effects and random-effects models for all our analyses. Besides reporting a limited number of comparisons of subgroups of studies ("blocking"), which are likely to be confounded with multiple other variables coded (or not coded), we employed (hierarchical) meta-regression techniques (Konstantopoulos & Hedges, 2009; Lipsey & Wilson, 2001; Pigott, 2012) to test for a limited number of moderator variables.

We also introduced and discussed newer methods to assess publication bias which are considered more appropriate than Rosenthal's (1979) fail-safe *N* (Rothstein, Sutton, & Borenstein, 2005; Sporer & Goodman-Delahunty, 2011; Sutton, 2009). Thus, one of the goals of the current meta-analytic review is to introduce newer meta-analytic procedures to the eyewitness domain (see also Meissner et al., 2008). Last but not least, we attempted to compare rival theoretical approaches by conducting separate meta-analyses for traditional WFE studies and newer studies focusing on the unusualness of an object as a rival theoretical explanation of attentional focus. These analyses shed light on possible explanatory mechanisms of the WFE.

In the following, we try to answer the following questions: (1) Is the WFE a robust phenomenon across a variety of situations? (2) How can the effect be explained? (3) What practical implications does the effect have for crime investigators and for court experts? (4) To what extent can a publication bias be ruled out as an alternative explanation?

Robustness of the Weapon Focus Effect

Overall, the existence of a medium to large WFE was demonstrated in the current meta-analysis only for accuracy of descriptions of the perpetrator but not for identification decisions. For lineup identifications, there was neither a significant effect for correct identifications in target-present lineups, nor for overall correct decisions. This general pattern of results which is based on a larger database than previous reviews, is only partially consistent with the earlier meta-analysis by Steblay (1992). Steblay (1992) had similarly observed a significant WFE for descriptions as a dependent variable. However, she had also concluded that there is a significant effect for person identifications which we could not confirm. Although we used odds-ratios rather than *h* as an effect size indicator, the reader should note that the odds-ratio is considered a more sensitive measure.

Somewhat different from our conclusions, the recent meta-analysis on the WFE by Fawcett Russell, Peace, and Christie (2011) produced a small, significant effect on identification accuracy, and, similar to our results, a medium size effect on description accuracy. Fawcett et al. (2011) included laboratory and simulation studies as well as archival analyses in calculating effect sizes. We decided not to include results of archival analyses here because it believe it not possible to determine accuracy of person identification or descriptions in these studies. However, we do discuss outcomes of archival studies below. Differences from Fawcett et al.'s and our results might arise from aggregations across different dependent measures and/ or study types without testing for differences within the set of studies. For example, these authors calculated a weighted effect size for identification accuracy including all possible identification outcomes. We consider these and other aggregations across target-presence, and across recognition and recall measures as problematic.

One of the main problems in this as well as other meta-analyses in the eyewitness domain is the wide variation between studies included. While *Q* tests may not always detect these discrepancies due to lack of power (Pigott, 2012), a notably large heterogeneity between studies remained for description accuracy, even after systematically accounting for some of the variance by moderator analyses. One possible explanation for this heterogeneity could be different methods employed to assess target descriptions accuracy. Some studies used free recall whereas others used multiple-choice formats. However, our preliminary analyses showed that type of measure used was not a reliable moderator (except when using correct details only without correcting for incorrect details[5]). Moreover, the content of the questions asked could also have differed between studies. We had to pool such diverse operationalizations, for example, the number of (correct) descriptors and proportion

---

[5] Note that Pickel (2009) reported both results for correct and for incorrect details but we only used the former in our analyses.

measures of accuracy (usually defined as number of correct plus number of incorrect divided by total number of descriptors provided; see Sporer, 1996, 2008) to have a large enough data set to analyze (cf. also Meissner, Sporer, & Susa, 2008, for a similar method). The different methods used in measuring description accuracy may also differ in terms of reliability of measurement (unfortunately, most authors did not report any indices of inter-coder reliability). Higher inter-coder reliabilities have been shown to moderate the association between description accuracy and identification accuracy in a recent meta-analysis by Meissner et al. (2008). Perhaps stronger effects would be obtained here as well if completeness and accuracy of descriptions had been measured with more elaborate methods.

Another reason for the variation in outcomes could be that authors were investigating other variables besides the WFE in their experiments. For example, some studies manipulated level of arousal (Hulse & Memon, 2006), while others simply presented a target with a weapon (Kramer et al., 1990). In the studies by Cutler and colleagues, context reinstatement procedures (Cutler, Penrod, & Martens, 1978b) were introduced, or the robber's disguise (O'Rourke et al., 1989) was varied. These different manipulations make it difficult to isolate the emergence of the WFE. Christianson (1992) criticized the variation across WFE laboratory studies, and this variation may have contributed to the heterogeneity of findings. On the other hand, real cases also differ with respect to many different situational factors, so variation in laboratory simulations should not necessarily be considered a bad thing. It is considered a strength of meta-analyses to demonstrate an effect across a variety of conditions.

Theoretical Implications

Focus of attention. Better recall results for weapons compared to neutral objects carried by the target person may be a consequence of more frequent and longer fixation times on the former (for direct evidence on this point, see E. F. Loftus et al., 1987). Thus, we can infer that a weapon is a central detail of the scene that attracts the attention of an eyewitness whereas peripheral details like the face of the perpetrator are paid less attention to. Also in support of this conclusion, when a weapon (or other unusual object) was present, a few studies indicated that the weapon (or object or hand area) was more accurately described than the face.

Another possible interpretation of this finding might be that weapons are more interesting objects and therefore less forgettable compared to neutral things like a wallet (Pickel, 1998). As Christianson and Loftus (1991) pointed out arousing events as well as distinctive or unusual events often produce *Von Restorff*-like data where an item, which stands out, is more likely to be remembered than other items. Whether or not a weapon is either arousing or unusual is discussed below.

Arousal. Only one of the studies included (with police officers as participants) measured participants' heart rate while observing the crime (Hulse & Memon, 2006). Although the authors found significant differences in self-reported arousal for the shooting scenario compared to the non-shooting scenario, d = 0.61, no such differences were observed for heart rate, d = 0.14. This null-finding for heart rate could be ascribed to the use of a weapon in both conditions. Another possible explanation could be that police officers who served as participants are used to weapons, thus not being affected as much as non-police officers (unfortunately, there was no civilian control group). Moreover, the results of other studies demonstrate that self-report may not be an appropriate measure for arousal levels experienced by observers (of videotaped crimes). No other physiological indices were used in any of the studies. One possible explanation might be that we are more used to guns presented in the media. Future studies should look into media consumption and being accustomed to weapons as potential moderators.

Because arousal rates were too infrequently reported to be used as a possible moderator we tried to indirectly gauge the likelihood of arousal; viz., we expected the type of weapon employed to invoke different levels of arousal. Specifically, we expected a gun to lead to a larger WFE compared to other weapons (e.g., a knife or a meat cleaver) because the former appears more dangerous as one can be injured from a larger distance. Contrary to expectations, the WFE was smaller in the gun condition compared to other types of weapons (like a knife).

Post hoc one may speculate that injuries caused by a knife or meat cleaver might be expected to end more bloodily and therefore could have been experienced as more scary by an observer. Due to these expectations participants may potentially feel more tense in anticipation of a bloody event. This tension, in turn, might lead to a larger WFE. Here it is important to distinguish between experienced threat and experienced tension. Presumably, both are leading to heightened arousal levels but in real cases arousal should be more intense. In experimental studies, only the actor in the film is threatened but not the participant viewing the scene who may only feel slightly aroused by the action displayed. The few studies that assessed arousal levels via self-report only observed low to moderate levels. Apparently, participants are not placing themselves in the role of the threatened victim in the film. Rather than using self-report, future studies should attempt to measure arousal levels more directly with physiological measures like heart rates or Galvanic skin responses.

Several studies not part of this current analysis due to differing dependent variables used an electroencephalogram as a physiological measure and demonstrated that the sharpness of knives lead to shorter peak latencies of P300 and faster reaction times at recognition compared to more blunt stimulus knives (Hamamoto & Hira, 2007, 2009; Oue, Hakoda, & Onuma, 2008). These authors argue that the sharpness of a weapon may be one of the factors contributing to the occurrence of the WFE. Extrapolating from this assumption, studies using a knife as a weapon should produce a stronger WFE compared to those using a handgun. This appears to be in line with our meta-analytical results.

Last it should be clear that due to obvious legal and ethical constraints all simulation studies will be a far cry from the potentially extreme levels of arousal and fear real crime victims are likely to experience (see also Deffenbacher et al., 2004, for a thorough discussion of stress and arousal regarding eyewitness testimony, in particular the role of orienting and defensive responses). Hence, in our view, none of the studies included in this review could really test main effects due to arousal, nor interactions between arousal and the presence of a weapon. To the extent that the WFE and arousal are additive effects, or interact with arousal functioning as a catalyst exacerbating the effect, the laboratory studies reviewed here are likely to underestimate the effect the presence of a weapon may have on real witnesses whose lives are threatened (but see Hulse & Memon, 2006). Archival studies, despite their shortcomings regarding the ground truth of identification accuracy and potential selection effects, should be consulted as complementary evidence (e.g., Sporer, 1992, 2008; Tollestrup, Turtle, & Yuille, 1994).

For example, in Behrman and Davey's (2001) archival study, the suspect was chosen in a lineup approximately equally often in the weapon-present as in the weapon-absent group, even in cases where there was additional evidence incriminating the suspect. Similar null-findings were observed for field showups. No differentiated analysis of lineup rejections and foil identifications was possible as many reports did not make the distinction sufficiently clear. Valentine, Pickering, and Darling (2003) observed fewer filler identifications (15.9%) when a weapon was present than when no weapon was involved (23.7%). Suspect identifications were as high (42.0%) when a weapon was present than when when no weapon was present (40.2%).These data would also support our conclusion that there is little evidence both in the laboratory and the field that the presence of a weapon negatively affects identification evidence. In the analysis of police records by Sporer (1992), person descriptions were nonsignificantly longer when a weapon was present than in cases without a weapon.

Unusualness. The unusualness approach follows the assumption that the WFE is not a function of level of threat but a consequence of the fact that a weapon is unexpected or out of context in a certain situation. Results of the current synthesis demonstrated an effect comparable in magnitude to studies investigating the traditional WFE. Studies manipulating both the presence of a weapon and unusualness did not differ from each other. Thus, the results would seem to indicate that the effect of the presence of a weapon can be explained by it being unexpected or unusual. Further support for this explanation comes from studies where expectancy violation per se was manipulated which showed larger effects (e.g., when a woman as opposed to a man was carrying a gun) than when the object (gun) was more in line with expectations. However, the psychological situation in real crime situations involving a weapon may still be different for a potential victim (or bystander) than would be invoked by the presence of unusual objects.

Here, it would be important to further investigate if unusual objects are also better remembered than neutral objects as an explanation for attentional focus. For example, Pickel (1998) was able to demonstrate a better memory for unusual (98%) compared to neutral objects (45%) but memory for the weapon was just as high (100%). Unfortunately, there were too few studies to calculate appropriate effect size models to test this notion in our meta-analysis.

From a theoretical point of view, in our opinion one also needs to distinguish between unusual and unexpected events (see also Boyce, Pollatsek, & Rayner, 1989; DeGraef et al., 1990; Friedman, 1979). A crime is unexpected rather than unusual in the sense that we are aware of the frequencies with which robberies occur in banks or stores but a crime would still be unexpected in a concrete situation in this context (Pickel, 1998). In support of this notion, Hope and Wright (2007) found that participants viewing a crime involving a weapon were impaired more than those in an unusual and a control condition, despite interpreting the situation to be in line with a robbery script. On the other hand, returning to Christianson's (1992) definition of a negative emotional event, these types of events are not only threatening but *new* and *unexpected*. Thus, it seems that more than one factor is responsible for the WFE due to the multi-dimensionality of negative events of which a crime is only a specific instance.

Practical Implications

One of the major purposes of meta-analyses is to inform policy makers and decision makers to evaluate eyewitness identifications to arrive at better evidence-based decisions. While other estimator variables like the duration of an event are difficult to estimate ex post facto (cf., Sporer, 1996; Sauerland & Sporer, 2009), witnesses are unlikely to misreport the presence of a weapon (as shown by the ceiling effects for object naming in most studies reviewed here). Therefore, it is important for fact finders like police officers, attorneys, juries and judges, as well as psychologists who testify on eyewitness issues in court, to know whether the presence of a weapon will make witnesses less reliable than those involved in crimes with no weapon present.

In terms of identification accuracy, the studies reviewed demonstrate that eyewitnesses observing a crime with a weapon are *not* less likely to correctly identify the perpetrator in a perpetrator-present lineup than when no weapon is present (*OR* = 1.29), nor are overall identification decisions affected by the presence of a weapon (*OR* = 1.10).

When witnesses did arrive at an incorrect decision, they either incorrectly rejected the lineup or picked a filler person. The latter decision generally has no consequences as in police lineups fillers are presumably innocent. In both cases, the real perpetrator remains at large. While police and state attorneys need to be aware of this to plan further investigations, witnesses rejecting a lineup or identifying a filler are not likely to testify in court. Hence, it is unlikely that an expert for the defense will be called upon in court when a witness rejects a lineup or identifies a filler.

As many experts testify for the defense to prevent a (presumably) innocent defendant from being wrongfully convicted, the crucial question appears to be what effect the presence of a weapon has on perpetrator-absent lineups. Unfortunately, most studies did either not use target-absent lineups, or simply reported results averaged across both TP and TA lineups ("correct decisions"). There were only four hypothesis tests of studies that reported false identifications in TA lineups as a function of manipulated WFE (see Table 3). The results were completely contradictory, with two tests reporting an increase, and two a decrease in false identifications, with odds ratios ranging from 0.13 to 2.34, making the calculation of an average effect size not only meaningless but statistically inappropriate. Even if we ignore the unpublished study by Bothwell (1991), who used showups rather than lineups, the remaining two studies completely contradict each other (Hulse & Memon, 2006; Maass & Koehnken, 1989). Consequently, no conclusions can be drawn regarding the WFE on perpetrator-absent lineups, and future research is clearly needed to remedy this deficit. This lack of knowledge also implies that there seems not to be sufficient evidence for experts testifying for the defense on the WFE with respect to identification decisions, as this type of expert testimony typically focuses on factors contributing to the likelihood of false identifications.

However, the fact that the WFE for descriptions of the perpetrator was much larger than for identification outcomes has additional implications beyond the ones discussed with respect to identifications. For one thing, being less able to correctly describe details of a perpetrator could mislead crime investigators as incorrect information could possibly lead to an arrest of an innocent person who happens to match the details described. Police detectives may also prematurely stop searching for alternative suspects once they have found one who matches the description given. Consequently, the true perpetrator may remain at large and escape prosecution.

Being better able to name the weapon correctly (as opposed to a control object) appears to be a rather trivial effect which should only be interpreted as a manipulation check. On the other hand, being able to describe a weapon with more details (and perhaps more accurately) has practical implications which as far as we know have never been discussed before. If a suspect gets arrested and a weapon is found that matches the witness's description, this may provide important incriminating evidence in a case (see Behrman & Davey, 2001). Furthermore, a weapon lineup based on the weapon description could be conducted which could add valuable incriminating (or exonerating) evidence in the case (see also Sauerland & Sporer, 2008, and Sauerland, Sagana, & Sporer, 2011, on the diagnosticity of object identifications).

Finally, we briefly address the possibility to prevent a WFE. Perhaps, particular professional groups like police officers, military, war veterans or security personnel as well as special interest groups accustomed to weapons (e.g., members of a shooting club, hunters) may be less susceptible to a WFE. It may also be possible to train people not to be affected as much by the presence of a weapon. A few studies have demonstrated that the WFE can be reduced when witnesses are warned beforehand (Pickel et al., 2003)--although these studies do not really simulate the threat witnesses may experience in real crime situations. Due to obvious ethical and legal constraints, there are no studies where witnesses were personally threatened. Furthermore, practically all studies used a (camera) perspective of a bystander witness, not that of a victim being attacked.[6]

Perhaps, information booklets for store owners, bank clerks or other persons likely to be victimized could be designed to instruct these high risk groups not only to attempt to minimize conflict but also about the WFE and the necessity to attempt to gain information about the perpetrator that may later help to pursue him or her.

Conclusion

In conclusion, the reported medium to large effect sizes for perpetrator/target descriptions appear reliable evidence for the existence of a WFE. Data from the unusualness studies indicate that this effect may be sufficiently explained by witnesses refocusing attention away from the perpetrator to an unusual object/weapon. The large heterogeneity of outcomes in both the WFE and unusualness studies clearly indicates that these robust effects may be affected by situational circumstances that have to be taken into consideration. There is no evidence, however, that perpetrator identifications are affected by the WFE. In

[6] We are indebted to Colin G. Tredoux for making us aware of this distinction.

particular, there is a dearth of studies testing the WFE with both perpetrator-present and perpetrator-absent lineups. We urge researchers to compare outcomes of laboratory studies with data from archival analyses, which may or may not yield comparable results. Archival analyses also did not find a WFE for suspect identifications, thus supporting the simulation studies reviewed here.

References

* = primary study included in the current meta-analysis; a = correct identification of the target in WFE studies; b = correct identification decision across target-present and target–absent lineups (WFE); c = false identification in target-absent lineups (WFE); d = description accuracy of the target (WFE); e = incorrect description of the target (WFE); f = correct naming of the object held by the target (WFE); g = correct object description (WFE); h = correct identification of the target in unusualness studies; i = description accuracy of the target in unusualness studies.

Becker, B. J. (2005). Failsafe N or file-drawer number. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis. Prevention, assessment and adjustments* (pp. 111-125). New York: Wiley.

Begg, C. B. (1994). Publication bias. In H. Cooper, & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 399-409). New York: Russell Sage Foundation.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50,* 1088-1101. doi: 10.2307/2533446

Behrman, B. W., & Davey, S. L. (2001). Eyewitness identification in actual criminal cases: An archival analysis. *Law and Human Behavior, 25,* 475-491. doi:10.1023/A:1012840831846

Behrman, B. W., & Richards, R. E. (2005). Suspect/foil identification in actual crimes and in the laboratory: A reality monitoring analysis. *Law and Human Behavior, 29,* 279-301. doi:10.1007/s10979-005-3617-y

Borenstein, M. (2005). Software for publication bias. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis. Prevention, assessment and adjustments* (pp. 193-220). New York: Wiley.

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 221-236). New York, NY: Russell Sage Foundation.

Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, R. (2009). *Introduction to meta-analysis.* Chichester, UK: Wiley.

*Bothwell, R. K. (1991). *Eyewitness description accuracy.* Paper presented at the Meeting of South-Eastern Psychological Association, New Orleans, LA. [a, b, c]

Boyce, S. J., Pollatsek, A., & Rayner, K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 556–566.

Christianson, S.-A. (1992). Emotional stress and eyewitness memory: A critical review. *Psychological Bulletin, 112*, 284-309. doi: 10.1037/0033-2909.112.2.284.

Christianson, S.-A., & Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Cognition and Emotion, 5*, 91-108. doi: *10.1080/02699939108411027*

Christianson, S.-A., Loftus, E. F., Hoffman, H., & Loftus, G. R. (1991). Eye fixation and memory for emotional events. *Journal of Experimental Psychology: Learning Memory and Cognition, 17*, 693-701. doi: *10.1037/0278-7393.17.4.693*

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

*Cutler, B. L., & Penrod, S. D. (1988). Improving the reliability of eyewitness identification: Lineup construction and presentation. *Journal of Applied Psychology, 73,* 281-290. doi: *10.1037/0021-9010.73.2.281* [b]

*Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987a). The reliability of eyewitness identification: The role of system and estimator variables. *Law and Human Behavior, 11,* 223-258. doi: 10.1177/0146167290161008 [b]

*Cutler, B. L., Penrod, S. D., & Martens, T. K. (1987b). Improving the reliability of eyewitness identification: Putting context into context. *Journal of Applied Psychology, 72,* 629-637. doi: 10.1037/0021-9010.72.4.629 [b]

*Cutler, B. L., Penrod, S. D., O'Rourke, T. E., & Martens, T. K. (1986). Unconfounding the effects of contextual cues on eyewitness identification accuracy. *Social Behavior, 1,* 113-134. [b]

Davies, G. M., Smith, S., & Blincoe, C. (2008). A "weapon focus" effect in children. *Psychology, Crime & Law, 14,* 19-28. doi: 10.1080/10683160701340593

De Graef, P., Christiaens, D., & Dydewalle, G. (1990). Perceptual effects of scene context on object identification. *Psychologische Forschung* [Psychological Research]*, 52,* 317-329.

Deffenbacher, K. A. (1983). Identification evidence: A psychological evaluation. *American Journal of Psychology, 96,* 591-595. doi: 10.1007/BF00868064

Deffenbacher, K. A. (1991). A maturing of research on the behavior of eyewitnesses. *Applied Cognitive Psychology, 5,* 377-402. doi: 1 0.1002/acp.2350050502

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28,* 687-706. doi: 10.1007/s10979-004-0565-x

Detterman, D. K. (1975). The Von Restorff effect and induced amnesia: Production by manipulation of sound intensity. *Journal of Experimental Psychology: Human Learning and Memory, 1,* 614-628. doi: *10.1037/0278-7393.1.5.614*

Duval, S. J. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments* (pp. 127–144). Chichester, UK: Wiley.

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,* 89-98. doi: 10.2307/ 2669529

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,* 455-463. doi: 10.1111/j.0006-341X.2000.00455.x

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66,* 183-201. doi: *10.1037/h0047707*

Egger, M., Smith, G. D., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *BMJ, 315,* 629-634. doi: 10.1136/bmj.315.7109.629

Fleiss, J. L., & Berlin, J. A. (2009). Effect sizes for dichotomous data. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 237-256). New York, NY: Russell Sage Foundation.

Fleiss, J. L., Levin, B., & Paik, M. C. (2003). *Statistical Methods for Rates and Proportions* (3rd ed.). New York: John Wiley.

Friedman, A. (1979). Framing pictures: The role of knowledge in automatized encoding and memory for gist. *Journal of Experimental Psychology: General, 108,* 316-355. doi: *10.1037/0096-3445.108.3.316*

Gleser, L. J., & Olkin, I. (2009). Stochastically dependent effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 357-376). New York: Russell Sage.

Greenhouse, J. B., & Iyengar, S. (2009). Sensitivity analysis and diagnostics. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 417-434). New York, NY: Russell Sage Foundation.

Grissom, R. J., & Kim, J. J. (2005). *Effect sizes for research: A broad practical approach*. Mahwah, NJ: Lawrence Erlbaum.

Hamamato, Y., & Hira, S. (2007, October). *The effect of the sharpness of a knife on weapon focus: A P300 study.* Paper presented at the 47th Annual Meeting of the Society for Psychophysiological Research in Savannah, Georgia, USA.

Hamamato, Y., & Hira, S. (2009, July). *The relationship between weapon focus effect and the physical features of weapons: A P300 study.* Paper presented at the 8th Biennial meeting of the Society for Applied Research in Memory and Cognition in Kyoto, Japan.

Harwell, M. (1997). An empirical study of Hedges's homogeneity test. *Psychological Methods, 2,* 219-231.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V., & Pigott, M. A. (2001). The power of statistical tests in meta-analysis. *Psychological methods, 6,* 203-217. doi: *10.1037/1082-989X.6.3.203*

Henderson, J. M., Week, P. A., Jr., & Hollingworth, A. (1999). The effects of semantic consistency on eye movements during complex scene viewing. *Journal of Experimental Psychology: Human Perception and Performance, 25,* 210-228. doi: 10.1037/0096-1523.25.1.210

Hepburn, L., Miller, M., Azarel, D., & Hemenway, D. (2007). The US gun stock: Results from the 2004 national firearms survey. *Injury Prevention, 13,* 15-19. doi: 10.1136/ip. 2006.013607

Heuer, F., & Reisberg, D. (1990). Vivid memories of emotional events: The accuracy of remembered minutiae. *Memory & Cognition, 18,* 496-506. doi: 10.3758/BF03198482

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539-1558. doi: 10.1002/sim.1186

*Hope, L., & Wright, D. (2007). Beyond unusual? Examining the role of attention in the weapon focus effect. *Applied Cognitive Psychology, 21,* 951-961. doi: 10.1002/acp.1307 [d, f, g, i]

*Hulse, L. M., & Memon, A. (2006). Fatal impact? The effects of emotional arousal and weapon presence on officers' memories for a simulated crime. *Legal and Criminological Psychology, 11,* 313-325. doi: 10.1348/135532505X58062 [c, d]

Ibabe, I., & Sporer, S. L. (2004). How you ask is what you get: On the influence of question form on accuracy and confidence. *Applied Cognitive Psychology, 18,* 711-726. doi: 10.1002/acp.1025

Ikeda, R. M., Dahlberg, L. L., Kresnow, M.-J., Sacks, J. J., & Mercy, J. A. (2003). Studying "exposure" to firearms: Household ownership access. *Injury Prevention, 9,* 53-57. doi: 10.1136/ip.9.1.53

Jackson, D. (2006). The power of the standard test for the presence of heterogeneity in meta-analysis. *Statistics in Medicine, 25,* 2688-2699. doi: 10.1002/sim.2481

*Johnson, C., & Scott, B. (1976). *Eyewitness testimony and suspect identification as a function of arousal, sex of witness, and scheduling of interrogation.* Paper presented at the American Psychological Association Annual Meeting, Washington, DC. [a]

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,* 469-484. doi: 10.1023/A:1024871622490

Kassin, S. M., Tubb, V. A., Hosch, H. M., & Memon, A. (2001). On the "general acceptance" of eyewitness testimony research: A new survey of the experts. *American Psychologist, 56,* 405-416. doi: 10.1037//0003-066x.56.5.405

Kenrick, M., & Mallard, D. (2006). Impact of weapon focus and context reinstatement on eyewitness identifications. *Australian Journal of Psychology, 58,* 152. doi: 10.1080/00049530600940019

Konstantopoulos, S., & Hedges, L. V. (2009). Analyzing effect sizes: Fixed-effects models. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 279-293). New York, NY: Russell Sage Foundation.

*Kramer, T. H. (1990, September). [Personal Communication; as cited in Steblay, 1992]. [a (effect size *OR* calculated from Appendix B in Steblay, 1992]

*Kramer, T. H., Buckhout, R., & Eugenio, P. (1990). Weapon focus, arousal and eyewitness memory. *Law and Human Behavior, 14,* 167-184. doi: 10.1007/BF01062971 [a, d, f]

Latour, P. L. (1962). Visual threshold during eye movements. *Vision Research, 2,* 261-262. doi: *10.1016/0042-6989(62)90031-7*

Lipsey, M. W. (2003). Those confounded moderators in meta-analysis: Good, bad, and ugly. *The Annals of the American Academy of Political and Social Science, 587,* 69-81. doi: 10.1177/0002716202250791

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. *American Psychologist, 48,* 1181-1201. doi: 10.1037/0003-066X.48.12.1181

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-Analysis*. Thousand Oaks, CA: SAGE Publications.

Loesel, F., & Schmucker, M. (2005). The effectiveness of treatment for sexual offenders: A comprehensive meta-analysis. *Journal of Experimental Criminology, 1,* 117-146.

Loftus, E. F., & Burns, T. (1982). Mental shock can produce retrograde amnesia. *Memory & Cognition, 10,* 318-323. doi: 10.3758/BF03202423

*Loftus, E. F., Loftus, G. R., & Messo, J. (1987). Some facts about "weapon focus". *Law and Human Behavior, 11,* 55-62. Retrieved from http://www.jstor.org/stable/1393529 [a, d, f]

Loftus, G. R. (1972). Eye fixations and recognition memory for pictures. *Cognitive Psychology, 3,* 525-551. doi: 10.1016/0010-0285(72)90021-7

Loftus, G. R., & Mackworth, N. H. (1978). Cognitive determinants of fixation location during picture viewing. *Journal of Experimental Psychology, Human Perception and Performance, 4,* 565-572. doi: *10.1037/0096-1523.4.4.565*

Lorenz, J., & Yan, S. (2008, July). *Eyewitness identification: Cultural differences in suspectibility to weapon focus*. Paper presented at the The XXIX International Congress of Psychology in Berlin, Germany.

*Maass, A., & Köhnken, G. (1989). Eyewitness identification: Simulating the "weapon focus effect". *Law and Human Behavior, 13*, 397-408. doi: 10.1007/BF01056411 [c]

*Mansour, J. K., Lindsay, R. C. L., & Munhall, K. M. (2008, March). *The flamingo focus effect.* Paper presented at the American Psychology-Law Society, Jacksonville, FLA, USA. [h]

Sanchez-Meca, J., Chacón-Moscoso, S., & Marín-Martínez, F. (2003). Effect-size indices for dichotomized outcomes in meta-analysis. *Psychological Methods, 8,* 448-467. doi: 10.1037/1082-989X.8.4.448

Meissner, C. A., Sporer, S. L., & Susa, K. J. (2008). A theoretical review and meta-analysis of the description-identification relationship in memory for faces. *European Journal of Cognitive Psychology, 20,* 414-455. doi: 10.1080/09541440701728581

*Mitchell, K. J., Livorsky, M., & Mather, M. (1998). The weapon focus effect revisited: The role of novelty. *Legal and Criminological Psychology, 3,* 287-303. doi: 10.1111/j.2044-8333.1998.tb00367.x [d, i]

Neiss, R. (1988). Reconceptualizing arousal: Psychobiological states in motor performance. *Psychological Bulletin, 103,* 345-366. doi: *10.1037/0033-2909.103.3.345*

*O'Rourke, T. E., Penrod, S. D., Cutler, B. L., & Stuve, T. E. (1989). The external validity of eyewitness identification research: Generalizing across subject populations. *Law and Human Behavior, 13,* 385-395. Retrieved from h*ttp://www.ncjrs.gov/App/Publications/abstract.aspx?ID=121550* [b]

Orwin, R. G., & Vevea, J. L. (2009). Evaluating coding decisions. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 177-203). New York, NY: Russell Sage Foundation.

Oue, W., Hakoda, Y., & Onuma, N. (2008). The pointed shape of a knife influences eyewitness perception. *Asian Criminology, 3,* 193-200. doi: 10.1007/s11417-007-9043-y

*Pickel, K. L. (1998). Unusualness and threat as possible causes of "weapon focus". *Memory, 6,* 277-295. doi: *10.1080/741942361* [a, d, f, h, i]

*Pickel, K. L. (1999). The influence of context on the "weapon focus" effect. *Law and Human Behavior, 23,* 299-311. doi: 10.1023/A:1022356431375 [a, d, h, i]

*Pickel, K. L. (2009). The weapon focus effect on memory for female versus male perpetrators. *Memory, 17,* 664-678. doi: 10.1080/09658210903029412 [d, e, f, i]

*Pickel, K. L., French, T. A., & Betts, J. M. (2003). A cross-modal weapon focus effect: The influence of a weapon's presence on memory for auditory information. *Memory, 11,* 277-292. doi: 10.1080/09658210244000036 [b, d, f]

*Pickel, K. L., Narter, D. B., Jameson, M. M., & Lenhardt, T. T. (2008). The weapon focus effect in child eyewitnesses. *Psychology, Crime & Law, 14,* 61-72. doi: 10.1080/10683160701391307 [d, i]

*Pickel, K. L., Ross, S. J., & Truelove, R. S. (2006). Do weapons automatically capture attention? *Applied Cognitive Psychology, 20,* 871-893. doi: 10.1002/acp.1235 [d, e, g, i]

Pigott, T. D. (2009). Handling missing data. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 399-416). New York, NY: Russell Sage Foundation.

Pigott, T. D. (2012). Advances in meta-analysis. New York: Springer. doi:10.1007/978-1-4614-2278-5

Pirelli, G., Gotdiener, W. H., & Zapf, P. A. (2011). A meta-analytic review of competency to stand trial research. *Psychology, Public Policy, and Law, 17,* 1-53. doi: 10.1037/a0021713

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research* *synthesis and meta-analysis* (2nd ed., pp. 295-315). New York, NY: Russell sage Foundation.

Rothstein, H. R., Sutton, A. J., & Borenstein, M. (Eds.). (2005). *Publication bias in meta-analysis. Prevention, assessment, and adjustments.* Chichester, UK: Wiley.

Read, J. D., & Desmarais, S. L. (2009). Knowledge of eyewitness issues: A Canadian evaluation. *Applied Cognitive Psychology, 23,* 301-326.

Rosenthal, R. (1979). The "File Drawer Problem" and tolerance for null results. *Psychological Bulletin, 86,* 638-641.

Sauerland, M., & Sporer, S. L. (2008). The application of multiple lineups in a field study. *Psychology, Crime & Law, 14,* 549-564. doi: 10.1080/10683160801972519

Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology: Applied, 15,* 46-62. doi: 10.1037/a0014560

Sauerland, M., Sagana, A., & Sporer, S. L. (2011, in press). Assessing nonchoosers' eyewitness identification accuracy from photographic showups by using confidence and response times. Law and Human Behavior. doi: 10.1037/h0093926

Scheck, B., Neufeld, P., & Dwyer, J. (2001). *Actual Innocence*. Garden City, NY: Doubelday.

Shadish, W. R., & Haddock, C. K. (2009). Combining estimates of effect sizes. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 257-277). New York, NY: Russell Sage Foundation.

*Shaw, J. I., & Skolnick, P. (1999). Weapon focus and gender differences in eyewitness accuracy: Arousal versus salience. *Journal of Applied Social Psychology, 29,* 2328-2341. doi: 10.1111/j.1559-1816.1999.tb00113.x [d, g]

Sporer, S. L. (1982). A brief history of the psychology of testimony. *Current Psychological Reviews, 2,* 323-339.

Sporer, S. L. (1992, March). *Person descriptions in an archival analysis of criminal cases.* Paper presented at the Biennial Meeting of the American Psychology-Law Society in San Diego, California.

Sporer, S. L. (1996). Describing others: Psychological issues. In S. L. Sporer, R. S. Malpass, & G. Koehnken (Eds.), *Psychological issues in eyewitness identification* (pp. 53-86). Hillsdale, NJ: Lawrence Erlbaum.

Sporer, S. L. (2008). Lessons from the origins of eyewitness testimony research in Europe. *Applied Cognitive Psychology, 22,* 737-757. doi: 10.1002/acp.1479

Sporer, S. L., & Cohn, L. D. (2011). Meta-analysis. In B. Rosenfeld, & S. D. Penrod (Eds.), *Research methods in forensic psychology*. New York, NY: Wiley.

Sporer, S. L., & Goodman-Delahunty, J. (2011, March). Publication bias in meta-analyses in psychology and law. Paper presented at the Biennial Meeting of the American Psychology-Law Society in Miami, FL.

*Stanny, C. J., & Johnson, T. C. (2000). Effects of stress induced by a simulated shooting on recall by police and citizen witnesses. *American Journal of Psychology, 113,* 359-386. Retrieved from http://www.jstor.org/stable/1423364 [d]

Sterne, J. A. C., Becker, B. J., & Egger, M. (2005). The funnel plot. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis. Prevention, assessment and adjustments* (pp. 75-98). Chichester, UK: Wiley.

Sterne, J. A., & Egger, M. (2001). Funnel plots for detecting publication bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology, 54,* 1046-1055.

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein

(Eds.), *Publication bias in meta-analysis. Prevention, assessment and adjustments* (pp. 99-110). New York: Wiley.

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16,* 413-424. Retrieved from http://www.jstor.org/stable/1394272

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 435-452). New York, NY: Russell Sage Foundation.

Sutton, A. J., & Higgins, J. P. (2008). Recent developments in meta-analysis. *Statistics in Medicine, 27,* 625-650. doi: 10.1002/sim.2934

Thompson, S. G., & Higgins, J. P. T. (2002). How should meta-regression analyses be undertaken and interpreted? *Statistics in Medicine, 21,* 1559-1573. doi: 10.1002/sim.1187

Tollestrup, P. A., Turtle, J. W., & Yuille, J. C. (1994). Actual victims and witnesses to robbery and fraud. In D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), *Adult eyewitness memory* (pp. 144-160). Cambridge, UK: Cambridge University Press.

Tooley, V., Brigham, J. C., Maass, A., & Bothwell, R. K. (1987). Facial recognition: Weapon effect and attentional focus. *Journal of Applied Social Psychology, 17,* 845-859. doi: 10.1111/j.1559-1816.1987.tb00294.x

Tulving, E. (1969). Retrograde amnesia in free recall. *Science, 164,* 88-90. doi: 0.1126/science.164.3875.88

Tulving, E. (1985). How many systems are there? *American Psychologist, 40,* 385-398.

U.S. Department of Justice (2009). *Criminal victimization, 2008.* Washington, DC: Bureau of Justice Statistics Bulletin.

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology, 17,* 969-993. doi: 10.1002/acp.939

Van Koppen, P. J., & Lochun, S. K. (1997). Portraying perpetrators: The validity of offender descriptions by witnesses. *Law and Human Behavior, 21,* 661-685. Retrieved from http://www.jstor.org/stable/1394218

Von Restorff, H. (1933). Über die Wirkung von Bereichsbildungen im Spurenfeld [The effects of field formation in the trace field]. *Psychologische Forschung, 18,* 299-334. doi: 10.1007/BF02409636

Wallace, W. P. (1965). Review of the historical, empirical, and theoretical status of the Von Restorff phenomenon. *Psychological Bulletin, 63,* 410-424. doi: 10.1037/h0022001

Wells, G. L., Malpass, R. S., Lindsay, R. C., 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,* 581-598. Retrieved from *www.ncbi.nlm.nih.gov*

Wells, G. L., & Olson, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54,* 277-295. doi: 10.1146/annurev.psych.54.101601.145028

White, H. D. (2009). Scientific communication and literature retrieval. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 51-71). New York, NY: Russell Sage Foundation.

Wilson, D. B. (2009). Systematic coding. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of reearch synthesis and meta-analysis* (2nd ed., pp. 159-176). New York, NY: Russell Sage Foundation.

Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18,* 459-482. doi: 10.1002/cne.920180503

Appendices

Appendix A

Table A1
Outcomes of Lineup Decisions

Target Presence	Witness Choice		
	No Choice	Choice	
Absent	Correct Rejection	False ID of Suspect (False Alarm)	Foil ID (Known error)
Present	False Rejection	Correct ID (Hit)	Foil ID (Known error)

Appendix B

Correspondence with authors

Authors contacted for sending unpublished material:

Elizabeth Loftus, University of California, Irvine

Karen Mitchell, Yale University, New Haven, Conneticut

Marilyn Livosky, Mercyhurst College, Erie, Pennsylvania

Kerri Pickel, Ball State University, Muncie, Indiana

Graham Davies, University of Leicester, UK

Terry Beehr, Central Michigan University, Michigan

Lorraine Hope, University of Portsmouth, UK

Jamal. Mansour, Queens University, Kingston, Canada

Appendix C

Weapon Focus Effect Coding Sheet

Authors ____

Title ____

date ____________
year of publication

testnumber ____________

totaln ____________

PubState
1 = yes
2 = no

Country
1 = US
2 = UK
3 = Germany
4 = Canada
5 = combo
9 = unreported

WeapComp
weapon comparison design: experimental and control group

1 = weapon present vs. absent
2 = weapon present, vs. present but not visible (implied)
3 = weapon present vs. neutral object
4 = weapon present (longer vs. shorter duration)
5 = weapon present (pointed at witness vs. ground)

WeapCont
weapon in control condition
1 = no weapon in control condition
2 = weapon in control condition

UnusComp
unusual comparison design: experimental and control group

1 = unusual object present vs. absent
2 = unusual object vs. neutral object
3 = unusual object vs. weapon
4 = weapon in an unusual context vs. in a non-unusual context
5 = weapon used by unusual target (e.g., priest) vs. non-unusual target (e.g., police officer)
6 = unusual weapon type vs. neutral object
7 = weapon in an unusual appearance vs. neutral object

DepVar
dependent measure

1 = lineup ID
2 = face description accuracy
3 = both
4 = object recall
5 = object description

sSession ________________
time session/video/slides total in seconds

Time visible

sVisTarg ________________
time target was visible in seconds

sVisCat
time visible target categorized

1 ≤ 10 seconds
2 = 11-20 seconds
3 ≥ 20 seconds

sVisWeap ________________
time weapon was visible in seconds

sNoWeap ________________
time neutral object was visible in seconds

Mode
mode of presentation of stimulus material

1 = slides; with or without sound track
2 = video
3 = live

TypeCrim
type of crime/event presented

0 = no crime
1 = robbery
2 = assault
3 = threat (to inject with syringe)
4 = mixed; varied across conditions
5 = exchange overheard from next room
6 = classroom intruder
7 = intruder/stalker
9 = unreported
10 = domestic dispute

Arousal
arousal level reported

1 = low
2 = mod
3 = high
4 = mixed (crossed across conditions)
9 = unreported

Complex
complexity of the scene presented

1 = Complex crime scene or event
2 = stark environment

Delay
retention interval in hrs/days

0 = immediate
1 = brief retention interval < 1 hour
2 = 1 week (7 days)
4 = mixed across conditions

Sample
participant population

1 = undergraduates
2 = community adults; may include students
3 = children
4 = both children and adults
5 = police

Gender
participant gender

1 = both
2 = female
3 = male

WeapType
weapon type used in study

0 = no weapon
1 = knife
2 = gun
3 = bottle
4 = meat cleaver
5 = syringe
6 = pistol or machete
7 = scissors
8 = screwdriver

WeapDang
"dangerousness" of weapon

1 = gun
2 = others

ObjUnus
unusual object used

0 = no unusual object
1 = chicken
2 = toy
3 = feather duster
4 = celery
5 = space cones
6 = conch shell
7 = stethoscope
8 = wooden snake
9 = semiautomatic pistol with 2 bands of fluorescent yellow tape wrapped around barrel
10 = 19-th century single-shot percussion pistol
11 = knitting needle

NeutrObj
neutral object used in control condition

0 = no neutral object
1 = magazine/book
2 = check
3 = pen
4 = can of soup or bag of potato chips
5 = cellular phone
6 = bottle of soda or water
7 = wallet
8 = CD

LConst
method used to construct lineup

1 = matched to photo/target
2 = matched to description
9 = unreported

SimTarg
similarity between target and filler persons

1 = high similarity between target and fillers
2 = moderate similarity between target and fillers
3 = low similarity between target and fillers
4 = mixed across conditions
9 = unreported

LType
lineup type used

1 = simultaneous
2 = sequential
3 = showup
4 = mixed (varied across conditions)
5 = mugshot book

LSize ______________
number of persons in lineup

LMode
mode lineup was presented

1 = photo
2 = live
3 = video
4 = mixed (varied across conditions)
5 = slides

WarnTest
participants knew of memory test ahead

1 = yes
2 = no
3 = instructed to avoid weapon focus
4 = mixed across conditions
9 = unreported

FairInst
participants received cautionary instruction
"target might or might not be in the lineup"

1 = yes (or option to reject)
2 = no, a biased instruction
3 = mixed (varied across conditions)
9 = unreported

LBias
biased lineup used

1 = yes: how so? ______________________
2 = no
3 = mixed across conditions
9 = unreported

Comment ____
Special characteristics of the study

____

____

Target Identification

N *Total sample*	________________
nEG *weapon*	________________
nCG *control group*	________________
TPWPHits	Proportion of Correct Identification of target – Target present - Weapon *present*
TPWAHits	Proportion of Correct Identification of target – Target Present - Weapon *absent*
TPWPFA	Proportion of False alarms - Target Present - Weapon present
TPWAFA	Proportion of False alarms - Target Present - Weapon absent
TAWPCR	Proportion of Correct rejections - Target Absent - Weapon present
TAWACR	Proportion of Correct rejections - Target Absent - Weapon absent
WPDeccor	Proportion of correct decisions - Weapon present
WADeccor	Proportion of correct decisions - Weapon absent
WPDecInc	Proportion of incorrect decisions - Weapon present
WADescInc	Proportion of incorrect decisions - Weapon absent
IDFvalue	F- value: weapon vs. no weapon
IDchisq	chi^2- value: weapon vs. no weapon
IDdf1	degrees of freedom numerator

IDdf2	degrees of freedom denominator
IDpF	p-value

Feature accuracy for target
sometimes involves subtraction of errors from correct

MWPDesc	Correct details – weapon present
SDWPDesc	Standard deviation – weapon present
MWADesc	Correct details – weapon absent
SDWADesc	Standard deviation – weapon absent
CorrDesF	F- value: weapon vs. no weapon
CorrDest	t- value: weapon vs. no weapon
CorrDesdf1	degrees of freedom numerator
CorrDesdf2	degrees of freedom denominator
CorrDesp	p-value
Desetasq	eta^2
MWPDeIn	Incorrect details – weapon present
SDWPDeIn	Standard deviation – weapon present
MWADeIn	Incorrect details – weapon absent
SDWADeIn	Standard deviation – weapon absent
DescInF	F-value: weapon vs. no weapon
DescInt	t-value: weapon vs. no weapon
DesIndf1	degrees of freedom numerator
DesIndf2	degrees of freedom denominator
DesInp	p-value

DesIneta eta^2

Object recall for weapon

WCorName Proportion Correct naming of weapon

OCorName Proportion Correct naming of neutral object

Feature accuracy for weapon

sometimes involves subtraction of errors from correct

MWDesc Correct details – weapon present

SDWDesc Standard deviation – weapon present

MODesc Correct details – weapon absent/ neutral object

SDODesc Standard deviation – weapon absent/ neutral object

DesWOF F-value: weapon vs. neutral object

DesWOt t-value: weapon vs. neutral object

DesWOdf1 degrees of freedom numerator

DescWOdf2 degrees of freedom denominator

DescWOp p-value

Comment Comments to the codings

Meta-analysis 2

Identification Evidence by Elderly Eyewitnesses: A Meta-analysis

Elderly are the fastest growing age group in industrial countries. Improved living conditions, including better medical care, predominantly account for these developments. In 2010, 12.8% of the population in the United States will be 65 years and older (U.S. Bureau of the Census, 2011), and this proportion will increase to an estimated 20% by 2030.

As the number of older people in the overall population increases, the number of elderly who become involved in the criminal justice system in some manner will grow alike, even if crime rates remain stable (Dunlop, Rothman, & Hirt, 2001).

Although victimization rates for violent crime in the US decline with victim age (e.g., 16-19 years = 30.3%, and 65 years and older = 3.2%, respectively; Snyder, 2009), victimization per se is age specific. Whereas the majority of people younger than 65 years (78.1%) experienced crimes like aggravated or simple assault, burglary, and larceny, even more older people (> 65 years; 87.4%) were victims of crimes like robbery, aggravated or simple assault, intimidation, burglary, larceny and vandalism (McCabe & Gregory, 1998).

Because in most crimes reliable DNA-rich biological traces are not available, eyewitness evidence is particularly crucial for investigating and prosecuting crimes (Wells & Olson, 2003). Besides other factors, age of witness is one variable contributing to high false identification rates (e.g., Memon & Gabbert, 2003a, 2003b; Wilcock, Bull, & Vrij, 2007).

Definition of Old Age

Within the set of reviewed studies quite different cut-offs were used (e.g., Wright & Stroud, 2002: 35-55 years; Searcy, Bartlett, & Memon, 1999: 60-80 years; Scogin, Calhoon, & d'Enrico, 1994: 75-94 years).

Although the chronological age is not necessarily understood as a cause or source of theoretical explanations (Dixon, 2011), it seems to be the only way to classify studies due to its reliable measurability. Usually, authors assess psychological and functional variables to obtain a high homogeneity in their sample. Despite the wide variation in cut-off points, we included all studies involving a comparison of younger and older witnesses (except the study by Wright & Stroud, 2002).

Memory of the Elderly

Doubt concerning the accuracy of older eyewitnesses arises due to intensively investigated memory problems in the older age group. Behavioral research on aging has demonstrated robust declines in memory abilities such as encoding of new facts, working memory, inhibitory functions, long-term memory and in processing speed (for reviews see Light, 1991; Lustig, Hasher, & Zacks, 2007; Park & Reuter-Lorenz, 2009; Salthouse, 1996).

Nonetheless, not all facets of memory are affected in the same way by aging. There is evidence that short-term memory, autobiographical memory and semantic knowledge, for example, remain relatively stable (Hedden & Gabrieli, 2004). Moreover, age-related declines in recognition are smaller than in recall (Craik & McDowd, 1987; Ratcliff, Thapar, & McKoon, 2011), and differences are smaller in memory for content than those in context memory (Spencer & Raz, 1995).

However, Nilsson (2003) found episodic memory to be especially impaired in normal aging. Specifically, differences in memory for context appear consistent, presumably because older persons have more problems with remembering the source of information (e.g., Aizpurua, Garcia-Bajos, & Migueles, 2011; Bornstein, 1995; Cansino, Trejo-Morales, & Hernández-Ramos, 2010; Johnson, Hashtroudi, & Lindsay, 1993).

The inability to *use* contextual information is often described as the key mechanism of these age deficits in memory even though the information has been stored (Thomas & Blevich, 2006; Old & Naveh-Benjamin, 2008). Where information is not available individuals use heuristic strategies as they rely on assessing the vividness of perceptual detail and the familiarity of a stimulus (Bright-Paul, Jarrold, & Wright, 2005; Johnson et al., 1993).

Especially for eyewitnesses, remembering the source of information is essential as an identification in a lineup involves recognizing a person seen at a particular place at a particular time committing a criminal act.

Choosers vs. Nonchoosers

When confronted with a lineup, a witness may either positively identify somebody in the lineup (chooser), or reject it (nonchooser). Choosing behavior of an eyewitness has been demonstrated as an important moderator for post-dicting eyewitness accuracy from witnesses' confidence (Sporer, Penrod, Read, & Cutler, 1995). Another meta-analysis found that sequential testing was accompanied by lower choosing rates (Steblay, Dysart, & Wells, 2011).

Besides higher correct identification rates (hits), higher choosing rates also imply more frequent foil identifications. It seems as if the elderly pick more often an innocent person from a lineup (e.g., Memon & Bartlett, 2002; Wilcock, Bull, & Vrij, 2005, 2007). Thus, it is important to know if there is a response bias towards choosing in the elderly.

Eyewitness versus Facial Recognition Studies

To examine old-age effects on eyewitness identifications, two types of methodologies were usually used in research. Elderly participants' accuracy was either examined directly by comparing older and younger participants' performance in staged or filmed eyewitness scenarios, or indirectly, by investigating the effect of age on performance in face recognition tasks (Bornstein, 1995).

To study eyewitness identification, participants are usually exposed to an "event" (e.g., a staged theft) without their prior knowledge and often without warning that they will later have to identify the target person (e.g., Adams-Price, 1992; Memon & Gabbert, 2003a, 2003b; Wilcock et al., 2005, 2007). The event is experienced live or shown on slides or videotape. Typically, there is only one target. The delay between the event and the recognition test varies between a few minutes, hours or even days. At the identification phase, either a photospread or video lineup with the target (perpetrator) present or absent is shown. For target-present (TP) lineups, data are analyzed in terms of correct identifications (hits), incorrect rejections, or foil identifications. For target-absent (TA) lineups, correct rejections or false identifications (including foil identifications) are possible outcomes. Some studies only reported correct decisions across both TP and TA lineups. Unfortunately, most studies did not separate between the identification of the target replacement and that of a filler person in TA lineups. Especially in terms of source monitoring problems in the elderly, this would be important to investigate.

In contrast, in face recognition studies participants are exposed to a series of faces (e.g., 20 faces) on slides or on a computer screen (e.g., Anastasi & Rhodes, 2005; Bartlett, Leslie, Tubbs & Fulton, 1989; Bastin & Van der Linden, 2003, 2006). In the recognition phase, photographs of previously seen faces, interspersed with new faces, are shown to participants whose task is to designate the previously seen faces as "Old" or "New". Data are analyzed in terms of hits and false alarms, often combining them via signal detection theory parameters.

In the current meta-analysis only eyewitness identification studies were included due to their higher ecological validity. A separate meta-analysis concerning face recognition studies is conducted separately (Martschuk, Kocab, & Sporer, in preparation).

Own-age Bias

Findings investigating an own-age bias, that is, participants' ability to better recognize faces of their own age as compared to other-age faces, are mixed. Some eyewitness studies reported results supporting the effect (e.g., Memon, Bartlett, Rose, & Gray, 2003; Wright & Stroud, 2002), whereas other studies did not find it (e.g., Rose, Bull, & Vrij, 2003; 2005; Wilcock et al., 2005; 2007).

Only few studies (*k* = 9) in this set of eyewitness studies examined the own-age bias (of which most data are dependent from each other because participants had to view two videos with both a young and an old target). Also, using only a single target to-be-recognized for each age group poses problems of interpretation due to limited stimulus sampling (Wells & Windschitl, 1999).

Method

Search Strategies Used

Several complementary strategies were pursued to identify all potentially eligible study reports (Lipsey & Wilson, 2001; Sporer & Cohn, 2011; White, 2009): (a) an electronic literature search in the databases *Web of Science* (including the *Social Sciences Citation Index*, *SSCI*, since 1956, the *Science Citation Index* since 1945 and the *Arts and Humanities Index* since 1975), *PsycINFO* (since 1967), *MEDLINE* (since 1902) and *Dissertation Abstracts* (since 1980) from the first available year to January 2012. Furthermore, (b) a citation search in the *SSCI* for existing reviews (Yarmey, 1996; Bartlett & Memon, 2007), (c) a citation search in the *SSCI* with authors who published multiple studies in that area (Bartlett, Memon, Searcy, Yarmey), (d) a manual search in topically relevant journals (e.g., *Law and Human Behavior*, *Psychology and Aging*), (e) a complete search of relevant references cited in any of the previously located articles or reviews were conducted. Finally, (f) we contacted first authors of included studies by e-mail to send us any relevant work, published or unpublished.

To minimize the widely discussed problem of *publication bias*, that is, studies with nonsignificant findings are less likely to be submitted, or if submitted, are less likely to be published (Lipsey & Wilson, 1993; Sporer & Goodman-Delahunty, 2011; Sutton, 2009), we tried to include unpublished data as well to minimize this problem.

Inclusion versus Exclusion Criteria

For inclusion in the current analyses, studies had to meet the following inclusion criteria: (a) they examined eyewitness identification tasks as described above, (b) they investigated at least two age groups ("older" vs. "younger group"), (c) age of participants was at least 16 years (no children), (d) they reported hits, false alarms, correct rejections, false rejections and/or overall correct decisions, (e) they studied a healthy sample, and (f) reported the results in English, German, French or Russian.

Reasons for excluding studies involved (a) lack of sufficient data to compute an effect size and (b) the use of verbal recall measures only. A total of 24 eyewitness studies met these criteria.

Coding of the Studies

Two well-trained coders (first and second author) completed all coding independently by means of a complex coding manual (see Wilson, 2009), comprising information about study characteristics (e.g., sample properties, assessment of dementia, stimulus material used), as well as statistics reported like identification rates or tests of statistical significance (see Lipsey, 2009; Lipsey & Wilson, 2001). Moderator variables were coded separately to be used for accordant analyses.

Inter-coder reliability was estimated using Cohen's *kappa* for categorical variables as it controls for chance agreement (Orwin & Vevea, 2009; Sporer & Cohn, 2011) and the intraclass correlation coefficient (r_I) for continuous variables which also takes systematic differences between coders into account (see Orwin & Vevea, 2009). Overall, inter-coder agreement was highly satisfactory, with all coefficients indicating either perfect agreement, or $r_I \geq .979$ and *kappa* ≥ .938. Disagreements were resolved via discussion among all authors.

Effect Sizes

When testing differences between proportions the odds ratio is the effect size of choice (Fleiss & Berlin, 2009; Lipsey & Wilson, 2001). The odds ratio has the inconvenient form of being centered around 1. To compensate this peculiarity, it is customary to perform all analyses on the natural log of the odds ratio (*LOR*) which is centered around 0. When interpreting results, *LOR*s are backtransformed (Fleiss & Berlin, 2009; Lipsey & Wilson, 2001). We report results in a way that when younger adults perform better than the elderly, *LOR*s will be positive (and *OR*s greater than 1).

The current meta-analysis followed the approach recommended by Hedges and Olkin (1985), in which a weighted mean effect size for the sample of studies was initially calculated along with a homogeneity test (Borenstein, 2009; Fleiss & Berlin, 2009; Konstantopoulos & Hedges, 2009; Shadish & Haddock, 2009; Raudenbusch, 2009). Outlier analyses were conducted to detect extreme cases.

Fixed and random effects models. We used fixed and random effects models (Konstantopoulos & Hedges, 2009; Lipsey & Wilson, 2001; Raudenbush, 2009), where studies are weighted by the inverse of their variance. Large samples provide a more accurate estimate of population parameters than smaller ones in the sense that the variance of the estimate around the parameter will be smaller (Levine, Asada, & Carpenter, 2009; Shadish & Haddock, 2009).

The assumption using a *fixed effects model* (FEM) is that observed studies would yield effect sizes that differ from the true population effect size only by sampling error whereas the *random effects model* (REM) accounts for between-studies variance reflecting true underlying population differences. The homogeneity test statistic Q was used to assess the assumption of the FEM. Because the homogeneity test has little statistical power (Harwell, 1997; Hedges & Pigott, 2001; Jackson, 2006), Matt and Cook (2009) recommend the additional use of the descriptive statistic I^2 to quantify the proportion of total variation between effect sizes.

Theoretically driven moderator analyses were conducted to detect covariates moderating the old-age effect. To avoid confounding of predictor variables, meta-regression approaches were preferred over simple sub-group comparisons (Pigott, 2012).

Results

Overview of Analyses

Separate meta-analyses were conducted for correct decisions and choosing rates (across TP and TA lineups) as well as for each of the specific outcomes in TP (hits and filler identifications) and TA lineup identification tasks (filler or false identifications of an innocent suspect which were usually not reported separately). After checking for outliers, an overall test of age differences was calculated for each dependent measure averaged across all experimental conditions. However, as many studies included specific manipulations to test specific conditions (e.g., exposure time or delay; mode of testing) that may affect identification performance, we also discuss individual studies that may form exceptions from the general pattern.

As results for young and old targets were largely statistically dependent, separate meta-analyses for young and old target faces were calculated. However, to illustrate a potential own-age effect, we contrast results for young and old targets with the help of graphical methods. At the end, we address the possibility of publication bias.

To give an overview of the studies included the study characteristics will be described first.

Study Characteristics

Appendix A describes important variables of the studies included: type of presentation (photo, video, live), type of event/crime, encoding time in s, retention interval in min, lineup type (TP, TA, TP/TA), type of lineup presentation (simultaneous, sequential, both) and number of persons in the lineup (lineup size).

Overall, 24 studies could be identified investigating the old-age effect on eyewitness identification. Of these studies, 23 were published and one was presented at a conference. One study had to be excluded due to the exceptionally low age range in the older age group (Wright & Stroud, 2002; 35-55 years). When studies used mugshot presentations prior to the identification task only the control conditions were used (Perfect & Harris, 2003; Goodsell, Neuschatz, & Gronlund, 2009).

Within the remaining 22 studies the age distribution of all 2866 participants was as follows: The younger group (n = 1315) had a mean age of 21.93 years (SD = 3.29; Mdn = 21.1) and the older group (n = 1266) of 69.88 years (SD = 6.03; Mdn =

70.0). The range of individual participants was 16 to 40 years in the younger groups and 50 to 94 years in the older groups.

Most studies (95%) used video films as stimulus material. Encoding time of the face varied between 6 to 90 seconds with one outlying study presenting the face for 1200 seconds where the experimenter served as target (Searcy, Bartlett, Memon, & Swanson, 2001). The retention interval between encoding and identification task varied from 15 minutes to 1 week and was lower than one hour in 64% of all studies. A six-person lineup was used in 86% of all reports included, with three studies showing larger lineups (of 8, 9 or 15 persons).

Weighted Mean Effect Sizes for Young Targets

Table 1 displays the meta-analytical results for correct decisions overall, correct identifications and foil identifications in TP lineups, foil identifications in TA lineups and choosing rates. Single effects per study for each dependent measure are displayed in Appendix B. Figures C1 to C5 in Appendix C show these effect sizes (with 95% CIs), each sorted in ascending order. Single effects were averaged where more than one condition was reported by adding the frequencies across conditions to calculate *OR*s.Taken together, results demonstrated an advantage of the younger groups over the elderly for all dependent measures. As mentioned above, all analyses were conducted with fixed- and random-effects models (see Table 1) but only the former will be reported below as there were no substantial differences between the two models yielding similar conclusions.

Table 1

Mean Weighted Effect Sizes for all Dependent Variables for Young Targets in Fixed and Random Effects Models

						95% CI							
DV	Model	*k*	*N*	*LOR*	*OR*	*LL*	*UL*	*Z*	p_Z	*VC*	*Q*	p_Q	I^2
Correct Decision	FEM	13	1611	0.746	2.108	1.714	2.591	7.076	<.001		15.348	.223	.218
	REM	13	1611	0.767	2.154	1.698	2.733	6.318	<.001	.041			
Hits[a]	FEM	15	1165	0.474	1.607	1.251	2.065	3.708	<.001		21.424	.091	.347
	REM	15	1165	0.489	1.694	1.218	2.355	2.973	<.01	.133			
Filler ID (TP)[b]	FEM	13	905	0.897	2.453	1.794	3.356	5.616	<.001		18.218	.109	.341
	REM	13	905	0.974	2.649	1.768	3.969	4.721	<.001	.178			
Foil ID (TA)[c]	FEM	14	964	1.123	3.074	2.310	4.091	7.705	<.001		21.583	.062	.398
	REM	14	964	1.150	3.158	2.166	4.605	5.975	<.001	.199			
Choosing	FEM	11	1437	0.815	2.259	1.799	2.838	7.011	<.001		20.347	.026	.510
	REM	11	1437	0.833	2.301	1.643	3.221	4.854	<.001	.158			

Note. LOR = logged odds ratio; *OR* = odds ratio; *CI* = confidence interval; *LL* = lower limit, *UL* = upper limit; *Z* = significance test; p_Z = significance level for *Z*-test; *VC* = variance component in REM; *Q* = heterogeneity test statistic; p_Q = significance level for *Q*-test; I^2 = description measure of heterogeneity.

[a] One outlier excluded (Kinlen et al., 2007); [b] One outlier excluded (Memon & Gabbert, 2003a); [c] One outlier excluded (Searcy et al., 2000).

Correct decisions. Collectively, 13 independent hypothesis tests could be identified examining age effects on correct decisions averaged across TP and TA lineups, that is, correct identifications of the target in a TP lineup and correct rejections of a TA lineup. Outlier analyses revealed no extreme cases. The homogeneity test failed to reach significance, $Q(12) = 15.35$, $p = .223$, $I^2 = .218$.

Results were in line with our predictions: Younger adults were 2.1 times more likely to make a correct decision compared to their older counterparts, *OR* = 2.108 [1.714, 2.591], $Z = 7.08$, $p < .001$, $k = 13$.

Correct identifications in TP lineups. For the correct identification of a target person 16 experiments were identified for inclusion in the model. Outlier analyses showed one hypothesis test to yield an LOR with more than 3 *SD*s below the mean (Kinlen, Adams-Price, & Henley, 2007, *LOR* = -0.810 [-1.596, -0.024]). In this study, participants had to visualize or verbalize the criminal scene prior to the identification task. When excluding this study the *Q*-test was no longer significant, $Q(14) = 21.42$, $p = .091$, $I^2 = .347$.

A significant but small weighted mean of *OR* = 1.607 [1.251, 2.065], $Z = 3.71$; $p < .001$, $k = 15$, was calculated. Younger participants were 1.6 times more likely to pick the correct person from the TP lineup compared to the elderly.

Foil identifications in TP lineups. Overall, 14 independent hypothesis tests were located examining age effects on foil identifications in TP lineups. Outlier analyses revealed one extreme case: Memon and Gabbert (2003b), *LOR* = 2.147 [1.360, 2.934]. In contrast to other studies, the authors presented the target in some conditions with a changed appearance (hairstyle) in the lineup. This could be something older adults have more problems with. The exclusion of this study led to a nonsignificant homogeneity test, $Q(12) = 18.22$, $p = .109$, $I^2 = .341$.

As expected, the elderly were 2.5 times more likely to choose a filler from a TP lineup compared to their younger counterparts, *OR* = 2.453 [1.794, 3.356], $Z = 5.62$; $p < .001$, $k = 13$.

Foil identifications in TA lineups. Of the 15 independent hypothesis tests, one extreme case with almost 3 *SD*s below the mean was detected; Searcy, Bartlett, & Memon, 2000, *LOR* = -0.169 [-0.976, 0.637]. Searcy and colleagues presented biased and unbiased lineup instructions to the participants which could make the study different from others. By excluding the study the test for homogeneity failed to reach significance, $Q(13) = 21.58$, $p = .062$, $I^2 = .398$. The older group was 3.1 times more likely to choose a wrong person from a TA lineup compared to younger eyewitnesses, *OR* = 3.074 [2.310, 4.091], $Z = 7.71$; $p < .001$, $k = 14$.

Choosing. Overall, 11 studies used TP as well as TA lineups as between-participants variable so that choosing rates could be calculated. An unweighted mean comparison between the two age groups showed that about half of the participants in the younger group (54%) chose someone from the lineup, compared to almost three-quarters (73%) in the old-age group.

There was a significant age effect on choosing, *OR* = 2.259 [1.799, 2.838], $Z = 7.01$, $p < .001$. This demonstrates a response bias where older persons were 2.3 times more likely to choose a person from a lineup than younger people. The homogeneity test was significant, $Q(10) = 20.35$, $p = .026$, $I^2 = .510$.

Moderator variables. Although homogeneity tests (Q and I^2) indicated no substantial variances, we explored whether theoretically and practically important variables were associated with the old-age effect.

Additionally to the mean age of the older group we investigated associations of estimator and system variables via hierarchical meta-regression models. We entered the mean age of the old-age group (characteristic of study sample) as a continuous variable first, followed by a block of additional variables: Exposure duration (1-29 s, ≥ 30 s), retention interval (< 24 hours, ≥ 24 hours) and lineup type (simultaneous only, simultaneous *and* sequential, sequential only). This hierarchical method allowed us to independently examine the contributions of these variables after the mean age of the elderly has been controlled for (see Sporer & Cohn, 2011;

Konstantopoulos & Hedges, 2009; Lipsey & Wilson, 2001; Pigott, 2012; See Table D1 to D5 in the Appendix D).

Overall, three models reached significance (for correct decisions, foil identifications in TP lineups and choosing). For correct decisions, the study conducted by O'Rourke, Penrod, Cutler, & Stuve (1989) was excluded because the mean age of the old-age group (60 years) was considered an outlier compared to the other studies. Overall, the model explained 63% of the existing variance, $Q_{between}(4) = 9.57$, $p = .048$; $Q_{total}(11) = 15.10$, $p = .178$; $Q_{within}(7) = 5.53$, $p = .596$, mostly by the mean age of the older age group (55%), *beta* = .741, $p = .004$.

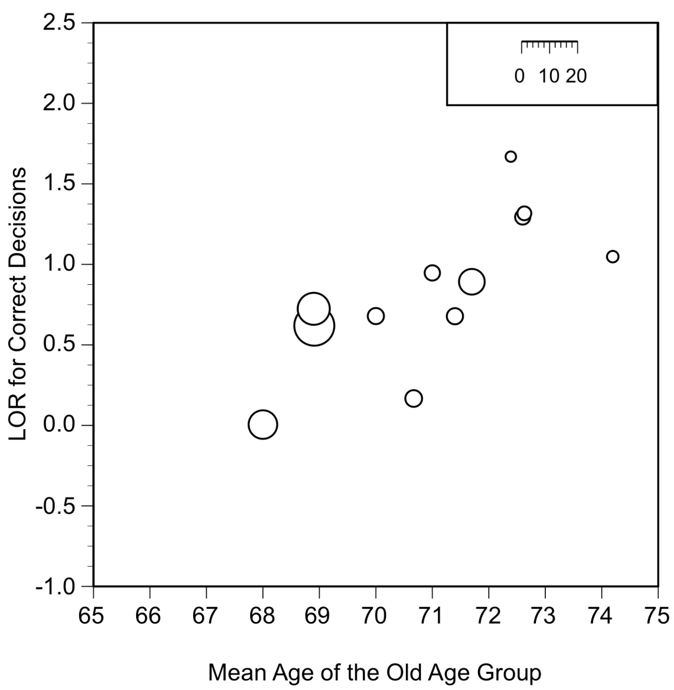

Figure 1. Effect sizes LOR for correct decisions as a function of the mean age of the old age group. The size of the bubbles reflects the inverse variance weights of the estimates.

The bubble graph in Figure 1 displays the association of correct decisions with the mean age of the old-age group. Bubble sizes represent the inverse variance weights. Overall, Figure 1 demonstrates that there are larger differences in correct decisions between young and old witnesses the older participants are in the old-age group. The size of the bubbles reflects the relative accuracy of estimates (i.e., larger bubbles correspond to smaller standard errors and variances). Results for choosing were similar (see Figure 2), with higher effect sizes for choosing the older the witnesses in the old-age group.

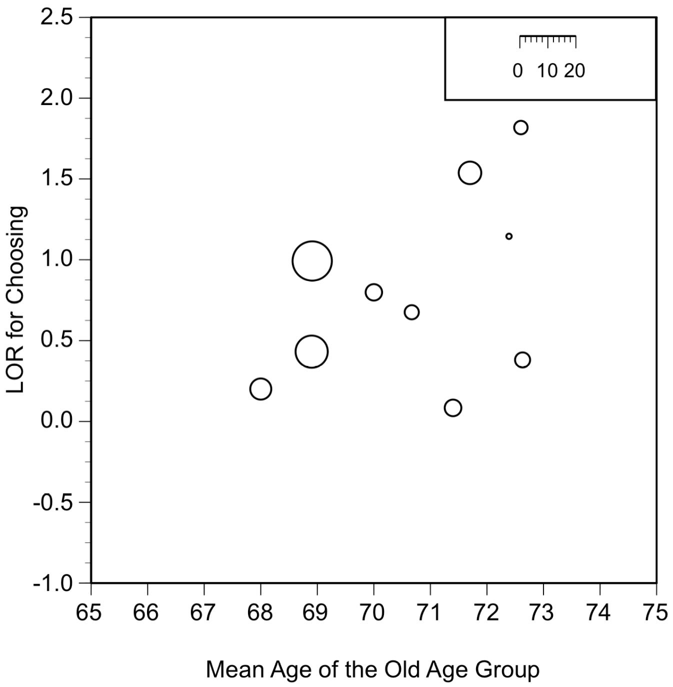

Figure 2. Effect sizes LOR for choosing as a function of the mean age of the old age group. The size of the bubbles reflects the inverse variance weights of the estimates.

For foil identifications in TP lineups, 76% of the existing variance could be explained by the model, $Q_{between}(4) = 13.88$, $p = .008$; $Q_{total}(12) = 18.22$, $p = .109$; $Q_{within}(8) = 4.33$, $p = .826$. The only significant predictor was the mean age of the old-age group, *beta* = .944, $p < .001$, which explained 65% of the variation.

Table 2

Hierarchical Weighted Fixed-Effect Meta-regression for Choosing for TP and TA Lineups for Young Targets (k = 11)

Study characteristics	*B*	*beta*	*p*	R^2
Step 1:				
Model $Q(1) = 4.91$, $p = .027$				.241
Residual $Q(9) = 15.44$, $p = .080$				
Mean age of the older group	.144	.491	.027	
Step 2:				
Model $Q(4) = 12.31$, $p = .015$				.605
Residual $Q(6) = 8.03$, $p = .236$				
Type of lineup				
(simultaneous only, both, sequential only)	-.480	-.575	.046	
Exposure Time				
(1-29 s, ≥ 30 s)	.569	.540	.070	
Delay				
(< 24 hours, ≥ 24 hours)	-.045	-.027	.930	
Regression constant	-11.155			

Note. Q statistics shown are from fixed-effects meta-regression models and test whether (1) mean age of the old-age group, and (2) the set of moderator variables in the model accounts for an additional portion of the variance in effect sizes.

Table 2 presents the results for choosing, where 61% of the between-studies variance could be explained by the model, $Q_{between}(4) = 12.31$, $p = .015$; $Q_{total}(10) = 20.35$, $p = .026$; $Q_{within}(6) = 8.03$, $p = .236$. Due to the small and largely homogeneous sets of studies a more liberal significance level seems to be appropriate. Three variables reached significance at least on a 10%-level: Mean age of the older age group (*beta* = .491, *p* = .027), type of lineup (*beta* = -.575, *p* = .046) and exposure time (*beta* = .540, *p* = .070). There were larger differences in choosing rates between the two age groups when the older participants were older, when the exposure of the target was *longer* and when a simultaneous lineup was presented.

In the model for correct identifications in TP lineups only exposure duration emerged as a significant predictor (see Table D2 in Appendix D), *beta* = -.841, *p* = .010. After longer exposure, the difference between the age groups was smaller compared to shorter exposure times. However, the overall model failed to reach significance, $Q_{between}(4) = 8.58$, $p = .073$; $Q_{total}(14) = 21.42$ $p = .091$; $Q_{within}(10) = 12.85$, $p = .232$.

Weighted Mean Effect Sizes for Older Targets

As no direct comparisons of studies presenting both young and old targets could be conducted due to statistical dependencies, separate analyses were calculated for studies using old targets (see Table 3; because models were almost identical only FEM results will be presented). Individual effects per study are listed in Appendix E; Figures C6 to C10 in Appendix C display the distributions for the different outcomes.

In none of the models were outliers identified. Homogeneity tests did not reach significance in any case. The descriptive statistic I^2 indicated little heterogeneity between studies. Again, results demonstrated a clear advantage of the younger groups over the elderly for all dependent measures.

Table 3

Mean Weighted Effect Sizes for all Dependent Variables for Old Targets in Fixed Effects Models

DV	*k*	*N*	*LOR*	*OR*	95% CI *LL*	95% CI *UL*	*Z*	p_Z	*Q*	p_Q	I^2
Correct Decision	6	619	0.950	2.587	1.845	3.628	5.508	<.001	5.485	.360	.088
Hits	7	366	0.940	2.560	1.630	4.022	4.081	<.001	6.019	.421	.003
Filler ID (TP)	7	366	1.008	2.740	1.702	4.409	4.151	<.001	3.147	.790	.000
Foil ID (TA)	6	308	0.876	2.400	1.489	3.871	3.592	<.001	1.657	.894	.000
Choosing	6	619	0.346	1.413	1.003	1.989	1.981	0.048	0.988	.964	.000

Note. LOR = logged odds ratio; *OR* = odds ratio; *CI* = confidence interval; *LL* = lower limit, *UL* = upper limit; *Z* = significance test; p_Z = significance level for *Z*-test; *Q* = heterogeneity test statistic; p_Q = significance level for *Q*-test; I^2 = descriptive measure of heterogeneity.

Correct decisions. We obtained a medium effect size for correct decisions, *OR* = 2.587 [1.845, 3.628], $Z = 5.51$, $p < .001$, $k = 6$. No significant heterogeneity was found, $Q(5) = 5.49$, $p = .360$, $I^2 = .088$. When the target was of older age the younger participants were 2.6 times more likely to make a correct decision across both TP and TA lineups compared to their older counterparts.

Correct identifications. A medium effect was calculated for correct identifications of older targets with very little heterogeneity across studies, $Q(6) = 6.02$, $p = .421$, $I^2 = .003$; *OR* = 2.560 [1.630, 4.022], $Z = 4.08$, $p < .001$, $k = 7$. Younger participants were 2.6 times more likely to choose the correct person from a TP lineup compared to the elders when confronted with a target of older age.

Foil identifications in TP lineups. For foil identifications in TP lineups *OR* was 2.740 [1.702, 4.409], $Z = 4.15$; $p < .001$, $k = 7$, with older participants being 2.7 times more likely to choose a foil from an old target TP lineup than the younger participants. Studies included appear homogeneous, $Q(6) = 3.15$, $p = .790$, $I^2 = .000$.[1]

Foil identifications in TA lineups. The effect size of *OR* = 2.400 [1.489, 3.871], $Z = 3.59$; $p < .001$, $k = 6$, demonstrates that older persons were 2.4 times more likely to choose a filler from a TA lineup compared to the younger group. Results were highly homogeneous, $Q(5) = 1.66$, $p = .894$, $I^2 = .000$.

Choosing. Overall, six studies reported TP as well as TA lineups so that choosing rates could be calculated. The set was highly homogeneous without any outlying effect sizes, $Q(5) = 0.988$, $p = .964$, $I^2 = .000$.

The mean age effect on choosing for older faces was *OR* = 1.413 [1.003, 1.989], $Z = 1.99$; $p = .048$, $k = 6$.

---

[1] By convention, $I^2 = 0$ is reported when calculations result in a negative I^2 (Shadish & Haddock, 2009).

Comparison of Young and Old Targets

As the effect sizes are dependent in the studies with younger and older targets, no significance test could be used to compare them. Figures F1 to F5 in Appendix F graphically compare the single effect sizes for young and old targets for each study, which used both young and old target faces. Apparently, there is no consistent pattern as some studies found the effect to be larger with young faces whereas others reported larger effects for older targets.

To control for other differences between studies, mean weighted effect size models were calculated only for those studies using young as well as old targets (See Figure 3). For both young and old targets, the respective age effects were comparable in size for correct decisions (k = 6, OR = 2.554 [1.834, 3.557] and OR = 2.587 [1.845, 3.628]), correct identifications (k = 7; OR = 1.915 [1.228, 2.987] and OR = 2.560 [1.630, 4.022]) and foil identifications in TP lineups (k = 7; OR = 3.314 [1.939, 5.665] and OR = 2.740 [1.702, 4.409]) for young and old targets, respectively. A substantially larger effect for young faces existed for foil identifications in TA lineups (k = 6; OR = 3.774 [2.268, 6.282] compared to OR = 2.400 [1.489, 3.871] for old targets), and for choosing (k = 6; OR = 2.560 [1.801, 3.638] and OR = 1.413 [1.003, 1.989] for young and old targets, respectively).

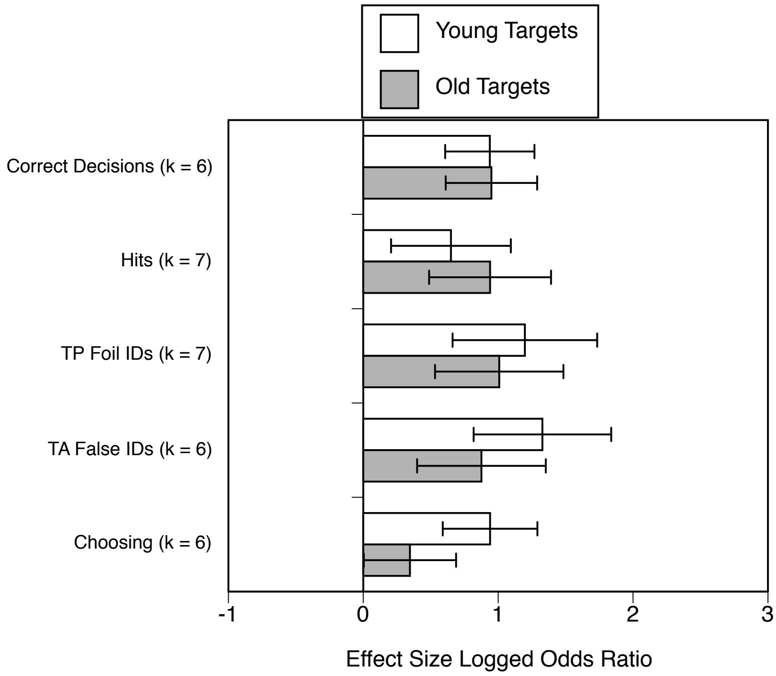

Figure 3. Weighted mean effect sizes (*LOR*) for studies using both young and old targets.

Diagnosticity

The so-called diagnosticity ratio (*DR*) presents the probative value of an eyewitness identification and explains how much more likely it is that the suspect, if identified, is the culprit rather than an innocent person (Wells & Lindsay, 1980; Wells & Turtle, 1986). It is calculated as the ratio of correct identifications (identification of the culprit) to false identifications in TA lineups (identification of an innocent suspect). As no study differentiated between suspect and foil identifications in TA lineups, suspect identification rates for TA lineups were estimated by dividing the filler identification rates in TA lineups by lineup size (Clark, 2012; Clark, Howell, & Davey, 2008).

When the diagnosticity ratio equals 1, the lineup is neither diagnostic of the guilt nor of the innocence of the suspect. To contrast the two age groups of eyewitnesses, a larger diagnosticity index reflects superiority of the respective age group. However, these analyses were only conducted for *unweighted* means of hits and false alarms.[2] When younger participants had to identify young targets the *DR* was 9.44 on average. Hence, when a young eyewitness picked a suspect from a lineup it is almost 10 times more likely that the suspect is guilty than that he or she is innocent. For the older group, *DR* = 3.99. A positive identification by a young eyewitness is 2.37 times more diagnostic than the identification by an older person. When older targets were used the average *DR* was also larger for the younger than for the older group (*DR* = 9.55 vs. 3.54, respectively).

Additionally, we used an index, which represents diagnosticity as a conditional probability (*CP*), defined as (correct identifications/(correct + false identifications)), again correcting for lineup size (Clark & Wells, 2008). Values can range from 0 to 1, indicating no diagnosticity for a probability of .50. Values > .50 indicate guilt, and values < .50 indicate innocence.

When younger participants had to identify young faces the *CP* was .87 on average, for older eyewitnesses it was *CP* = .76. Data for old targets were similar as the *CP* was .88 for younger adults and .73 for older eyewitnesses.

Publication Bias

For examining the likelihood of publication bias we first used funnel plots as a graphical method to detect asymmetry within the sets of studies. We therefore plotted the effect sizes (*LOR*) against the standard error (SE_{LOR}), as recommended in the specific literature (see Sterne & Egger, 2001; Sterne, Becker, & Egger, 2005; Sutton, 2009). In the presence of a publication bias the distribution would be expected to be asymmetrical with empty sections on the lower left side (for positive mean *LOR*s) where estimates from studies with small sample sizes and small effect sizes would be located.

For correct decisions via TP and TA lineups (*k* = 13) the asymmetry of the funnel plot in Figure 4 demonstrates that when the standard error was rather large

---

[2] To our knowledge there is no statistical theory that has investigated population sampling distributions of *DR*s.

(due to smaller sample sizes) there seems to be a lack of smaller effect sizes reported in the literature. Similar patterns were observed for the other dependent measures. The black dots in Figure 4 represent hypothetical study outcomes estimated by the trim and fill method to arrive at a more symmetric distribution of effect sizes (see below). Because conclusions from graphical methods are highly subjective (Sterne, Becker, & Egger, 2005; Sutton, 2009), we also employed more formal statistical tests of funnel plot asymmetry. Results for young and old targets are displayed in Appendix G.

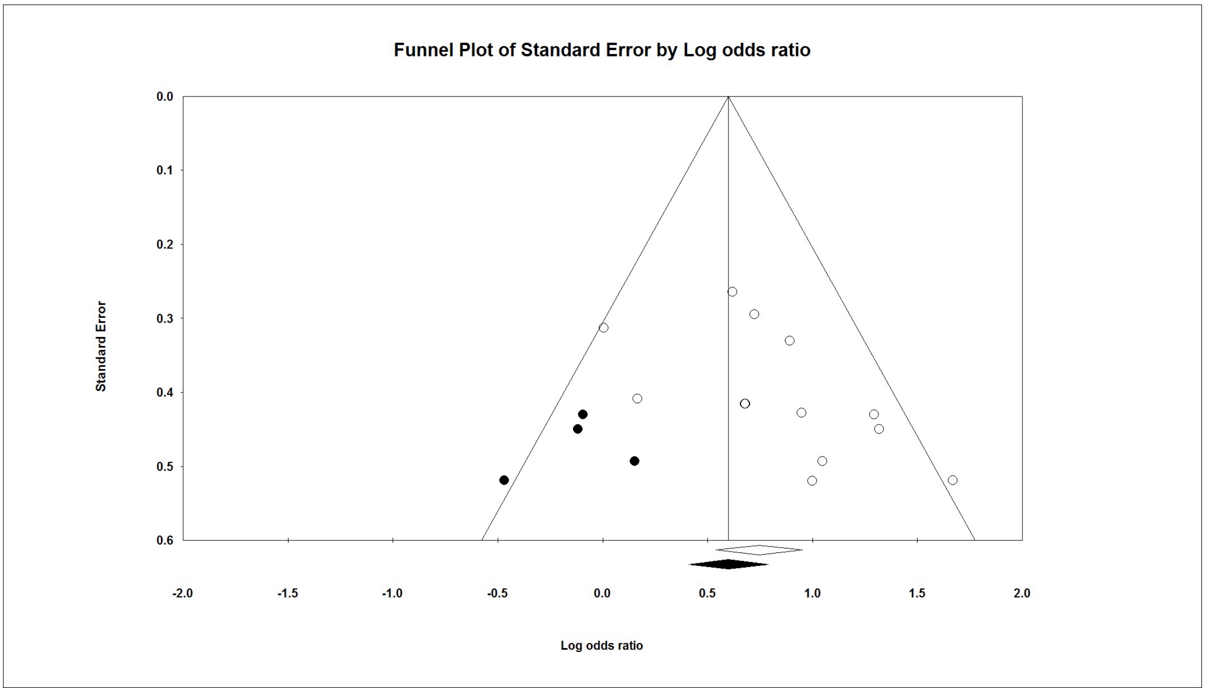

Figure 4. Funnel plot of effect sizes (*LOR*) against standard error. The effect sizes represent the magnitude of correct decisions for young targets (*k* = 13). The vertical line shows the weighted average effect size. The black dots represent the filled studies which would have been added to make the distribution more symmetrical; at the bottom the original and the reestimated mean weighted effect size when those studies would be included are displayed (trim and fill method by Duval & Tweedie, 2000a, 2000b).

The rank correlation test examines the association between effect size estimates and their variances (see Sutton, 2009; Begg & Mazumdar, 1994). For correct decisions, Kendall's *tau* (without continuity correction) = .557, $Z = 2.745$, $p = .006$ (two-tailed) reached significance. As this test has rather low power for meta-analyses that include few studies (Sterne, Gravagham, & Egger, 2000), we also applied the linear regression method (Egger, Smith, Schneider, & Minder, 1997) that regresses the standard normal deviate against precision (1/*SE*). This test yielded a marginally significant effect, intercept = 2.808 [0.020, 5.596], $SE = 1.266$, $t(11) = 2.22$, $p = .049$. It should be noted, that under certain circumstances Egger's test has inflated type I error rates when binary outcomes are considered (Sutton, 2009). Benefits from recently proposed alternative regression tests (e.g., Harbord, Egger, Sterne, 2006; Peters, Sutton, Jones, Abrams, & Rushton, 2006) are equivocal, especially when the number of studies is rather small.

A comparative evaluation of these modified methods is needed as all tests apparently have different advantages and disadvantages depending on the specific meta-analytical situation (Sutton, 2009).

As an alternative method to estimate the potential impact of publication bias, we conducted the trim and fill method (Duval & Tweedie, 2000a, 2000b; see Duval, 2005). For correct decisions this test resulted in four studies that would have to be added to the left side of the distribution to arrive at a more symmetrical distribution (see black dots in Figure 4). The trim and fill method reestimates the overall effect size by inclusion of these hypothetical effects, which yielded a more conservative weighted mean effect size for correct decisions, $OR = 1.822$, [1.174, 2.591]. This re-estimate is slightly lower than the original observed effect size, $OR = 2.108$, [1.509, 2.200].

For correct identifications in TP lineups the rank correlation test reached significance, Kendall's *tau* (without continuity correction) = .467, $Z = 2.425$, $p = .015$, and the trim and fill method re-estimated a nonsignificant weighted mean effect size with six additional studies included, $OR = 1.155$, [0.928, 1.438], which was considerably smaller than the originally observed *OR*.

For foil identifications in TP and TA lineups as well as for choosing the likelihood of publication bias was less strong. For old target faces the small sets of studies do not make these tests meaningful (see Appendix G).

Discussion

The meta-analysis reported examined the question whether younger eyewitnesses are better with identification tasks compared to older adults. Results showed that age of the person witnessing a crime indeed plays an important role with regard to the accuracy of eyewitness testimony. In the current review only identification performance was investigated rather than descriptions of the target or the event. A consistent advantage of younger participants over the elderly was found for all possible identification outcomes (correct decisions, correct identifications and foil identifications in TP and TA lineups). Small to medium effects were calculated for correct decisions and correct identifications of the target. Medium to large effects were obtained for false identifications in TP and TA lineups. The largest differences between the two age groups were found for false identification rates in TA lineups as the elderly were almost three times more likely to pick a wrong person from a TA lineup compared to their younger counterparts. Diagnosticity indices demonstrated a superiority of identifications by younger adults over those of older eyewitnesses. Another considerable effect was calculated for choosing rates demonstrating that older adults were 2.3 times more likely to choose someone from a lineup (either TP or TA) compared to younger eyewitnesses.

Although the sets of studies were largely homogeneous we tried to detect variables potentially moderating the old-age effect, using meta-regression to control for the mean age of the old-age group and the mutual dependencies among the predictor variables. We found the mean age of the older group to be associated with the size of the old-age effect, primarily for foil identifications in TP lineups and for choosing. The older an eyewitness, the more likely he or she was to choose someone from a lineup and the higher the risk of picking an innocent filler from a TP lineup. While filler identifications in TP lineups are considered forensically irrelevant, they do reflect both a tendency to choose even though the witness' memory does not seem to be good enough to discriminate between a perpetrator and an innocent foil.

The amount of time the face of the perpetrator was visible for the eyewitness as well as lineup presentation mode was also associated with the old-age effect. As one would expect from common sense and from the literature on encoding problems of the elderly, old witnesses were overall more likely to correctly identify the target in TP lineups when the opportunity to view a target was longer. On the other hand, the longer the target was in view the more likely older witnesses were to choose someone in the lineup, whether or not the target was present or not.

On the positive side, a sequential lineup was associated with smaller differences between the two age groups. Even though the number of studies using sequential lineups was limited, it seems as if sequential testing helps both age groups to adopt a stricter decision criterion. However, the elderly apparently tended to benefit more, which counteracts their higher choosing tendency.

Theoretical Implications

Two main approaches will be discussed as possible reasons for worse identification performance in the elderly: a *cognitive approach* focusing on memory restrictions and a *social approach* focusing on response bias of the older eyewitnesses.

Amongst others, cognitive limitations arise from impairments in episodic memory (Nilsson, 2003) where older adults have problems with context memory and are less able to remember the source of information (e.g., Aizpurua et al., 2011; Bornstein, 1995; Cansino et al., 2010; Johnson et al., 1993). There is evidence that people use heuristic strategies as they rely on the familiarity of a stimulus when contextual information (in this case, the source of a memory) is not available (e.g., Bright-Paul et al., 2005). Applied to the identification context, older eyewitnesses might draw the conclusion that the person appearing familiar has to be the one they saw at the scene of the crime. Consequently, they may rely on this familiarity and chose the person even though the source of this familiarity is improper for the identification decision.

Procedures like the think-aloud technique or self-reported lineup decision strategies could help to better understand reasons for picking a person. As Dunning and Stern (1994) stated accurate witnesses explained more often that their judgment resulted from *automatic processes* like "his face just 'popped out' at me" whereas inaccurate witnesses were more likely to report from a *process of elimination strategy* (e.g., "I compared the photos to each other to narrow the choices"). In this context, older adults may also need more time to allocate the familiarity of information to its source when they decide whether the person was seen during the crime or not. Response latencies in connection with introspective reports may help to answer these questions (see Palmer, Brewer, McKinnon, & Weber, 2010; Sauerland & Sporer, 2008, 2009).

On the other hand, a *social explanation* of older persons' identification performance focuses on their *choosing behavior*. In the current synthesis, results demonstrated that older persons were 2.3 times more likely to choose a person from a lineup compared to their younger counterparts. Consequently, people who more often pick someone from a lineup will receive higher correct identification rates and more foil identifications in TP lineups, as well as higher false identifications in TA lineups. However, differences between the age groups tended to be smaller for hit rates but larger for foil identifications in both TP and TA lineups. These results demonstrate a more liberal response bias of the elderly towards choosing but also a memory deficit of the elderly.

Reasons for higher choosing rates might be an inclination of the elderly to help the police. Gallagher, Maguire, Mastrofski, and Reisig (2001) observed a more positive general image of the police in the public with rising age. This positive attitude may increase older adults' intention to help solving a crime. Future studies should measure such attitudes and relate them to accuracy of identifications and choosing rates.

Another reason for higher choosing rates could be a restriction of awareness that the perpetrator may not be in the lineup. Biased (or the absence of warning) instructions increase choosing (Clark, 2005; Steblay, 1997). However, Rose et al. (2003, 2005) demonstrated that older adults had a less clear memory for unbiased lineup instructions compared to younger eyewitnesses, thus reducing their effectiveness. Additionally, Wilcock et al. (2005) showed that enhanced unbiased lineup instructions (including an additional instruction concerning mistaken identification) led to significantly better memory in the elderly for the instruction but did not improve their lineup performance. A clear signal is required that a positive identification decision is not the only response option and that a response such as *not there* or *not sure* is quite appropriate (Brewer & Palmer, 2011). Including a “Don’t Know” option on lineup forms may also help to reduce unwarranted choosing (Weber & Perfect, 2012).

A third explanation for increased choosing could be differences concerning the *need for cognitive closure* in older adults. *Need for closure* (NFC) describes a dimension of individual differences related to a person's motivation with respect to information processing and judgment (Webster & Kruglanski, 1994). NFC reflects the search for clarity and avoidance of ambiguity. According to this theory, NFC may be elevated where processing is seen as effortful and closure may serve as a means of escaping an unpleasant activity (Webster & Kruglanski, 1994). An identification task may constitute such an unpleasant situation eyewitnesses may want to leave rapidly, particularly older people.

However, using the *Need for Cognitive Closure* questionnaire Havard and Memon (2009) failed to find differences between younger and older adults or correlations of NFC with identification measures. However, there were also no differences between the two age groups in choosing behavior in that study. There is evidence, however, that older adults' social judgment biases are related to the degree to which they need quick and decisive closure. This is not so for younger age groups (Cavanaugh & Blanchard-Fields, 2010).

In summary, both *cognitive aspects* (relying on familiarity due to source monitoring problems) as well as a *response bias* (higher choosing rates due to a higher motivation to help or to seek clarity) could be jointly responsible for the reduced identification performance and the higher choosing rates of the elderly. Future research should focus on these approaches to clarify the reasons of the identification deficits in older adults.

Own-age Bias

The so-called own-age bias describes the phenomenon of people being better with faces of their own age compared to other age groups (Sporer, 2001; Wright & Stroud, 2002). We found that older eyewitnesses being confronted with an old-age target made more errors as they chose less often the right person or picked more often a foil compared to younger witnesses. Differences between the age groups were medium to large for all possible outcomes.

In the current analysis, the comparison of young and old target faces did not show a consistent pattern. When we graphically compared the effect sizes for young and old faces *within* studies that used both, some found larger effects for the young face and some for the old face. When we calculated separate meta-analyses for young and old faces we obtained larger effects for young faces for foil identifications in TA lineups and for choosing than for old face targets.

The meta-analysis by Rhodes and Anastasi (2012) which collapsed data across facial recognition and identification studies reported a small but significant own-age bias for hits (g = 0.23), false alarm rates (g = -0.23) and discriminability (g = 0.37). First results from a separately conducted meta-analysis of face recognition studies (Martschuk et al., in preparation) indicate superiority of younger participants over the elderly for all dependent measures. Large own age effects for both target age groups were found for false alarms (g_u = 1.002 vs. g_u = 0.784 for young and old faces, respectively). Effects for hits and sensitivity (d', A') were large for young faces (g_u = 0.567 and g_u = 1.353, respectively) but small or even nonsignificant for old faces (g_u = 0.126 and g_u = 0.472), suggesting a possible own-age effect.

But even if an own-age bias exists, it is not likely to have much practical importance. In general, offenders are rarely of old age. For example, the US arrest data for 2010 showed that only 0.75% of the individuals arrested were persons aged 65 and older (Bureau of Justice Statistics, 2012).

Practical Implications

Results demonstrated older eyewitness to be worse with identification tasks compared to their younger counterparts. Should older adults therefore not serve as eyewitnesses in court? Or should there even be an age restriction? Of course, this would overinterpret these data which only summarize average tendencies across large numbers of participants. Most importantly, old-age effects become increasingly larger with increasing age of the old-age group. But these moderator analyses with *mean* age ignore the large individual differences in cognitive aging.

First, we have to consider that older adults who served as participants in the current studies were screened for mental deficiencies. For example, adults having cognitive impairments due to a beginning Alzheimer disease were excluded from the samples. This implies that the data reported may actually underestimate the effects observed. On the other hand, we also obtained evidence for a publication bias which suggest that the effects found in the literature may be overestimates. Consequently, identifications from older eyewitness should be examined particularly carefully. Yet reliance on mean effect sizes cannot serve as an exclusion rule.

Besides inter-individual differences in the aging process, which should always be considered, results of our moderator analyses demonstrated that under certain conditions the differences between younger and older eyewitnesses may be reduced.

One factor, not under the control of the justice system, is the time a perpetrator is visible for an eyewitness. On the one hand, older adults did profit from longer exposure times for hits in TP lineups. It seems that the longer exposure helped the elderly to overcome their slower processing speed of information, allowing for sufficient encoding. If an older eyewitness was exposed to the perpetrator for more than 30 seconds he or she was more likely to correctly identify a target.

On the other hand, longer exposure times tended also to be associated with higher choosing rates, irrespective of target-presence. It seems as if the elderly feel capable to identify someone, hence will be more likely to choose someone, even though their memory may not be strong enough to do so. Consequently, even if hit rates may be higher when the perpetrator is present there may be more false identifications in TA lineups. An increase in foil identifications in TP lineups is less problematic as fillers are considered innocent.

There is some evidence that older adults may profit from sequential lineup presentations as they choose less often someone from the lineup (e.g., Memon & Gabbert, 2003a) by inducing a conservative shift in response bias (Clark, 2012). Maybe not being able to directly compare the faces leads to a stricter decision criterion (see above: *process of elimination strategy*). Steblay et al. (2011) as well as specific studies (e.g., Memon & Gabbert, 2003a, 2003b; Rose et al., 2005) in this review found the sequential-superiority effect for false identification rates in TA lineups for both age groups. When choosing is reduced in the elderly one would expect lower false identification rates in TA lineups for older adults. Our data indicate that sequential testing works comparably for both age groups for all lineup outcomes as mode of testing was not a significant moderator in our meta-regression models.

As noted above, there is evidence that performance gradually decreases with increased age. Hence, there is no definite age threshold for a decrease in identification performance, which may start as early as the fifties (Wright & Stroud, 2002) but may not be visible up to the late sixties.

Future Directions

On the basis of the theoretical explanations discussed, we think that new methods can improve identification performance of older eyewitnesses. If *cognitive aspects* are responsible for age differences in identification performance (especially, source monitoring problems) context reinstatement techniques may be promising.

But results are inconclusive (e.g., Wilcock et al., 2007; Rose et al., 2003). Within the current synthesis, too few studies were available to investigate this as a moderator. A recent meta-analysis on *recall* demonstrated benefits of the cognitive interview for older participants (Memon, Meissner, & Fraser, 2010). Furthermore, a modified version of the cognitive interview for the use with older witnesses is currently under development (Holliday, Ferguson, Milne, Bull, & Memon, 2009; see Memon et al., 2010) but effects on identification performance are still contradictory for both younger and older adults.

There is some evidence that older adults profit from visualization and verbalization tasks prior to identification (Kinlen et al., 2007). Possible reasons might be that visualizing the target and the criminal context may help the elderly to get a more accurate picture of the perpetrator in their minds and/or to reinstate the context for a better memory, which in turn may improve identification.

As noted above, from a more *social perspective* the high choosing of older eyewitnesses is one of the main concerns. Consequently, individual instructions should encourage older adults to adopt a stricter decision criterion (Memon, Hope, Bartlett, & Bull, 2002). It should be made clear to witnesses that they have the possibility not to choose someone from a lineup. In most studies included in the current analysis it was not obvious if the option not to choose was emphasized clearly enough in the experimental instructions.

Searcy et al. (2000) used biased as well as unbiased instructions and found somewhat (but not significantly) higher choosing rates for biased compared to unbiased instructions for both age groups. As the authors stated, a possible explanation for the lack of effectiveness could have been the 'none of them' option in that study.

However, there is controversy if unbiased instructions also decrease correct identifications besides false identifications in general (Clark, 2005; Fulero, 2009; Steblay, 1997; for a discussion see Clark, 2012). Older people who may be affected more strongly by instructions (given they understand and remember them) could become overcautious in their choosing behavior. New strategies are needed in line with the so-called *no-cost view*, delivering lower false identification rates with little or no reduction in hit rates (Clark, 2012). Adding a “Don’t Know” option may also prove beneficial (Weber & Perfect, 2012).

Finally, Wilcock and Bull (2010) demonstrated profits in TA lineups with the use of pre-lineup questions and practice lineups. Especially the latter could help to establish a stricter decision criterion in the elderly.

Limitations

One problem we faced with was missing information in many study reports. Often, procedures how experiments were conducted were not explained in enough detail (e.g., the use of biased or unbiased instructions, fairness or functional size of the lineup). This lack of information restricts the possibility to use these factors as moderators.

Second, statistical methods adopted from the medical literature delivered evidence of the possible existence of publication bias. With the trim and fill method we reestimated weighted effect sizes which were lower for almost all dependent measures than the originally observed effect sizes. For correct identifications the reestimated weighted mean *OR* even failed to reach significance. As the effects for foil identifications and choosing were already large a possible correction with 'missing' studies did not change conclusions. However, a major limitation of these sensitivity analyses is that different mechanisms may be responsible for publication bias (lack of significance, small effects, and/or small sample sizes). However, smaller (unpublished) studies may also differ in quality.

As noted in the introduction, the current meta-analysis only included lineup studies, and a separate analysis for face recognition studies is almost completed (Martschuk et al., in preparation). One main reason for separate analyses was that lineup studies are clearly superior with respect to ecological validity regarding criminal investigations compared to face recognition studies. On the other hand, face recognition studies better address the problem of stimulus sampling (Wells &

Windschitl, 1999). In particular, face recognition studies are more suitable to test for an own-age effect (Rhodes & Anastasi, 2012) than lineup studies in which young and old targets were usually operationalized by a single target in a single situation (film or staged event; Sporer & Martschuk, in press).

Conclusion

In conclusion, the reported large effect sizes for foil identifications in TP and TA lineups appear reliable evidence for the existence of an old-age effect. Results for choosing behavior indicate that older eyewitnesses are more likely to pick someone from a lineup compared to younger adults. Although this tendency may increase the possibility of choosing the right person in TP lineups it also leads to higher foil or false identifications in TA lineups. As in real life the police does not know if the suspect is the perpetrator this is problematic when older adults pick the target replacement. Unfortunately, primary studies did not differentiate between choices of target replacement and fillers. In general, identifications by younger eyewitnesses indicate stronger evidence for the proposition that the suspect is the culprit. Nonetheless, the elderly are not inevitably worse than their younger counterparts. For one thing, there are large individual differences along the age continuum. There are also several possibilities to reduce the differences between the two age groups, for example, by using sequential lineups, visualization or verbalization tasks prior to the lineup presentation or practice lineups. All these methods seem to encourage older adults to adopt a stricter decision criterion. Further research testing these and other methods are urgently needed.

References

* = primary study included in the current meta-analysis; a = correct decision across target-present and target-absent lineups (young target); b = correct identification of the target (young target); c = false alarms in target-present lineups (young target); d = false alarms in target-absent lineups (young target); e = correct decision across target-present and target-absent lineups (old target); f = correct identification of the target (old target); g = false alarms in target-present lineups (old target); h = false alarms in target-absent lineups (old target).

*Adams-Price, C. (1992). Eyewitness memory and aging - predictors of accuracy in recall and person recognition. *Psychology and Aging, 7*, 602-608. doi: 10.1037/0882-7974.7.4.602 [b]

Aizpurua, A., Garcia-Bajos, E., & Migueles, M. (2011). False recognition and source attribution for actions of an emotional event in older and younger adults. *Experimental Aging Research, 37*, 310-329. doi: 10.1080/0361073X.2011.568829

Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin & Review, 12*, 1043-1047. doi: 10.1080/17470218.2010.537926

Bartlett, J. C., Leslie, J. E., Tubbs, A., & Fulton, A. (1989). Aging and memory for pictures of faces. *Psychology and Aging, 4*, 276-283. doi: 10.1037/0882-7974.4.3.276

Bartlett, J. C., & Memon, A. (2007). Eyewitness memory in young and older adults. In R. C. L. Lindsay, Ross, D. F., Read, J. D. & Toglia, M. P. (Eds.), *Handbook of eyewitness psychology: Memory for people* (Vol. 2, pp. 309-338). Mahwah, NJ: Lawrence Erlbaum & Associates.

Bastin, C., & Van der Linden, M. (2003). The contribution of recollection and familiarity to recognition memory: A study of the effects of test format and aging. *Neuropsychology, 17*, 14-24. doi: 10.1037/0894-4105.17.1.14

Bastin, C., & Van der Linden, M. (2006). The effects of aging on the recognition of different types of associations. *Experimental Aging Research, 32*, 61-77.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088-1101. doi: 10.1080/03610730500326291

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 221-236). New York: Russell Sage Foundation.

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

Bright-Paul, A., Jarrold, C., & Wright, D. B. (2005). Age-approriate cues facilitate source-monitoring and reduce suggestibility in 3- to 7-year-olds. *Cognitive Develeopment, 20*, 1-18. doi: 10.1016/j.cogdev.2004.06.001

Cansino, S., Trejo-Morales, P., & Hernández-Ramos, E. (2010). Age-related changes in neural activity during source memory encoding in young, middle-aged and elderly adults. *Neuropsychologica, 48*, 2537-2549. doi: 10.1016/j.neuropsychologia.2010.04.032

Cavanaugh, J. C., & Blanchard-Fields, F. (2010). *Adult development and aging*. Den Haag: Thomson Learning.

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior, 29*, 395-424. doi: 10.1007/s10979-005-5690-7

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science, 7*, 238-259. doi: 10.1177/1745691612439584

Clark, S. E., Howell, R., & Davey, S. D. (2008). Regularities of eyewitness identification. *Law and Human Behavior, 32*, 406-422. doi: 10.1007/s10979-006-9082-4

Clark, S. E., & Wells, G. L. (2008). On the diagnosticity of multiple-witness identifications. *Law and Human Behavior, 25*, 406-422. doi: 10.1007/s10979-007-9115-7

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

Dixon, R. A. (2011). Enduring theoretical themes in psychological aging: Derivation, functions, perspectives, and opportunities. In K. W. Schaie, & S. L. Willis (Eds.), *The handbook of the psychology of aging* (Vol. 7, pp. 3-23). London, UK: Elsevier.

Dodson, C. S., Koutstaal, W., & Schacter, D. L. (2000). Escape from illusion: Reducing false memories. *Trends in Cognitive Sciences, 4*, 391-397. doi:10.1016/S1364-6613(00)01534-5

Dodson, C. S., & Krueger, L. E. (2006). I misremember it well: Why older adults are unreliable witnesses. *Psychonomic Bulletin & Review, 13*, 770-775.

Dunlop, B. D., Rothman, M. B., & Hirt, M. G. (2001). Elders and criminal justice: International issues for the 21stcentury. *International Journal of Law and Psychiatry, 24*, 285-303.

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

Duval, S. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment, and adjustments*. Chichester, UK: Wiley.

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*, 89-98. doi: 10.2307/ 2669529

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*, 455-

463. doi: 10.1111/j.0006-341X.2000.00455.x

Fleiss, J. L., & Berlin, J. A. (2009). Effect sizes for dichotomous data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 237-256). New York: Russell Sage Foundation.

Fulero, S. M. (2009). System and estimator variables in eyewitness identification: A review. In D. A. Krauss, & J. D. Lieberman (Eds.), *Psychological expertise in court*. Burlington, VT: Ashgate.

Gallagher, C., Maguire, E. R., Mastrofski, S. D., & Reisig, M. D. (2001). *The public image of the police*. Final report to The International Association of Chiefs of Police by The Administration of Justice Program, George Mason University. Retrieved from http://www.theiacp.org/PoliceServices/ProfessionalAssistance/ThePublicImageofthePolice/tabid/198/Default.aspx#ch2

*Goodsell, C. A., Neuschatz, J. S., & Gronlund, S. D. (2009). Effects of mugshot commitment on lineup performance in young and older adults. *Applied Cognitive Psychology, 23*, 788-803. doi: 10.1002/acp.1512 [b, f, g]

Harbord, R. M., Harris, R. J., & Sterne, J. A. C. (2006). Updated tests for small-study effects in meta-analyses. *The Stata Journal, 9*, 197-210. doi: 10.1002/sim.2380

Harwell, M. (1997). An investigation of the Raudenbush (1988) test for heterogeneity. *Journal of Experimental Education, 65*, 181-190. doi: 10.1080/00220973.1997.9943791

*Havard, C., & Memon, A. (2009). The influence of face age on identification from a video lineu-up: A comparison between older and younger adults. *Memory, 17*, 847-859. doi: 10.1080/09658210903277318 [a, b, c, d, e, f, g, h]

Hedden, T., & Gabrieli, J. D. E. (2004). Insights into the ageing mind: A view from cognitive neuroscience. *Nature Reviews Neuroscience, 5*, 87-96. doi:10.1038/nrn1323

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL:

Academic Press.

Hedges, L. V., & Pigott, M. A. (2001). The power of statistical tests in meta-analysis. *Psychological Methods, 6*, 203-217. doi: 10.1037/1082-989X.9.4.426

Holliday, R., Ferguson, J., Milne, R., Bull, R. H. C., & Memon, A. (2009). *Isolating the effect of the cognitive interview instructions on witness testimony*. Paper presented at the American Psychology Law Society Conference, San Antonio, TX.

Jackson, D. (2006). The power of the standard test for the presence of heterogeneity in meta-analysis. *Statistics in Medicine, 25*, 2688-2699. doi: 10.1002/sim.2481

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

*Kinlen, T. J., Adams-Price, C. E., & Henley, T. B. (2007). Verbal overshadowing and face recognition in young and old adults. *Educational Gerontology, 33*, 969-979. doi: 10.1080/03601270701632040 [b]

Konstantopoulos, S., & Hedges, L. V. (2009). Analyzing effect sizes: Fixed-effects models. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 279-293). New York: Sage Russell Foundation.

Levine, T., Asada, K. J., & Carpenter, C. (2009). Sample sizes and effect sizes are negatively correlated in meta-analyses: Evidence and implications of a publication bias against nonsignificant findings. *Communication Monographs, 76*, 286-302. doi: 10.1080/03637750903074685

Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology, 42*, 333-376. doi: 10.1146/annurev.ps.42.020191.002001

Lipsey, M. W. (2009). Identifying interesting variables and analysis opportunities. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 147-158). New York: Russell Sage Foundation.

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational and behavioral treatment: Confirmation from meta-analysis. *American Psychologist, 48*, 1181-1209.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage Publications.

Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent development in a "New view". In D. S. Gorfein, & C. M. MacLeod (Eds.), *The place inhibition in cognition* (pp. 145-162). Washington, DC: American Psychology Association.

Martschuk, N., Kocab, K., & Sporer, S. L. (in preparation). *Memory of faces in old age: A meta-analysis*.

Matt, G. E., & Cook, T. D. (2009). Threats to validity generalized inferences. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 537-560). New York: Russell Sage Foundation.

McCabe, K. A., & Gregory, S. S. (1998). Elderly victimization: An examination beyond FBI's index crimes. *Research on Aging, 20*, 363-372. doi: 10.1177/0164027598203005

*Memon, A., & Bartlett, J. C. (2002). The effects of verbalization on face recognition in young and old adults. *Applied Cognitive Psychology, 16*, 635-650. doi: 10.1002/acp.820 [b, c]

*Memon, A., Bartlett, J. C., Rose, R., & Gray, C. (2003). The aging eyewitness: Effects of age on face, delay, and source-memory ability. *Journal of Gerontology: Psychological Sciences, 58B*, 338-345. doi: 10.1093/geronb/58.6.P338 [a, b, c, d, e, f, g, h]

*Memon, A., & Gabbert, F. (2003a). Improving the identification accuracy of senior witnesses: Do prelineup questions and sequential lineup help? *Journal of Applied Psychology, 88*, 341-347. doi: 10.1037/0021-9010.88.2.341 [a, b, c, d]

*Memon, A., & Gabbert, F. (2003b). Unravelling the effects of sequential presentation in culprit-present lineups. *Applied Cognitive Psychology, 17*, 703-714. doi: 10.1002/acp.909 [b, c]

*Memon, A., Gabbert, F., & Hope, L. (2004). The ageing eyewitness. In J. Adler (Ed.), *Forensic psychology: Debates, concepts and practice* (pp. 96-112). Uffcolme, Devon: Willan, Forensic Psychology Series. [d]

*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. doi: 10.3758/BF03213404 *Memory & Cognition, 30*, 1219-1227. [d]

*Memon, A., Hope, L., & Bull, R. (2003). Exposure duration: Effects on eyewitness accuracy and confidence. *British Journal of Psychology, 94*, 339-354. doi: 10.1348/000712603767876262 [a, b, c, d]

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, & Law, 16*, 340-372. doi: 10.1037/a0020518

Nilsson, L.-G. (2003). Memory function in normal aging. *Acta Neurologica Scandinavica Supplement, 179*, 7-13. doi: 10.1034/j.1600-0404.107.s179.5.x

*O'Rourke, T. E., Penrod, S. D., Cutler, B. L., & Stuve, T. E. (1989). The external validity of eyewitness identification research: Generalizing across subject populations. *Law and Human Behavior, 13*, 385-395. doi: 10.1007/BF01056410 [a]

Old, S. R., & Naveh-Benjamin, M. (2008). Differential effects of age on items and associative measures of memory: A meta-analysis. *Psychology and Aging, 23*, 104-118. doi: 10.1037/0882-7974.23.1.1

Orwin, R. G., & Vevea, J. L. (2009). Evaluating coding decisions. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 177-203). New York: Russell Sage Foundation.

Palmer, M.A., & Brewer, N. (2012). Sequential lineup presentation promotes less-biased criterion setting but does not improve discriminability. *Law and Human Behavior, 36,* 247-255.

Palmer, M. A., Brewer, N., McKinnon, A. C., & Weber, N. (2010). Phenomenological reports diagnose accuracy of eyewitness identification decisions. *Acta Psychologica, 133*, 137-145. doi: 1*0.1016/j.actpsy.2009.11.002*

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: Aging and neurocognitive scaffolding. *Annual Review of Psychology, 60*, 173-196. doi: 10.1146/annurev.psych.59.103006.093656

Perfect, T. J., & Harris, L. J. (2003). Adult age differences in unconscious transference: Source confusion or identity blending? *Memory & Cognition, 31*, 570-580. doi: 10.3758/BF03196098

Pigott, T.D. (2012). *Advances in Meta-analysis*. Heidelberg: Springer Link.

Ratcliff, R., Thapar, A., & McKoon, G. (2011). Effects of aging and IQ on item and associative memory. *General Journal of Experimental Psychology, 140*, 464-487. doi: 10.1037/a0023810

Raudenbush, S. W. (2009). Handling missing data. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 399-416). New York: Russell Sage Foundation.

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

*Rose, R. A., Bull, R., & Vrij, A. (2003). Enhancing older witnesses' identification performance: Context reinstatement is not the answer. *Canadian Journal of Police and Security Services, 1*, 173-184. [a, b, c, d, e, f, g, h]

*Rose, R. A., Bull, R., & Vrij, A. (2005). Non-biased lineup instruction do matter: A problem for older witnesses. *Psychology, Crime, & Law, 11*, 147-159. doi: 10.1080/10683160512331316307 [a, b, c, d, e, f, g, h]

Salthouse, T. A. (1996). Constrains on theories of cognitive aging. *Psychonomic Bulletin & Review, 3*, 287-299. doi: 10.3758/BF03210753

Sauerland, M., & Sporer, S. L. (2008). The application of multiple lineups in a field study. *Psychology, Crime, & Law, 14*, 549-564. doi:

10.1080/09541440701728417

Sauerland, M., & Sporer, S. L. (2009). Fast and confident: Postdicting eyewitness identification accuracy in a field study. *Journal of Experimental Psychology-Applied, 15*, 46-62. doi: 10.1037/a0014560

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

*Searcy, J., Bartlett, J. C., & Memon, A. (1999). Age differences in accuracy and choosing in eyewitness identification and face recognition. *Memory & Cognition, 27*, 538-552. [a, b, c, d]

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

*Searcy, J., Bartlett, J. C., Memon, A., & Swanson, K. (2001). Aging and lineup performance at long retention intervals: Effects of metamemory and context reinstatement. *Journal of Applied Psychology, 86*, 207-214. doi: 10.1037/0021-9010.86.2.207 [a]

Shadish, W. R., & Haddock, C. K. (2009). Combining estimates of effect sizes. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 257-277). New York: Russell Sage Foundation.

Snyder, H. N. (2012). *Arrest in the United States, 1990-2010*. Report for the U.S. Department of Justice. Retrieved from http://bjs.gov/index.cfm?ty=pbdetail&iid=4515

Spencer, W. D., & Raz, N. (1995). Differential effects of aging on memory for content and context: A meta-analysis. *Psychology and Aging, 10*, 527-539. doi: 10.1037//0882-7974.10.4.527

Sporer, S.L. (2001). Recognizing faces of other ethnic groups: An integration of theories. *Psychology, Public Policy, and Law, 7*, 36-97.

Sporer, S. L., & Cohn, L. D. (2011). Meta-analysis. In B. Rosenfeld, & S. D. Penrod (Eds.), *Research methods in psychology*. New York: Wiley.

Sporer, S. L., & Goodman-Delahunty, J. (2011). *Publication bias in meta-analyses in psychology and law*. Paper presented at the Annual meeting of the American-Psychology - Law Society, Miami, FL.

Sporer, S. L., & Martschuk, N. (in press). The reliability of eyewitness identifications by the elderly: An evidence-based review. To appear in: M. Toglia, D. Ross, J.Pozzulo, & E. Pica (Eds.) The elderly witness in Court. New York: Psychology Press.

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

Steblay, N. K. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21*, 283-297. doi: 10.1023/A:1024890732059

Steblay, N. K., Dysart, J. E., & Wells, G. L. (2011). Seventy-two tests of sequential lineup superiority effect: A meta-analysis and police discussion. *Psychology, Public Policy, & Law, 17*, 99-139. doi: 10.1037/a0021650

Sterne, J. A. C., Becker, B. J., & Egger, M. (2005). The funnel plot. In A. J. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis. Prevention, assessment and adjustments* (pp. 75-98). Chichester, UK: Wiley.

Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting publication bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology, 54*, 1046-1055. doi: 10.1016/S0895-4356(01)00377-8

Sterne, J. A. C., Gavaghan, D., & Egger, M. (2000). Publication bias and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology, 53*, 1119-1129.

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 435-452). New York: Russell Sage Foundation.

Thomas, A. K., & Bulevich, J. B. (2006). Effective cue utilization reduces memory errors in older adults. *Psychology and Aging, 21*, 379-389. doi: 10.1037/0882-7974.21.2.379

U.S. Bureau of the Census (2011). *Population and Housing Unit Estimates.* Report for the U.S. Department of Commerce. Retrieved from http://www.census.gov/popest/data/index.html

*Vom Schemm, K., Kraus, U., & Köhnken, G. (2007). Tatverdächtigenidentifizierung im Altersvergleich. In C. Lorei (Ed.), *Polizei & psychologie 2006. Schriftreihe Polizei & Wissenschaft*. Frankfurt: Verlag für Polizeiwissenschaft. [d]

Weber, N., & Perfect, T.J. (2012). Improving eyewitness identification accuracy by screening out those who say they don't know. *Law and Human Behavior, 36,* 28-36.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67*, 1049-1062. doi: 10.1037/0022-3514.67.6.1049

Wells, G. L., & Lindsay, R. C. L. (1980). On estimating the diagnosticity of eyewitness non-identifications. *Psychological Bulletin, 88*, 776-784. doi: 10.1037/0033-2909.88.3.776

Wells, G. L., & Olsen, E. A. (2003). Eyewitness testimony. *Annual Review of Psychology, 54*, 277-295. doi: 10.1146/annurev.psych.54.101601.145028

Wells, G. L., & Turtle, J. W. (1986). Eyewitness identification: The importance of lineup models. *Psychological Bulletin, 99*, 320-329. doi: 10.1037/0033-2909.99.3.320

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25*, 1115-1125. doi: 10.1177/01461672992512005

White, H. D. (2009). Scientific communication and literature retrieval. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 51-71). New York: Russell Sage Foundation.

*Wilcock, R. A., & Bull, R. (2010). Novel lineup methods for improving the performance of older eyewitnesses. *Applied Cognitive Psychology, 24*, 718-736. doi: 10.1002/acp.1582 [a, b, c, d]

*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, & Law, 12*, 129-140. doi: 10.1375/pplt.2005.12.1.129 [a, b, c, d, e, f, g, h]

*Wilcock, R. A., Bull, R., & Vrij, A. (2007). Are older witnesses always poorer witnesses? Identification accuracy, context reinstatement, own age bias. *Psychology, Crime, & Law, 13*, 305-316. doi: 10.1080/10683160600822212 [a, b, c, d, e, f, g, h]

Wilson, D. B. (2009). Systematic coding. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 159-176). New York: Russell Sage Foundation.

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*, 641-654. doi: 10.1023/A:1020981501383

Yarmey, A. (1996). The elderly witness. In S. L. Sporer, R. Malpass, & G. Köhnken (Eds.), *Psychological issues in eyewitness identification*. Mahwah, NJ: Lawrence Erlbaum.

Appendices

Appendix A

Table A1

Study Characteristics of all Primary Studies

Authors	Year	Type of presentation	Type of crime	Encoding (s)	Retention interval (min)	Lineup type	Type of lineup presentation	# Persons in lineup	Mean age of older group
Adams-Price	1992	video	theft	30	10	TP	simultaneous	15	67.50
Goodsell et al. (No mugshot search)[a]	2009	video	theft	25	10080	TP / TA	simultaneous	6	70.20
Havard & Memon	2009	video	theft	90	40-60	TP / TA	sequential	9	72.63
Kinlen et al. (Verbalization)[b]	2007	video	robbery	30	25	TP	sequential	8	71.70
Kinlen et al. (Visualization)[b]	2007	video	robbery	30	25	TP	sequential	8	71.70
Kinlen et al. (Control)[b]	2007	video	robbery	30	25	TP	sequential	8	71.70
Memon & Bartlett (Sim+Desc)[d]	2002	video	theft	60	60	TP	simultaneous	6	69.90
Memon & Bartlett (Sim+No desc)[d]	2002	video	theft	60	60	TP	simultaneous	6	69.90

Table A1 (continued)

Authors	Year	Type of presentation	Type of crime	Encoding (sec)	Retention interval (min)	Lineup type	Type of lineup presentation	# Persons in lineup	Mean age of older group
Memon & Bartlett (Seq+Desc)[d]	2002	video	theft	60	60	TP	sequential	6	69.90
Memon & Bartlett (Seq+No desc)[d]	2002	video	theft	60	60	TP	sequential	6	69.90
Memon, Bartlett, et al. (35 min)[c]	2003	video	theft	43	35	TP / TA	simultaneous	6	71.70
Memon, Bartlett, et al. (1 week)[c]	2003	video	theft	43	10080	TP / TA	simultaneous	6	71.70
Memon & Gabbert (Sim)[e]	2003a	video	theft	60	60	TP / TA	simultaneous	6	68.91
Memon & Gabbert (Seq)[e]	2003a	video	theft	60	60	TP / TA	sequential	6	68.91
Memon & Gabbert (Sim+App changed)[f]	2003b	video	theft	15	60	TP	simultaneous	6	69.20
Memon & Gabbert (Seq+App changed)[f]	2003b	video	theft	15	60	TP	sequential	6	69.20
Memon & Gabbert (Sim+No changes)[f]	2003b	video	theft	15	60	TP	simultaneous	6	69.20

Table A1 (continued)

Authors	Year	Type of presentation	Type of crime	Encoding (sec)	Retention interval (min)	Lineup type	Type of lineup presentation	# Persons in lineup	Mean age of older group
Memon & Gabbert (Seq+No changes)[f]	2003b	video	theft	15	60	TP	sequential	6	69.20
Memon et al.	2004	video	no crime	na	10080	TA	na	na	69.00
Memon et al.	2002	video	theft	26	2880	TA	simultaneous	6	69.00
Memon et al. (12 sec)[g]	2003	video	robbery	12	35	TP / TA	simultaneous	6	68.00
Memon et al. (45 sec)[g]	2003	video	robbery	45	35	TP / TA	simultaneous	6	68.00
O'Rourke et al. (18-19 vs. 60+)[h]	1989	video	robbery	75	10080	TP / TA	sequential	6	66.00
O'Rourke et al. (20-29 vs. 50-59)[h]	1989	video	robbery	75	10080	TP / TA	sequential	6	54.50
Rose et al.	2003	video	violence	11	30	TP / TA	simultaneous	6	72.39
Rose et al.	2005	video	violence	11	30	TP / TA	both	6	70.67
Scogin et al.	1994	video	violence	6	na	TP / TA	simultaneous	6	74.20

Table A1 (continued)

Authors	Year	Type of presentation	Type of crime	Encoding (sec)	Retention interval (min)	Lineup type	Type of lineup presentation	# Persons in lineup	Mean age of older group
Searcy et al.	1999	video	robbery	60	15; 60	TP / TA	both	6	70.00
Searcy et al. (Sim)[i]	2000	video	violence	90	690	TA	simultaneous	6	69.30
Searcy et al. (Seq)[i]	2000	video	violence	90	690	TA	sequential	6	69.30
Searcy et al.	2001	live	no crime	1200	1 month	TP / TA	simultaneous	6	71.00
Vom Schemm et al.	2007	video	robbery	40	75	TA	sequential	6	69.35
Wilcock et al.	2005	video	violence	7	30	TP / TA	both	6	71.40
Wilcock et al. (Context)[j]	2007	video	violence	7	30	TP / TA	simultaneous	6	72.60
Wilcock et al. (Control)[j]	2007	video	violence	7	30	TP / TA	simultaneous	6	72.60
Wilcock & Bull (Prelineup questions)[k]	2010	video	theft	6	30	TP / TA	simultaneous	6	68.90
Wilcock & Bull (Practice lineup)[k]	2010	video	theft	6	30	TP / TA	simultaneous	6	68.90

Table A1 (continued)

Authors	Year	Type of presentation	Type of crime	Encoding (sec)	Retention interval (min)	Lineup type	Type of lineup presentation	# Persons in lineup	
Wilcock & Bull (Control)[k]	2010	video	theft	6	30	TP / TA	simultaneous	6	68.90
Excluded									
Wright & Stroud (Exp. 1; 1 day)[l]	2002	video	theft	67.5	1440	TP	simultaneous	6	45.00
Wright & Stroud (Exp. 1; 1 week)[l]	2002	video	theft	67.5	10080	TP	simultaneous	6	45.00
Wright & Stroud (Exp. 2)[l]	2002	video	theft	67.5	1440	both	simultaneous	6	47.50

Note. If only age ranges were reported for the groups, the central point was extracted as the mean age of the group.

[a] Goodsell et al. (2009): Mugshot search condition excluded

[b] Kinlen et al. (2007): Prelineup tasks manipulated: Verbalization = Verbalization group; Visualization = Visualization group; Control = Control group

[d] Memon & Bartlett (2002): Lineup presentation + description of suspect before ID: Sim+Desc = simultaneous lineup + description; Sim+No desc = simultaneous + no description; Seq+Desc = sequential + description; Seq+No desc = sequential + no description

[c] Memon, Bartlett, et al. (2003): Retention interval manipulated: 35 min = retention interval of 35 minutes; 1 week = retention interval of 1 week

[e] Memon & Gabbert (2003a): Lineup presentation manipulated: Sim = simultaneous; Seq = sequential

[f] Memon & Gabbert (2003b): Lineup presentation + appereance of suspect manipulated: Sim+App changed = simultaneous lineup + appereance of suspect changed; Seq+App changed = sequential + appereance changed; Sim+No changes = simultaneous + appereance not changed; Seq+No changes = sequential + appereance not changed

[g] Memon et al. (2003): Exposure time of target manipulated: 12 sec = target presented for 12 seconds; 45 sec = target presented for 45 seconds

[h] O'Rourke et al. (1989); Multiple age groups of participants: 18-19 vs. 60+ = 18-19 years vs. 60+ years; 20-29 vs. 50-59 = 20-29 years vs. 50-59 years

[i] Searcy et al. (2000): Lineup presentation manipulated: Sim = simultaneous; Seq = sequential

[j] Wilcock et al. (2007): Context reinstatement manipulated: Context = Context reinstatement group; Control = Control group

[k] Wilcock & Bull (2010): Prelineup tasks manipulated: Prelineup questions = prelineup instruction groups; Practice lineup = practice lineup groups; Control = control groups

[l] Wright & Stroud (2002). Excluded due to exceptional low age thresholds in the older group

Appendix B

Table B1

Single Effect sizes (OR) of the Individual Conditions in Primary Studies for Young Targets

Authors	Year	*N*	n_y	n_o	Correct decisions	Hits (TP)	Filler ID (TP)	Foil ID (TA)	Choosing
Adams-Price[a]	1992	80	40	40	na	1.5555	na	na	
Goodsell et al. (No mugshot search)[b]	2009	48	27	21	na	2.5000	1.0461	na	
Havard & Memon	2009	88	45	43	3.7333	4.0800	3.7502	3.9598	1.4623
Kinlen et al. (Verbalization)[c]	2007	37	20	17	na	0.0988	na	na	
Kinlen et al. (Visualization)[c]	2007	37	20	17	na	0.5926	na	na	
Kinlen et al. (Control)[c]	2007	37	20	17	na	0.9872	na	na	
Memon & Bartlett (Sim+Desc)[e]	2002	40	19	21	na	0.4334	4.0000	na	
Memon & Bartlett (Sim+No desc)[e]	2002	40	20	20	na	0.6581	1.2857	na	
Memon & Bartlett (Seq+Desc)[e]	2002	30	15	15	na	1.0000	1.8333	na	

Table B1 (continued)

Authors	Year	*N*	n_y	n_o	Correct decisions	Hits (TP)	Filler ID (TP)	Foil ID (TA)	Choosing
Memon & Bartlett (Seq+No desc)[e]	2002	30	15	15	na	1.0000	2.2858	na	
Memon, Bartlett, et al. (35 min)[d]	2003	84	43	41	1.3886	0.8308	2.3540	2.8126	4.7427
Memon, Bartlett, et al. (1 week)[d]	2003	87	41	46	4.5240	1.5256	32.3075	17.0628	7.1256
Memon & Gabbert (Sim)[f]	2003a	120	60	60	2.2133	1.0000	0.8571	7.8751	2.5000
Memon & Gabbert (Seq)[f]	2003a	120	60	60	1.6000	0.8000	8.1052	6.0002	4.6772
Memon & Gabbert (Sim+App changed)[g]	2003b	45	25	20	na	3.5001	4.0000	na	
Memon & Gabbert (Seq+App changed)[g]	2003b	45	25	20	na	2.6666	61.7380	na	
Memon & Gabbert (Sim+No changes)[g]	2003b	45	25	20	na	6.3751	7.6668	na	
Memon & Gabbert (Seq+No changes)[g]	2003b	45	25	20	na	2.7693	7.3331	na	
Memon et al.	2004	63	32	31	na	na	na	1.5722	
Memon et al.	2002	169	84	85	na	na	na	3.9287	

Table B1 (continued)

Authors	Year	*N*	n_y	n_o	Correct decisions	Hits (TP)	Filler ID (TP)	Foil ID (TA)	Choosing
Memon et al. (12 sec)[h]	2003	82	42	40	0.6203	0.8000	1.0909	0.4474	1.1093
Memon et al. (45 sec)[h]	2003	82	42	40	1.5408	3.3528	2.2222	1.5000	1.3182
O'Rourke et al. (18-19 vs. 60+)	1989	51	39	12	3.1579	na	na	na	
O'Rourke et al. (20-29 vs. 50-59)	1989	29	15	14	2.1876	na	na	na	
Rose et al.	2003	72	36	36	5.3079	10.0001	13.6004	10.8179	3.1528
Rose et al	2005	96	48	48	1.1819	1.1819	1.8000	2.3333	1.9656
Scogin et al.	1994	84	27	57	2.8519	2.6666	13.7494	3.4199	4.4674
Searcy et al.	1999	134	76	58	1.9721	1.5714	2.3375	2.4305	2.2218
Searcy et al. (Sim)[i]	2000	48	24	24	na	na	na	1.0000	
Searcy et al. (Seq)[i]	2000	50	25	25	na	na	na	0.7159	

Table B1 (continued)

Authors	Year	N	n_y	n_o	Correct decisions	Hits (TP)	Filler ID (TP)	Foil ID (TA)	Choosing
Searcy et al.	2001	94	45	49	2.5782	na	na	na	
Vom Schemm et al.	2007	80	40	40	na	na	na	8.7548	
Wilcock et al.	2005	96	48	48	1.9697	1.9697	3.26667	1.1819	1.0880
Wilcock et al. (Context)[j]	2007	47	24	23	6.2451	0.8750	41.6666	6.9999	3.8183
Wilcock et al. (Control)[j]	2007	49	25	24	2.1778	3.2001	8.5725	15.0007	10.7995
Wilcock & Bull (Prelineup questions)[k]	2010	66	33	33	2.1251	3.0557	12.4448	1.8668	1.4480
Wilcock & Bull (Practice Lineup)[k]	2010	66	33	33	1.2894	-0.9000	0.6429	1.5400	1.2821
Wilcock & Bull (Control)[k]	2010	64	32	32	3.2858	1.8000	1.0000	6.6002	2.1428

Table B1 (continued)

Studies excluded

Authors	Year	N	n_y	n_o	Correct decisions	Hits (TP)	Filler ID (TP)	Foil ID (TA)	Choosing
Wright & Stroud (Exp. 1; 1 day)[l]	2002	59	30	29	na	2.7500	1.2187	na	
Wright & Stroud (Exp. 1; 1 week)[l]	2002	54	26	28	na	1.6297	0.7211	na	
Wright & Stroud (Exp. 2)[l]	2002	180	90	90	1.5371	2.0119	1.9523	1.1034	

[a] Adams-Price (1992): Results averaged across film A and B because manipulated as within-variable.

[b] Goodsell et al. (2009): Mugshot search condition excluded.

[c] Kinlen et al. (2007): Prelineup tasks manipulated: Verbalization = Verbalization group; Visualization = Visualization group; Control = Control group.

[e] Memon & Bartlett (2002): Lineup presentation + description of suspect before ID manipulated: Sim+Desc = simultaneous lineup + description; Sim+No Desc = simultaneous + no description; Seq+Desc = sequential + description; Seq+No Desc = sequential + no description.

[d] Memon, Bartlett, et al. (2003): Retention interval manipulated: 35 min = retention interval of 35 minutes; 1 week = retention interval of 1 week

[f] Memon & Gabbert (2003a): Lineup presentation manipulated: Sim = simultaneous; Seq = sequential.

[g] Memon & Gabbert (2003b): Lineup presentation + appereance of suspect manipulated: Sim+App changed = simultaneous lineup + appereance of suspect changed; Seq+App changed = sequential + appereance changed; Sim+No changes = simultaneous + appereance not changed; Seq+No changes = sequential + appereance not changed.

[h] Memon et al. (2003): Exposure time of target manipulated: 12 sec = target presented for 12 seconds; 45 sec = target presented for 45 seconds.

[i] Searcy et al. (2000): Lineup presentation manipulated: Sim = simultaneous; Seq = sequential

[j] Wilcock et al. (2007): Context reinstatement manipulated: Context = Context reinstatement group; Control = Control group

[k] Wilcock & Bull (2010): Prelineup tasks manipulated: Prelineup questions = prelineup questions group; Practice lineup = practice lineup group; Control = control group.

[l] Wright & Stroud (2002). Excluded due to exceptional low age thresholds in the older group

Appendix C

Distribution of Single Effect Sizes and their Confidence Intervals for each Dependent Variable

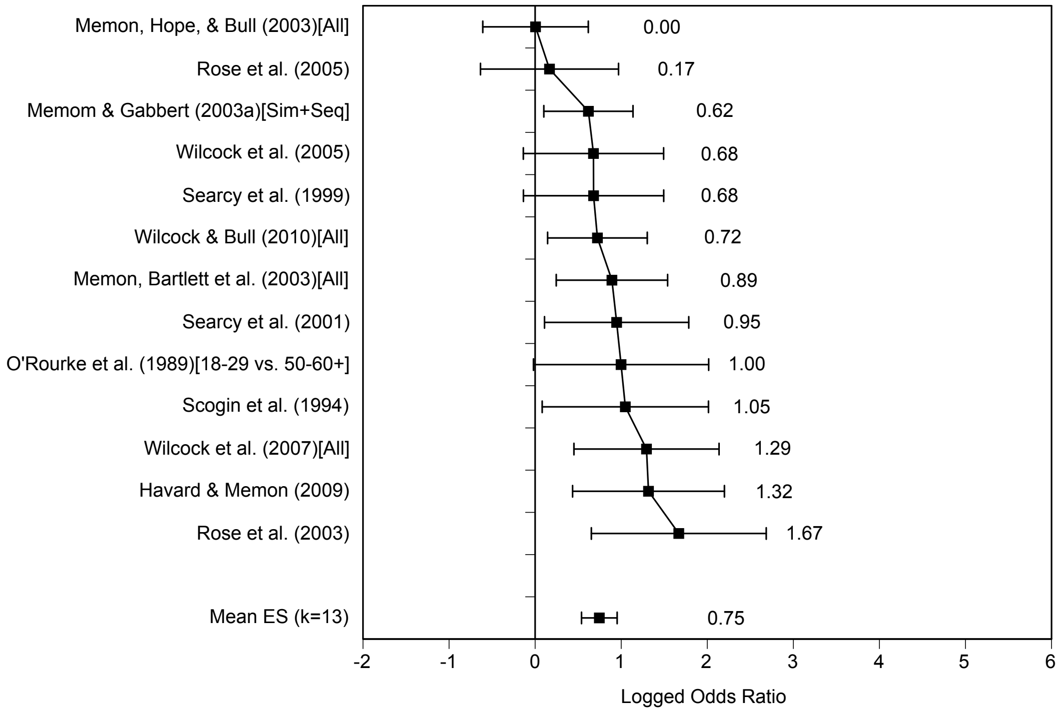

Figure C1. Distribution of effect sizes of the age effect for correct decisions via TP and TA lineups for young targets.

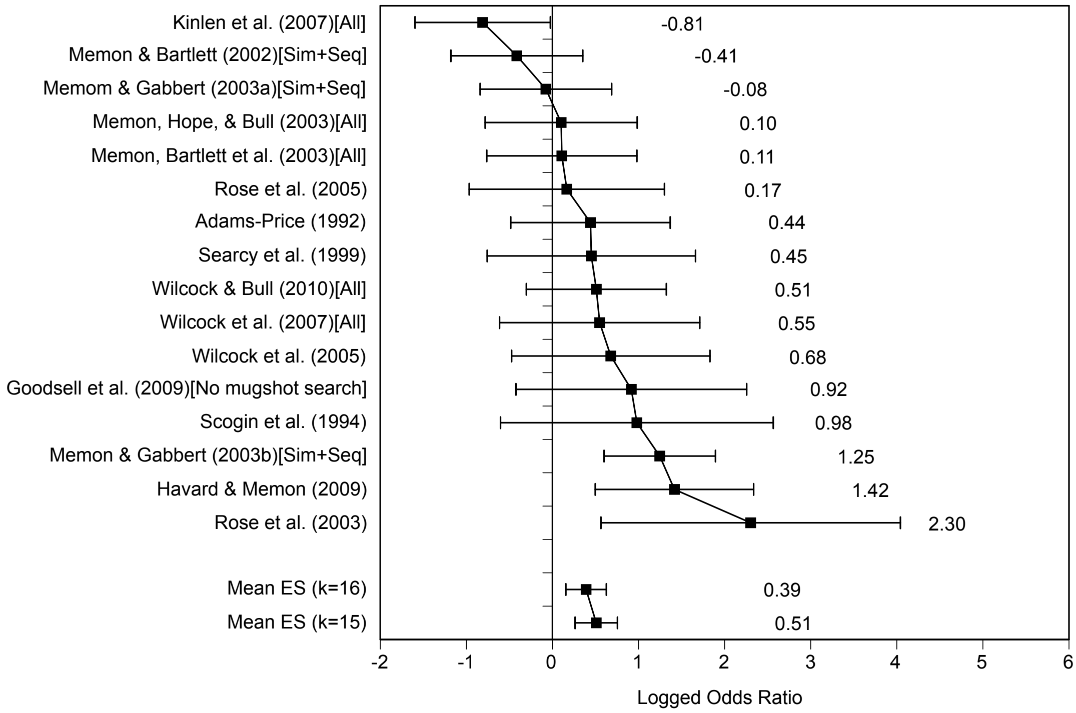

Figure C2. Distribution of effect sizes of the age effect for correct identifications in target-present lineups for young targets. Weighted mean effect size with (*k*=16) and without outlier (*k*=15).

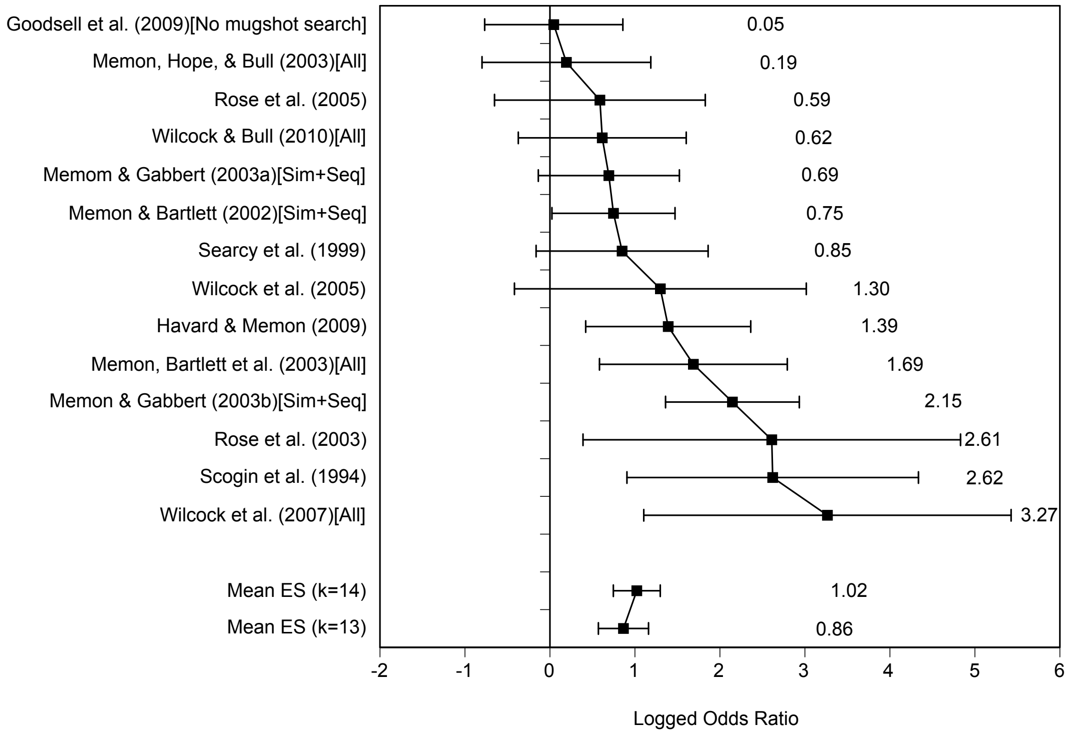

Figure C3. Distribution of effect sizes of the age effect for false identifications in target-present lineups for young targets. Weighted mean effect size with (*k*=14) and without outlier (*k*=13).

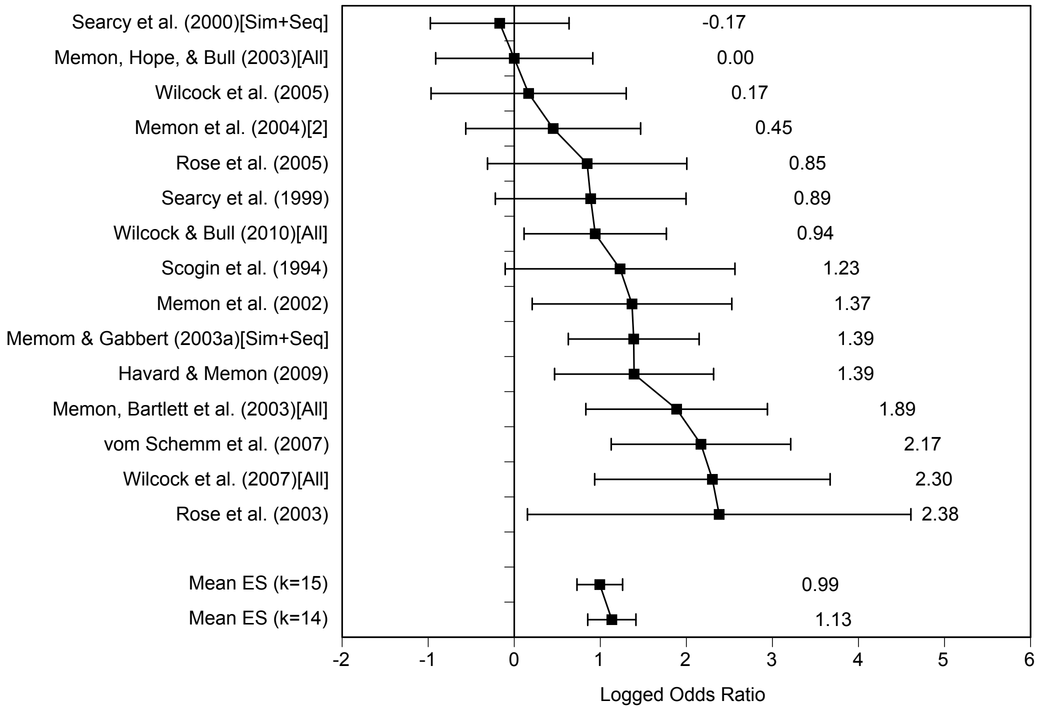

Figure C4. Distribution of effect sizes of the age effect for false identifications in target-absent lineups for young targets. Weighted mean effect size with (*k*=15) and without outlier (*k*=14).

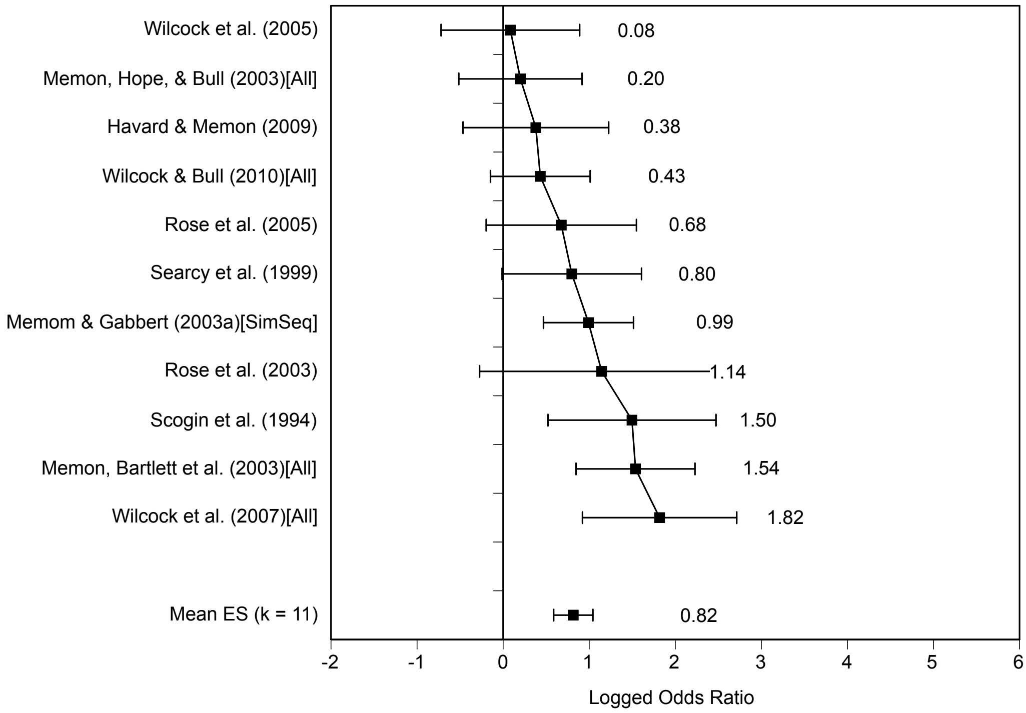

Figure C5. Distribution of effect sizes of the age effect for choosing for young targets.

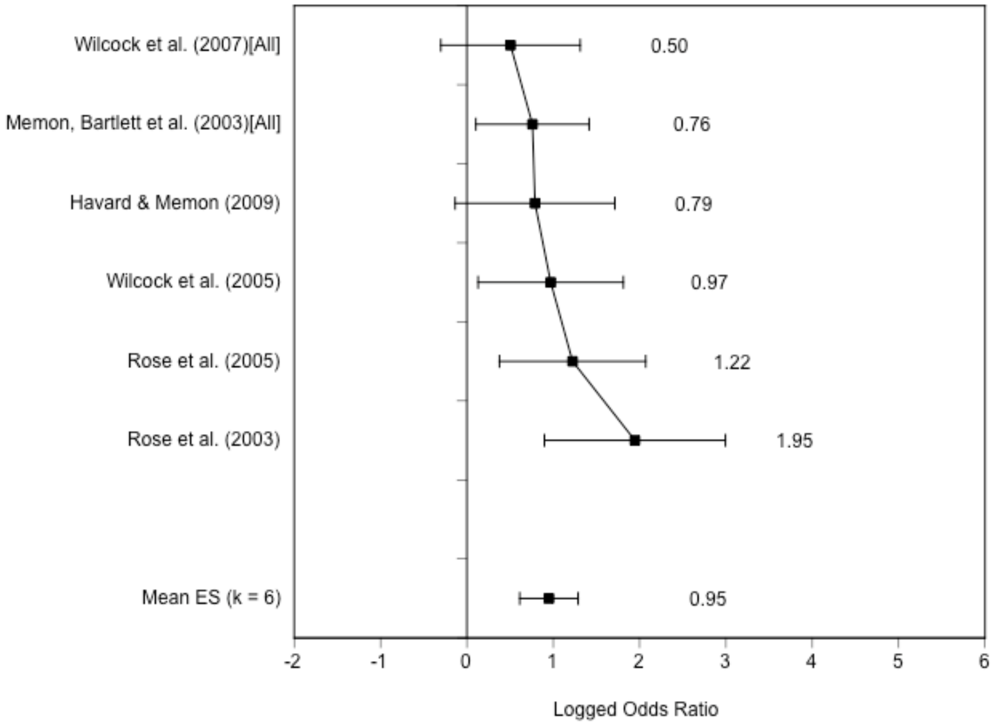

Figure C6. Distribution of effect sizes of the age effect for correct decisions via TP and TA lineups for old targets.

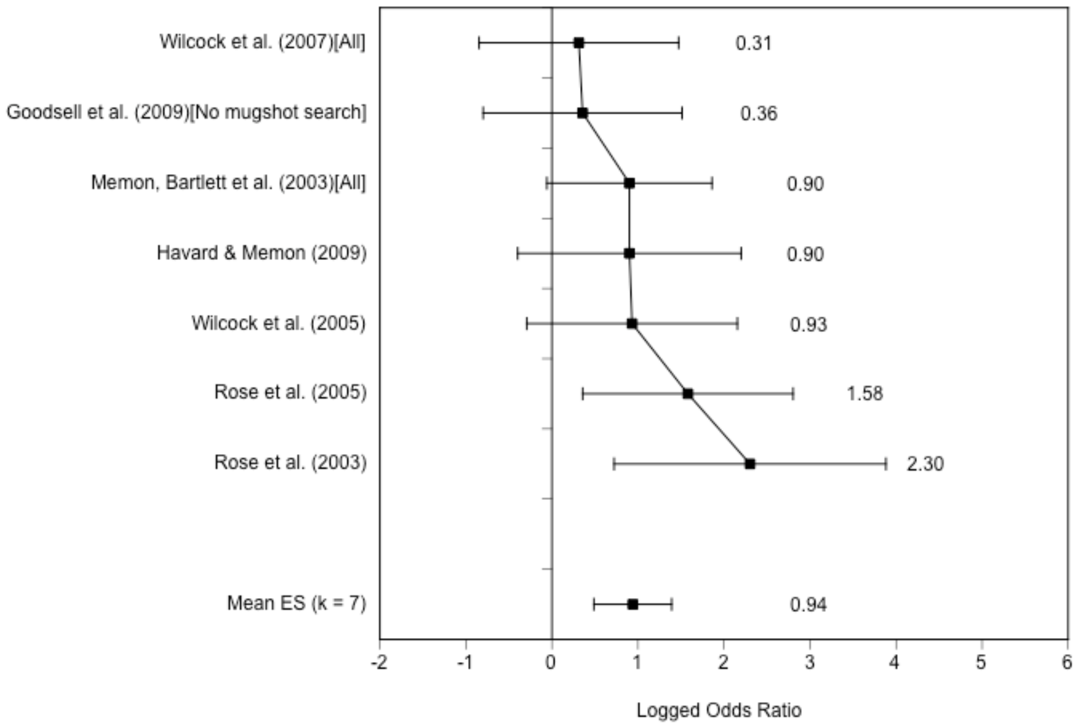

Figure C7. Distribution of effect sizes of the age effect for correct identifications in target-present lineups for old targets.

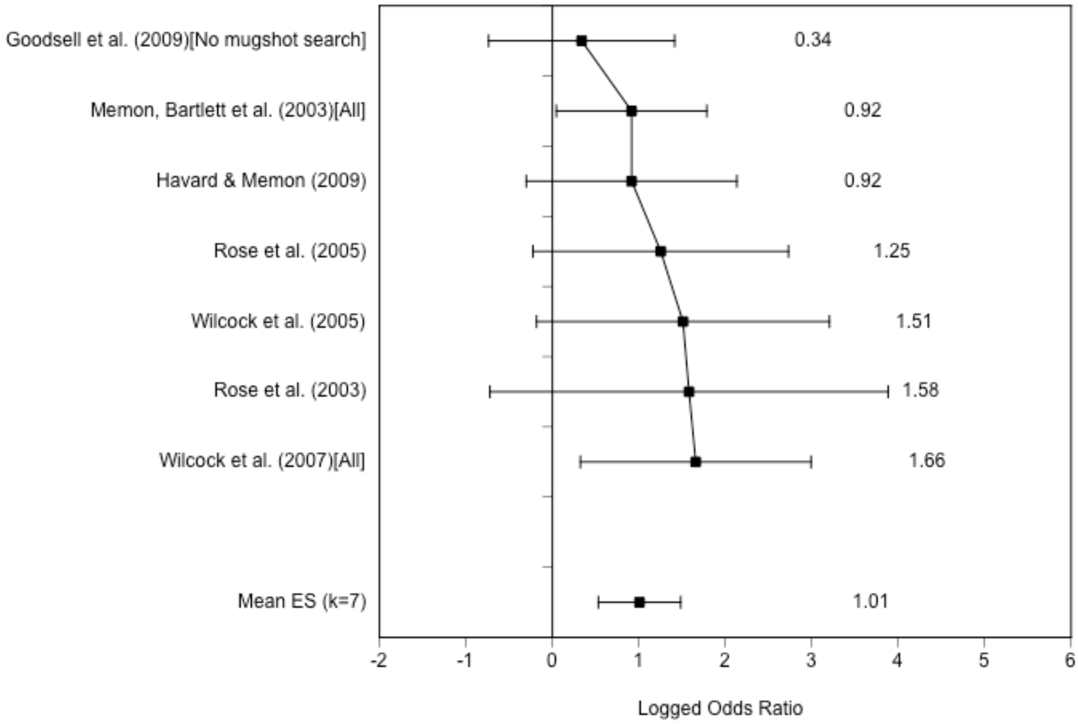

Figure C8. Distribution of effect sizes of the age effect for false identifications in target-present lineups for old targets.

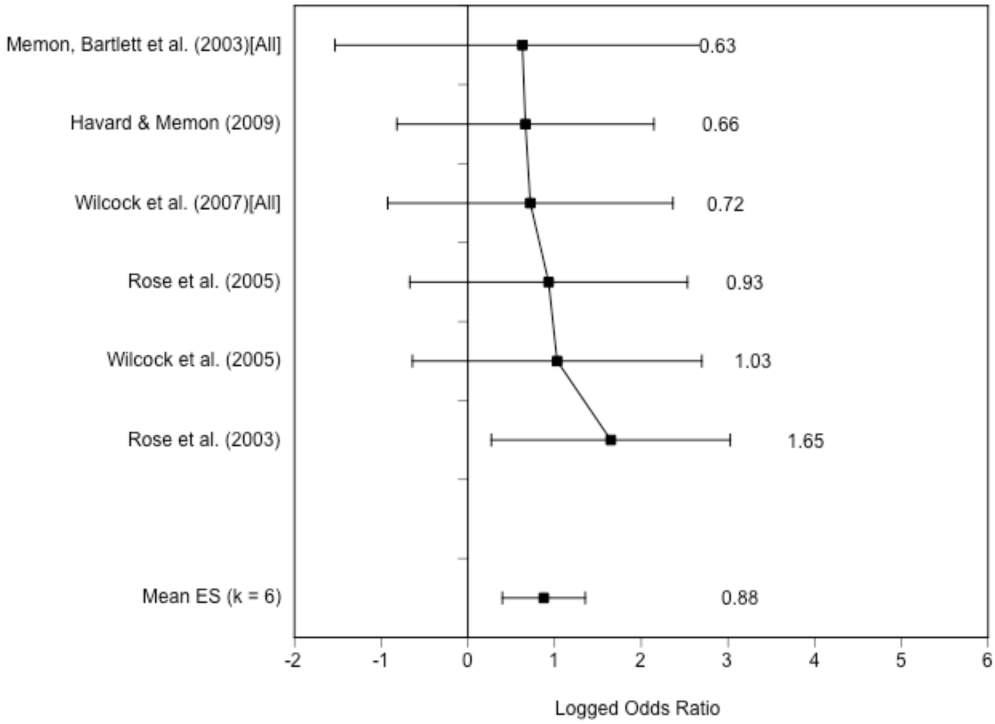

Figure C9. Distribution of effect sizes of the age effect for false identifications in target-absent lineups for old targets.

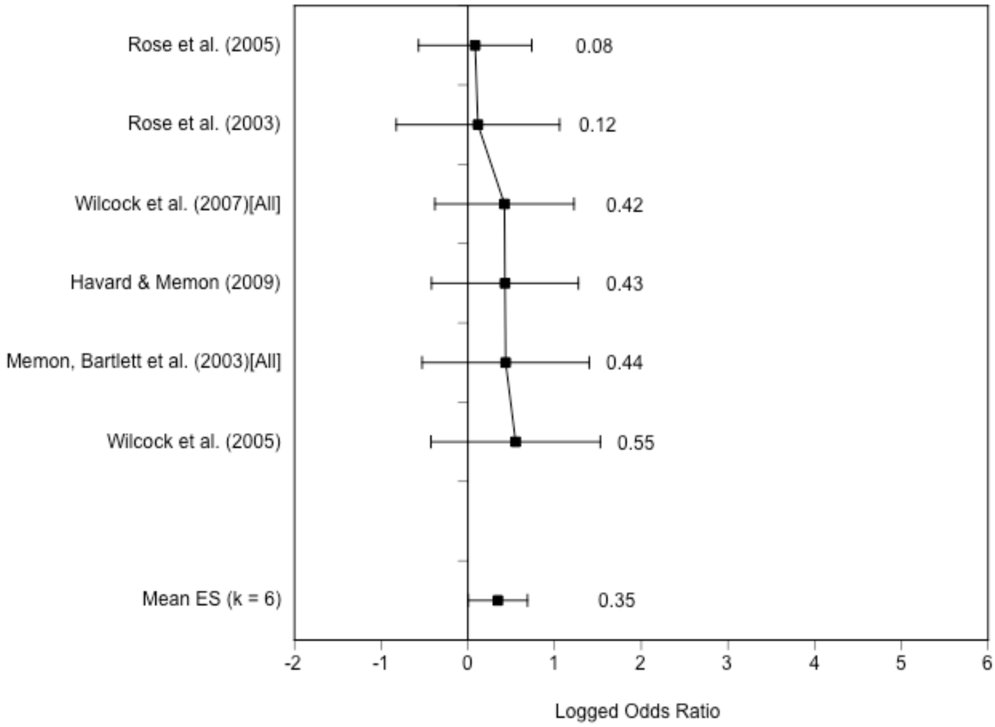

Figure C10. Distribution of effect sizes of the age effect for choosing for old targets.

Appendix D

Table D1

Hierarchical Weighted Fixed-Effect Meta-regression for Correct Decisions via TP and TA Lineups for Young Targets (k = 12)

Study characteristics	*B*	*beta*	*p*	R^2
Step 1:				
Model $Q(1) = 8.28$, $p = .004$				.548
Residual $Q(10) = 6.82$, $p = .743$				
Mean age of the older group	.177	.741	.004	
Step 2:				
Model $Q(4) = 9.57$, $p = .048$				.634
Residual $Q(7) = 5.53$, $p = .596$				
Type of lineup (simultaneous vs. sequential)	-.186	-.265	.434	
Exposure Time (1-29 s, ≥ 30 s)	.347	.413	.260	
Delay (< 24 hours, ≥ 24 hours)	-.344	-.309	.424	
Regression constant	-13.109			

Note. Q statistics shown are from fixed-effects meta-regression models and test whether (1) mean age of the old-age group, and (2) the set of moderator variables in the model accounts for an additional portion of the variance in effect sizes; the study conducted by O'Rourke, Penrod, Cutler, & Stuve (1989) was excluded due to an exceptional low mean age of the older group.

Table D2

Hierarchical Weighted Fixed-Effect Meta-regression for Correct Identifications in TP Lineups for Young Targets (k = 15)

Study characteristics		B	beta	p	R^2
Step 1:					
Model	$Q(1) = 1.38$, $p = .239$				.065
Residual	$Q(13) = 20.04$, $p = .094$				
Mean age of the older group		.093	.254	.240	
Step 2:					
Model	$Q(4) = 8.58$, $p = .073$				.400
Residual	$Q(10) = 12.85$, $p = .232$				
Type of lineup (simultaneous vs. sequential)		.316	.303	.255	
Exposure Time (1-29 s, ≥ 30 s)		-.841	-.699	.010	
Delay (< 24 hours, ≥ 24 hours)		.454	.247	.404	
Regression constant		-6.572			

Note. Q statistics shown are from fixed-effects meta-regression models and test whether (1) mean age of the old-age group, and (2) the set of moderator variables in the model accounts for an additional portion of the variance in effect sizes.

Table D3

Hierarchical Weighted Fixed-Effect Meta-regression for Foil Identification in TP Lineups for Young Targets (k = 13)

Study characteristics	*B*	*beta*	*p*	R^2
Step 1:				
Model *Q*(1) = 11.86, *p* = .001				.651
Residual *Q*(11) = 6.360, *p* = .848				
Mean age of the older group	.371	.807	.001	
Step 2:				
Model *Q*(4) = 13.88, *p* = .008				.762
Residual *Q*(8) = 4.33, *p* = .826				
Type of lineup (simultaneous vs. sequential)	-.404	-.352	.267	
Exposure Time (1-29 s, ≥ 30 s)	.139	.098	.744	
Delay (< 24 hours, ≥ 24 hours)	-.729	-.383	.234	
Regression constant	-28.269			

Note. Q statistics shown are from fixed-effects meta-regression models and test whether (1) mean age of the old-age group, and (2) the set of moderator variables in the model accounts for an additional portion of the variance in effect sizes.

Table D4

Hierarchical Weighted Fixed-Effect Meta-regression for Foil Identifications in TA Lineups for Young Targets (k = 13)

Study characteristics	*B*	beta	*p*	R^2
Step 1:				
Model $Q(1) = 2.33$, $p = .127$				.118
Residual $Q(11) = 17.43$, $p = .096$				
Mean age of the older group	.135	.344	.127	
Step 2:				
Model $Q(4) = 8.20$, $p = .085$				.415
Residual $Q(8) = 11.57$, $p = .172$				
Type of lineup (simultaneous vs. sequential)	.058	.062	.840	
Exposure Time (1-29 s, ≥ 30 s)	.651	.482	.131	
Delay (< 24 hours, ≥ 24 hours)	.168	.088	.777	
Regression constant	-10.439			

Note. Q statistics shown are from fixed-effects meta-regression models and test whether (1) mean age of the old-age group, and (2) the set of moderator variables in the model accounts for an additional portion of the variance in effect sizes; the study conducted by Memon, Gabbert, & Hope (2004) had to be excluded from this analyses due to missing information.

Appendix E

Table E1

Single Effect Sizes (LOR) of the Primary Studies for Old Targets

Authors	Year	*N*	n_y	n_o	Correct decisions	Hits (TP)	Filler ID (TP)	Foil ID (TA)	Choosing
Goodsell et al. (No mugshot search) [a]	2009	55	29	26	na	0.3567	0.3383	na	na
Havard & Memon	2009	88	45	43	0.7885	0.9008	0.9163	0.6642	0.4283
Memon, Bartlett, et al. (35 min)[b]	2003	44	22	22	0.3424	0.1919	0.1854	0.5441	0.3194
Memon, Bartlett, et al. (1 week)[b]	2003	42	20	22	1.2133	2.4720	1.7272	0.7314	0.5816
Rose et al.	2003	36	18	18	1.9459	2.3026	1.5805	1.6487	0.1155
Rose et al.	2005	48	24	24	1.2238	1.5805	1.2528	1.2528	0.0837
Wilcock et al.	2005	48	24	24	0.9707	0.9316	1.5106	0.9316	0.5508
Wilcock et al. (Context)[c]	2007	24	12	11	0.7802	1.7918	2.7362	0.1823	1.2452
Wilcock et al. (Control)[c]	2007	24	13	12	0.2697	-0.8473	-0.4445	1.9279	-0.0606

Table E1 (continued)

Excluded								
Wright & Stroud (Exp. 1; 1 day)[l]	2002	59	30	29		-0.4776	-0.0741	
Wright & Stroud (Exp. 1; 1 week)[l]	2002	54	26	28		-0.1358	-0.1823	
Wright & Stroud (Exp. 2)[l]	2002	180	90	90	-0.2381	-0.3264	-0.2353	-0.984

[a] Goodsell et al. (2009): Mugshot search condition excluded

[b] Memon, Bartlett, et al. (2003): Retention interval manipulated: 35 min = retention interval of 35 minutes; 1 week = retention interval of 1 week

[c] Wilcock et al. (2007): Context reinstatement manipulated: Context = Context reinstatement groups; Control = Control groups

[l] Wright & Stroud (2002). Excluded due to exceptional low age thresholds in the older group

Appendix F

Single Effect Sizes (LOR) for all Dependent Variables of Studies using Young and Old Targets

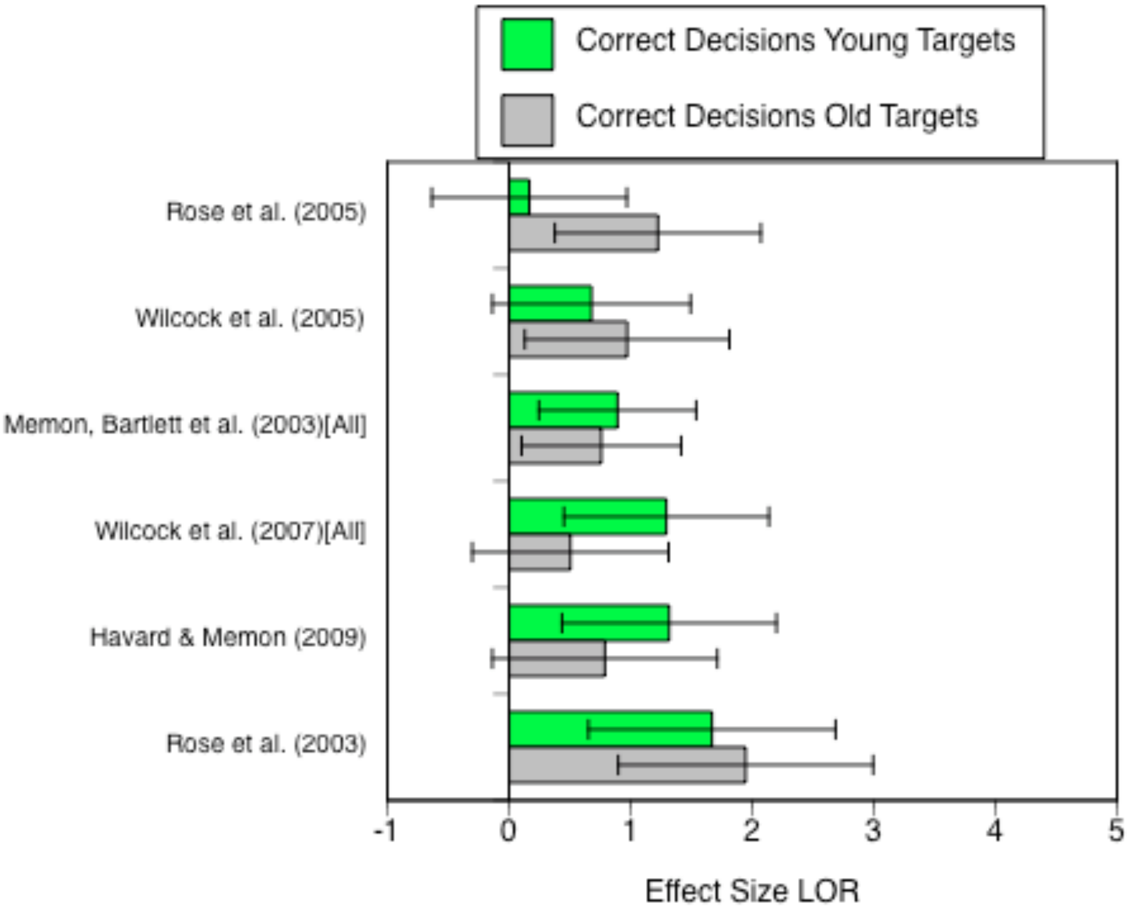

Figure F1. Effect sizes (LOR) for correct decisions across TP and TA lineups of studies using young and old targets.

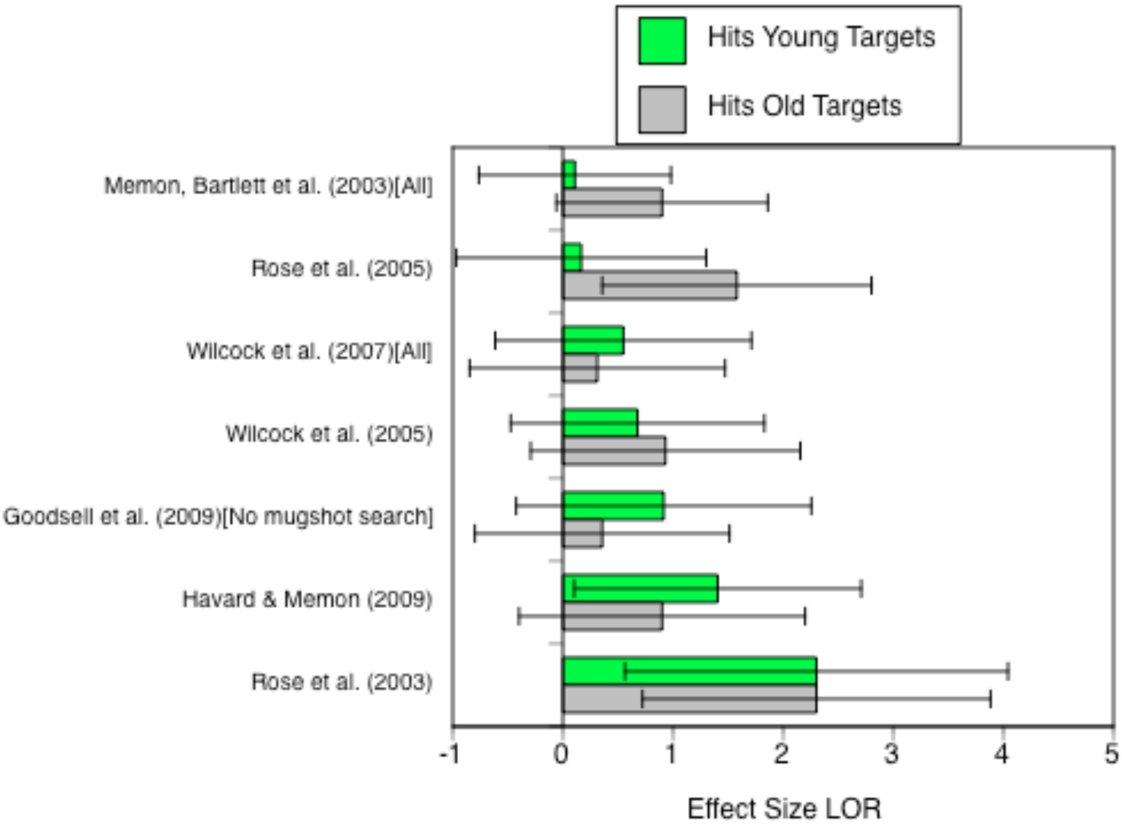

Figure F2. Effect sizes (LOR) for correct identifications in TP lineups of studies using young and old targets.

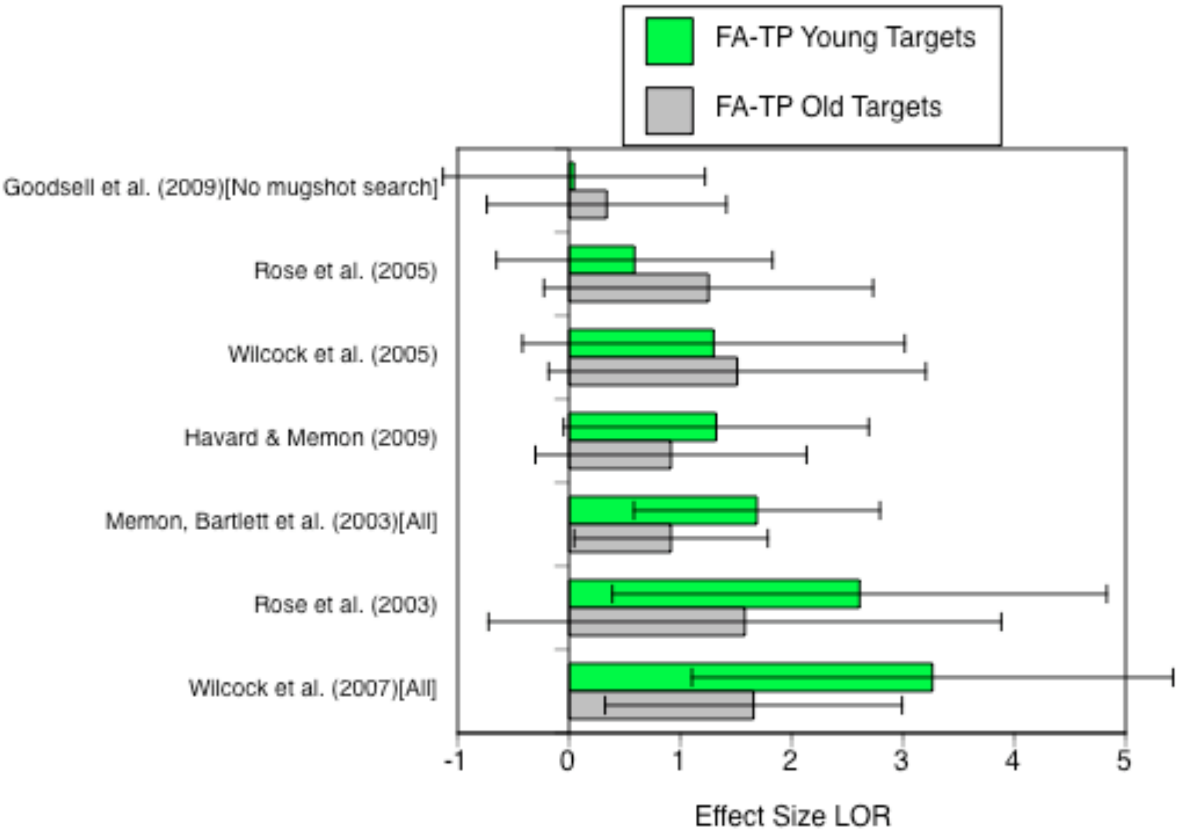

Figure F3. Effect sizes (LOR) for foil identifications in TP lineups of studies using young and old targets.

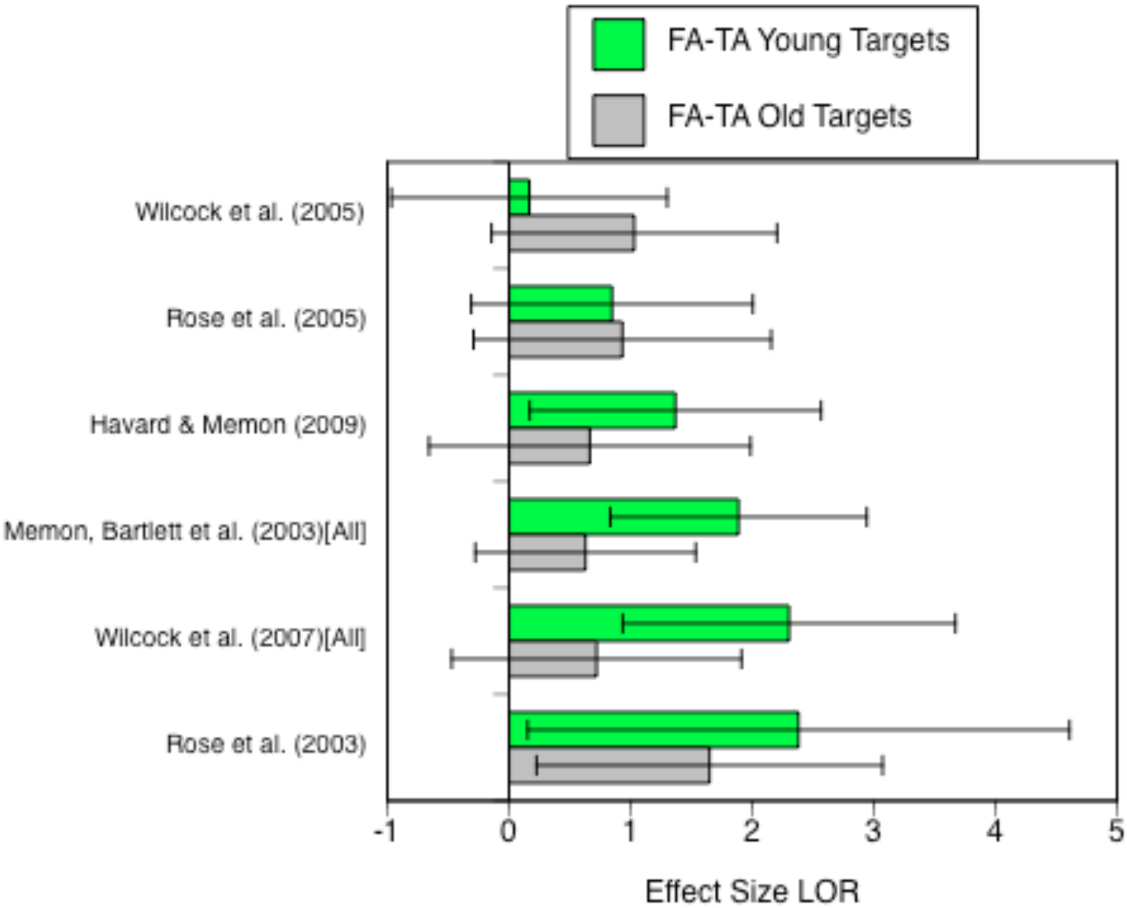

Figure F4. Effect sizes (LOR) for foil identifications in TA lineups of studies using young and old targets.

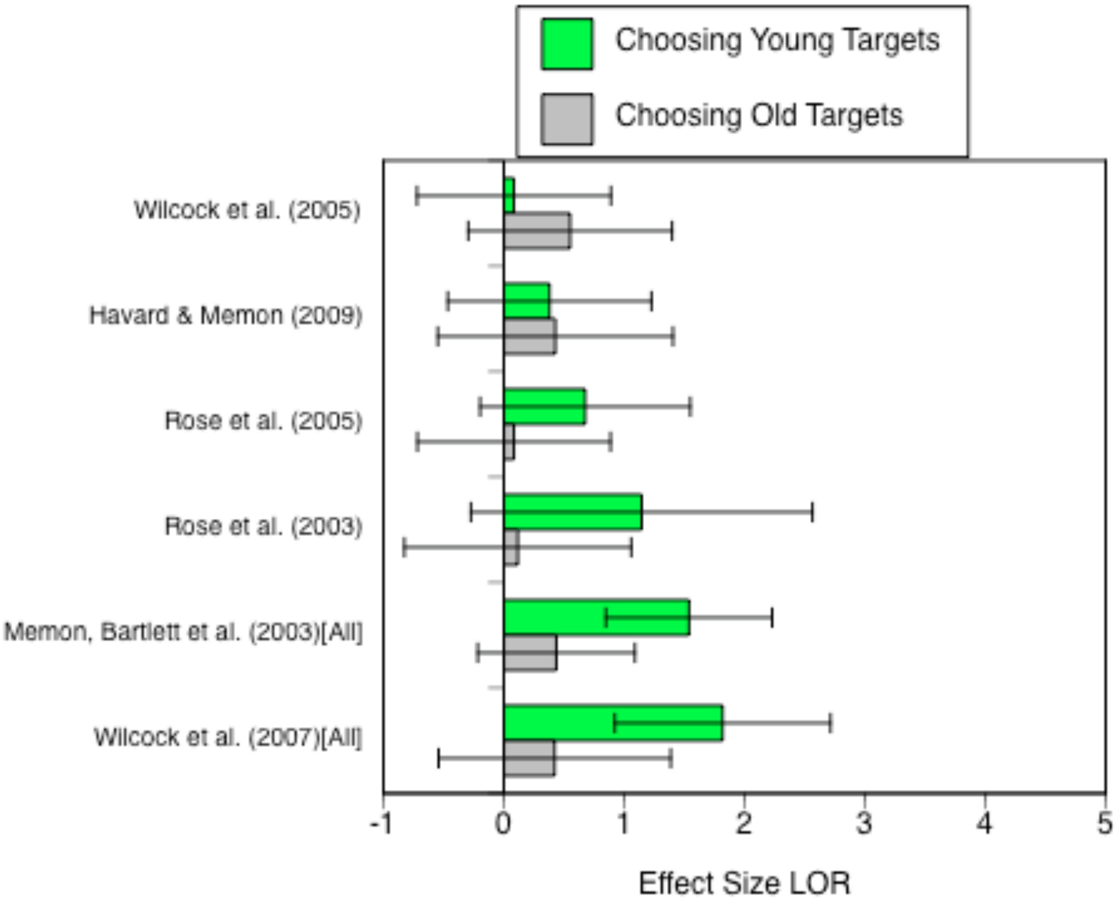

Figure F5. Effect sizes (LOR) for choosing of studies using young and old targets.

Appendix G

Table G1

Publication Bias Indices for all Dependent Measures for Young Targets

	Rank correlation			Regression						Trim and fill			
					95% CI						New estimates		
												95% CI	
DV	*tau*	*Z*	*p*	Intercept	*LL*	*UL*	*SE*	*t*	*p*	*k* added	*OR*	*LL*	*UL*
Correct decisions *k* = 13	.577	2.745	.006	2.808	0.020	5.596	1.266	2.217	.049	4	1.822 (2.108)	1.174 (1.509)	2.591 (2.200)
Correct identification (TP) *k* = 15	.467	2.425	.015	1.600	-0.949	4.149	1.180	1.356	.198	6	1.155 (1.607)	0.928 (1.251)	1.438 (2.065)
Foil identification (TP) *k* = 13	.462	2.196	.028	2.682	0.827	4.536	0.843	3.183	.009	2	2.191 (2.453)	1.615 (1.794)	2.972 (3.356)
Foil identification (TA) k = 14	.209	1.040	.298	1.762	-1.653	5.178	1.567	1.124	.283	2	2.847 (3.074)	2.157 (2.310)	3.758 (4.091)
Choosing *k* = 11	.200	0.856	.392	0.967	-3.262	5.197	1.870	0.517	.617	0	2.259 (2.259)	1.799 (1.799)	2.838 (2.838)

Note. Rank correlation method by Begg and Mazumdar, 1994; Regression method by Egger, Smith, Schneider, and Minder, 1997; Trim and fill method by Duval & Tweedie, 2000a, 2000b; all p-values are two-tailed, k added are the adjusted values for the trim and fill method, observed values in parentheses; estimated values are identical to the observed values when no values were adjusted within the trim and fill method.

Appendix H

Correspondence with authors

Authors contacted for sending unpublished material:

Tim Perfect, University of Plymouth, UK

Jennifer Ryan, University of Toronto, Canada

Jeffrey Neuschatz, The University of Alabama in Huntsville

Rachel Wilcock, London South Bank University, UK

Joanne Fraser, University of Aberdeen, UK

Christine Bastin, University of Liège, Belgium

Ray Bull, University of Leicester, UK

Dimitra Filippou, University of Plymouth, UK

Rachel Rose, University of Kingston, UK

Isabelle Boutet, University of Ottawa, Canada

Armina Memon, University of Aberdeen, UK

Daniel Wright, Florida International University, Florida

Jeffrey Anastasi, Sam Houston Present State University, Texas

James Bartlett, Yale University, Dallas

Appendix I

Elderly Eyewitnesses Coding handbook

Important! Insert -999 if no information on the accordant variable is available!

Notice! The coding handbook was used to code eyewitness identification studies as well as facial recognition studies.

General information

StudyID: study identification, code of study

Insert the code of the study for definite identification of the study. The studies have to be numbered serially from 1 – x. If a study contains two or more independent experiments, insert 15.1, 15.2, 15.3, etc., for example. The number behind the point stands for the experiment number.

Paradigm: Paradigm the study included used

Of which type ist he study included?

1 = Facial Recognition Study

2 = Eyewitness Identification Study

In facial recognition studies participants are confronted with a set of faces and their recognition task is to identify the faces they have seen within a larger amount of 'old' and 'new' faces. Within eyewitness identification studies participants usually watch an event/ a crime and they have to identify the perpetrator or another target face, respectively.

ArtAuth: names of the authors

Insert the name of all authors.

Year: Year of publication

Insert the year the article/ study was published.

ArtTitle: Title of article

Insert the complete title of the article/ study.

ArtCit: Citation of article

Insert the complete reference of the article/ study with journal, volume, issue and pages.

Pub: Type of publication

1 = Journal

2 = Dissertation

3 = Conference Presentation

4 = Unpublished study

Coder: Number of the coder

1 = Kerstin

2 = Natalie

Participants

NPart: Number of participants

Insert the number of all subjects who participated in the study.

NYoung: Number of participants in the younger age group

Insert the number of participants of the younger age group.

NMiddle: Number of participants in the middle age group (if available)

Insert the number of participants of the middle age group.

NOld: Number of participants in the older age group

Insert the number of participants of the older age group.

MAgeY: Average age of the younger age group

Insert the mean age of the younger age group. If age is not reported insert:

1 = children

2 = Highschoolers / students

MAgeMid: Average age of the middle age group

Insert the mean age of the younger age group. If age is not reported insert:

3 = adults

MAgeOld: Average age of the older age group

Insert the mean age of the younger age group. If age is not reported insert:

4 = elderly

SDAgeY: Standard deviation of age in the younger age group

Insert the standard deviation of age of the younger age group.

SDAgeMid: Standard deviation of age in the middle age group

Insert the standard deviation of age of the middle age group.

SDAgeOld: Standard deviation of age in the older age group

Insert the standard deviation of age of the older age group.

PGender: Gender of participants

1 = male

2 = female

3 = both

PGenderY: Percentage of male participants in the younger age group

Insert the percentage (%) of male participants in the younger age group

PGenderM: Percentage of male participants in the middle age group

Insert the percentage (%) of male participants in the middle age group

PGenderO: Percentage of male participants in the older age group

Insert the percentage (%) of male participants in the older age group

PGetY: Recruiting of younger participants

Where were the younger participants recruited?

1 = University/ School

2 = Working place

3 = Retirement organization

4 = Club/ church

5 = Community

6 = Newspaper/ television

7 = mixed

PGetM: Recruiting of middle-aged participants

Where were the middle-aged participants recruited?

1 = University/ School

2 = Working place

3 = Retirement organization

4 = Club/ church

5 = Community

6 = Newspaper/ television

7 = mixed

PGetO: Recruiting of older participants

Where were the older participants recruited?

1 = University/ School

2 = Working place

3 = Retirement organization

4 = Club/ church

5 = Community

6 = Newspaper/ television

7 = mixed

EduY: Average education of the younger age group

How many years of education (including vocational training) the younger age group had on average?

EduM: Average education of the middle age group

How many years of education (including vocational training) the middle age group had on average?

EduO: Average education of the older age group

How many years of education (including vocational training) the older age group had on average?

Educsdyou: Standard deviation of period of education within the younger age group

Educsdmid: Standard deviation of period of education within the middle age group

Educsdeld: Standard deviation of period of education within the older age group

Matching: Were the groups matched?

1 = No, the groups were not matched

2 = yes, the groups were matched on the basis of their education level

3 = yes, the groups were matched on the basis of their health status

4 = yes, the groups were matched on the basis of their intelligence quotient

5 = Yes, the groups were matched in another way. Please describe!

AgeDiff: Did the groups differ significantly?

1 = no, they did not differ

2 = yes, they differed a bit, but not significantly

3 = yes, they differed significantly

Insert 1 only, when it is reported explicitly! Otherwise, insert -999.

Diffvar: Variable the groups differed in.

Dementia: Was a test of dementia conducted?

1 = no

2 = yes, the MMSE

3 = yes, another test, fill in name!

If the name of the test is not reported, insert 3.

PathTest: Were other psychological or psychiatric tests conducted?

1 = no

2 = yes, to test for depression

3 = yes, to test for intellectual impairments

4 = yes, to test for anxiety

5 = yes, another test, fill in name!

If the type of the test is unknown, insert 5.

IQ: Was a intelligence test conducted?

1 = no

2 = yes, the WAIS and HAWIE, respectively

3 = yes, the IST

4 = yes, another intelligence test, fill in name!

If the name of the test in unknown, insert 4.

Perform: Was another performance test conducted?

1 = no

2 = yes, to test reading abilities

3 = yes, a memory test

4 = yes, another performance test, fill in name!

If the type of the test is unknown, insert 4.

PathSelf: Could diseases be excluded?

1 = no or not reported

2 = yes, through self report

3 = yes, through testing

This implies neurological or psychiatric diseases, which could impair cognitive functions.

VisionT: Was vision tested?

1 = no

2 = yes

Insert 1, if not reported.

Vision: Data of vision

Data of vision as reported in the study (mostly via snellen test)

Reward: Did the participants receive compensation?

1 = no

2 = yes

Insert 1, if not reported.

Method: Procedure / material

NEncod: Number of faces while encoding

Insert the number of faces participants had to memorize in the encoding phase.

NRecog Number of faces while recognition

Insert the number of faces, the participants had to recognize in the recognition phase.

NBystand: Number of bystanders while encoding

Insert the number of faces present in the encoding phase but not relevant as target person. Bystanders are presented peripheral in the encoding phase, but had not to be remembered in a later recognition phase. They could be misleadingly mixed up with the target person.

NFoilRec: Number of foils while recognition

Insert the number of foils within the recognition phase/ the lineup.

Foils are persons similar to the suspect. If target-present (TP) as well as target-absent (TA) lineups were used, insert the number of foils used in the TP lineup.

ExpoTime: Exposure time of target person while encoding

1 = as long as needed (not exactly reported)

2 = as long as needed (exactly reported)

If possible insert the amount of seconds the target person was visible to the participants. If more than one face has to be recognized later, insert the amount of time per face.

TestTime: Time available while recognition

1 = as long as needed (not exactly reported)

2 = as long as needed (exactly reported)

If possible insert the amount of seconds the participants had to make a decision while the recognition phase. If more than one face has to be recognized, insert the amount of time they had per face.

DelayMin: Retention interval/ delay

1 = immediately (not exactly reported)

2 = immediately (exactly reported)

3 = up to 60 minutes (but not immediately)

4 = 61 – 1440 minutes

5 = 1441 – 2880 minutes

6 = more than 2880 minutes

Insert here the time between encoding phase and recognition phase.

ModeEnc: Mode of presentation while encoding

1 = Photo

2 = Video

3 = Live

This variable describes, how the faces were presented within the encoding phase.

ModeRec: Mode of presentation while recognition

1 = Photo

2 = Video

3 = Live

This variable describes, how the faces were presented in the recogniton phase.

PoseEnc: Pose of target person while encoding

1 = Profile

2 = _ views of the target person

3 = Frontal view

4 = All three views / Live / Video

5 = Profile & _ views

6 = Profile & frontal views

7 = _ views & frontal view

8 = three-quarter view

This variable describes in which views the target person was presented while encoding.

PoseRec: Pose of target person while recognition

1 = Profile

2 = _ views of the target person

3 = Frontal view

4 = All three views / Live / Video

5 = Profile & _ views

6 = Profile & frontal views

7 = _ views & frontal view

8 = three-quarter view

This variable describes in which views the target person was presented while recognition.

ViewEnc: Visibleness of target person while encoding

1 = Head and shoulders visible

2 = Whole person visible

This variable describes how much could be seen of the target person while the encoding phase. If a video was presented, insert 2.

ViewRec: Visibleness of target person while recognition

1 = Head and shoulders visible

2 = Whole person visible

This variable describes how much could be seen of the target person while the recognition phase. If a video was presented, insert 2.

AppCha: Was anything changed from encoding to recognition?

1 = No

2 = Facial Expression

3 = Pose

4 = Appereance

5 = Mixed

It is meant, if anything on the target person was modified from encoding phase to recognition phase.

Datasep: Were the data reported seperately?

1 = No

2 = Yes

It is meant, if data for 'old', 'modified' and 'new' faces were reported seperately.

TAge1: Age of target person 1

1 = Baby/ toddler (0-5 years)

2 = Children (6-13 years)

3 = Highschooler (14-17 years)

4 = Students (18-30 years)

5 = Adults (30-50 years)

6 = Elderly (50+ years)

If a target person can not be classified exactly into one of these categories, write down the description reported in the study.

TAge 2: Age of target person 2

1 = Baby/ toddler (0-5 years)

2 = Children (6-13 years)

3 = Highschooler (14-17 years)

4 = Students (18-30 years)

5 = Adults (30-50 years)

6 = Elderly (50+ years)

If a target person can not be classified exactly into one of these categories, write down the description reported in the study.

TAge 3: Age of target person 3

1 = Baby/ toddler (0-5 years)

2 = Children (6-13 years)

3 = Highschooler (14-17 years)

4 = Students (18-30 years)

5 = Adults (30-50 years)

6 = Elderly (50+ years)

If a target person can not be classified exactly into one of these categories, write down the description reported in the study.

TAge 4: Age of target person 4

1 = Baby/ toddler (0-5 years)

2 = Children (6-13 years)

3 = Highschooler (14-17 years)

4 = Students (18-30 years)

5 = Adults (30-50 years)

6 = Elderly (50+ years)

If a target person can not be classified exactly into one of these categories, write down the description reported in the study.

TGender: Gender of the target person

1 = Male

2 = Female

3 = Both

FillTask: Filler task

1 = No

2 = Yes

Did the participants receive any filler task between the encoding and the recognition phase; presentation of other faces/ mugshots included?

NFaceFil: Amount of faces presented within the filler task

Insert the number of faces presented to the participants within the interference task.

CrimeTyp: Was a crime scenario presented?

1 = No

2 = Yes, a theft

3 = Yes, a robbery

4 = Yes, a violent felony

5 = Yes, other than the named above, fill in!

NLineups: Number of lineups a participant had to accomplish

In some studies, participants had to watch more than one lineup. Insert the number of lineups each participant had to accomplish.

SimSeq: Type of lineup presentation

1 = simultaneous presentation of lineup

2 = sequential presentation of lineup

3 = both

In seequential lineups every face/ person is presented by otself, whereas in simultaneous lineups all faces are presented at the same time. In most facial recognition studies faces are presented one at a time.

TPres: Target presence

1 = Target person present in a lineup

2 = Target person absent in a lineup

3 = Both

4 = Facial recognition study

FunctSz: Functional size of the lineup measured?

1 = No

2 = Yes

Was the Functional size, a quantative index of lineup fairness, measured? It is defined as the total number of moch witnesses divided by the number of mock witnesses who chose the suspect (Wells et al., 1979).

SimilPil: Was a simulation/ pilot study conducted to test lineup fairness?

1 = No

2 = Pilot study without data reported

3 = Yes

Insert if a pilot or simulation study was conducted to test the lineup fairness.

NFaceExp: Expected amount of faces within the recognition phase

Insert the amount of faces the participants were told to be confronted with in the recognition phase. This value can differ from the actual amount of faces presented.

FacTrain: Did the participants receive a training?

1 = No

2 = Yes

Did the participants receive a facial recognition training?

TestExp: Knowlede of task beforhand

1 = No

2 = Yes

Did the participants knwo that they had to recognize/ identify the target person in a later task?

LInstru: Lineup instruction

1 = No

2 = Yes

Were there an option (explicitly pointed out) that they had the option not to choose someone from the lineup?

Extras

Comment: Comments to the studies

Insert here any relevant information or noticeable problems which you noted while coding procedure.

If studies assessed confidence of the participants, code this information seperately.

ANOVA

StudyId: Study identification, code of study

Insert the code of the study for definite identification of the study. The studies have to be numbered serially from 1 – x. If a study contains two or more independent experiments, insert 15.1, 15.2, 15.3, etc., for example. The number behind the point stands for the experiment number.

Paradigm: Paradigm the study included used

Of which type ist he study included?

1 = Facial Recognition Study

2 = Eyewitness Identification Study

In facial recognition studies participants are confronted with a set of faces and their recognition task is to identify the faces they have seen within a larger amount of 'old' and 'new' faces. Within eyewitness identification studies participants usually watch an event/ a crime and they have to identify the perpetrator or another target face, respectively.

ArtAuth: names of the authors

Insert the name of all authors.

Year: Year of publication

Insert the year the article/ study was published.

Subjectage: Main effect 'Age of participants'

AgesubF: F-value of the main effect 'Age of participants'

Agesubdf1: First degree of freedom of the main effect 'Age of participants'

Agesubdf2: Second degree of freedom of the main effect 'Age of participants'

Agesubp: p-value of the main effect 'Age of participants'

AgesubMSe: Mean sum of squares of the main effect 'Age of participants'

Agesubeff: Effect size of the main effect 'Age of participants'

Typeeff: Type of effect (eta, d, etc.)

FaceAge: Main effect 'Age of target'

AgesfaceF: F-value of the main effect 'Age of target'

Agefacedf1: First degree of freedom of the main effect 'Age of target'

Agefacedf2: Second degree of freedom of the main effect 'Age of target'

Agefacep: p-value of the main effect 'Age of target'

AgefaceMSe: Mean sum of squares of the main effect 'Age of target'

Agefaceeff: Effect size of the main effect 'Age of target'

Interaction: Interaction between 'Age of participants' and 'Age of target'

AgesubfacF: F-value of the interaction

Agesubfacdf1: First degree of freedom of the interaction

Agesubfacdf2: Second degree of freedom of the interaction

Agesubfacp: p-value of the interaction

AgesubfacMSe: Mean sum of squares of the interaction

Agesubfaceff: Effect size of the interaction

Discrimination Index: Index of sensitivity (d')

meand'you: Mean index of sensitivity (d') of the younger age group

meand'mid: Mean index of sensitivity (d') of the middle age group

meand'eld: Mean index of sensitivity (d') of the older age group

sdd'you: Standard deviation of d' of the younger age group

sdd'mid: Standard deviation of d' of the middle age group

sdd'eld: Standard deviation of d' of the older age group

Decision criterion C

meanCyou: Mean C-value of the younger age group

meanCmid: Mean C-value of the middle age group

meanCeld: Mean C-value of the older age group

sdCyou: Standard deviation of C of the younger age group

sdCmid: Standard deviation of C of the middle age group

sdCeld: Standard deviation of C of the older age group

Rates

StudyId: Study identification, code of study

Insert the code of the study for definite identification of the study. The studies have to be numbered serially from 1 – x. If a study contains two or more independent experiments, insert 15.1, 15.2, 15.3, etc., for example. The number behind the point stands for the experiment number.

Paradigm: Paradigm the study included used

Of which type ist he study included?

1 = Facial Recognition Study

2 = Eyewitness Identification Study

In facial recognition studies participants are confronted with a set of faces and their recognition task is to identify the faces they have seen within a larger amount of 'old' and 'new' faces. Within eyewitness identification studies participants usually watch an event/ a crime and they have to identify the perpetrator or another target face, respectively.

ArtAuth: names of the authors

Insert the name of all authors.

Year: Year of publication

Insert the year the article/ study was published.

NYoung: Number of participants in the younger age group

Insert the number of participants of the younger age group.

AgeYou: Age category of the younger age group

1 = Baby/ toddler (0-5 years)

2 = Children (6-13 years)

3 = Highschooler (14-17 years)
4 = Students (18-30 years)
5 = Adults (30-50 years)
6 = Elderly (50+ years)
Otherwise, insert the definition used in the article.

MAgeY: Mean age of the younger age group

SDAgeY: Standard deviation of age in the younger age group

Ageyourange: Age span of the younger age group

NOld: Number of participants in the older age group
Insert the number of participants of the older age group.

AgeOld: Age category of the older age group
1 = Baby/ toddler (0-5 years)
2 = Children (6-13 years)
3 = Highschooler (14-17 years)
4 = Students (18-30 years)
5 = Adults (30-50 years)
6 = Elderly (50+ years)
Otherwise, insert the definition used in the article.

MAgeOld: Mean age of the older age group

SDAgeOld: Standard deviation of age in the older age group

Ageeldrange: Age span of the older age group

Hitmyou: Proportion of hits within the younger age group

Hitsdyou: Standard deviation of hits in the younger age group

FaTPyou: False alarms in TP lineups in the younger age group

FaTPsdyou: Standard deviation of false alarms (TP) in the younger age group

Incorrmyou: Incorrect rejection in TP lineups within the group of younger participants

Incorrsdyou: Standard deviation of incorrect rejections (TP) in the group of younger participants

Corrmyou: Correct rejection in TA lineups in the group of younger participants

Corrsdyou: Standard deviation of correct rejections in TA lineups in the group of younger participants

FaTAmyou: False alarms in TA lineups in the group of younger participants

FaTAsdyou: Standard deviation of false alarms (TA) in the group of younger participants

Correcmyou: Corrected recognition of the younger age group (Hits – false alarms)

Correcsdyou: Standard deviation of the corrected recognition of the younger age group

CorRespY: Correct decision of the younger age group via TP and TA lineups (Hits + Correct rejection / 2)

Hitmmid: Proportion of hits within the middle age group

Hitsdmid: Standard deviation of hits of the middle age group

FaTPmid: False alarms in TP lineups in the middle age group

FaTPsdmid: Standard deviation of false alarms (TP) in the middle age group

Incorrmmid: Incorrect rejection in TP lineups within the group of middle-aged participants

Incorrsdmid: Standard deviation of false rejections (TP) in the group of middle-aged participants

Corrmmid: Correct rejection in TA lineups in the group of middle-aged participants

Corrsdmid: Standard deviation of correct rejections in TA lineups in the group of middle-aged participants

FaTAmmid: False alarms in TA lineups in the group of middle-aged participants

FaTAsdmid: Standard deviation of false alarms (TA) in the group of younger participants

Correcmmid: Corrected recognition of the middle age group (Hits – false alarms)

Correcsdmid: Standard deviation of the corrected recognition of the younger age group

CorRespM: Correct decision of the middle age group via TP and TA lineups (Hits + Correct rejection / 2)

Hitmeld: Proportion of hits within the older age group

Hitsdeld: Standard deviation of hits in the older age group

FaTPeld: False alarms in TP lineups in the older age group

FaTPsdeld: Standard deviation of false alarms (TP) in the older age group

Incorrmeld: Incorrect rejection in TP lineups within the group of younger participants

Incorrsdeld: Standard deviation of false rejections (TP) in the group of older participants

Corrmeld: Correct rejection in TA lineups in the group of older participants

Corrsdeld: Standard deviation of correct rejections in TA lineups in the group of older participants

FaTAmeld: False alarms in TA lineups in the group of younger participants

FaTAsdeld: Standard deviation of false alarms (TA) in the group of younger participants

Correcmeld: Corrected recognition of the older age group (Hits – false alarms)

Correcsdeld: Standard deviation of the corrected recognition of the older age group

CorRespO: Correct decision of the younger age group via TP and TA lineups (Hits + Correct rejection / 2)

Moderator variables

ExpoTime: Mean exposure time of the target person while encoding

1 ≤ 30 seconds

2 > 30 seconds

Delay: Retention interval between encoding and recognition

1 ≤ 60 minutes

2 > 60 minutes

ModeExp: Mode of presentation while encoding

1 = Photo

2 = Video
3 = Live

TPres: Target presence while recognition
1 = Target present
2 = Target absent
3 = Both
4 = Face Recognition

SimSeq: Lineup type
1 = Simultaneous
2 = Sequential
3 = Both

AppCha: Was anything changed/ modified?
1 = No
2 = Expression
3 = Pose
4 = Appereance
5 = Expression and pose

DescPerp: Had the participants to describe the target before identification task?
1 = No
2 = Yes
3 = Mixed

PreTask: Any relevant task before identification?
0 = No
1 = Description of perpetrator
2 = Context reinstatement
3 = Prelineup questions
4 = Practice lineup
5 = Free recall of event
6 = Audio narrative of event heard

7 = Mixed

Discussion

This dissertation reported two meta-analyses that are concerned with variables affecting eyewitness' testimony. Following Wells' (1978) distinction there are two groups of variables, *estimator* and *system variables*. Whereas the latter is usually under control of the criminal system (e.g., lineup instructions, presentation of the lineup, type of interview) the former cannot be manipulated in real cases so that its influence can only be estimated post-hoc (e.g., distance between perpetrator and eyewitness, exposure duration, ethnicity of the perpetrator). The current meta-analyses investigated two estimator variables, *presence of a weapon* and *age of eyewitnesses*. There are numerous studies assessing the effect of these variables showing inconsistent findings. Meta-analyses pursue the goal to shed light on the dark as they summarize and quantify the results of studies included.

Meta-analysis 1 focused on the question whether a weapon carried by the perpetrator during a crime decreases eyewitness' performance to describe and to identify the suspect in a lineup (*weapon focus effect, WFE*). A previous meta-analysis supported the effect for identification and description accuracy (Steblay, 1992). Today, the number of studies investigating the effect has almost doubled demanding an update of Steblay's meta-analysis.

In meta-analysis 2, the *age of eyewitnesses* was investigated as older witnesses are often supposed to be worse in identification tasks compared to younger adults (e.g., Havard & Memon, 2009; Memon, Bartlett, Rose, & Gray, 2003; Rose, Bull, & Vrij, 2003; 2005). To our knowledge, there is no meta-analysis published yet comparing the identification performance of younger and older witnesses.

In the following, the main results of each meta-analysis will be discussed. For a more comprehensive discussion, see discussion sections of each meta-analysis.

Weapon focus effect (WFE)

Meta-analysis 1 included 23 studies investigating the effect of weapon presence on identification and description accuracy. Measures of identification performance were correct identifications in target-present (TP) lineups and correct decisions (across TP and TA lineups). Unfortunately, data from target-absent (TA) lineups alone were too sparse (k = 3) to calculate a weighted effect size. We found a WFE for description accuracy (g_u = 0.568, [0.490, 0.647]) but not for correct identifications (*OR* = 1.287 [0.971, 1.706]), and correct decisions (*OR* = 1.101 [0.795, 1.525]). As others already stated, recall seems to be a more sensitive measure for eyewitness performance than identification accuracy (Pickel, 2009; Steblay, 1992) so that larger effects for description accuracy were to be expected.

That Steblay (1992) found a significant effect on identification accuracy but we did not might arise from a larger set of studies included in the current work and/ or the use of different effect sizes, although *OR* is considered a more sensitive measure. A more recent meta-analysis on WFE published during our investigations included laboratory and simulation studies (as in the current work) as well as actual crimes (Fawcett, Russell, Peace, & Christie, 2011). Fawcett et al. (2011) found a small effect on identification accuracy (g = 0.22) and a large effect on feature accuracy (g = 0.75). Overall effects via identification *and* feature accuracy were largest for simulation studies (using staged events; g = 0.82, k = 6), and smallest for actual crimes (g = 0.15, k = 7). Differences to our findings might result from the aggregation of identification outcomes across TP and TA lineups. But with regard to theoretical and practical implications it seems important to conduct separate analyses for different lineup outcomes.

Within the current meta-analysis theoretical assumptions were explored mainly with the help of moderator analyses. Three main reasons are discussed in the literature to explain the WFE: *stress and arousal (e.g.,* Deffenbacher, 1983; Deffenbacher, Bornstein, Penrod, & McGorty, 2004; Easterbrook, 1959), *unusualness* (e.g., Pickel, 1998, 2009), and *attentional fixation (e.g.,* Christianson & Loftus, 1991).

The *stress hypothesis* assumes that the perceived arousal while observing a crime where the perpetrator carries a weapon is responsible for the worse eyewitness performance. Because levels of individual arousal were rarely assessed, variables apparently associated with higher levels of stress were expected to moderate the WFE (presence of a crime, dangerousness of a weapon and mode of stimulus presentation). Supporting this approach, two moderators (presence of a crime and dangerousness of the weapon) were associated with poorer description accuracy of the target person. The presentation of a crime scene yielded larger effects compared to neutral settings. Surprisingly, knifes produced higher effect sizes as guns. Therefore, it was concluded that people are more used to guns through TV (crime thriller). Moreover, an assault with a knife could be more bloodily and therefore more frightening for the observers compared to a shooting, for example. In any case, it would be important to assess the amount of contact eyewitnesses have with weapons (e.g., shooting competitors, video games, police officers) to understand underlying processes and formulate practical advices (see below).

However, for most eyewitness studies it is not possible to reach a maximum of realism due to ethical reasons. Feelings of threat and arousal will therefore be much higher when being confronted with a weapon in real life than watching a video showing a gun in an experimental setting. Certainly, studies included in this meta-analysis are not able to explore stress and arousal as it is probably related to the WFE in real life. Larger effects are therefore expected in real cases. Nevertheless, previous archival studies were not able to confirm the existence of the WFE neither for identification accuracy nor for description accuracy (see Fawcett, et al., 2011; Tollestrup, Turtle, & Yuille, 1994; Valentine, Pickering, & Darling, 2003; Wagstaff, MacVeigh, Scott, Brunas-Wagstaff, & Cole, 2003). Probably, longer exposure times of the perpetrator in real crimes might be responsible for the missing effect as there is enough time for an eyewitness to study the face of the target (see below *attentional approach*). Mastrobuoni (2012) could demonstrate that 90% of the bank robberies that took place in Italy between 2005 and 2007 last at most 9 minutes. Within this time eyewitnesses are surely able to turn their eye away from the central details (weapon) to more peripheral details like the face of the target. Finally, archival studies have to be interpreted cautiously due to the lack of external proof of guilt of the perpetrator.

The *unusualness* hypothesis assumes the WFE to appear due to feelings of surprise as a weapon is not expected in certain situations. Studies examining the effect of unusualness were therefore expected to produce similar effects as traditional WFE studies. Separate analyses had to be conducted because of statistical dependencies between the studies when testing weapon and unusualness conditions against one control group. Supporting this approach, unusual objects (like a raw chicken or a pink plastic flamingo) led to similar effects like weapons do as they reduce the ability to later describe the perpetrator. Differences between the two groups (unusual vs. control) were also not significant for identification accuracy. Moderator analyses revealed a significant effect for expectancy violation (e.g., a priest vs. police officer with a gun) but not for weapon presence. Feelings of surprise due to schema incongruent experiences are therefore playing a main role within the explanation of WFE.

Further results of the current meta-analysis indicated a refocusing of attention away from the perpetrator to the weapon supporting the *attentional focus* hypothesis where eyewitnesses focus on central details (weapon) rather than on peripheral (face or clothing of the perpetrator). In this regard a weapon was better remembered than a neutral object.

In summary, following results are outlined: a) there is a moderate WFE for description accuracy but not for identification accuracy; b) stress and arousal may be responsible but are difficult to evaluate within laboratory studies; c) unusualness of the object carried by a perpetrator plays a significant role when explaining the WFE; d) in any case a refocusing to a weapon takes place.

Age of eyewitness

Meta-analysis 2 included 22 studies comparing at least one younger and one older group in their identification performance. Dependent measures under investigation were: Correct decisions (across TP and TA lineups), correct identifications, filler identifications in TP lineups, and foil identifications in TA lineups (target replacement and filler identification). Consistent with our hypotheses, every comparison demonstrated a clear advantage of the younger age group over the elderly. Weighted mean effect sizes were moderate to large with the largest differences found for false identifications in TP and TA lineups (*OR* = 2.453 [1.794, 3.356] and *OR* = 3.074 [2.310, 4.091], respectively). Older adults were therefore 2.5 to 3 times more likely to choose a wrong person from a lineup compared to younger eyewitnesses. Another considerable effect was found for choosing rates demonstrating that the elderly were 2.3 times more likely to choose someone from a lineup (either TP or TA) compared to their younger counterparts. Homogeneity tests as well as sensitivity analyses supported the robustness of the age effect on identification performance. Unfortunately, description accuracy could not serve as dependent variable as most studies did not investigate person descriptions. As mentioned before, to describe a person apparently is a more difficult task due to its feature-based character. If cognitive deficits are responsible for larger differences between younger and older witnesses effects for description accuracy might even be larger because of its higher demand on cognitive abilities. Future research should focus on this.

Within moderator analyses, exposure time was found to moderate choosing behavior. Surprisingly, results were in the opposite direction as expected in that the difference between the two age groups was larger when exposure time was longer. A reason might be, that older witnesses feel capable to identify the target person when confronted for a longer time, even though their memory is not strong enough.

Additionally, older adults seem to profit from sequential lineup testing as they choose less often someone from a lineup. Maybe the missing option to compare all persons in the lineup with each other rather than to compare them separately with the picture of the perpetrator in their minds lead to a more conservative decision criterion.

Finally, analyses could demonstrate a moderate to large association between the mean age of the older group and all dependent measures under investigation. Our findings therefore suggest that performance gradually decreases with increased age. Hence, the older an eyewitness the poorer is his or her lineup performance.

As many researchers are interested in the own-age bias (e.g., Goodsell, Neuschatz, & Gronlund, 2009; Harvard & Memon, 2009; Wilcock, Bull, & Vrij, 2007), differences between younger and older witnesses were expected to be smaller for old age faces compared to younger targets. When examining studies with older targets as stimulus material the elderly were still worse in identifying the suspect from a lineup. However, there is meta-analytical evidence that identification performance of the elderly is less erroneous when confronted with older faces especially for correct identifications indicating an own-age effect for older eyewitnesses (Martschuk, Kocab, & Sporer, in preparation; Rhodes & Anastasi, 2012). Because perpetrators are rarely of old age these findings are rather of theoretical interest (Bureau of Justice Statistics, 2012).

When explaining the age effect two main approaches seem to be helpful, a *cognitive* and a *social approach*. From a *cognitive perspective* it is evident that older adults have problems with episodic memory and especially with the source of information (e.g., Aizpurua, Garcia-Bajos, & Migueles, 2011; Bornstein, 1995; Cansino, Trejo-Morales, & Hernández-Ramos, 2010; Johnson, Hashtroudi, & Lindsay, 1993). When this contextual information is not available people rely on familiarity of a stimulus (e.g., Bright-Paul, Jarrold, & Wright, 2005). An older eyewitness could therefore feel familiar with a person from a lineup but cannot locate where this familiarity results from. Possibly, the person looks similar to the postman or a friend long time not seen and the witness concludes that it has to be the perpetrator observed in a criminal act.

From a *social perspective* choosing behavior of the elderly seem to play an important role when explaining the poorer identification performance. Our findings indicated that older adults were 2.3 times more likely to choose someone from a lineup compared to their younger counterparts. Reasons might be that they feel some kind of forced to help the police to find a suspect, or are not absolutely aware of the option that the perpetrator could also be absent in the lineup presented. Previous research could demonstrate that biased (or the absence of warning) instructions increase choosing (Clark, 2005; Steblay, 1997). Finally, a higher *Need for Closure* might also be responsible for higher choosing rates in the elderly. *Need for Closure* reflects an aversion to uncertainty as well as a preference towards definite answers to questions (Webster & Kruglanski, 1994). Particularly older adults may therefore want to rapidly leave the unpleasant situation of an identification task while they reach clarity trough choosing.

In sum, these are the main results: a) older people are worse in identification tasks compared to their younger counterparts; b) largest differences were found for false identifications in TP and TA lineups; c) the elderly show a more liberal choosing behavior; d) a longer exposure time is associated with higher choosing rates in the elderly; e) sequential lineup testing seem to be beneficial for older witnesses; f) the older an eyewitness the poorer his or her identification performance; f) cognitive *and* social aspects seem to be responsible for the age effect.

Practical Implications

By now it is widely known that eyewitness misidentification is one of the main causes of wrongful conviction in the USA, playing a role in more than 75% of convictions overturned through DNA testing (see www.innocenceproject.org). In the current dissertation practical advice to prevent such eyewitness errors is restricted because both variables investigated are not under control of the criminal system (*estimator variables*). However, there are methods apparently helping to reduce the influence of both factors.

Special WFE trainings could be effective as people would be better able to avoid focusing on the weapon. First, high risk groups (e.g., employees of banks, convenience stores or petrol stations) should be informed about the WFE. The knowledge about the phenomenon seems to be helpful with regard to a higher description accuracy of the target than without special instructions (Pickel, Ross, & Truelove, 2006). Providing information paired with practical trainings should be promising. Within such practical sessions people could be trained in situations where they are confronted with a weapon. Stanny and Johnson (2000) demonstrated that police officers showed no differences in heart rate when being directly confronted with a gun compared to officers not being confronted. One could therefore assume that they are used to weapons and are therefore better able to control the situation as they pay attention to peripheral details.

Second, in terms of police investigations a lineup consisting of different weapons could also be helpful as current analyses could demonstrate eyewitnesses having a good memory for the weapon the perpetrator carried with. This could be additional evidence.

With regard to the *witness' age*, older adults are not inevitably poor witnesses. Besides inter-individual differences to consider in the individual case there seem to be methods reducing the differences on eyewitness identification between younger and older eyewitnesses. Sequential testing (e.g., Memon & Gabbert, 2003a) combined with distinct advices that the perpetrator could also be *not* part of the lineup (e.g., Memon, Hope, Bartlett, & Bull, 2002) seem to encourage older eyewitnesses to adopt a stricter decision criterion. There is evidence that older people have problems remembering non-biased lineup instructions (Rose, et al., 2003; 2005), but Wilcock, Bull and Vrij (2005) could demonstrate that enhanced instructions where additional information concerning mistaken identifications were given led to better memories in the elderly. Additionally, pre-identification procedures (practice lineup, pre-lineup questions) seem to be beneficial for achieving a more conservative decision criterion (Dysart & Lindsay, 2001; Wilcock & Bull, 2010). Against the background of the *no-cost view* it should be investigated more intensively how these methods affect both, false identifications as well as correct identifications, as there is some evidence that hit rates might also be reduced through the strategies (Memon & Gabbert, 2003a; Wilcock & Bull, 2010).

From a cognitive perspective, context reinstatement strategies could help to reduce source monitoring problems in the elderly because older persons seem to have problems with identifying the source of the remembered information. But research is inconclusive (e.g., Wilcock, et al., 2007; Rose, et al., 2003) delivering conflicting results. There is evidence that older adults profit from visualization and verbalization tasks prior to identification as they choose more often the correct person from the lineup (Kinlen, Adams-Price, & Henley, 2007). Possibly, such strategies lead to a better memory for the source of information as eyewitnesses have to visualize the crime scene. This could help them to achieve a more distinct picture of the perpetrator in their minds.

Methodological Limitations

The two meta-analyses presented combined theoretical interests with current meta-analytical procedures. Although, meta-analyses have remarkable advantages (e.g., Cooper & Hedges, 2009; Hedges & Pigott, 2001; 2004) there are also limitations to consider (Cafri, Kromrey, & Brannick, 2010; Cooper, 2009; Cooper & Hedges, 2009; Shercliffe, Stahl, & Tuttle, 2009).

First, for some analyses the set of independent hypothesis tests was rather small. This reduces the statistical power or prevented entirely the accomplishment of some syntheses (e.g., meta-regression analyses). Nevertheless, the statistical power of meta-analysis is usually higher than in primary studies (Cohn & Becker, 2003; Hunter & Schmidt, 1990; Matt & Cook, 1994).

Second, there is always the possibility to oversight studies especially those which are not published for some reasons. In most cases, the findings of those unpublished studies are non-significant (*publication bias*: see Lipsey & Wilson, 1993; Sporer & Goodman-Delahunty, 2011; Sutton, 2009). An extensive literature search should prevent us from missing these findings. However, multiple analyses indicated the possibility of *publication bias*. Results should therefore be interpreted with caution.

Third, the information reported by the authors of the primary studies was often sparse. Missing data is always a problem in meta-analysis (see Pigott, 2009). Consequently, some interesting moderators could not be included in accordant meta-regression models. Sometimes information were not reported, sometimes they were ambiguous. Overall, if information were not given in some studies a calculation of moderating effects was useless and would be misleading.

Fourth, meta-analyses can only provide evidence about correlations between variables (Cooper & Hedges, 2009; Sporer & Cohn, 2011). Causal relations can only be generated by primary studies highlighting its necessity and not its competing alternative to meta-analyses (Cooper, 2009).

Finally, it is likely that the results of these two meta-analyses differ from real life cases. One reason could be higher stress levels while encoding the information in a real crime, the other could be more severe consequences in real life cases compared to experimental situations. These circumstances could therefore not only lead to smaller effect sizes, but also to an underestimation of the effect as would be expected in real life. Further archival analyses could give additional advice.

Conclusion

This dissertation demonstrated a moderate WFE for description accuracy and a robust age effect for identification performance with the use of meta-analytical methods. Results of the WFE for lineup identification remained non-significant.

Although both variables account to the group of *estimator variables* there are methods apparently reducing their impact on eyewitness testimony. To avoid the WFE it seems promising to provide information to high risk groups, and to train them focusing on important details. Additionally, a lineup consisting of weapons could deliver helpful information in police investigations as the eyewitnesses have good memory for these objects. For *older people* who serve as eyewitnesses in criminal investigations it might be helpful to confront them with a sequential lineup and/ or pre-identification procedures (pre-lineup questions, practice lineups). Through these methods older eyewitnesses seem to adopt a stricter decision strategy (Dysart & Lindsay, 2001; Wilcock & Bull, 2010). More obvious warnings that the perpetrator might also be *not* part of the lineup could also help to narrow their liberal choosing behavior.

Concluding, there is some proposal and demand for future research: In terms of WFE, more research is needed for TA lineups since the few studies published so far delivered ambiguous results. This is important as experts in court usually testify for a defendant to prevent a (presumably) wrongful conviction.

Additionally, methods are needed to better induce and assess arousal during the observed crime scenes within ethical constraints of human research. Direct and objective assessment techniques of stress (e.g., heart rate or Galvanic Skin Response) should be more helpful in examining levels of arousal compared to subjective self-evaluations. Information from archival analyses could also deliver adjuvant information.

For studies investigating the age effect it is necessary to evaluate more intensively methods for improving the ability of older witnesses within identification tasks, as there is evidence that differences between younger and older individuals can be reduced. In this regard, it is important to examine decision processes of older eyewitnesses. Furthermore, it is crucial to assess if the elderly rather pick the target replacement from a TA lineup or a filler person who is usually known to be innocent. In contrast, the police do not know if the suspect is really the perpetrator.

In general, it is necessary to investigate methods apparently improving eyewitness evidence against the background of the *no-cost view* (Clark, 2012). Few studies either showed a reduction in both, hit rates and false identifications (Memon

& Gabbert, 2003a; Wilcock & Bull, 2010) or solely examined TP outcomes (Kinlen, Adams-Price, & Henley, 2007). Both eyewitness decisions have dramatic consequences and need to be considered when expressing those strategies. Finally, a decrease of correct identifications implies that guilty criminals do not get arrested whereas a high level of false identifications means that innocent people get arrested for a crime they did not commit.

Finally, for both factors there is the need to combine archival findings with laboratory results more extensively. Even though, results from actual criminal cases have to be cautiously interpreted due to missing external proof of guilt of the perpetrator, both approaches could mutually be beneficial. For example real eyewitnesses could retrospectively describe what they focused while observing a crime (Fawcett et al., 2011), what they felt (e.g., surprise, fear, curiosity), or the amount of contact they had with weapons.

References

Aizpurua, A., Garcia-Bajos, E., & Migueles, M. (2011). False recognition and source attribution for actions of an emotional event in older and younger adults. *Experimental Aging Research, 37*, 310-329. doi: 10.1080/0361073X.2011.568829

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

Bright-Paul, A., Jarrold, C., & Wright, D. B. (2005). Age-approriate cues facilitate source-monitoring and reduce suggestibility in 3- to 7-year-olds. *Cognitive Develeopment, 20*, 1-18. doi: 10.1016/j.cogdev.2004.06.001

Cafri, G., Kromrey, J.D., & Brannick, M.T. (2010). A meta-meta-analysis: Empirical review of statistical power, Type I error rates, effect sizes, and model selection of meta-analyses published in psychology. *Multivariate Behavioral Research, 45,* 239-270. doi: 10.1080/00273171003680187

Cansino, S., Trejo-Morales, P., & Hernández-Ramos, E. (2010). Age-related changes in neural activity during source memory encoding in young, middle-aged and elderly adults. *Neuropsychologica, 48*, 2537-2549. doi: 10.1016/j.neuropsychologia.2010.04.032

Christianson, S.-A., & Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Cognition and Emotion, 5*, 91-108. doi: 10.1080/02699939108411027

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior, 29*, 395-424. doi: 10.1007/s10979-005-5690-7

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform:

Psychological science and public policy. *Perspectives on Psychological Science, 7*, 238-259. doi: 10.1177/1745691612439584

Cohn, L. D., & Becker, B. J. (2003). How meta-analysis increases statistical power. *Psychological Methods, 8*, 243-253. doi:

Cooper, H. (2009). Hypotheses and problems in research synthesis. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 237-256). New York: Russell Sage Foundation.

Cooper, H., & Hedges, L.V. (2009). Potentials and limitations. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2, pp. 237-256). New York: Russell Sage Foundation.

Deffenbacher, K. A. (1983). Identification evidence: A psychological evaluation. *American Journal of Psychology, 96,* 591-595. doi: 10.1007/BF00868064

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28*, 687-706. doi: 10.1007/s10979-004-0565-x

Douglass, A. B., & Steblay, N. M. (2006). Memory distortion in eyewitnesses: A meta-analysis of the post-identification feedback effect. *Applied Cognitive Psychology, 20*, 859-869.

Dysart, J. E., & Lindsay, R. C. L. (2001). A preidentification effect: Serendipitously increasing correct rejections. *Law and Human Behavior, 25*, 155-165. doi:

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66,* 183-201. doi: 10.1037/h0047707

Fawcett, J. M., Russell, E. J., Peace, K. A., & Christie, J. (2011). Of guns and geese: A meta-analytical review of the 'weapon focus' literature. *Psychology, Crime & Law, 19*, 1-32. doi: 10.1080/1068316X.2011.599325

Goodsell, C. A., Neuschatz, J. S., & Gronlund, S. D. (2009). Effects of mugshot commitment on lineup performance in young and older adults. *Applied Cognitive Psychology, 23*, 788-803. doi: 10.1002/acp.1512

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*, 847-859. doi: 10.1080/09658210903277318

Hedges, L. V., & Pigott, M. A. (2001). The power of statistical tests in meta-analysis. *Psychological Methods, 6,* 203-217. doi: *10.1037/1082-989X.6.3.203*

Hedges, L.V., & Pigott, M.A. (2004). The power of statistical tests for moderators in meta-analysis. *Psychological Methods, 9*, 426-445. doi: 10.1037/1082-989X.9.4.426

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings*. Newbury Park, CA: Sage.

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

Kinlen, T. J., Adams-Price, C. E., & Henley, T. B. (2007). Verbal overshadowing and face recognition in young and old adults. *Educational Gerontology, 33*, 969-979. doi: 10.1080/03601270701632040

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment. *American Psychologist, 48*, 1181-1201.

Martschuk, N., Kocab, K., & Sporer, S. L. (in preparation). *Memory of faces in old age: A meta-analysis*.

Mastrobuoni, G. (2012). Optimal criminal behavior and the disutility of jail: Theory and evidence on bank robberies. Unpublished submitted manuscript. Retrieved from https://sites.google.com/site/giovannimastrobuoni/

Matt, G. E., & Cook, T. D. (1994). Threats to validity of research synthesis. In H.

Cooper & L. V. Hedges (Eds.), *The handbook of research synthesis* (pp. 503-520). New York: Russell Sage Foundation.

Memon, A., Bartlett, J. C., Rose, R., & Gray, C. (2003). The aging eyewitness: Effects of age on face, delay, and source-memory ability. *Journal of Gerontology: Psychological Sciences, 58B*, 338-345. doi: 10.1093/geronb/58.6.P338

Memon, A., & Gabbert, F. (2003a). Improving the identification accuracy of senior witnesses: Do prelineup questions and sequential lineup help? *Journal of Applied Psychology, 88*, 341-347. doi: 10.1037/0021-9010.88.2.341

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*, 1219-1227. doi: 10.3758/BF03213404

Pickel, K. L. (1998). Unusualness and threat as possible causes of "weapon focus". *Memory, 6,* 277-295. doi: 10.1080/741942361

Pickel, K. L. (2009). The weapon focus effect on memory for female versus male perpetrators. *Memory, 17*, 664-678. doi: 10.1080/09658210903029412

Pickel, K. L., Ross, S. J., & Truelove, R. S. (2006). Do weapons automatically capture attention? *Applied Cognitive Psychology, 20*, 871-893. doi: 10.1002/acp.1235

Pigott, T. D. (2009). Handling missing data. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (2nd ed., pp. 399-416). New York, NY: Russell Sage Foundation.

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

Rose, R. A., Bull, R., & Vrij, A. (2003). Enhancing older witnesses' identification performance: Context reinstatement is not the answer. *Canadian Journal of Police and Security, 1*, 173-184.

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

Shercliffe, R.J., Stahl, W., & Tuttle, M.P. (2009). The use of meta-analysis in psychology: A superior vintage or the casting of old wine in new bottles? *Theory Psychology, 19,* 413-430. doi: 10.1177/0959354309104161

Sporer, S. L., & Cohn, L. D. (2011). Meta-analysis. In B. Rosenfeld, & S. D. Penrod (Eds.), *Research methods in forensic psychology*. New York, NY: Wiley.

Sporer, S. L., & Goodman-Delahunty, J. (2011). *Publication bias in meta-analyses in psychology and law*. Paper presented at the Annual meeting of the American-Psychology - Law Society, Miami, FL.

Stanny, C. J., & Johnson, T. C. (2000). Effects of stress induced by a simulated shooting on recall by police and citizen witnesses. *American Journal of Psychology, 113*, 359-386. Retrieved from http://www.jstor.org/stable/1423364

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16*, 413-424. Retrieved from http://www.jstor.org/stable/1394272

Steblay, N. K. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21*, 283-297. doi: 10.1023/A:1024890732059

Sutton, A. J. (2009). Publication bias. In H. Cooper, L. V. Hedges & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (Vol. 2). New York, NY: Russell Sage Foundation.

Tollestrup, P. A., Turtle, J. W., & Yuilee, J. C. (1994). Actual victims and witnesses to robbery and fraud: An archival analysis. In D. Ross, D. Read & C. Ceci (Eds.), *Adult eyewitness testimony: Current trends and developments*. New York, NY: Press Syndicate of the Universita of Cambridge.

Tooley, V., Brigham, J. C., Maass, A., & Bothwell, R. K. (1987). Facial recognition: Weapon effect and attentional focus. *Journal of Applied Social Psychology, 17*, 845-859. doi:

Valentine, T., Pickering, A., & Darling, S. (2003). Characteristics of eyewitness identification that predict the outcome of real lineups. *Applied Cognitive Psychology, 17*, 696-993. doi: 10.1002/acp.939

Wagstaff, G. F., MacVeigh, J., Scott, L., Brunas-Wagstaff, J., & Cole, J. (2003). Can laboratory findings on eyewitness testimony be generalized to the real world? An archival analysis of the influence of violence, weapon presence, and age on eyewitness accuracy. *Journal of Psychology, 137*, 17-28. doi: 10.1080/00223980309600596

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

Wilcock, R., & Bull, R. (2010). Novel lineup methods for improving the performance of older eyewitnesses. *Applied Cognitive Psychology, 24*, 718-736. doi: 10.1002/acp.1582

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, & Law, 12*, 129-140. doi: 10.1375/pplt.2005.12.1.129

Wilcock, R. A., Bull, R., & Vrij, A. (2007). Are older witnesses always poorer witnesses? Identification accuracy, context reinstatement, own age bias. *Psychology, Crime, & Law, 13*, 305-316. doi: 10.1080/10683160600822212

Deutsche Zusammenfassung

Die Bedeutsamkeit von Augenzeugen bei der strafrechtlichen Verfolgung von Tätern ist nach wie vor unumstritten. Im Rahmen vieler Verbrechen ist es nicht möglich, auf DNA-Beweise zurückzugreifen, so dass der Wert von Personen, die ein Verbrechen entweder als Opfer oder als Augenzeuge beobachtet haben, sehr hoch ist. Neben dem Geständnis eines Tatverdächtigen zählen die Augenzeugenberichte zu den wichtigsten Ermittlungshinweisen (Kassin & Neumann, 1997). Jedoch hat sich gezeigt, dass die Zuverlässigkeit dieser Augenzeugenberichte durch den Einfluss verschiedener Faktoren beeinträchtigt sein kann. Wells (1978) unterteilte jene Variablen, welche mittlerweile im Rahmen vieler Studien untersucht wurden, in zwei Kategorien, die sogenannten *Schätzvariablen* und die *Kontrollvariablen*. Zu der ersten Gruppe zählen Faktoren, deren Einfluss nicht durch das Rechtssystem kontrolliert werden kann. Zu nennen wären zum Beispiel Lichtverhältnisse am Tatort, der Abstand zwischen Täter und Augenzeuge, sowie die Dauer des Verbrechens. Das tatsächliche Ausmaß des Einflusses auf die Qualität der Zeugenaussage lässt sich daher nur post hoc schätzen. Im Gegensatz liegen die sogenannten *Kontrollvariablen* in der Regel unter der Kontrolle des Rechtssystems wie zum Beispiel genutzte Interviewtechniken bei der Vernehmung von Zeugen, die Zusammensetzung der Gegenüberstellung oder die Instruktion der Augenzeugen vor einer solchen.

Die vorliegende Dissertation beschäftigt sich mit dem Einfluss zweier *Schätzvariablen*, welche bereits im Rahmen von Primärstudien intensiv untersucht wurden, dem sogenannten *Waffenfokuseffekt* (Metaanalyse 1) sowie dem *Alter eines Augenzeugens* (Metaanalyse 2). Die zu den beiden Variablen vorliegenden

Primärstudien wurden demnach zusammengefasst und auf quantitativer Ebene metaanalytisch untersucht.

Metaanalyse 1: Waffenfokuseffekt

Der sogenannte Waffenfokuseffekt beschreibt das Phänomen, dass Augenzeugen, die ein Verbrechen beobachtet haben, bei welchem der Täter eine Waffe mit sich geführt hat, später schlechter in der Lage sind, den Täter zu beschreiben und wiederzuerkennen, als Augenzeugen welche nicht mit einer Waffe konfrontiert wurden. Eine erste Metaanalyse aus dem Jahr 1992 konnte einen kleinen Effekt der Waffenpräsenz auf die Zuverlässigkeit der Personenidentifizierung (h =.13) sowie einen mittleren Effekt auf die Genauigkeit der Personenbeschreibung (h =.55) nachweisen (Steblay, 1992). Steblay schloss insgesamt 12 Studien in ihre Analysen ein.

Es gibt drei theoretische Ansätze, um das Phänomen zu erklären. Die *Stresshypothese* geht davon aus, dass Augenzeugen aufgrund der erlebten Erregung durch die Konfrontation mit einer Waffe schlechtere Augenzeugenleistungen zeigen (Deffenbacher, 1983; Deffenbacher, Bornstein, Penrod, & McGorty, 2004; Easterbrook, 1959). Ein neuerer Ansatz, die *Ungewöhnlichkeitshypothese*, wiederum postuliert, dass der Effekt entsteht, da generell nicht erwartet wird, mit einer Waffe konfrontiert zu werden (Pickel, 1998, 2009). Demnach rechnet man beim Betreten einer Bank in der Regel nicht mit einem bewaffneten Überfall. Studien, welche jene Hypothese zu ihrem Forschungsgegenstand gemacht haben, nutzten vor allem ungewöhnliche Objekte, welche der Täter mit sich trug (rosa Plastikflamingo oder ein Huhn), aber auch Waffen in ungewohnten Kontexten (Polizeibeamter vs. Priester mit einer Waffe). Ein dritter Ansatz, die *Aufmerksamkeitshypothese*, geht davon aus, dass es zu einer

Fokussierung von zentralen Details einer Szene kommt und peripheren Informationen, wie dem Gesicht des Täters, weniger Aufmerksamkeit geschenkt wird (Christianson & Loftus, 1991). Letzterer Zugang kann auch als Bestandteil der beiden ersten Erklärungsmodelle angesehen werden, wenn er streng genommen jedoch ein eigenes Modell darstellt.

In der ersten Metaanalyse dieser Dissertation wurde demnach die Forschungslage um den Waffenfokuseffekt auf den neusten Stand gebracht, indem neuere Studien mit eingeschlossen und Erklärungsmodelle überprüft wurden. Insgesamt wurde angenommen, die Effekte der Waffenpräsenz auf die Beschreibungs- und Wiedererkennensleistung von Versuchspersonen replizieren zu können (Steblay, 1992).

Hinsichtlich der Überprüfung der *Stresshypothese* wurde erwartet, dass das Ausmaß des Waffenfokuseffekts vom Stresserleben eines Augenzeugen abhängt. Hier ergaben sich jedoch Schwierigkeiten in der Operationalisierung, da nur wenige Studien das empfundene Stressausmaß mit erhoben haben und dies im Allgemeinen auf subjektiver Ebene exploriert wurde. Aufgrund ethischer Beschränkungen ist natürlich in experimentellen Designs nie von einem vergleichbaren Erregungsniveau auszugehen, wie man es bei dem Beobachten eines tatsächlichen Verbrechens vermuten würde. Im Rahmen von Moderatoranalysen wurden demnach Variablen untersucht, die augenscheinlich in Verbindung mit einem höheren Stressempfinden stehen könnten (Darstellung eines Verbrechens, Gefährlichkeit der Waffe, Präsentationsformat des Versuchsmaterials).

Im Hinblick auf die *Ungewöhnlichkeitshypothese* wurde ein vergleichbarer Effekt wie bei traditionellen Waffenfokuseffektuntersuchungen postuliert. Ein ungewöhnliches Objekt oder eine Waffe, präsentiert in einem ungewöhnlichen

Kontext, sollte demnach zu ähnlichen Effekten führen wie die traditionelle Präsentation einer Waffe in einem Verbrechensszenario.

Bezüglich der *Aufmerksamkeitshypothese* wurde angenommen, dass eine Waffe (zentrale Information) besser erinnert werden kann als ein anderes (neutrales) Objekt, welches von einem Täter mitgeführt wurde.

Auf der Basis einer umfangreichen Literatursuche konnten letztlich 23 Studien identifiziert werden, welche den vorher formulierten Einschlusskriterien entsprachen. Als abhängige Variablen fungierten korrekte Entscheidungen in der Lichtbildvorlage (über beide Gegenüberstellungsformen hinweg: mit und ohne Täterpräsenz), sowie korrekte Identifizierungen des Täters in Lichtbildvorlagen mit Täterpräsenz, korrekte Benennung des mitgeführten Objekts und die Genauigkeit einer Personenbeschreibung. Ein entsprechender Effekt im Rahmen von Falschidentifizierungen bei Lichtbildvorlagen ohne Täterpräsenz konnte leider nicht untersucht werden, da nur wenige Studien (k = 3) hierfür Daten lieferten.

Die Einzeleffekte der Primärstudien wurden schließlich in einem ursprünglich von Hedges und Olkin (1985) empfohlenen Modell gewichteter Effekte integriert (siehe auch Borenstein, 2009; Fleiss & Berlin, 2009; Shadish & Haddock, 2009; Konstantopoulos & Hedges, 2009; Raudenbusch, 2009). Im Hinblick auf die Identifizierungsleistung (binäre abhängige Variable) wurde das sogenannte Chancenverhältnis (*Odds ratio*: *OR*; Fleiss & Berlin, 2009; Lipsey & Wilson, 2001; Sporer & Cohn, 2011) berechnet. *OR* ist eine Effektstärke, welche zwei Gruppen hinsichtlich ihrer relativen Chance des Eintretens eines Ereignisses miteinander vergleicht. Sofern für beide Gruppen dieselbe Chance für ein Ereignis besteht (hier zum Beispiel der korrekten Identifizierung eines Tatverdächtigen) so hat *OR* die Größe 1. Bei der Synthese der Effektstärken sollte allerdings der natürliche

Logarithmus der *OR* verwendet werden (*LOR*), dessen Werte sich um 0 zentrieren. Bei der Interpretation werden jene Werte schließlich zurücktransformiert (Fleiss & Berlin, 2009; Lipsey & Wilson, 2001). Die Effektstärken wurden so berechnet, dass ein OR größer 1 eine bessere Identifizierungsleistung der Kontrollgruppe darstellt. Hinsichtlich der kontinuierlichen Variablen der Personen- und Objektbeschreibungen wurde Hedges g_u berechnet, eine standardisierte Effektstärke zum Vergleich zweier Mittelwerte und deren Standardabweichungen.

Bevor die berechneten Effekte der einzelnen Studien im Rahmen des *Modells fester Effekte* (fixed effects model, FEM) und des *Modells zufallsvariabler Effekte* (random effects model, REM) gewichtet wurden, wurde anhand von Ausreißeranalysen überprüft, ob sich unter den Studieneffekten Extremwerte befanden (Hedges & Olkin, 1985). Hierfür wurden sowohl graphische Darstellungen (Verteilungen der Einzeleffektstärken mit ihren Konfidenzintervallen) genutzt als auch die Ausreißeranalyse nach Hedges und Olkin (1985) berechnet (siehe auch Sporer & Cohn, 2011).

Die Einzeleffekte wurden schließlich anhand ihrer inversen Varianz in einem Modell gewichtet (FEM). Im REM wurde zusätzlich eine Varianzkomponente berechnet, da in diesem Modell davon ausgegangen wird, dass die Varianz nicht nur auf Stichprobenfehler der einzelnen Studien zurückzuführen ist, sondern ebenfalls auf systematische Varianz zwischen jenen Studien. Neben der *Q* Statistik, die generell über eine geringe statistische Power verfügt, wurde als Homogenitätsmaß außerdem die deskriptive Statistik I^2 herangezogen (Shadisch & Haddock, 2009; Hedges & Pigott, 2001, Higgins & Thompson, 2002; Sutton & Higgins, 2008).

Moderatorenanalysen wurden vor allem im Rahmen von hierarchisch sequentiellen Metaregressionen berechnet, bei denen Abhängigkeiten zwischen den

Prädiktorvariablen berücksichtig werden. Sofern geblockte Analysen Anwendung fanden, wurden Kreuzauswertungen sowie Interkorrelationen berechnet, um potentielle Abhängigkeiten zwischen Prädiktorvariablen und/ oder geringe Zellbesetzungen zu entdecken (siehe Sporer & Cohn, 2011).

Die Ergebnisse hinsichtlich der korrekten Identifizierung eines Tatverdächtigen durch den Zeugen erreichten keine Signifikanz (*OR* = 1.287). Demnach war es für Teilnehmer, welche nicht mit einer Waffe konfrontiert wurden, lediglich 1.29 mal wahrscheinlicher den Täter bei einer Lichtbildvorlage mit Zielperson korrekt zu identifizieren im Vergleich zu den Versuchspersonen in den Waffenbedingungen. Der Effekt für richtige Entscheidungen über beide Gegenüberstellungsarten (mit und ohne Täterpräsenz) war vergleichbar gering (*OR* = 1.101).

Im Gegensatz ergab sich ein moderater Effekt für Beschreibungen des Täters, indem Personen, die ein Szenario ohne Waffenpräsenz beobachteten die Zielperson genauer beschreiben konnten, als jene die mit einer Waffe konfrontiert wurden (g_u = 0.568).

Demnach konnten die Ergebnisse einer ersten Metaanalyse aus dem Jahr 1992 nur teilweise repliziert werden, indem auf der Basis einer größeren Stichprobe der Effekt für die Identifizierungsleistung ausblieb.

Bezüglich der Überprüfung der *Stresshypothese* zeigte sich im Rahmen von Moderatoranalysen, dass sowohl die Waffenart als auch die Darbietung eines Verbrechens signifikant mit der Genauigkeit der Täterbeschreibung assoziiert waren. Während ein größerer Waffenfokuseffekt unter Darbietung eines Verbrechens beobachtet werden konnte, zeigte sich ein nicht erwarteter Effekt im Hinblick auf die Tatwaffe. Demnach ließ sich ein größerer Effekt bei einem Messer (oder einem

Fleischerbeil) als Tatwaffe (g_u = 0.773) im Vergleich zu einer Pistole (g_u = 0.400) berechnen. Es ist möglich, dass heutzutage durch TV-Thriller eine Gewöhnung an die Darbietung einer Pistole stattgefunden hat. Eine andere Erklärung könnte sein, dass durch die Präsentation eines Messers bei den Beobachtern eine höhere Erregung ausgelöst wird, da ein blutigeres Szenario erwartet wird. Ob tatsächlich ein höheres Stresserleben mit den untersuchten Variablen einhergeht, konnte letztlich im Rahmen der vorliegenden Arbeit nicht eindeutig geklärt werden.

Zur Überprüfung der *Ungewöhnlichkeitshypothese* zeigten sich vergleichbare Effekte wie bei der Untersuchung des traditionellen Waffenfokuseffekts. Demnach blieben die Ergebnisse hinsichtlich der korrekten Identifizierung eines Tatverdächtigen insignifikant (*OR* = 1.329), während moderate bis große Effekte für die Personenbeschreibung berechnet werden konnten (g_u = 0.761). Dies spricht dafür, dass bei der Entstehung des Phänomens Gefühle des Erstaunens und der Überraschung mit verantwortlich sind (Pickel, 1998, 2009).

Die *Aufmerksamkeitshypothese* konnte anhand von Studien untersucht werden, die ein Abfragen des mitgeführten Objekts beinhalteten. Hier zeigte sich, dass im Durchschnitt 93.8% der Studienteilnehmer korrekt berichteten, eine Waffe gesehen zu haben, während lediglich 72.0% in den Vergleichsgruppen das entsprechende Kontrollobjekt benannten. Dies spricht dafür, dass es bei der Anwesenheit einer Waffe zu einer Fokussierung zentraler Informationen kommt, während die Personen in den Vergleichsbedingungen mehr Kapazitäten für andere Details zur Verfügung haben.

Es bleibt festzuhalten, dass, auch wenn kein signifikanter Effekt der Waffenpräsenz auf die Identifizierungsleistung von Augenzeugen nachgewiesen werden konnte, eine fehlerhafte Personenbeschreibung Hinweise auf eine unschuldige Person liefern kann. Jene könnte letztlich im Rahmen einer Gegenüberstellung identifiziert und verhaftet werden.

Praktische Implikationen gehen in die Richtung, gewisse Hochrisikogruppen, wie zum Beispiel Mitarbeiter einer Bank oder einer Tankstelle entsprechend über die Existenz des Effekts zu schulen. So konnten Pickel, French und Betts (2003) zeigen, dass allein das Wissen um das Phänomen die Stärke des Effekts reduzieren kann. Des Weiteren könnten sie in entsprechenden Trainings den Umgang mit entsprechenden Situationen üben. Stanny und Johnson (2000) konnten zum Beispiel zeigen, dass Polizeibeamte, die in einer Simulationsstudie mit einer Waffe konfrontiert wurden, keine Unterschiede bezüglich ihrer Herzfrequenz zeigten im Vergleich zu Polizeibeamten, die nicht mit einer Waffe konfrontiert wurden. Man könnte daher schlussfolgern, dass sie eher an den Umgang mit einer Waffe gewöhnt sind, wodurch sie mehr Kapazitäten für periphere Details verfügbar hätten.

Auf der anderen Seite könnte das detaillierte Wissen über die Waffe des Täters für Ermittlungsarbeiten genutzt werden. Denkbar wäre zum Beispiel eine Lichtbildvorlage mit unterschiedlichen Waffen (siehe auch; Sauerland & Sporer, 2008; Sauerland, Sagana, & Sporer, 2012, zur Diagnostizität von Objektidentifizierungen).

Schließlich sollten Forscher ermutigt werden, Untersuchungen mit Gegenüberstellungen ohne Täterpräsenz durchzuführen, da über die Auswirkungen hier kaum etwas bekannt ist und Sachverständige bei Gericht häufig hinzugezogen werden, um einen Tatverdächtigen zu entlasten.

Metaanalyse 2: Ältere Menschen als Augenzeugen

Aufgrund verbesserter Lebensbedingungen ist heutzutage mit einer höheren Lebensdauer zu rechnen, was mit sich bringt, dass die Gruppe älterer Menschen zu der schnellst wachsenden Population in Industrienationen zählt. Aufgrund dieses Anstiegs wird es immer mehr ältere Personen geben, die in irgendeiner Weise in das Strafjustizsystem involviert sind, sei es als Augenzeuge eines Verbrechens oder gar als Opfer. Im Verlauf der vergangenen Forschungsdekaden haben sich jedoch deutliche Hinweise auf schlechtere Gedächtnisleistungen älterer Menschen im Gegensatz zu Jüngeren ergeben. Zum Beispiel zeigte sich eine Abnahme der Gedächtnisfertigkeiten im Bereich des Enkodierens neuer Informationen, des Arbeitsgedächtnisses, von Inhibitionsfunktionen, des Langzeitgedächtnisses und der Verarbeitungsgeschwindigkeit (für einen Überblick, siehe Light, 1991; Lustig, Hasher, & Zacks, 2007; Park & Reuter-Lorenz, 2009; Salthouse, 1996). Deutliche Einbußen zeigen sich auch im Hinblick auf das episodische Gedächtnis und hier inbesondere für kontextuelle Informationen. Ältere Menschen zeigen demnach Schwierigkeiten im Erinnern der Quelle einer Information (Aizpurua, Garcia-Bajos, & Migueles, 2011; Bornstein, 1995; Cansino, Trejo-Morales, & Hernández-Ramos, 2010; Johnson, Hashtroudi, & Lindsay, 1993). Ist jene Quelle nicht verfügbar, nutzen Menschen in der Regel heuristische Strategien, indem sie zum Beispiel die Vertrautheit eines Reizes als Informationsquelle heranziehen (Bright-Paul, Jarrold, & Wright, 2005; Johnson et al., 1993). Vor einer Identifizierungsaufgabe ist es jedoch wichtig, sich seiner Erinnerung und vor allem der Quelle dieser bewusst zu sein. Aufgrund der dargestellten wissenschaftlichen Befunde wurde postuliert, dass ältere Augenzeugen generell weniger korrekte Identifizierungen, mehr Falschidentifizierungen und insgesamt weniger korrekte Entscheidungen im Rahmen einer Identifizierungsaufgabe treffen im Vergleich zu jüngeren Vergleichspersonen.

Nach intensiven Recherchen konnte keine Metaanalyse identifiziert werden, welche die Identifizierungsleistung älterer Augenzeugen im Vergleich zu jüngeren

Personen untersucht hat. Aufgrund einer umfangreichen Literaturrecherche konnten schließlich 22 Identifizierungsstudien mit in die Analysen eingeschlossen werden. Das methodische Vorgehen bei der Integration der Effekte entspricht dem der ersten Metaanalyse. Als Effektstärke wurde *OR* berechnet.

Insgesamt zeigte sich hinsichtlich aller untersuchten abhängigen Variablen ein signifikanter Vorteil der jüngeren Augenzeugen über die Älteren. Das Chancenverhältnis, eine richtige Entscheidung zu treffen (über beide Gegenüberstellungsformen hinweg; mit und ohne Täterpräsenz), war für junge Erwachsene 2.1 Mal so hoch wie für ältere Personen (*OR* = 2.11). Die größten Unterschiede zwischen den Altersgruppen zeigte sich im Hinblick auf Falschidentifizierungen. Hier war das Chancenverhältnis, eine falsche Person aus der Gegenüberstellung zu wählen, für ältere Versuchspersonen 2.5 bis 3 Mal so hoch wie für jüngere (Falschidentifizierungen in Gegenüberstellungen mit Täter: *OR* = 2.45; Falschidentifizierungen in Gegenüberstellungen ohne Täter: *OR* = 3.07).

Als bedeutsam für die Interpretation der Ergebnisse erwies sich das Wahlverhalten der älteren Versuchspersonen. Hier zeigte sich, dass für ältere Augenzeugen das Chancenverhältnis, eine Person aus einer Gegenüberstellung zu wählen, 2.3 Mal so hoch war wie für junge Erwachsene (*OR* = 2.26). Es ist folglich von einer *Antwortneigung zum Wählen* bei älteren Personen auszugehen (*sozialer Erklärungsansatz*). Dieses Verhalten erhöht zwar auf der einen Seite die Trefferquote, führt jedoch auf der anderen Seite zu mehr Falschidentifizierungen (Clark, 2005; Fulero, 2009; Steblay, 1997; siehe auch Clark, 2012, für eine allgemeine Diskussion). Gründe für die Neigung zum Wählen könnten sein, dass sich ältere Augenzeugen besonders berufen fühlen, der Polizei behilflich zu sein. Eine andere Erklärung könnte sein, dass ihnen die Möglichkeit, dass sich der Täter nicht in der Gegenüberstellung befindet, unzureichend bewusst ist. Dass ältere Augenzeugen jene Instruktionen weniger erinnern, konnten Studien bereits belegen (Rose, Bull, & Vrij, 2003, 2005). Demnach sollte Augenzeugen vor einer Identifizierungsaufgabe sowohl in schriftlicher als auch in mündlicher Form deutlich gemacht werden, dass sich ein Täter auch nicht in der Gegenüberstellung befinden kann. Das zusätzliche Verdeutlichen der Konsequenzen einer Falschidentifizierung führt außerdem zu einem besseren Erinnern der gegebenen Instruktionen (Wilcock, Bull, & Vrij, 2005).

Auf der anderen Seite könnten auch Probleme hinsichtlich der Identifizierung der Quelle einer erinnerten Information für höhere Fehlerraten bei älteren Augenzeugen verantwortlich sein (*kognitiver Erklärungsansatz*). Hier könnte eine dargebotene Person in der Gegenüberstellung zum Beispiel bei älteren Zeugen Gefühle der Vertrautheit wecken. Da sie die Quelle jedoch nicht eindeutig identifizieren können, schlussfolgern sie, dass es der Täter sein muss.

Als bedeutsamer Moderator erwies sich unter anderem die Darbietungszeit der Zielperson. Entgegen der vorher postulierten Erwartungen zeigte sich jedoch, dass eine längere Darbietungszeit zu größeren Unterschieden zwischen den beiden Altersgruppen führte. Möglicherweise fühlen sich die älteren Augenzeugen durch die lange Beobachtungsdauer besonders befähigt, den Täter in einer Gegenüberstellung zu identifizieren. Außerdem zeigte sich, dass fast alle Identifizierungsentscheidungen (abhängige Variablen) signifikant mit dem durchschnittlichen Alter der älteren Versuchspersonen assoziiert waren. Es scheint demnach: Je älter ein Augenzeuge, desto schlechter die Identifizierungsleistung.

Als eine vom Rechtssystem während Ermittlungsarbeiten beeinflussbare Moderatorvariable wurde die Darbietungsform der Gegenüberstellung variiert. Hier profitieren ältere Zeugen von einer sequentiell dargebotenen Gegenüberstellung, bei welcher die Personen der Reihe nach präsentiert werden und der Augenzeuge bei jeder Person entscheiden muss, ob es sich um den beobachteten Täter handelt oder nicht. Eventuell wenden ältere Augenzeugen hierdurch ein strengeres Entscheidungskriterium an, indem sie jede dargebotene Person mit dem erinnerten Bild des Täters in ihrem Gedächtnis vergleichen müssen und nicht die Personen in der Gegenüberstellung untereinander.

Um schließlich weitere praktische Hinweise geben zu können, sollte das Wahlverhalten älterer Augenzeugen detaillierter untersucht werden. Welche Gedanken vollziehen sich während des Entscheidungsprozesses? Wen genau wählt ein Augenzeuge fälschlicherweise aus einer Gegenüberstellung? Die Ersatzperson der Zielperson oder lediglich eine Füllerperson, deren Unschuld in der Regel bekannt ist. Hier sind vor allem Studien mit Gegenüberstellungen ohne Täterpräsenz vonnöten.

Eventuell wäre es sinnvoll, in der Praxis Strategien anzuwenden, mit denen älteren Personen das Erinnern der benötigten Informationen erleichtert werden würde, wie zum Beispiel sogenannte Kontext-Wiederherstellungsinstruktionen (Wilcock, Bull, & Vrij, 2007; Rose, et al., 2003) oder das Visualisieren des Tathergangs (Kinlen, Adams-Price, & Henley, 2007). Hier sind jedoch die Ergebnisse teils wiedersprüchlich, so dass intensivere Forschung Aufschluss liefern könnte. Fragen vor der Gegenüberstellung, z.B. hinsichtlich des Aussehens des Täters oder der Zuversicht in das eigene Wahlverhalten (prelineup questions; O'Rourke et al., 1989; Scogin et al., 1994; Searcy et al., 2001; Wilcock & Bull, 2010) sowie Trainingslichtbildvorlagen (Wilcock & Bull, 2010) könnten ältere Menschen unterstützen, ein vorsichtigeres Entscheidungskriterium anzuwenden.

Zusammenfassend kann aufgrund der vorliegenden Daten beider Metaanalysen festgehalten werden, dass die Präsenz einer Waffe mit einer schlechteren Beschreibung des Täters und das höhere Alter eines Augenzeugen mit schlechteren Identifizierungsleistungen einhergeht. Es ist jedoch davon auszugehen, dass die Ergebnisse aus Laborstudien von denen realer Fälle abweichen, so dass die vorliegenden Metaanalysen die Effekte noch unterschätzen. Mögliche Gründe könnten Unterschiede im Stresserleben, aber auch hinsichtlich der Konsequenzen einer Identifizierungsentscheidung sein. Hier wäre es sicherlich sinnvoll Laborergebnisse stärker mit Ergebnissen archivarischer Analysen zu verbinden.

Auf der anderen Seite zeigte sich jedoch, dass das Vorhandensein einer Publikationsverzerrung im Hinblick auf die dargestellten Metaanalysen nicht ausgeschlossen werden kann. Dies bedeutet, dass nicht signifikante Ergebnisse häufig nicht eingereicht oder veröffentlicht wurden und daher in den durchgeführten Analysen unterrepräsentiert sind. Auch wenn durch den Kontakt zu Autoren versucht wurde, unpubliziertes Material zu erhalten, könnten die dargestellten Ergebnisse jene realer Ereignisse überschätzen. Dennoch konnten vielversprechende Methoden aufgezeigt werden, um den Einfluss der untersuchten Variablen auf die Augenzeugenleistung zu verringern, welche in weiteren Studien vertiefend untersucht werden sollten.

Literatur

Aizpurua, A., Garcia-Bajos, E., & Migueles, M. (2011). False recognition and source attribution for actions of an emotional event in older and younger adults. *Experimental Aging Research, 37,* 310-329. doi: 10.1080/0361073X.2011.568829

Borenstein, M. (2009). Effect sizes for continuous data. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 221-236). New York, NY: Russell Sage Foundation.

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

Bright-Paul, A., Jarrold, C., & Wright, D. B. (2005). Age-approriate cues facilitate source-monitoring and reduce suggestibility in 3- to 7-year-olds. *Cognitive Development, 20,* 1-18. doi: 10.1016/j.cogdev.2004.06.001

Cansino, S., Trejo-Morales, P., & Hernández-Ramos, E. (2010). Age-related changes in neural activity during source memory encoding in young, middle-aged and elderly adults. *Neuropsychologica, 48,* 2537-2549. doi: 10.1016/j.neuropsychologia.2010.04.032

Christianson, S.-A., & Loftus, E. F. (1991). Remembering emotional events: The fate of detailed information. *Cognition and Emotion, 5,* 91-108. doi: 10.1080/02699939108411027

Clark, S. E. (2005). A re-examination of the effects of biased lineup instructions in eyewitness identification. *Law and Human Behavior, 29,* 395-424. doi: 10.1007/s10979-005-5690-7

Clark, S. E. (2012). Costs and benefits of eyewitness identification reform: Psychological science and public policy. *Perspectives on Psychological Science, 7,* 238-259. doi: 10.1177/1745691612439584

Deffenbacher, K. A. (1983). Identification evidence: A psychological evaluation. *American Journal of Psychology, 96,* 591-595. doi: 10.1007/BF00868064

Deffenbacher, K. A., Bornstein, B. H., Penrod, S. D., & McGorty, E. K. (2004). A meta-analytic review of the effects of high stress on eyewitness memory. *Law and Human Behavior, 28,* 687-706. doi: 10.1007/s10979-004-0565-x

Easterbrook, J. A. (1959). The effect of emotion on cue utilization and the organization of behavior. *Psychological Review, 66,* 183-201. doi: 10.1037/h0047707

Fleiss, J. L., & Berlin, J. A. (2009). Effect sizes for dichotomous data. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 237-256). New York, NY: Russell Sage Foundation.

Fulero, S. M. (2009). System and estimator variables in eyewitness identification: A review. In D. A. Krauss, & J. D. Lieberman (Eds.), *Psychological expertise in court*. Burlington, VT: Ashgate.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Orlando, FL: Academic Press.

Hedges, L. V., & Pigott, M. A. (2001). The power of statistical tests in meta-analysis. *Psychological methods, 6,* 203-217. doi: 10.1037/1082-989X.6.3.203

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21,* 1539-1558. doi: 10.1002/sim.1186

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

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,* 469-484. doi: 10.1023/A:1024871622490

Kinlen, T. J., Adams-Price, C. E., & Henley, T. B. (2007). Verbal overshadowing and face recognition in young and old adults. *Educational Gerontology, 33,* 969-979. doi: 10.1080/03601270701632040

Konstantopoulos, S., & Hedges, L. V. (2009). Analyzing effect sizes: Fixed-effects models. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 279-293). New York, NY: Russell Sage Foundation.

Light, L. L. (1991). Memory and aging: Four hypotheses in search of data. *Annual Review of Psychology, 42,* 333-376. doi: 0.1146/annurev.ps.42.020191.002001

Lipsey, M. W., & Wilson, D. B. (2001). *Practical Meta-Analysis*. Thousand Oaks, CA: SAGE Publications.

Lustig, C., Hasher, L., & Zacks, R. T. (2007). Inhibitory deficit theory: Recent development in a "New view". In D. S. Gorfein, & C. M. MacLeod (Eds.), *The place inhibition in cognition* (pp. 145-162). Washington, DC: American Psychology Association.

O'Rourke, T. E., Penrod, S. D., Cutler, B. L., & Stuve, T. E. (1989). The external validity of eyewitness identification research: Generalizing across subject populations. *Law and Human Behavior, 13,* 385-395. doi: 10.1007/BF01056410

Park, D. C., & Reuter-Lorenz, P. (2009). The adaptive brain: Aging and neurocognitive scaffolding. *Annual Review of Psychology, 60,* 173-196. doi: 10.1146/annurev.psych.59.103006.093656

Pickel, K. L. (1998). Unusualness and threat as possible causes of "weapon focus". *Memory, 6,* 277-295. doi: 10.1080/741942361

Pickel, K. L. (2009). The weapon focus effect on memory for female versus male perpetrators. *Memory, 17,* 664-678. doi: 10.1080/09658210903029412

Pickel, K. L., French, T. A., & Betts, J. M. (2003). A cross-modal weapon focus effect: The influence of a weapon's presence on memory for auditory information. *Memory, 11,* 277-292. doi: 10.1080/09658210244000036

Raudenbush, S. W. (2009). Analyzing effect sizes: Random-effects models. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 295-315). New York, NY: Russell sage Foundation.

Rose, R. A., Bull, R., & Vrij, A. (2003). Enhancing older witnesses' identification performance: Context reinstatement is not the answer. *Canadian Journal of Police and Security Services, 1,* 173-184.

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

Salthouse, T. A. (1996). Constrains on theories of cognitive aging. *Psychonomic Bulletin & Review, 3,* 287-299. doi: 10.3758/BF03210753

Sauerland, M., & Sporer, S. L. (2008). The application of multiple lineups in a field study. *Psychology, Crime & Law, 14,* 549-564. doi: 10.1080/10683160801972519

Sauerland, M., Sagana, A., & Sporer, S. L. (2012). Assessing nonchoosers' eyewitness identification accuracy from photographic showups by using confidence and response times. *Law and Human Behavior, 36,* 394-403. doi: 10.1037/h0093926

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

Searcy, J., Bartlett, J. C., Memon, A., & Swanson, K. (2001). Aging and lineup performance at long retention intervals: Effects of metamemory and context reinstatement. *Journal of Applied Psychology, 86,* 207-214. doi: 10.1037/0021-9010.86.2.207

Shadish, W. R., & Haddock, C. K. (2009). Combining estimates of effect sizes. In H. Cooper, L. V., Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis and meta-analysis* (pp. 257-277). New York, NY: Russell Sage Foundation.

Sporer, S. L., & Cohn, L. D. (2011). Meta-analysis. In B. Rosenfeld, & S. D. Penrod (Eds.), *Research methods in forensic psychology*. New York, NY: Wiley.

Stanny, C. J., & Johnson, T. C. (2000). Effects of stress induced by a simulated shooting on recall by police and citizen witnesses. *American Journal of Psychology, 113,* 359-386. Retrieved from http://www.jstor.org/stable/1423364

Steblay, N. M. (1992). A meta-analytic review of the weapon focus effect. *Law and Human Behavior, 16,* 413-424. Retrieved from http://www.jstor.org/stable/1394272

Steblay, N. K. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21,* 283-297. doi: 10.1023/A:1024890732059

Sutton, A. J., & Higgins, J. P. (2008). Recent developments in meta-analysis. *Statistics in Medicine, 27,* 625-650. doi: 10.1002/sim.2934

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

Wilcock, R. A., & Bull, R. (2010). Novel lineup methods for improving the performance of older eyewitnesses. *Applied Cognitive Psychology, 24,* 718-736. doi: 10.1002/acp.1582

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, & Law, 12,* 129-140. doi: 10.1375/pplt.2005.12.1.129

Wilcock, R. A., Bull, R., & Vrij, A. (2007). Are older witnesses always poorer witnesses? Identification accuracy, context reinstatement, own age bias. *Psychology, Crime, & Law, 13,* 305-316. doi: 10.1080/10683160600822212

Erklärung

Ich erkläre: Ich habe die vorgelegte Dissertation selbständig und nur mit den Hilfen angefertigt, die ich in der Dissertation angegeben habe. Alle Textstellen, die wörtlich oder sinngemäß aus veröffentlichten oder nicht veröffentlichten Schriften entnommen sind, und alle Angaben, die auf mündlichen Auskünften beruhen, sind als solche kenntlich gemacht.

Gießen, im Mai 2013

Kerstin Kocab